UNITED STATES
NUCLEAR WASTE TECHNICAL REVIEW BOARD

STRUCTURAL GEOLOGY & GEOENGINEERING and
HYDROLOGY & GEOCHEMISTRY JOINT PANEL MEETING

Update on Major Task Force Efforts

Stouffer Concourse Hotel
Ballroom - Lower Level
3801 Quebec Street
Denver, Colorado

March 6, 1991

BOARD MEMBERS PRESENT

Dr. Don U. Deere - Chairman, NWTRB
Dr. Clarence Allen - Chairman, SG&G Panel
Dr. Patrick Domenico - Co-Chairman, Hydrology & Geochemistry Panel
Dr. Donald Langmuir - Co-Chairman, Hydrology & Geochemistry Panel
Dr. Warner North - Member

SENIOR PROFESSIONAL STAFF

Dr. Leon Reiter
Mr. Russell McFarland

CONSULTANTS

Dr. Roy E. Williams
Dr. Edward J. Cording

I N D E X

P R O C E E D I N G S

DR. DEERE: Good morning. I'm Don Deere, Chairman of the Nuclear Waste Technical Review Board, member of the panel on Structural Geology & Geoengineering.

This is the first meeting of the year, and of the panel on Structural Geology & Geoengineering. It is in the form of an exchange with the DOE regarding exploratory shaft facility, and the surface-based testing program and other topics, as you will see in a moment.

Our Technical Review Board and our panel has had an interest in the exploratory shaft facility since it was first presented to us in March of '89 in Washington, and in June in Las Vegas. The subject captured our attention at that time in three ways: One, the construction was to have started in October, 1989, and we had only a short time to evaluate the project and to make any comments. Number two, the construction was to be by traditional drill and blast methods. Three, there was considerable engineering and scientific effort to take into account the disturbance of the rock by the drilling, drilling water, and the blast-generated shock waves.

We believe that the newer methods of raised boring, V-moling, or shaft boring could reduce the problems and give better technical results, and also the potential for greater advance rates to lower costs.

The DOE has accepted our recommendations to reexamine the problem, and to evaluate our other recommend- ations: including crossing the Ghost Dance Fault at least twice; excavating an east/west drift completely across a Yucca Mountain block to intercept known faults and define any unknown north/south striking faults; as well as, seeing more of the Tonopah Springs formation over a larger area and greater extent; and of considering an incline tunnel or ramp from the surface into the Yucca Mountain block to cross many of the known faults at some depth below the surface, so that their mechanical characteristics could be observed. Also, the incline would allow most of the tuff units to be crossed. Alcoves or short drifts could be excavated at any position of interest off to the side for mapping or testing.

The Board has a desire to remain in contact with the DOE and its contractors as the study goes forward. Dr. Clarence Allen is the Chairman of the Panel on Structural Geology/Geoengineering, and he has taken the lead on the aspects of volcanism, faulting, and seismicity. He has allowed me to take the lead on the geoengineering aspects; that is, the shaft facility, rock mechanics, et cetera.

Our professional staff members who have helped Clarence and me with our work on this panel and related panels have had a great input, and we certainly want to thank them for helping us organize this meeting.

The Panel has also had the benefit of other Board members, and they will be introduced shortly by Dr. Allen, Chairman of this committee. We have interest from the Panel of Risk & Performance Analysis in all of the work that we do as well, and before turning the meeting over to Clarence, I would like to thank all of you for coming to this meeting.

Now, what I have just read has the date January 31st, 1990, Denver. This is just to remind us that that's where we were one year ago, and I just thought it'd be of interest to see if we have progressed, and I believe that we will find by the presentations that are made that, indeed, a lot of work has been done. We're going to get summary and final conclusions on some of the work, and we're going to get some results of semifinals on others, and sort of a status report on others.

Actually, this meeting today is not a technical exchange. It is a joint meeting of two panels, and so we are expanding our scope a little bit primarily because of the items on testing that will be brought in, and our panel on hydrogeology and geochemistry is going to have a real interest in the testing. And the group from our particular panel is going to have a lot of interest in the testing priorities and the rock mechanics testing as well.

So with this little introduction, I would like to-- well, I have been very favorably impressed eight or ten days ago when I was on the Channel Tunnel, and I wanted to bring a couple words on that. They are essentially breaking records almost every day, and the machine this month will have--one machine will have advanced one mile in one month, and so it's going along at the rate of about a quarter of a mile a week. The other machines are maybe 50 or 100 feet or 200 feet behind that in their rate of advance. So they, who were several months behind schedule in June of last year because of running into faults or false ground or lots of water--which was, incidentally, fresh to brackish when they first hit it, and in a question of days, it was salt water going right straight down, so you know you had reached percolation. The depths are about 40 meters of rock above you, and 40 meters of water above that.

In the French crossover, which is a cavern about seven miles offshore where the two tunnels come together, so that in case there's an accident or you have one reason to take the trains from one tunnel into the other, there is a place they can cross over. Well, this is certainly one of the largest cavities ever built, and certainly the largest one ever built underground, as far as I know, below the sea. It's about the length of two football fields, and 80 feet in diameter, about 55 feet high.

The reason I'm mentioning it, to get access to it, they came off the surface tunnel and then drove two inclining ramps, and since we're talking about ramps, it's of interest to us. The ramps are inclined at 10.4 per cent, and they start at the bottom of the cavern and over the length of about 400 feet they rise up to the top. Now, to handle the--to drive these, they're using a road header, no blasting. It's all with a road header-type of excavator. The muck comes off the conveyor belt, and then drops into a conveyor belt that is really a trough about four feet wide, and perhaps two feet high, with a moving belt or chain in the bottom. So all of the muck is brought down off to the side, right along the edge of the wall, down to the bottom of the ramp, and then transferred over by another conveyor into the train, and then taken out.

Since so many of these--the advance, the use of a road header for going off to the side and doing alcoves and such things is a consideration that is being looked at and has been looked at, I thought it was of interest to tell you this. Now, the road header is not advancing at these rates. These high rates I'm telling you about are the TBM's, and you must remember that the rock has a strength just about like our nonwelded tuff, just about like that. It's in the range of 800 to about 1500 psi, with occasional zones, since this is cyclic sedimentation, of thin limestone that's much harder than that every one meter or meter and a half.

Well, so we know that these procedures can advance at a fairly good rate. I turned my head away from the microphone, and I had a note here, given to me at the very beginning of the meeting, "Please speak into the mike." So I will ask the other members to please obey me and try to speak into the mike.

Well, Dr. Clarence Allen, as I mentioned, is the Chairman of our Committee on Structural Geology and Geoengineering. I would like to turn it over to him for as long as he'd like it.

DR. ALLEN: Well, let me simply welcome you on behalf of the Panel, the Structural Geology and Geoengineering Panel. Although Don stated this was our first meeting of the year, some of you with long memories will remember we met last Friday in Tucson.

DR. DEERE: It wasn't in my 1990 notes.

DR. ALLEN: I see. And I should also emphasize, this is a joint meeting with the Hydrogeology and Geochemistry Panel. The two co-chairs of that Panel are with us today; Pat Domenico, second on my right, and Don Langmuir, down at the end of the table. In addition, other members of the Board here in addition to Don include Warner North. We have two consultants, Ed Cording and Roy Williams, and two members of the senior professional staff, Russ McFarland and Leon Reiter.

As I understand it, since the primary emphasis of this meeting is not on volcanism or earthquakes or related features, you are going to chair the program from here on out, so I turn it back over to you, Don.

DR. DEERE: Thank you very much.

So we are all interested and ready and wiling to hear from DOE, the DOE and their contractors. We want to thank Max Blanchard very much for coming to the meeting and leading off for the DOE.

MR. BLANCHARD: Thank you, Don.

I'd like to introduce our team for the next day and a half. I'm Max Blanchard. I'm Acting Deputy for the project. Carl sends his regrets. He'd like to be here and was planning on coming, but things that developed last week; tours of the test site with the Under Secretary and his staff, Carl's in Washington and, as you know, he lives about 50 per cent of the time in Washington and 50 per cent of the time in Las Vegas.

So the first topic we'll be covering today is the Calico Hills Risk/Benefit Analysis, and as you recall from previous discussions, we've done two parts. One is the value of information model, the second part was a multiattribute utility analysis. The--as we call it, the CHRBA is a document that's finished now. It's been completed by the team. It was a document that the Department agreed to prepare as a consequence of receiving some comments from the Nuclear Regulatory Commission on the consultation draft of the SCP. This document fed some important input, and that is the input on conducting a test program in one of the barriers that's beneath the Topopah Spring rock, and that is the Calico Hills rock. That's the barrier beneath the repository, and above the water table.

And so, with that report being finished, the feed of that technical analysis has gone into the ESF-AF analysis, which we'll be discussing tomorrow.

So supporting Dave, describing the summary and reviewing the VOI part model will be John Lathrop and Jack Robertson. John will describe the details of the MUA, and Jack will discuss some of the hydrologic technical aspects that the expert panel considered, and in particular, some views about the benefits of the unsaturated zone and some things that are related to unsaturated zone ideas for further testing. That will run us through lunch.

After lunch, we are presenting an activity that started this last year, called test prioritization. It started off in our first conversation with you as surface-based test prioritization, but it became obvious that it really needed to be combined with the underground test program, and so we've done that. Yesterday, the results of Phase I, which is approximately a nine months-one year effort, is finished and it was handed in to our office, having undergone its appropriate reviews, so it's ready for distribution now, and Russ Dyer will describe the approach that's been taken, and Bruce Judd will explain the details of the analysis and the test priorities that came out of that effort.

Then, in the late afternoon, we'll describe a new initiative that we've just gotten started a few months ago; early site suitability evaluation. It was a request to Carl by Dr. John Bartlett, the Director of OCRWM, and Steve will describe the background that is associated with the start-up of a task, and Jean Younker will describe the evaluation as it's gone so far, explain the plan, and identify what we've accomplished as of this date.

Then tomorrow morning, we'll pick up on the final stages of the exploratory shaft alternative study. As you know, the team has given you previous briefings on that. Lee Merkhofer will be here tomorrow to describe the sensitivity analysis that was done and and the decision analysis that was applied to that study, and then Larry Costin and Al Stevens will describe some of the design and technical details, programmatic insights and enhancements that may occur along with the options that were preferred, and Ted Petrie, from our office, will explain our general approach for resumption of design activities, finishing off Title I and getting prepared to start Title II design for the exploratory shaft.

Okay. That being the agenda, what I'd like to do is ask Dave to start with Calico Hills, unless there's some questions or changes to the agenda.

Don, is that all right?

DR. DEERE: Fine.

DR. DOBSON: What I'm going to do today is present a quick overview of what we did and why we did it, bring everybody back up to speed, hopefully, with a common understanding. Most everything that I'm going to say, the Board has heard before in several of the forums--fora--that we have had, so I will try and do this relatively quickly so that we can get to John Lathrop's presentation.

DR. ALLEN: You're using the same view graphs as last year?

DR. DOBSON: They've been modified. If you look down in the lower right-hand corner, you'll see that we always update the date.

Max did go over the--kind of the agenda, and I just wanted to note that John Lathrop is going to present the MUA. We have not really presented the results of the MUA, so we thought--I'm going to try and give John as much time as I can to go through that in some amount of detail, and there was a specific request for some information regarding the saturated zone modeling that we did as part of the study, and so we've made some time for Jack Robertson to talk about that.

Again, just because I feel I need to give at least one geologic slide in every talk that I give, for general background, the Calico Hills hydrogeologic units consist of the unwelded tuff units below the repository horizon, including the Calico Hills lithologic units, and pieces of the Bullfrog and Prow Pass units that are exposed in certain parts of the repository block that occur mostly in the southwest part of the repository block.

And the focus of this entire exercise was on characterization of the unsaturated rocks and how best to test in the unsaturated rocks, and that's--that may come up later on, but there have been a number of questions in the past regarding properties of the Calico Hills, for example, in the saturated zone. Many of these things were discussed and considered, but when it came time to sort of move the task force forward, we were obviously concentrating on what we wanted to do in terms of the testing program in the unsaturated zones.

The reason I wanted to say that is because there have been some questions, and I'm sure we'll have some discussions during Jack's presentation about the relative amounts of testing in the unsaturated zone and saturated zone. The purpose of this study was not to determine whether to do more testing in the saturated zone, it was to address how best to test the unsaturated Calico Hills. So, let's see, I'm going to skip the next slide in the view graphs, which is another cross-section.

When we set out to do this study, we obviously recognized, or we recognized that we thought it was a potentially very important study in terms of how it could have a significant impact on the overall site characterization program, so we set up a set of rules for ourselves in terms of how we were going to do business.

The first thing that we decided was that we were going to conduct the study in accordance with the requirements of our quality assurance program, so that we could rely on it in the future. We'd have all the documentation we needed.

We decided to use a decision analysis method, actually, and eventually ended up deciding--using two separate ones, but the rationale for that was that we wanted to be very clear about what our basis was for the recommendations that we were making, and so in order to attempt to make the recommendations as transparent as we could, we decided to do it in a very systematic way, using decision analysis principles, and I think--and we were sensitive to criticism that some past decisions by the Department had not always been clearly documented and didn't have a clear basis, and so we wanted to avoid that problem.

As you have heard in the past, we did use two different methodologies. We started with a value of information method, which Hollis Call, from Applied Decision Analysis, who is here today in the audience in case there are questions that come up under the VOI model, developed for us and helped us implement over time, and then we followed that up with a multiattribute utility analysis, and you've heard extensive presentations on the VOI model today, which is why we're not doing those again, but we haven't presented all the results of the MUA, so those will be heard today.

And finally, the task group was instructed basically to keep the focus as narrowly as they could on the criteria that were specified in the NRC objection. This was especially important early on, in that the NRC objection on the consultation draft of the site characterization plan, which led to us doing this activity, was very focused on whether the benefit from the testing program was going to be greater than the net detriment that was potentially caused by excavating in the site, and therefore, impacting performance. When we initially scoped the study, we did not want to--we specifically did not include all of the criteria that one might use to develop a testing program. We were trying to be specific to the NRC objection.

As you have heard in the past and will hear some more today, when we transitioned from the VOI technique to the MUA technique, we did expand the criteria by which we were judging the value of the testing program, and you'll hear some of the reasons why we did that and how that impacted the overall recommendation in the presentation.

I didn't put in the view graph package a summary of all the meetings we've had. I didn't think that we needed to go through that in any kind of systematic way, but I did want to point out that we have--in the original objection, we've promised to consult with the NRC, and we have done that on numerous occasions, and we've also had several meetings--this is the fourth, I think, meeting with the Board on this task force, so we have had an extensive program of interaction with the Nuclear Regulatory Commission, and also with the Technical Review Board.

The current status of where we are with the NRC is that we have given them the final report. We've presented the results of the final report to them a month or so ago, and they are basically reviewing it now. We haven't heard anything formal from them. We've had lots of interactions, as we've noted, and I think we have a pretty good feeling for where they are in terms of how, you know, what aspects of the analysis they like and what they don't like, and we'll hear in a formal sense when they're finished reviewing the report. We don't have, to my knowledge, a schedule for--we don't know when, precisely, they're going to send comments back, but we fully anticipate that the testing program in the Calico Hills, as well as in the main test level, will be something that we continue to discuss throughout the Title II design process.

Okay, who did all this work? The Calico Hills

Risk/Benefit Analysis was a relatively small group of people. The chairman, as everybody here knows from past meetings, was Ernest Hardin from Science Application, and Ernie is also in the audience today, and we put together a group of basically--a small group of scientists, engineers, and regulatory staff who we thought covered the bases, covered the major program elements and could provide some good input on what aspects of the criteria that we're considering might affect the test program.

The group was not designed to be all-encompassing. We did not go out and form sub-panels at all of the participant organizations and things like that. We did, however, when we felt it was appropriate, bring in some expertise, so in certain areas--examples I've used before include things like some of the waste package performance assessments, source term assessments that we did, and some of the retardation kind of estimates that we did. We didn't feel like we had all of the expertise we could use on our small panel, so we went out and got help when we thought we needed it.

Okay. What I'm going to do now is basically briefly go through the analysis, or the structure of the analysis, and then, hopefully, come to the summary and conclusions, and we can go through that.

As we have noted, this presentation will follow a couple of paths: First, an overall picture of the overall structure of the Calico Hills Risk/Benefit Analysis, and then brief summaries of the structure of the value of information model, and of the multiattribute utility analysis.

Now, this is a view graph that we have shown before, probably starting in March or April at the--almost a year ago now. The original analysis that we put together was composed of several key components.

In order to develop a recommendation for a testing program, we had to go through a series of steps, and the first step that we thought we had to go through was defining our information needs, and I might point at this box here, which is called supporting information. That was something that was done at the same time, and it involved things like deciding what kinds of conceptual models might apply to unsaturated flow at the site, and therefore, there were a lot of feedback loops between doing things like defining information needs and defining testing programs.

And we--our group proceeded with a kind of a hypothesis testing sort of perspective on how to develop a test program, so we thought of different hydrology programs that you might conduct, and if you were going to conduct that one, what information needs would you have, and so we kind of looped back and forth between how we thought the site was behaving and what testing would tell us something about that.

So stage one was basically defining information needs.

The second step in the analysis was given a list of information needs, basically--and this was developed largely by Jack Robertson and members of the U.S. Geological Survey; Bill Wilson originally. Bill is another member who's not here with us at this point, but given that list of information needs, we then took the step of trying to define what kinds of test techniques could provide the information, and so we went through a fairly extensive program of addressing all the different techniques; including: underground in situ exploration; borehole-based testing programs--both angle borehole and vertical borehole programs;--geophysical programs; analog programs. We thought of basically every kind of testing technique we could conceive, and we've put them all down on the list. These are all possibilities, ways we might go about acquiring the information.

We then generated a finite number of strategies from those lists of testing techniques. We wanted to keep the number of strategies relatively small, but we needed to make sure that they were--they encompassed a broad range of combinations of the appropriate techniques, so we started with 60-some different strategies, which we then screened by various techniques, and the strategies--I guess I'm kind of getting ahead of myself here, because we'll come back to that in a minute.

So following the generation of the strategies and the screening, we came up with a list of alternative strategies, which we then could analyze, and we chose two techniques, as I've noted, for analyzing them. The first cut, we did a value of information technique in which we ranked the strategies in terms of their testing benefit and in terms of their net detriment to performance. That is this part of the diagram, the developing the decision-aiding methodology, which went on in parallel with the development of the strategies, and then conducting the analysis. And then review the results, develop recommendations, and as you note from about six months ago, after we got to the end here, we decided to go back through this part one more time, and we conducted the multiattribute utility analysis. You'll hear that today.

The value of information model was originally employed for several reasons, and the most significant was we wanted to build a model that captured as best we could the quantitative data that we could get our hands on, the quantitative scientific models and data, and combine that with expert judgment, and we thought that the VOI model allowed us the mechanism for doing that. We were able to basically collect scientific information and data and put it into our-- into a quantitative setting, and combine it with the judgments about what the ranges of variables might be. And so that was the technique that was selected.

The objective, as we've discussed before, was to compare the benefits of testing, and in this case, it was measured--in the value of information model--by the improvement in your decision making, which was due to increased understanding of site performance, and then compare that to the adverse impacts of performance, and that is measured relatively straightforward--in a relatively straightforward fashion in curies released. So we were trying to measure--we were trying to compare improved decision making, if you will, with releases.

And again, another one of the view graphs that you've seen before and that we've talked about extensively in the past, one from one of Hollis's past presentations; the framework for the value of information model included these components that are shown on this view graph, most of which I've already talked through. Basically, this explains how we constructed the quantitative part of the value of information model, what kind of expert judgments were elicited, and how those judgments were folded into the analysis of how the site was going to perform and what the probability was that some result from testing would result in a change to your program.

Now, the multiattribute utility analysis was invoked because, basically, when we did the VOI analysis, we found a very small probability that the testing program was going to change any programmatic decisions, and that came about as a function, basically, of the fact that the--in order--the way that we defined a problem, in order for a testing program to change a decision, it was going to have to tell you something you didn't expect about the performance of the site, and that you were going to have to learn something about the performance that was so significant that it would cause you to do something different.

So basically, what that means is, that it anticipated that in order to change a decision: you would have to believe, at the current, time that the performance is at a given level; but that your testing program had some likelihood of telling you that you were very wrong about that; and that the performance was actually much worse.

When we did the expert elicitations and the judgments, we found that in the judgment of our panel, at least, the probability estimates of the group, that the testing program was powerful enough, first, and that we were wrong enough about performance, second, the probability of those two things happening and significantly changing a program decision downstream was very low, and that's why the value of information model indicated no VOI.

But we recognized--in fact, Hollis and the Decision Analysis team recognized that this result was likely going to happen very early on in our elicitations by looking at the kind of numbers that were coming out of the panels in terms of expected performance and the power or the robustness of the testing program. We wanted to finish the analysis, though, because, as I stated earlier on, we wanted to limit the criteria of judgment in this first phase to specifically what was in the NRC objection, and we also recognized that that VOI equals zero result did not mean that there was no value to testing. It meant that, given the definitions that we used, it was unlikely to change a decision. And so, given that definition, the VOI model indicated no value.

As I've noted on this view graph, the two most critical reasons why the VOI model--well, the results of the VOI model were not necessarily consistent with the intuition of the panelists in terms of the value of testing were one of a couple of factors. One is that we place a very high value on confidence and low uncertainties, even when you're at very low levels of releases; or, alternatively, there are other values that we didn't capture. And so, we initiated the MUA and you're going to hear a copious description of that over the next couple of hours.

We thought that it was appropriate for what we were trying to do because when we expanded the number of criteria by which we were going to judge the value of testing, we recognized that the different strategies varied along the different criteria axes, if you will, or varied independently on the different performance measures, and so when you're trying to add--when you're trying to combine the value of being able to go wherever you want in the repository block and compare that to time saved in the schedule during the license application process, there's no common denominator, and what you have to do is develop those common denominators so that you can essentially add up and say seven oranges and five apples is better than six bananas. You'll hear John explain how we did that and how we normalized the axes, in effect.

Also, again, for the same reasons that I stated when we started out this analysis, we think that the technique provides a very structured and, we hope, defensible way of combining the subjective judgments that we made with--well, of combining all the subjective judgments into a final recommendation.

Okay, and this shows very briefly--and I won't go into it in detail, but the major criteria that were considered in the multiattribute utility analysis, and you will hear John describe them, but the principal attributes of the model were postclosure risk, which we had also considered in the VOI model, and our results in the MUA were simply lifted from the VOI model for that attribute. Scientific confidence, which was defined somewhat differently than measures of value of testing that we did in the VOI model, and we have the principal components of what we considered to be the scientific confidence; that is, maximizing characterization, detecting the need for and characterizing alternative conceptual models, and supporting the performance confirmation program. All of those things, in our view, contributed to building scientific confidence, and that provided a value of testing that we hadn't measured previously.

Phasing potential was viewed as something which could potentially add value to a test program, because some people, given the choice of a fixed program or a phased program, will say the phased program was better. Other people might not, but it's possible to develop a ranking on that basis.

Service date is more or less a schedule criteria that's driven from two directions. One is, how long is it going to take to do the program you say you need to do? And the other is, what is the possibility of unexpected delays? So that includes both planned schedule and unplanned schedule delays in that criteria, and finally, cost, which is a relatively straightforward calculation, not accurate to many significant figures, but very useful in a relative range.

And this one I won't talk about at all, but it shows the flow chart of the complete analysis, and that's what you'll be hearing about as soon as I get out of here.

So back to the overall structure of the analysis, I think I've actually already talked through several of these things. As I noted, I kind of got ahead of myself, and so we'll just kind of skip quickly through the view graphs, but I had put in the presentation here the steps that we took prior to doing the analyses, the two decision-aiding analyses, including the development of the information needs, identification of the testing techniques, and the generation of strategies.

By the way, if anybody has any questions at any time, feel free to pipe up. We have lots of backup information and we're prepared to talk about virtually any part of the analysis that there's interest in.

This does summarize--this view graph summarizes in a little more detail than I previously talked about what kinds of information needs we looked at in terms of we went, essentially, right to the parameter level. What sort of numbers are we going to be measuring? Are we going to be measuring translucivities and porosities and permeabilities? And we put together a list of all of those parameters, and sort of statements of the need for this information in a spatial sense, or spatial and geographic sense, and also information on the spatial correlation of the parameters.

As I noted, the testing techniques were then developed, and I wanted to point out a couple of things. One was that when we constituted the strategies, the eight strategies that you've heard about, we included explicitly in each one of them both surface-based testing techniques and underground testing techniques. All of the strategies used the baseline SCP surface program that we have now as a starting point, and all of the strategies added something to that, so that the borehole programs, the site vertical borehole programs and the systematic drilling programs that are in the current SCP were assumed as part of our program, with some additional increments of testing, and those additional increments of testing ranged from simply a few more boreholes and surface facilities away from the site, to extensive drifting programs.

The techniques also included ones that were invasive; in other words, borehole drilling and underground exploration, and ones that were non-invasive; in other words, things like analog studies off the site, or geophysical studies which would not have the potential of impacting the performance of the site. And all of the techniques were ranked qualitatively internally prior to forming the strategies that we eventually used. We did sort of a best guess on whether a borehole gave you better information on a certain parameter like porosity than a geophysical test did. So we ranked all the tests in terms of their ability to provide information.

We then combined the strategies and, as I have noted, and each of them combine varying amounts of drilling, underground in situ exploration and surface-based studies, and particularly, analog studies.

Just a few comments about our view of the strategies that we put together. We were not intending to present a final design configuration when we did this. It was recognized early on that we were developing these for the purposes of comparing different--how well different testing techniques and strategies could provide certain kinds of information. We did not go to the level of detail in saying we're going to use an 18-foot in diameter road header and it's going to go precisely to these coordinates, and there will be four alcoves in this place, and somewhere else. We were concerned more with defining a range of options, and we fully anticipate that during the design process, all of that information will be developed.

We made a set of fairly simple assumptions. We did assume, for example, that the underground work in the Calico Hills would be machine-bored, presumably Alpine liners or road headers, some sort of technique like that in the relatively low-strength rocks of the Calico Hills. We assumed that they would be mined, to the extent possible, with as little water as possible, but we didn't go into any great amounts of detail in terms of defining mining specifications or anything.

We didn't talk specifically about means of access; in other words, we did not specify how to get to the Calico Hills. We were concerned with the quality of the testing program within the Calico Hills, but we did have a number of interactions with the ESF group, since they obviously needed to consider that very explicitly when they incorporated our recommendations, and so we did have some input as to the relative value of different ways of getting into the Calico Hills; in particular, shafts versus ramps, and we had some input in terms of how both of those things could affect, first, performance; and second, what kind of value they might add to your testing program, and I'll come back to that in a little while.

I wanted to--I had, originally, all eight strategies in here and I decided, well, that takes too long, so what I did was, I cut it down to the four top-ranked strategies, just so everybody has somewhat fresh in their mind what the different strategies that we're talking about look like so that when John starts reeling off numbers, 7 and 1 and 2 and 5, it's not too dark in our memory and it's at least in the view graph package.

Let me start by saying the other strategies, which would be 3, 4, 6, and 8, were less--involved less testing, in effect. Strategy 6 involved only surface-based testing, no underground excavation in the repository area. Strategies 3, 4, and 8 involved small amounts of testing either within or outside the repository block and I'm not going to describe them in detail. Strategies 1, 2, 5, and 7, which consistently came fairly highly ranked in our various analyses, vary in terms of the amount of drifting inside the repository block, principally, and also, to a lesser extent, in terms of the amount of analog testing conducted off site. Those may be the two critical differences.

Strategy 7, which is the first one I'm going to describe--I kind of do these in a mixed order, but Strategy 7 was a program of extensive excavation, but it was all outside of the repository block. We didn't do any drifting in this strategy inside the block, so we put in an extensive facility down here that allowed us to get a lot of information on both the vitric and zeolitic tuffs, on structural features--since most of you may recall that there are a fair number of faults that run outside the southeast end of the repository block--but there was no excavation in the block.

We did supplement the surface-based drilling program with an additional angle hole in the Solitario Canyon Fault, and another one on the Ghost Dance Fault, and we deepened the currently planned MPBH's in the northeast end of the repository through the Calico, and in the current SCP, they're not drilled through the Calico Hills.

The reason for developing a strategy like this was we wanted to see, we wanted to be able to compare whether, if you have a lot of drifting outside the block and a lot of drifting inside the block, the scientific community--at least our small scientific community--would value the information from inside the block more highly than the valuation of the information outside of the block. So we attempted to maximize the amount of test accuracy, without going inside the block, and I've already talked through the other aspects of the strategy.

Strategy No. 1 built on Strategy No. 7 by going with that extensive program outside the block in the southeast, and then put a small program of drifting inside the block, more or less as a confirmatory program so that you could say that you had done drifting inside the block and you were confident that what you had measured outside the block was not inconsistent with the properties inside the block.

This is the biggest, most expensive strategy that we considered. It's a grand total of twenty-some odd thousand feet of drifting, including 5,000 feet or so inside the block, and 15-20,000 feet outside the block. It would require, of course, two separate facilities. Basically, the way we configured it, you'd have to have surface facilities outside the block and another set inside the block, and it provided the most information because it measured more things in different places than--this one was also supplemented, you might note here, by a surface facility called the Prow Pass Test Facility, which was another area of testing within the Calico Hills, so we really, in this program, considered--this was our maximum program in terms of, you know, feet of drifting and number of tests. It was not the maximum program inside the repository block, but in terms of dollars spent, it was.

The Prow Pass facility was a surface-based facility in the zeolitic tuffs, about five kilometers--three kilometers north of the repository block, up in Yucca Wash, I think, and it would be an opportunity do whatever sorts of testing programs that we felt needed to be done without having to worry about proximity of the block.

This summarizes the remarks I just made, basically. It's to attempt to achieve high test accuracy while still limiting the excavation within the block, and it built on No. 7, and it basically was designed to be able to explore everything you could explore without going into the block.

Strategies 2 and 5, as many of you will recall, are essentially identical, with access in different places. Strategies 2 and 5 were our maximum exploration strategies inside the repository block. They included access to all of the structural features near--in and very near the repository block, including the Calico Hills Fault and Drill Hole Wash structures, Solitario Canyon, the imbricate fault zone on the east side, and for the purposes of comparing and so we could try and get some sense of how much--how important we felt like information was inside the block versus outside the block, this strategy was not supplemented with all the bells and whistles, if you will, that the other ones had. We did not add any additional surface-based boreholes to this strategy. We did not add a surface-based testing facility up in the north in Prow Pass. It's--this one is a program of extensive exploration, but that's basically it. This is the current program, plus an extensive underground exploration program.

Anyway, this basically summarizes that, and Strategy 2 I will show, but as I noted, it's essentially identical to Strategy 5, except that in this case, the ESF access was in the south end. I might note that we added Strategy--I don't remember which one we added, I think Strategy 2--late in the game because in our interactions with the ESF alternatives group, we learned that they were considering both north and south accesses for the main testing facilities, and we didn't want to get into a situation where one of our strategies got eliminated because it had an access that was inconsistent with the access that was chosen for the ESF, when it really didn't matter, and when you have an extensive drifting program, it didn't matter to us whether you got in there from the north or the south, from a testing perspective. So we accommodated that variation of the possibility of different access locations by adding an additional strategy.

Okay, and just to summarize, Strategies 2 and 5 basically have as much as up to, as you will see, 19,000 feet of drifting, and that includes exploration of all the key features, including the faults of note, and the vitric zeolitic transition in the Calico Hills lithologic unit.

Okay. Well, I came to the conclusions and recommendations. How are we doing?

What I'm going to give you now are basically straight out of the conclusions and recommendations section of the Calico Hills Risk/Benefit Analysis, and these were the group consensus conclusions and recommendations, and feel free, if anybody has any questions and would like us to expand on them a little bit, to go ahead and talk about it.

The record memorandum, which is what we called the final report of this exercise, contains seven conclusions and five recommendations, and we're going to go through those right now.

First--and this is one that you've heard before, it basically came out of the value of information study--the potential impacts from characterization on postclosure aqueous releases from the total system are expected to be low for all of the strategies and would not preclude extensive underground exploration in the Calico Hills below the repository block, and we have had several presentations on why we think that, but basically, we were unable to--well, we modeled a whole bunch of different scenarios in terms of the likely impacts on waste isolation, and we don't believe that there's a large impact; in other words, that any of these strategies is likely to lead to a situation where, because of exploration, we have threatened the EPA standard in terms of the performance of the repository.

DR. NORTH: Dave, given that you were using the NRC criteria, could you describe the discussions with NRC up to this point on this Conclusion No. 1?

DR. DOBSON: Yeah. I guess I wouldn't say that we have had any final consensus established, but we have had a lot of discussions on how conservative our analyses were and how broadly they captured the uncertainty in the performance-related--in the impacts on performance based on our model. I guess I--well, I guess my feeling is I'm kind of waiting to see what their comments are in writing, you know.

I think we've had very good discussions with them. I think we've presented our case and Ernie Hardin and Charlie Voss have done several presentations for the NRC, as well as this group, and we think that the analyses are reasonably conservative, but they have not made any commitment, which is, I think, appropriate, without having the ability to sit down with the report. And basically, that section of the report is about 40 pages. From a technical perspective, it's the thickest chunk of the report, and they're in the process of reviewing it now and we expect to hear relatively soon what they say.

We have had some indications, I guess, from certain members of the staff that they would have preferred more quantitative performance assessments than we did, in place of some of the elicitations. We have tried to respond basically that we did quantitative assessments where we could, and we tried to supplement those with judgments.

I might note, just for the record, that one of the recommendations that came out of this--it's not something, I don't think, that's in the report, but it's something that we realized was a potentially significant thing--is that we're going to be continuing to do waste isolation impact assessments throughout site characterization and throughout the design process.

We asked Pacific Northwest Labs to start doing some of these additional ones in December and, in fact, we've received at least one, and we have a representative of PNL here today, Mark Freshley, who is sitting by Ernie Hardin over there, who has finished one of those calculations, also. So there will be more supporting the overall design process than simply the Calico Hills analysis. We didn't intend it to be the final word, and I guess all I can say is that the PNL analysis that Mark has done is basically consistent with the results that we got out of the more limited quantitative analyses we did, combined with the judgments, and we think that continuing that kind of program, of analyzing possible impacts, is really important to our success with the NRC, and we think it's a prudent thing to do as you're going through design. So as we learn more, we continue to reiterate on our analyses, too.

DR. CORDING: Dave, are you looking at specific engineering designs at this point, or methods of, say, backfill seals, cutoffs, the sort of engineering modifications you would make to a facility after it's--and to the drifts to provide the postclosure--

DR. DOBSON: Yeah. We're not in a final design phase, but we have considered a whole bunch of things in the context of our analyses. I'm not sure if you heard the presentation that Charlie Voss did in the last round, but Charlie did talk about some of the different techniques that we considered in our analysis for mitigating adverse impacts, and those involve various combinations of sealing materials with properties that are engineered for the local environment, and combining with seals and plugs and things like that. So I guess I can say we've considered it, but I couldn't say that I have a design package that I can hand you that has detailed analysis in it.

DR. CORDING: But built into this conclusion, No. 1, is that--is it that you feel that there's an array of techniques that are available that would offer what you need?

DR. DOBSON: Oh, absolutely; absolutely.

DR. CORDING: Or are you looking at certain specific techniques and saying, with this technique we can achieve such and such?

DR. DOBSON: Well, no, I wouldn't say that. I'd say--there's two parts to the conclusion. One is that the analyses that we've done, using certain sets of assumptions about the properties of the backfill materials and things like that, would suggest that there's not a significant performance problem. But the--I think we could go one step further than that, and say that the analyses would also suggest that whatever detriments to performance we identify--in other words, if we use a backfill material that's not perfect, or something like that, or, excuse me, if we engineer--if the local environment does turn out to be a little worse than we thought, we do have means, engineering means to mitigate the consequences that are of concern.

So the first part of the analysis is, even if you don't use the engineering means, we don't think the impacts are large. The second part is, even if there are some unanticipated adverse impacts that we don't currently foresee, there are means available to mitigate those. I mean, I would want to make it clear that our conclusion was not that there's a significant adverse impact, but we can mitigate that. That's not what I'm trying to say. That's not what we said.

Okay, second, Testing Strategies 1, 2, 5, and 7 all include extensive underground exploration either in or very near the repository block, and in our view--and you'll see this in the multiattribute analysis--they provide a significant improvement in scientific confidence, as we defined it, relative to the small strategies, 3, 4, 6, and 8.

When all of the objectives from the MUA were considered--the confidence, risk, cost, delay, and phasing potential--Strategies 2 and 5 were preferred to Strategy 1 by a small margin. We have proposed some modifications to Strategies 2 and 5, which are consistent with their definition and would provide greater scientific confidence than Strategy 1. I noted that when we set up 2 and 5, we purposely did not attack, sort of, all of the possible additional features that we could have, to those strategies. And the small modifications that we're suggesting here include: extending a drift in the southeast part of the repository, probably a few hundred feet or a thousand feet, to get structural information from the Abandoned Wash fault zone; and, also, providing for the possibility of a testing facility outside the block. So that's what the modifications are, for instance.

Now, here's one that you won't know what I mean because I haven't explained it as we've gone along now, but in the purpose--in the process of developing our definition of scientific confidence, we defined 12 key features of the site about which we wanted information, and Strategies 2 and 5 investigate each of the 12 features.

The relative benefit of early access to the Calico Hills--this may be an obvious statement, but in our view, the relative benefit of the early access to the Calico Hills is directly related to how many of those features you can get to. So if your key concern is to Ghost Dance Fault and you want a strategy that gets to the Ghost Dance Fault first, then you ought to set it up to go that way, but in order to get to all the features that we've defined, it takes a fair amount of drifting.

Number six, I noted earlier that we had a lot of interchange on the access leads, and it was the view of the Calico Hills Risk/Benefit Analysis that a ramp from the east of the repository could provide significant information, which would be potentially very beneficial in characterization of the Calico Hills unit. You get a lot of information in an area which is nominally more complex structurally than the repository block. So if you're interested in characterizing the potential for fracture versus matrix flow; you have an area which, it appears on the cross sections, at least, to have more fractures. It's a good place to get that kind of information.

And finally, the relative importance of the Calico Hills unit as a barrier depends, not surprisingly, on the relative important of the other barriers, both natural and engineered. This conclusion is stated because of some of the questions we've had about the saturated zone and we're going to talk about that a little later.

Essentially, I think, one thing that we found in our analyses was that we think that each of the barriers at Yucca Mountain is likely to perform very well, and the greatest increment of confidence that we see in terms of overall performance comes when you combine all the barriers together. Each of the barriers, in and of itself, has what appears likely to be fairly good performance, but with fairly broad ranges of uncertainty. When you combine three separate barriers that have those characteristics and you do it analytically--as we did in the model--you end up with a performance range that appears to be very good.

And so, another way of stating this is basically that the multiple barrier concept that was envisioned in the original WAS seems to be a very good idea, and having those multiple barriers really does tend to lend confidence to the overall performance of the repository system.

One last note--and again, we'll hear more about this--for our analysis, the estimated performance of the engineered barriers and the host rock that we have in our models are, we believe, conservative. We made, we think, very conservative assumptions about how engineered barriers would perform and how the unsaturated zone would perform.

For the saturated zone model--which was not directly coupled to the other two, it was done as a separate component of our overall model--we did not intentionally try to be overly conservative. We tried to be realistic, but we didn't skew the results toward the lower end of the scale, as we probably did in the other two models in terms of the way the elicitations, the expert judgments were obtained.

DR. LATHROP: Conservative meaning risk overstating.

DR. DOBSON: Pardon me?

DR. LATHROP: Conservative means risk overstating.

DR. DOBSON: Right; overstating of potential risk. We tried consciously, more or less, to do that in the engineered barrier model and the unsaturated zone model. We really didn't try intentionally to do that in the saturated zone model.

Okay. What were the recommendations of the group? That's all the conclusions we could agree on. We recommended using extensive drifting within the block, something similar to Strategies 2 or 5. Well, as I noted, we don't intend that as a final design configuration.

We also recommended that the strategies--as I also discussed--were--should be modified to include a drift to explore the Abandoned Wash fault zone on the southeast end of the repository block, and a possible underground access outside the block for what we referred to as aggressive testing, and what we meant by aggressive testing was testing programs where we might want to use, for example, large amounts of water, and we thought that it would be prudent to do that outside the block instead of doing it inside the block, because of potential impacts on waste isolation.

The modifications that we recommended would make the decision more robust. As you will note, the preference for Strategies 2 and 5, over 1, in the model that John will describe was small. When you add these modifications, the preference would get bigger. So that means, in our view, that the decision would be more robust.

The recommendations are potentially dependent on the sensitivity to differences in risk; in other words, one of our issues is waste isolation impact and, in our view, because of the amount of drifting within the block, it is likely that the net adverse impacts of Strategies 2 and 5 are slightly higher than for 1, because there's more drifting in the block. So if you, therefore, magnify that difference by a large amount, then you could change the recommendations, and the amount that you magnify the difference by, how sensitive you are to that difference in risk is--would lead you to a different conclusion. It might lead you to Strategy 1 if you magnify it by a certain amount. It might lead you to Strategy 7 if you magnify it by even more than that, or it might lead you to Strategy 6 if you magnify it so much that no underground exploration is recommended.

The last bullet on this one, I think, is relatively important. We wanted to say that the view of the Calico Hills Risk/Benefit Analysis is that we need to design a facility that can go to all these places, that has the ability to go to all these places, but we're not certain that a final commitment of excavation of all 19,000 feet is required. The reason for that is, we may learn things--we may basically realize that we know enough at some point. We're hoping that at some point during site characterization, we conclude that we know enough about the site to make conclusions about its overall performance, and it's possible that we might learn that amount before finishing 19,000 feet of drifting.

Alternatively, it's also possible that we might discover something that we didn't know about waste isolation impacts, and therefore, we might choose a different strategy. So we wanted to say that while the recommendation is clearly, from our perspective, that you need to be prepared to do this drifting, it's not, at this point, obvious that we will have to finish it.

The planning for the exploration program should focus, at least early on, in providing access to the 12 features that we have defined and you'll hear more detail about in a little bit.

We think that it would be a good idea, based on several trips that the Calico Hills Risk/Benefit Analysis took to the Rainier Mesa area, to undertake some collection of what we called observational data. We think that there's a lot of information that potentially is useful to characterization of unwelded tuffs in higher flux environments that we could probably get in a relatively simple way by going out to Rainier Mesa. As many people here are familiar, Rainier Mesa is a couple thousand feet higher than Yucca Mountain, and has been suggested by some people as a possible analog for the Yucca Mountain site, an alluvial environment, so--it is fractured, too, quite extensively.

We think, as I noted, that waste isolation impacts should not--we don't intend the analyses in the Calico Hills Risk/Benefit study to be final. We think that the evaluation of impacts on waste isolation should be done -- throughout the site characterization program, and as I noted, we will start to undertake that kind of a program. And we have one publication that will be available shortly that's evidence of that. And we think that at each major step of the design, the issue should be readdressed.

And finally, certain of the assumptions and criteria that were made in the Calico Hills analysis about where we would go drift are, we think, important to keep, to tag along with the design, because if you change that assumption, it might change your design, and the particular ones of interest are we made certain assumptions about how far drifts outside the block would be outside the block for Strategies 1 and 7, and we made, basically, an arbitrary assumption that we would keep them at least 2,000 feet outside the block.

There's no magical reason for that. It was just done because that was an amount that we thought was adequate to move us outside of the area where we perceived potential adverse waste isolation impacts, but if we were to go with Strategies 1 and 7 or something like that, then you would want to, you know, kind of keep that assumption along with you so you could document why you were doing it that way.

Another important one is the water table standoff. We essentially made assumptions that we would go no closer to the water table in the underground program than, I think it was 70 meters, but again, there is nothing magical about that number. It's a number that was chosen for certain performance reasons, and we figured that would leave us a minimum thickness of Calico Hills, but it's important to take the number and the -- number and carry it as part of your design so you know why you chose to stop 70 meters from the water level.

And that's about all I have to say, so we'll take any more questions that you want now, and otherwise, we're about ready to go with John.

MR. SHAW: Dave, I had a few questions back towards the beginning when you referred to information needs, and the first question was, how were those information needs determined? And my second question is: Were the information needs a fixed set that you defined at the beginning, and didn't change as you went through your analysis? And my third question is: Considering where you are now, if you looked at things like information needs and some of the criteria that you put together, how sensitive are your conclusions and recommendations on that particular set? In other words, if you went back and revisited and you changed those, do you think that would have much influence on your conclusions and recommendations?

DR. DOBSON: Okay, a three-part question, if I can remember all three parts.

The information needs were developed by a sub-group of the overall panel, including, as I noted, Jack and Bill Wilson and Scott Sinnock, and they basically started with the SCP program. What information do you need there? And then expanded on that by, as I noted, sort of using a hypothesis testing approach to developing what information you need to resolve certain conceptual models. That set of information needs was fixed throughout the VOI analysis that we did, and it was revisited, basically, when we did the MUA analysis, and you will note that kind of a--we didn't really redo the information needs, but we took a different cut on it. When we came back to the MUA analysis, we had all the information needs that we had defined in the value of information model, and then we looked at them from a different perspective, which was, we identified the key features that you've heard me talk about several times, and then we defined how you would go about obtaining the right information to characterize those key features.

So it was a slightly different spin, not really a new set of information needs, but a different cut and a different perspective on what information was needed. So I guess you could sort of look at it as that the MUA treated the information needs slightly differently than the VOI did.

As to your last question, I'm pretty confident that we did not change in any significant way that would have led to a result in--a change in the results. I think basically in each of the models, in each of the decision models, we considered a program to get an extensive amount of information that we defined fairly carefully, and it's in the report, and that in each case, our final recommendation provides the information we think is needed.

DR. DOMENICO: Dave, Item 7, can you tell me what you mean by Item 7?

DR. DOBSON: Item 7 of the conclusions?

DR. DOMENICO: The conclusions, right.

DR. DOBSON: I think this was put in here mainly as a result of a lot of discussions that we've had, and some with the Board and some with the NRC, and some internal to our group. When you look at the relative performance of the various barriers that we defined, there was some question as to whether the unsaturated zone was being over-emphasized. I think kind of what we're trying to say here is that, if you take look at our results and say, well, the unsaturated zone isn't a very good barrier. Why are you calling it the primary barrier?

It's kind of a, as I said, I think the point we're trying to make is that the multiple barrier system is where you start adding a lot of confidence to overall performance, and I think it may be that in the past we have done ourselves a little bit of a disservice by saying--by even calling the Calico Hills a primary barrier. It's a very important barrier and likely performs very well, but there are several other barriers in the system as well, and when you combine the performance.

DR. DOMENICO: When you say, but not necessarily conservative, do you mean that you may have given it more credit than the saturated zone?

DR. DOBSON: The saturated zone.

DR. DOMENICO: That's what you mean?

DR. DOBSON: You should ask Jack that question, but yeah, I think that's what we mean. We were--we didn't intentionally skew the results toward better, or toward worse performance, as we kind of tried to do with the unsaturated zone. With the unsaturated zone, we were very good at imagining models that would result in a lesser performance.

In other words, you know, we assumed climactic conditions that would leave large amounts of fracture flow, and we postulated relatively large amounts of water moving through the block, and the assumptions that we made about the amount of water that would see waste, and the amount of waste that would dissolve and be transported, and all of those things in the unsaturated zone, we think are pretty conservative.

With the saturated zone, it was more like, okay, what's your best estimate of the range of properties, best estimate of the range of flow times, best estimate of the range of transport absorption properties, and sort of multiply those out and see what it looks like. It wasn't nearly as scenario-driven.

MR. McFARLAND: Dave, in the recommended testing strategy that was passed over the ESF alternative study, did that include modification to Option 2,5 that went into Abandoned Wash Fault?

DR. DOBSON: Well, it turns out that the recommendation that was sent to them in June of last year did not. It also turns out that Strategy 30 has a drift into Abandoned Wash Fault anyway, so they--I think they anticipated this. I'm not sure.

But the final recommendation that goes in the Calico Hills Risk/Benefit Analysis that went to the--that came to the DOE, does have the recommended modifications, and the current recommended option can easily accommodate those recommendations. It's not a problem there.

DR. DEERE: Thank you very much. We may well want to come back with questions after we've had the other presentations, and I'm sure you will be available.

Now it's time for a 15-minute break.

(Whereupon, a brief recess was taken.)

DR. DEERE: We are ready to continue with the second--third presentation this morning. Doctor?

DR. LATHROP: Okay, thank you. I will apologize. Given the schedule, I have a lot of ground to cover, so I might be talking a little too fast, and if I do, please let me know.

I'm going to be talking about the multiattribute utility analysis part of the CHRBA. That's our word for it, CHRBA, so the obvious first question to ask and answer is: What is MUA? And it simply--it's really a very simple methodology, frankly. It's a methodology to evaluate alternative actions--in this case, test strategies--by how well each of those actions satisfies a set of several objectives where the degree to which it satisfies that objective is measured by a performance measure.

So really, all you're doing is describing each of the alternatives, each of the test strategies, in terms of a set of performance measures, and building a scoring function to evaluate that and coming out with a single number score of overall desirability. So there's nothing about this that's terribly advanced.

I can describe the MUA in terms of seven steps. Actually, I tried to do the 12 steps, so it looked like you were recovering from alcoholism, but I couldn't come up with the last five steps. And that's the thing about it, there's a lot of similarities between alcoholism and decision analysis, but that's another story.

But basically, what you do--and I'll be stepping through these steps in the course of the presentation--is you define the objectives and the performance measures. You identify the people whose opinions are to be incorporated into the evaluation. It is methodology that's specifically designed to incorporate subjective evaluation judgments into a defensible analysis.

Then you ask a set of value elicitation questions of those people, and those questions are specifically designed to break the evaluation question down into its component parts, particular tradeoffs between each pair of objectives, how you evaluate each of the levels of performance among each of the objectives, so the questions are simpler, although I think if you talked with any of the respondents, you'll find that the questions were actually quite hard. They were always A versus B. Which would you prefer: this site characterization; or that one. Where they would vary on just one or two dimensions in particular ways.

And those questions were designed so that the answers would give us the information we needed to basically fit the scoring function to those answers; simply a function-fitting sort of exercise. Then we have the scoring function which we apply to the data set. We see which test strategy would rank the highest. We do some sensitivity analyses, and we're done.

The key features of MUA that apply here is it allows you to evaluate action alternatives or test strategies in terms of subjective performance measures when you need to, and to look at the subjective value tradeoffs between the different objectives, and the subjective evaluations of the relative worth of different levels of performance. For instance, we measured a service date here, and if we happened to have measured that--as they have in some other occupations --in terms of months of delay, it might not necessarily be the case that 20 months of delay is twice as bad as ten months. We do that subjective evaluation.

The bottom line is the methodology allows you to build a very systematic process of structured, expert judgment to--and build that into a formally correct and defensible evaluation analysis.

Why we use it here is, very obviously--and you'll see in the course of the presentation--several of the performance measures that measure the desirability of the different test strategies are, of necessity, subjective measures and we have to use very structured judgments to get a handle on the relative value of different test strategies.

Now, we have heard a lot about the VOI analysis, and you've heard presentations about that. One of the most interesting parts of this project, to me, is the comparison of the MUA versus the VOI. And on the surface of it, they have seemed to have come out with different conclusions. That's not true, but superficially, the VOI says: Well, gee, there's no VOI-type benefits to these test data, and the MUA, as you've heard, has found there are benefits to the tests, and clearly there's benefits other than the benefits as characterized by the VOI analysis.

Basically, the two analyses are measuring different aspects of the test strategies. The VOI evaluates the test strategies in terms of how the test data would affect performance of the repository by affecting decisions made in the design, construction, and operation of that repository. So you're looking at how the test data would help DOE make better decisions; whereas, the MUA evaluates the test strategies in terms of several performance measures--release risk, cost, scientific confidence, delay, and phasing potential, and those are the five, in fact, that is evaluating in a way not tied directly to how the data would affect specific decisions.

So a way of restating that is the two methods actually have different paradigms of learning, or paradigms of value of information. With the VOI, you analyze the test accuracy and the decision outcomes--and by the way, you'll see some of these same things in the test prioritization task, which you'll be hearing about later in these meetings--and you identify the best decision for each test outcome, and you set up the system so that you anticipate the different outcomes the tests may have and how you would react to those. And then, basically, you go to the rock and you conduct the tests and you decide on the action based on that test. And before the fact, you model out how that process will happen and you see how well you will do with the tests and how well you will do without the tests. You subtract the two, and the difference is the test value.

So the information, then, in the VOI context has a value to the extent that it results in better performance through better decisions, and you value each test strategy in terms of its value to the extent that it results in better decisions.

The MUA, on the other hand, takes a different cut at the same problem. It basically says, well, we'll go to the rock, we'll collect data, we'll learn from that data in ways that we can't reliably anticipate; that we can't anticipate just tremendously well, we can't anticipate completely what we're going to learn and how we're going to learn it, but we are going to go to the rock and collect the data and learn from the data in some ways, and we're going to evaluate that simply in a way that gives credit to a test strategy for providing some information. So information, in this case, is valued to the extent that it improves site characterization; that is, the understanding and the confidence, not necessarily related to its improvement in the decisions made.

So each strategy, each test strategy now has value simply because it exposes the rock. It provides you an opportunity to learn. This is intriguing to me, because it gets at the whole question of why you're doing site characterization. Are you doing it to improve the performance of the repository? Are you doing it to establish, in an external mind, a "reasonable expectation" of compliance in the performance of the repository?

Warner?

DR. NORTH: John, I'm going to slow you down a little bit.

DR. LATHROP: Okay, I need to be.

DR. NORTH: There are three words I think I would like to have you define a little bit more. First of all, you've been using "we." Who is the "we"? Is it the Department of Energy? Is it the American public, or is it somewhere in between?

Then how about understanding and confidence? Can you tell us what you mean there, in a way where there is an operational decision? In other words, how can I measure it other than very, very subjectively? Is there a way we can define it in such a way that it's a little less subjective when we're comparing different people and different situations?

DR. LATHROP: Okay. In fact, one of the ways to look at the MUA is in doing the MUA, you do the definition, and over the next few slides, you'll see how we've built a multiattribute utility function to represent scientific confidence. In fact, that's the best, and most completely can describe what we mean by scientific confidence, and it has a--in the philosophy of science sense, an objectivity in the sense that I am confident that if you convened another set of panels, you'd come up with strategically similar measures, or strategically similar multiattribute utility function; that is to say, it might differ in some of the details, but give you the same results.

That's speculation, but based upon my experience in doing this sort of thing and the conduct of the Panel to any elicitations which were done, we were able to put together a multiattribute utility index of scientific confidence, which obeys the rigors of the methodology, and the index itself has a definition of scientific confidence. So we can revisit that question after the next several slides.

And the "we," the definition of scientific confidence was developed through a process of convening a technical panel and a regulatory/management panel as part of the CHRBA task force, but the "we" in terms of establishing a reasonable expectation of attainment of compliance, that "we" is a very different "we". That "we" includes the regulatory process, because we can characterize the whole mission of site characterization--in fact, this is interesting. I only came across this the end of last week when I was looking at some of these slides and I added a couple of lines and said, gee, how can I add those lines without getting to the bottom of this? And I dug back through the regs and I came up with the reason why we do site characterization, and I'll have to read off my crib sheet here.

We do it to establish reasonable expectation that compliance will be achieved in an external mind; not necessarily a person who is against you, but a person from outside the particular technical community which has done the measurement. And this is intriguing, then, because we have--not we--the CHRBA task force and the people I've been intending to be talking with have been working hard on getting that CCDF, the complimentary cumulative distribution function tied down in a reasonable way in terms of the performance assessment models, in terms of a probabilistic description of releases from the repository, and quite naturally, as a result of that orientation, the VOI analysis tended to look at how well the testing enables you to improve that CCDF, when, in fact--maybe I'm saying this too many times--perhaps we should be orienting the testing toward this demonstration of compliance, providing a reasonable expectation in an external mind.

When you do that, you come up with a different orientation toward the testing which has to do with one of the primary benefits of the testing is improving scientific confidence. Great. Now, how do we do that? How do we define that? You'll see the definition in the course of the next several slides.

So the two analyses came up with different conclusions but they're, in fact, not in conflict. The conclusions can be summarized by saying that there are differences in net benefits--there are net benefits and differences in net benefits among the eight test strategies. Those benefits do not happen to include improved performance due to improved decisions. There are other benefits.

So we have to talk about, now, what are we after when we do these tests? And as Dave Dobson put it, the first thing a good decision analyst does is develop an objectives hierarchy, or simply a way of organizing the goals that you're after. Now, you see the top goal here is appropriate site characterization, and that involves five sub-goals; that is, well, let's get a handle on residual risk, scientific confidence--which we will be defining--phasing potential, which is actually a tricky one, because that phasing potential, strictly speaking, we might have to call it a proxy variable. It's a variable that stands in for the significance in terms of cost, schedule and performance of the fact that some of the test strategies have a greater potential for phasing than other strategies do. Service date and cost, and these have been foreshadowed by Dave Dobson.

On the residual risk, we basically lifted the results from the VOI study. Scientific confidence is what I'll be going into in more detail later, and in fact, I'll be going into, in the course of this presentation, all the definitions of all five of these sub-objectives. They are, in fact--they wind up being the arguments toward an overall multiattribute utility function, and now it's a good time to show you what such a function looks like.

The math is really fairly simple. What I have shown you here is an additive function. Actually, in some parts of the analysis, I explicitly tested the function for the appropriateness of the additive multiattribute utility function. There are multiplicative and slightly more complicated versions which could be used. I tested to see the appropriateness of those, and found that they were not necessary to represent the values here.

Basically, what you have is a set of terms where each term represents the performance of each of the sub-objectives, and in each case you have the performance on the dimension; for instance, months of delay, or dollars of cost, or curies of release. That is then put through a utility function which represents the relative value of different levels of that performance measure; for instance, like I said, 20 months of delay might be greater or less than twice as bad as ten months. And then the k's are the importance weights, which represent the relative value of these different performances.

What this does for you is it allows you to compare apples and oranges. What you have to do is prepare test strategies, because they vary on how long they're going to take, how much they're going to cost, the residual risk that results, the scientific confidence. These are all different dimensions of performance and they are, in fact, apples and oranges. What this does is, it allows you to take the number of apples and transform it, through the utility function and the importance weight, into the Utiles, and then this term takes the bananas, or whatever, and transforms those into Utiles. Now they're all in Utiles and you can add them up. You can only add up on the commensurate measure, and that's what you're doing here.

Now, the result that you get is in Utiles, and nobody knows what a Utile is, so you use the utility function to transform the performance on the several dimensions down to some equivalent, single dimension of performance, and you'll see exactly how we do this in future slides.

So here's an overall flow chart of the analysis, and people tell me I write my flow charts upside down. Most flow charts go from top to bottom; mine go from bottom to top, and that's simply because I like the flow chart to look like the objectives hierarchy, and if you look carefully at the center of this flow chart, you'll see basically the objectives hierarchy I had about four slides ago; the risk, confidence, phasing potential, service date, cost, and the first steps are basically involved in developing the performance of each of the eight test strategies on each of these five performance measures or attributes. So, in fact, all we're doing here is building in a 40-cell table; eight strategies times five performance measures for each strategy, and I'll be going into these in detail over the next several slides.

Once we have that 40-number table--and you'll see that, I think, about Slide No. 28 or something--then you build a multiattribute utility function to combine those into a single measure.

The one thing I wanted to point out from this point of view, this overview of the analysis, is we look at scientific confidence in two different ways. We evaluate it in terms of the technical community's perspective of scientific confidence as, in particular, one of the five sub-objectives. We also use a slightly different cut at scientific confidence, that that scientific confidence isn't characterized by the regulatory/management panel as one of the inputs into the measure of a regulatory delay. Supposedly, the idea is that the test strategy that provides more scientific confidence would result in less regulatory delay, so we actually take this measure into account twice, two different ways, and I'll show you how we do that.

For now, we're going to concentrate on this part of the flow chart here--which I'll have a blow-up of in a couple of slides--how we measure scientific confidence. In fact, you'll see we use the multiattribute utility analysis a couple of times here at two different sorts of levels. One is we build an MUA of scientific confidence, and then we build an MUA combining the five dimensions down to one overall performance dimension.

So we had to come up with a definition of scientific confidence, but I don't--I can't--I'm sorry, I can't spend much time on this slide. But basically, we talked about all the different things that go into scientific confidence; things like the degree to which the CCDF is apt to remain relatively stable, in the sense that it shifts in the ways that you would imagine that it would as the data comes in, as opposed to some data coming in which was not adequately represented in the models that were behind the CCDF, and resulted in a discontinuous jumbling of what that CCDF is.

Another aspect of scientific confidence is simply demonstrating the ability to predict the behavior of the system; for instance, on the basis of the site characterization, before we look at it we are going to tell you that we expect the permeability here to be such and such, and then measure it and find that to be true.

Two things having to do with understanding: the ability to interpret the data that's coming in from the characterization under a consistent conceptual framework, and not having to evoke different conceptual models for each subset of performance of the repository. Maybe the most straightforward definition is simply the ability to answer questions that may be raised in licensing. That may be the most pragmatic and operational sort of definition of scientific confidence. That gets tricky. Well, what about answering nonsensical questions, or questions that aren't technically astute, and so forth and on.

Another angle of scientific confidence is simply that you have involved recognized sorts of expertise, and reasonable assurance. Actually, we should all read the regs, and I've read the regs about three times, but I didn't read for the third time until last Thursday, and I realized we shouldn't be talking about reasonable assurance here, we should talk about reasonable expectation, according to 191.13.

We boiled those down into very operationally saying that scientific confidence--let's see if I have the definition here--is the basis for establishing a reasonable expectation that compliance will be achieved. I probably have said that five times so far. What that means, then, operationally, in terms of our evaluation here, scientific confidence is increased by collecting data, not just any data, but data that addresses any of a list of specific sorts of issues which we defined in the course of developing the MUA.

Here are the issues, and I don't have time to go over them in detail, but suffice to say they fall into three categories; maximization of characterization, detecting and characterizing the need for an alternative conceptual model, and support for performance confirmation.

Looking at these--I'll spend just a little bit of time on this--when we first started defining these--and of course, it came up to be a list of 28, then 24, then 18, then 21, and finally we're down to 15--it turned out to be quite a trick, frankly, to get a set of issues which are reasonably mutually exclusive, and they--and we haven't--we didn't quite achieve that, but we achieved--we developed the set of these 15 issues, which is a way to describe all of those things, all of those ways in which scientific confidence can be improved by data, and there were some surprises here.

I expected them all to be things, for instance, in terms of like statistical characterization. Yes, the better degree to which you can statistically characterize the unit, the better off you are. But then flexibility was a funny one to me when I first thought about it, but that basically is giving credit to a test strategy for getting you to many different places within Calico Hills, because if you get to many places, then if you find some surprises in the data, you're in better position to react to those surprises.

So, really, the issues aren't so much points of scientific debate. They are ways in which a test strategy can be good or not so good in terms of giving you--putting you in a position to improve your level of scientific confidence.

DR. DOMENICO: Hold it, John. The acronyms on the left, the difference between the ACM's and the MC-1's and the SPC-1's, can you--

DR. LATHROP: Sure, yeah. The MC's are that set of issues which address the sub-objective of maximized characterization, and this has to do with; you like a test strategy that gives you more characterization more than one that gives you less characterization, simply in terms of describing the unit and the geohydrology and the geology.

A different cut on that is there are some alternative conceptual models. In fact, a large set of them were defined in the site characterization plan, and one of the things that these tests should be doing for us is testing to see if those alternative conceptual models are more appropriate than our current conceptual model, and if so, collect the data necessary to exercise that alternative conceptual model.

Differently, again, is we like test strategies that put us in the best position to monitor performance in the long term, in a few different ways. We like a test strategy that gets us to features that are particularly appropriate for long-term testing. We like the strategy that gets us places which gives us the best handle on the baseline data from which we will perform the long-term characterization, and, gee, we even like test strategies that get us to places that when a third party comes in saying, "Hi, I'd like to look at such and such," we can say, "Well, we have a room for you down here at this particular part of Calico Hills. Please make yourself at home and set up your equipment." So we like strategies which can do that, and that's the--so there are--but these are, in fact, three different ways a test strategy can be of service.

DR. NORTH: What's SPC again, John?

DR. LATHROP: Support performance confirmation. I'm cryptic in so many ways, I don't mean to be. Sorry.

DR. DOMENICO: Are the MC's truly issues?

DR. LATHROP: Excuse me?

DR. DOMENICO: Are the four MC characteristics you have up there, are they really issues? Do you consider those issues?

DR. LATHROP: Well, maybe--you know, we started out calling them issues, and maybe that was an unfortunate term. Again, they are ways in which a test strategy can excel or not excel, and the test strategy that gets you a lot of places puts you in a good position in terms of flexibility in a very different way than a test strategy that gets you to all of the physical bounds of the block. That test strategy does good for you because it allows you to do a good job on the boundary conditions for your models, because you've physically gone to basically the six planes that define the block that you'd like to model.

So MC-4 is just--it's a very different thing than MC-2, but they are both ways in which you would want to evaluate a test strategy, and in fact, you can imagine a test strategy that goes many places but doesn't happen to hit all six planes bounding the block, and that would score high on MC-2 and low on MC-4.

DR. DOMENICO: Thank you.

DR. LATHROP: Then, as people have mentioned, we defined features of the site, and a feature, just like an issue, is an opportunity for a test strategy to be of service. A feature is an opportunity to learn something about the site at a particular physical location or through a particular means, and I'll just go straight to the next slide, which lists the features.

And here are the 12, and it's interesting because, you know, the first five are actual--this is what I thought would be features, you know; faults, fault zones, and so forth. That seems pretty good. Then unknown features. This is a little strange, and I said, "Gee, guys, what do you mean by unknown features?" Well, it's the perched water and dikes, and so forth, so I said, "How are we going to model that?" Well, we'll look at each particular case and talk about the relative benefit a test strategy will give you for uncovering an unknown feature.

What this does is basically build in a credit to the evaluation function for those strategies that do a lot of drifting. Yeah, the strategy that drifts more is more apt to stumble across the unknown features. You say, well, that's--well, yeah. No, actually, that makes a lot of sense, and it does make a lot of sense. You should have as part of your meter that measures how good a strategy is, some credit to the strategies that just get you through a lot of places, because they're the ones that are going to allow you to discover some unknown features.

Now, again, remember the orientation hearing. There are some people here in this room that says, "Gee, there's not going to be anything of significance there," and maybe there's not. I'm not addressing that question. I'm addressing the question of--how do those words go again--establishing a reasonable expectation that what you're--how you're saying things will behave will behave that way. If you've drifted for 19,000 feet through the block, you're further ahead in some particular ways than if you've drifted through 3,000 feet, simply because you've looked at another 16,000 feet. It gives you that much more of an opportunity to stumble across some of these unknown features.

The stratigraphic features are--make a lot of sense. Site hydrochemistry, I said, "Hey, wait a minute, guys. What's that doing in there? That's not a physical feature." Ah, but it's an opportunity, again. It's an opportunity to learn about some of these 15 issues, so that is, in fact, a feature even though it doesn't have a specific physical location, and then L, similar conditions outside the block, as was mentioned, go a place where you can do analogy studies using aggressive testing technique; for instance, with a lot of water, or radiologic tags, things that you would find difficult to do for either technical or political or regulatory reasons, to do within the block itself.

So I've probably thoroughly confused you by now, talking about strategies and features and issues. They do all tie together. This is how they tie together. In fact, this is a representation of how the actual Test Strategy 1 tells us about Issue 1, and basically, we have 12 features, and we're saying that each test strategy gives you 12 looks at each of the 15 issues. So the dimensionality here is eight test strategies, each looking through 12 different lenses, at each of 15 issues, and the whole process I'll be going through, simply collapsing all that down to one final measure of how well a test strategy looks at all 15 issues.

And what you do here is; we have a rating system where we rate how well a test strategy gives you access to a feature. You have the maximum feasible access, down to a limited access, down to access where you can, in a sense, you can do some analogy studies, but not actually there, or access that's no better than baseline. And then once you're at a feature and you have access, the features do vary in terms of how well they inform you about the issue, and that goes through three levels, of it doesn't inform you any more than baseline, toward an intermediate level, toward it gives you a significant increase in scientific confidence about that sort of issue.

So the different thicknesses of the arrows are meant to represent that, so we can see that, of course, there is a maximum look, which is the maximum access at a feature that gives you the best information about an issue, and this we'll call our max-strong look; max access, strong increase in scientific confidence, and that's the peg upon which we'll evaluate these, and obviously, the other combinations of qualities of linkages here would be less good looks at each issue through each of the features.

So we elicited from the technical panel how well each of the alternative strategies accessed each of the features, and you can read for yourself what the nomenclature is. These are four levels of access.

One of the interesting things here has been mentioned, and that is that Strategies 2 and 5 don't get you any look at all at similar rock conditions outside the block, and don't get you to the Abandoned Wash Fault, just because of the way that it was defined, and some of the other strategies get you to particular features much better than others, and it's all there.

In addition, with the same technical panel, we filled in another table of how well each of the 12 site characteristics or site features address each of the 15 issues, which we've talked about, again, on a three-level scale. This is just--we systematically went through each one with the panel, discussing how we would do that evaluation. We took votes. We saw how people would differ in their voting. We asked the people who would differ to debate with each other and we took a second vote, and so forth and so on; a very methodical, but really very simple process. If anybody would be interested, we have stacks and stacks of ballots of how well each particular member--or how each member voted.

DR. NORTH: John, before you leave that, I'm confused about the five entries under MC. You've got MC-2 down twice. Now, you only had MC-1 to 4 in your original list.

DR. LATHROP: That's something--that's an aberration, yes. That's an interesting one, but it's the same--

DR. NORTH: So is that just a mistake, or am I confused?

DR. LATHROP: --as MC-2. I think somehow MC-2 just popped up twice, I don't know how. How about that? Well, the best laid plans oft--and all that.

DR. DOMENICO: Is that called system redundancy, John?

DR. LATHROP: That's completely what it is, yes. We need it every place we can get it, even on the slides.

So then the flow chart is, we've laid the groundwork. We identified the 12 features and the 15 issues. We have listed those tables, which I just showed you, then we had to elicit a utility function that gave us the relative value of those different combinations of looks; that is, what if you have max access, but an--maximum access to a feature which gives you an intermediate level of confidence, or an intermediate access to a feature which gives you the most confidence. How do you get all of those combinations? We reduced those down to a single measure on the utility function, which now tells us how well each of the test strategies informs us about an issue of each of the features.

If you do the utility function right, you can add it up over the 12 features. Now you have how well a test strategy will inform you about an issue through all of its 12 looks at that issue, and then we developed a multiattribute utility function which collapsed over those 15 issues down to a single overall measure of scientific confidence, and obviously, I don't have time to go through all of those steps in detail here, but I'll go through roughly what the tables look like.

Your next three slides are all on this one slide. I'm doing this to save time again. We elicited a utility function that simply put a number, an equivalence number, or a utility number on each combination of your feature/issue link, which had one of three levels, and your test/feature link, which had one of four levels. Then we moved those utility functions to fill in these tables, and this, in fact, is the-- the top one is for Test Strategy 1, and below it is for Test Strategy 2. We have them for all eight, of course, and in each case, we have the 12 features and the 15 issues, and how good a look you were given at each of those. And then you can sum them across, because the way in which I've listed the utility functions, it makes sense to sum it across. So now, for each test strategy, you have a 15-number vector of how well that strategy gives you a look at each of the 15 issues.

So you can now leaf through that, but one thing is-- the thing of significance there is, again, on Test Strategy 2, 5, you see they have a column of zeros for the two features which are outside the block, the Abandoned Wash Fault that's D, and similarly, rock outside the repository, which is L.

So then we elicited weights for scientific confidence, like I said, two different ways: one with a technical panel and one with a regulatory/management panel. I'm leaving out a lot on this talk. There's a very systematic procedure for eliciting these weights that doesn't simply say, "Give me a number for a weight." We actually asked a set of A versus B questions. Which would you prefer, site characterization A or B? Where A might put a very good look at Issue 1 and a poor look at Issue 3, and B would give you a very good look at Issue 3, but a very poor look at Issue 1.

So in answering the A versus B questions, I force you to reveal the relative importance of the different issues, and we asked this of the technical panel to tell us, from a scientific and technical community's perspective, what the relative importance of these 15 issues are. And we asked this of the regulatory/management panel to give us a regulatory/management community cut at this, and if you'll look at this, you'll say, "Gee, they're all over the board," but they're not actually all over the board. For 11 of the 15, they agreed to within two placements on the ranking. This could get too complicated, but what matters is not the absolute number, but the ratio of importance weights between any pair of dimensions. To actually do the comparison, that's very involved.

A reasonable, proxy way to look at these is simply look at the relative ranking of the issues in the two cuts at scientific confidence, and all I'll say about that is relatively good agreement. The two places where they didn't agree in a big way are ways that are perhaps predictable. Flexibility was rated much higher by the regulatory/management panel than the scientific panel.

DR. DOMENICO: What do you mean by flexibility?

DR. LATHROP: Flexibility is this. Again, this is one of the odder ones. This is the ability to respond to unexpected data by giving you access to several different features. And so, in fact, the regulatory/management people thought that was much more important benefit than the scientific people did, and you can imagine why. Perhaps they've been through enough regulatory battles that they know people are always asking for more data. We're always surprised by what data we wish we had. So we like the idea of simply getting different places, so we have the flexibility later on in the process of collecting unexpected data, of data that we don't expect to have to collect.

The other thing which is interesting, actually, the highest level of disagreement in terms of relative rank is water table instability; that is, fluctuations in the water table, and this is rated much, much higher by the regulatory/ management panel than the scientific panel. Again, at this point, I'll just say I'm just a meter builder. I build the meter. I don't ask questions about why, but it makes sense, as a meter builder, that this would be the case, that the regulatory/management panel would be more sensitive to this than the scientific panel.

DR. NORTH: Now that that question has been raised, is there a story? I mean, did you stop there, or, given that there's quite a difference on flexibility in water table instability, is there a story written down now as to what the issue is and why these groups differed?

DR. LATHROP: Yeah. We didn't have a chance--the ideal way to do this study is to do it a couple of times, and once you've seen this, go back and reconvene the panels and ask them to do that. We didn't have a chance to do that.

DR. NORTH: So you're telling me that there has only been one iteration of that, and--

DR. LATHROP: Right.

DR. NORTH: --these questions that come up from looking at the differences in the assessments really haven't been addressed?

DR. LATHROP: That's true at this level of comparison. Within each panel, when there were differences between the people, I asked the people to debate with each other. Well, you know, Scott, talk with Jack, and Scott, you tell me why you said that side and Jack said that side, and vice versa, and they would talk for awhile. As it happens, by the way, very few people change their minds very much after you've done that discussion, but at least you've gone through the discussion to make sure that all members of the panel are working from the same level of information.

DR. NORTH: But did the argument get written down? Did each of the contending parties state their position in terms of why they felt that way, as opposed to just giving you a number?

DR. LATHROP: Yes. We had a Court reporter, in fact, recording the transcripts, and we have sort of--it's one of these, it's too much data and it's that thick.

DR. NORTH: Has that data been mined?

DR. LATHROP: Well, a little bit, and Ernie and Dave will be able to tell you where the data is, and it has been partially mined, but it's almost a question of too much data.

It's been partially mined and put into the report. It's a plethora, a wealth of data. It's very interesting. I don't want to over-represent to you what it is. We didn't go through a debate for every question. We went through those debates when there were significant differences among the panel members, and we went through more debates at the beginning than at the end, because it was at the beginning of all these processes that we were defining what the scales meant.

As it is, what is it, 800 pages? I don't know. It's a very large set of transcripts, and it can be mined, but it's relatively medium-grade ore, I'd have to say. There's an awful lot of talk in there that's just gone, you know.

DR. NORTH: What about the high-grade veins?

DR. LATHROP: That's right. That would be fun to dig through. It would be fun to dig through.

DR. NORTH: I mean, your chart shows some very obvious ones.

DR. LATHROP: Um-hum. Yes, it does. Those two questions in particular, yes.

DR. DOBSON: This is Dave Dobson. I guess I would like to make one remark, and that is, I think we tried to deal with those issues as they affected our overall goal, which was, you know, whether there was something that was going to come out of this that was going to lead to a fundamental difference in the opinions of how you might go about testing the Calico Hills. I mean, so I--but, I mean, we did not--we didn't stop to take time to explore the difference between the regulatory perspective and the technical perspective on the potential for water table rises for itself.

We do have--I mean, I think--I don't know how much we'll get into them specifically, but that is one example of one of the kinds of things that we learned about why we're doing testing programs and what we learn from the testing programs as we go through the whole process of site characterization, and, you know, you could use that example, you could use--Clarence would probably tell you the volcanism meeting last week was another example of a difference between --the regulatory/management panels are more sensitive to perceived problems, perhaps, than the technical panels, who tend to address things from a narrower perspective.

And so I think we have learned a lot of things from this process, but I agree with John, and he was perhaps even a little gracious in characterizing the grade of the ore, you know. There's a lot of--

DR. LATHROP: There is low grade ore there.

DR. DOBSON: There's a lot of information there, but we did not stop to address questions like that, except if there had been a fundamental difference from the two panels that, you know, the technical panel said, "Test this way," and the regulatory panel said, "Test a different way," then we certainly would have taken the time to explore that, but that didn't really happen.

DR. NORTH: Well, you're giving me an intriguing hint with respect to a pattern in the ore body, and that pattern is that on the regulatory/management side, you were looking at high value to situations where there is a perceived issue, let me say, out there in the world of public opinion. The scientific technical panel may say, "Well, we think that perception may be way off base. It's not really of scientific interest," but it is of interest from the regulatory/ management point of view.

Is that accurate?

DR. DOBSON: Well, I guess I would agree with that statement. I look at it and I wouldn't say that the technical panel said, "This issue is not of interest," but, "In the number of issues that we defined, here is where we would rank them." So you can see on that list, you know. I mean, the-- obviously, on that list, the technical panel was more concerned with essentially unsaturated zone flow processes in the list of things that we were considering, relatively speaking, on that list, and on the regulatory/management panel, the water table instability was rated somewhat higher than it was on the other. It doesn't mean that there was zero concern, though, I guess, and I want to clarify that.

DR. DOMENICO: This is all for Calico Hills; strictly, exclusively?

DR. DOBSON: This is all for Calico Hills, yes.

DR. LATHROP: But there are some intriguing philosophical questions here, which I want to make sure we pin down, and that is that if you look at the objectives hierarchy here, the sub-objective that we viewed as one of the five sub-objectives for appropriate site characterization is the technical perspective on scientific confidence, not the regulatory/management. We looked at regulatory/management task on scientific confidence for its significance for a regulatory delay, but not as a direct component of appropriate site characterization.

This is an important point, that we're not saying that we defined appropriate site characterization in terms of sensitivity toward non-technical perceptions of the relative importance of the issues. Where non-technical perception of the relative importance of the issues comes into play is in our anticipation of the delay. I haven't--I should have a special slide for that, and I don't. I trust you and the Board will hear what I say.

DR. DOMENICO: The previous slide, the statistical characterization rated very high from the managerial as well as the technical perspective. Statistical characterization of what are we talking about here?

DR. LATHROP: Of the unit itself.

DR. DOMENICO: Just the properties within the unit?

DR. LATHROP: Of simply being able to describe mathematically, statistically, spatially the Calico Hills unit.

DR. DOMENICO: In terms of the properties?

DR. LATHROP: In terms of its rock properties, in terms of its geohydrologic and geologic rock properties. I mean, which, going into this, I thought that that was the whole thing, and I thought, well, that's what you're doing, right? But it turns out to be only one of 15. Now, it's highly ranked of the 15, but it's one of the 15 things which you're doing. I learned a lot in this process. The nice thing about being a decision man was that you learned an awful lot about this.

DR. DOBSON: John, I might add just one comment, and that is that there's a slight difference in the things that are labeled MC and the things that are labeled ACM there, and MC is maximized characterization, and that essentially entails understanding both the distribution and variability of properties. And the ACM ones, which are also very important, are sort of focused differently, you know, ability to figure out how appropriate certain conceptual models are.

So they're defined in slightly different ways, but there is, as John noted earlier, a little bit of overlap between some of them.

DR. LATHROP: Yeah. We did not do a perfect job of making them mutually exclusive. I don't want to represent that to you.

The next slide I don't want to spend much time on. I'm trying to shorten the talk down. It looks like a zooming Thunderbird, and it's basically the relative weights that the two panels gave to the different dimensions, and I allowed the different ranking to show up there, so this--basically, it simply tells you that yes, with both panels, there is significant drop-off in weight from the most important to the least important dimension.

I put up slides like this, because what a lot of people don't realize with these MUA studies, is you do find that some of the attributes are much, much, much more important than other attributes, and this simply is a way of representing that. They do drop off in a slightly different way, and what I've noticed from other applications, they didn't test--and in particular, with this one, is when they drop off a little more slowly, it's more likely to be an indicator of greater level of disagreement among the panel members, as I average out over the panel members, than it is a substantial, within each panel member, a consensus on a lower rate of drop-off from the most important to the least important. I don't want to get sidetracked about that.

So now, with those importance weights, we can look for instance, at respondents. There were six respondents from the technical panel, and we can take this eight strategy by 15-issue matrix, with a utility function for each one. I actually had to elicit a utility function for each of the 15 issues. I listed three generic ones and applied them to different issues by asking a set of structured questions, like, Which of these issues would be represented by this type of function or that type of function? And these importance weights, and then, although I did test a multiplicative form, the additive form did about the same thing as the multiplicative form and it's a lot easier.

We simply take an additive weighted sum of these utilities to get an overall utility score for Respondent A for each of the eight test strategies. That is in Utiles. I then ran the utility function backwards to get another measure, which is called a Uniform Look Equivalent. It's basically if this wound up being a utility of 1.04, simply a weighted average of these 15 numbers, then I could say that is equivalent in value to a hypothetical strategy which would give you 8[illegible] max/strong looks at each of the 15 issues. That would also score 1.04.

So this is a more meaningful way to represent scientific confidence. Test Strategy 1 gives you as much scientific confidence as a hypothetical strategy that would give you these 8 max/strong looks at every one of the 15 issues. And that, in fact, is the index which we used to define scientific confidence.

This is a little confusing, I'm sorry, because what is up here are the utilities. What I should have had is a table that showed you the number of max/strong looks for each of these 15, then showed you the equivalent number here as a measure of those.

That's our index of scientific confidence, and here's the numbers, and frankly, I'm always--I'm very conscious of the problem with MUA, that people always think, well, you're modeling these so carefully with respect to the preferences of the people you're working with. Aren't you going to give the decision-maker what he wants to hear?

Well, I can assure you that in almost every one of the applications I've done--I've done it here, I've done it at siting powerplants, on public policy things--you always come up with results which are, in fact, different than what you expected to hear. Now, I never talked with Dave or Ernie about what they expected to hear, but I think they probably expected Strategies 2 or 5 to score highest in scientific confidence, and, in fact, it didn't. Strategy 1 did. In fact, let's just go to the next slide, which is the rank, rank orders in scientific confidence.

In fact, all six respondents uniformly gave us a utility function which corresponded with Test Strategy 1, giving you the highest level of scientific confidence. 2,5 was second for four of the six, and third for the others. Those people thought that Strategy 7 should, in fact, score second, and then there's relatively good agreement among the other strategies.

Now, in fact, this is where we were thinking, oh, well, why is that? We went back to the tables, and the nice thing about this is it shows you why you get to any answer. Well, why it is, is because 2,5 didn't get you out to features D or L, the Abandoned Wash Fault, or the similar rock outside the repository, and Test Strategy 1 did. And, in fact, we'll see later that it matters how many features you get to. The features give--each feature gives you about the same level of increments in scientific confidence. So if a test strategy gets you to more features than another one, it's going to score higher, generally speaking, in terms of scientific confidence.

DR. REITER: John?

DR. LATHROP: Yes.

DR. REITER: Just a point of information. The max/strong look, again, is something which relates the test/feature and the feature/issue link; those two links?

DR. LATHROP: That's true. That is the two links, the maximum access to a feature, which gives you the strongest increment of scientific confidence.

DR. REITER: So you had that little matrix beforehand, which gives weights that you showed?

DR. LATHROP: Um-hum.

DR. REITER: Okay. Now, so a single max/strong look is a one?

DR. LATHROP: Is a one, right. So when, here, we say that 8.5, well, what does that mean? That means that test--well, let's go to the average, the same thing. Test Strategy 1, 8.6; what's that mean? That means that that strategy gives you this whole vector of things. I mean, you know, a test strategy, gee, it looks through 12 features and 15 issues. That's too much to match up. What does it mean? It means that it's equivalent in scientific confidence to a test strategy that gives you about 8 max/strong looks at every one of the 15 issues. For every one of those issues, it's equivalent to getting full access to a feature that gives you the maximum level of information about scientific confidence.

So--and it was tough to come up with this matrix, and I sort of would hope that I had a--I would come up with a more intuitive measure than this one, but this is the best that we could do given the high dimensionality for the problem. And, in fact, it came out with results which, in retrospect, made sense, although it was a little bit of a surprise, but it all made sense.

DR. ALLEN: But we heard earlier that Strategy 2 could be easily modified to take advantage of the things that were giving the preference to Strategy 1.

DR. LATHROP: Exactly. In fact, when we do that, we come up with a 9.7. So 2,5 can have a higher level of scientific confidence. Why I didn't generate a slide for that is we get on a little bit of a slippery slope there, because we didn't happen to have the risk or the cost measures for that hypothetical strategy. Now, in fact, it wouldn't--it shouldn't have higher risk, because what you're adding is outside of the block, not much higher risk. And cost, I have a sensitivity analysis up in my hotel room, which shows that you can add--it would still remain a superior strategy upon adding those even if you could add--even if it would cost you $50 million to add those, and that's reasonable.

So we went into that. I didn't want to bring that into this discussion because, in fact, we did not have a chance to fully measure that hypothetical strategy 2,5 extended to the outside in terms of risk and cost.

So where are we? Well, all that was defining what scientific confidence is. Now, looking at the other measures, residual risk was brought right out of the VOI study. Cost was assessed in a relatively straightforward manner. This is the direct cost of doing the characterization. So the other two are phasing potential and service date.

Phasing potential, we basically, you know--in an ideal world, we would have time to build a complete phasing tree. Okay, we'd show, well, this strategy allows you to branch here and there, and so forth and so on. We didn't have time for that, so we built a very simple sort of phasing tree that basically said: Strategy 1 gives you four ways to end up, where each of those ways can be characterized by one or the other of the eight strategies. So one can be designed so that you can stop here, and it looks like Strategy 3; you can stop here, it looks like Strategy 8; stop here, it looks like Strategy 7; go all the way and it looks like Strategy 1. Conversely, Strategy 3 basically only has one way to go.

So we have a very simple measure of what phasing potential is. It's simply the number of different ways that the thing can end, and it ranges from one, very low phasing potential, to four, very high. And then we evaluated the utility function on that measure, and you will see that.

Delay? Again, in fact, in some other studies--I do a lot of siting studies for large facilities, and have these delay models, and the delay models are, here's all the different things that could happen for permit acquisition delay. Here's all the highs and lows, extra months delays. Here's which ones are parallel, which ones are serial, build a PERT chart, do it all. We didn't have that level of data for this, as you might imagine.

But what we did do is we developed a delay scale based on the scientific confidence, regulatory perspective--it's the regulatory perspective, not the technical one--and two other things. And again, one of these was a surprise to me. It was felt to be important that if you have a strategy which drifted inside the block, that put you in a better position to respond to later requests for data; that is, if you didn't drift inside the block and later on somebody said, "Gee, we'd really like to see some rock from such and such," we said, "That's fine, we'll get you that rock. Come back in a year and a half." If you've done the drifting, you can say, "Fine, we'll get you that rock. Is Tuesday okay?" All right, so this can be important.

It's important--I should have stated at the beginning, the delay, the particular delay we're looking at is docketing delay, that delay that occurs post-characterization between submittal of the license application and the actual formal docketing of that application, which would start the clock, and this particular delay was selected because it is the delay that the test strategies would be seen as affecting differently. You want to pick these measures to be diagnostic, to discriminate among the test strategies, and it was felt that this is the area that the test strategies make the biggest difference, in this post-characterization docketing delay.

Now, on these--and then, of course, there are two things. One is this idea of your ability to respond for later requests for data, and also, delay due to concerns about the residual risk, because you have--you do have--it's quite small, according to the assessment, but you do have a slightly higher level of residual risk when you've drifted inside the block. At a technical level, it doesn't seem to be something you'd be sensitive to, but in the regulatory spirit, you may be.

We measured these, again, by, again, a proxy variable, which is simply the expected release, the assessed increase of expected release in the "R" measure, which I hope we're familiar with, as a proxy for these two, because this measures the degree of intrusiveness within the block. The more you're inside the block, the higher that number is, and also, of course, the better your ability to respond to later requests for rock from inside the block, and also, the greater your concerns about residual risk.

It's important to keep in mind, though, that because this is post-characterization, this is not regulatory delay due to concerns that, gee, maybe you'd better not do that. Maybe you'd better not drift. Let's think about it some more. It's more, well, you've done it. It's a fact. Given that you've done it, we have some more concerns about what we want to do with your license application before we docket it.

Looking at that, we looked at the test strategies. We ranked them. We presented it to the regulatory/management panel, and this was the assessment by the regulatory/ management panel, by their potential for a delay. We grouped them into five groups, and you see Test Strategies 3, 7, and 4 are grouped into the middle not because they're the same--they're quite different on scientific confidence and quite different on intrusiveness--but they happened to be at about the same level on the scale, and we couldn't reliably rate them higher or lower, and this gave us a five-level scale of delay potential based on scientific confidence, and intrusiveness as a proxy for in-block flexibility and risk concern. Again, like I said, we could have done a more elaborate measure. This one captured the important issues.

So now we're ready to actually do the five-attribute, multiattribute utility function that combines these five measures down to one. Basically, take the data I'm about to show--I should have had that slide first, I guess--this 40-number data table, how the eight strategies perform in each of the five measures. We built single-attribute utility functions, one for each performance measure. We assessed the relative weights, and we actually assessed three different sets of weights; a DOE perspective, particular risk-averse measure--which I'll talk about--and another one. Look at the ratings, do some sensitivity analyses and contrasts, and come up with conclusions.

Let's look at this 40-number measure. Actually, maybe it's a little more. What I like about the methodology, very simple, very systematic, here it is. In fact, supposedly, we could have stopped here and said, well, we've evaluated the eight test strategies. Just look at it and see what conclusions you might make. As it happens, people have a hard time. It's confounded several experimental psychology sorts of experiments, and so people have a hard time comparing things when they vary on five dimensions, and so we did the MUA.

But here's the data: Here's scientific confidence. You see that, actually, 2,5 and 7 scored the same, but two sig figs--and by the way, I wouldn't trust this--I'm sorry--significant figures. I wouldn't trust it past 1[illegible]. Okay, so one is higher, and then 2,5 and 7, and then the others. On residual risk, it put it in in three ways because this is one of those things that it may depend upon what format you use how you look at that risk.

The three measures--first of all, the fraction of increment, how much of an increase in risk over the surface-based testing alternative results from the test strategy? The highest was Test Strategy 2,5, which gave you a 13 per cent increase, and the lowest was Test Strategy 8, which gave you way down there, okay. And so, that shows you, well, gee, in fact, looking at that--and I use that as a format for some of the tradeoffs and I'm a little sorry I did, because it tends to emphasize the fact that they are different in risk, without telling you the magnitude of the risk. Hence, on this table anyway--perhaps a little too late--I put in two other measures.

One is residual risk in terms of increased expected fatalities. This is simply based on taking at faith the idea that the risk measure presented in 191.13 is based on if you do--if you have this risk measure of one, or "R" of one, it corresponds with a thousand fatalities over 10,000 years. Taking that on faith, with a large swallow, I then say, well, here is the expected fatalities. Oh, now, I see. Well, the difference between the worst and the best is a 50th of an expected fatality. Now, as we've heard in some military briefings, every fatality matters, and that is certainly true, but at 1/50th, we're not talking a large change in expected fatalities.

Residual risk can also be measured in terms of the fraction of EPA limit, and this is hard to read. What it basically says is, at best, you're operating at 15/100,000 of the EPA limit; at worst, you're up to 17/100,000. So just a way of expressing the assessed risk impacts of the testing, which was lifted completely and totally from the VOI study, the best strategy puts you basically the same as no test strategies in terms of risk at 15/100,000. The worst would lift you up to 17/100,000, not a big change.

Delay, that five-level scale. Cost is in terms of millions of dollars relative to the cheapest one, brings us over 174 million; and phasing potential, as I mentioned, is simply the number of different ways that you can end the project in a way that looks like one of the eight strategies.

This is utility functions from the regulatory/ management panel. Scientific confidence, yes, there's some concavity downward there. Going from four to five or five to six max/strong matters somewhat more than going from six to seven, seven to eight, or eight to nine. Not a big difference, but some difference. The utility functions on the individual measures were--tended to be a little sharper than that, but on the overall measure, it looked like that.

Delay and phasing potential are discrete functions, so I didn't draw lines. Don't draw lines on discrete functions, just put dots, and there they are.

The interesting thing there is on phasing potential, you see you get almost all your benefit going from one way to the end of the project, to two, it was felt by the regulatory/ management panel. Again, this is a result of a structured set of elicitation questions with questions about, if you had your choice between this type of site characterization plan and that one, which would you prefer? Pair comparisons in a structured way. It's all down. I can give you the names of the books to look at to see how it works, and I can show you the elicitation protocols if you'd like. It turns out you get most benefit going to two, and almost no incremental benefit going to three and four ways to end the project.

Residual risk and cost were linear. The old adage, over a small enough range relative to assets or relative to the world at large, every function is linear, and that is true for these two.

We elicited weights, and again, through a procedure which I have gone into, and here it is, and I can just imagine a funny sort of scenario that somewhere in the back of the room is a PR person for the DOE, and he looks at this slide, he gasps, he comes running up and rips the slide out. "You can't put that up there. Look at that, look at that. The relative importance of residual risk by the DOE is .06. It's the least important. We can't let people see that." And, of course, I could see his point, okay, that oh, that doesn't look good.

Well, what this does is, is this illustrates one of what I view the biggest advantages of the MUA approach, it forces a logical consistency to what you do. One of the things it forces is the idea that the relative importance of an attribute is not simply a function of the intrinsic importance of that attribute. It's a function of the attribute and the swing rate from the best to the worst among the alternative set.

Now, the DOE can quite legitimately say--and I'm sure it's true--that risk is paramount, that the intrinsic--that the level of intrinsic value, whatever that is--and actually, that is something I made up. That doesn't come out of the methodology. The methodology doesn't say anything about intrinsic value. We can say, as people, at the level of intrinsic value residual risk is by far the most important dimension to DOE.

Now, looking at the range over the alternatives, ranging from 15/100,000 of the EPA limit to 17/100,000, oh, okay, on that range, it's not very important, and that's what we mean by this .06. This, in fact, is the difference in overall utility from the worst level to the best level on residual risk, and .06--in fact, frankly, I think the way we elicited it, we probably over--through an artifact in the elicitation, we over-represented what that relative risk is. Highest level is confidence, and then delay, cost, and phasing potential rank down below that. .06 is the level for residual risk.

Now, I say DOE perspective. We elicited these importance weights from the regulatory/management panel and we said to take the DOE perspective when they gave us the answers to all these A/B questions. They insisted on pointing out that they were not representing DOE as it exists in a vacuum. Any good decision analyst, the first thing you ask is, who's the decision maker? Well, the DOE's the decision maker. Yes, but DOE not on Mars, okay, DOE living in Washington, D.C., with the regulatory agency a few blocks away, all right? That's important.

So it's a constrained decision maker, and when we asked the DOE for--the people from the regulatory and management panel to give you the relative importances of these different things--and you don't ask them directly, you ask them all these A/B comparisons, they kept in mind the fact that, yes, it's the DOE but, of course, we have to keep in mind that we work in the regulatory milieu, and given that, we actually gave this particular weight.

Now, they didn't say so and I didn't ask them to say so, but I suspected from the way that they were wording their answers that if it had been DOE living on Mars or something, it would have been a lower rate. But, so this weight represents both the fact that the swing range goes over 1/50th of an expected fatality, and the fact that even though that's a small range, it is still the case if Test Strategies 2 or 5 do score the highest on risk relative to the eight strategies.

And an external person looking at that might say, well, I know it's a very small change in risk, but still, for institutional or whatever reasons, I have a hard time being comfortable with a site that happens to be the riskiest of the eight you looked at. And I say, well, yeah, well, that violates an axiom of decision analysis. That violates the independence from irrelevant alternatives axiom of decision analysis, but I can understand how you might feel that way, given the environment you work in, and given that environment, you give it that weight.

We also asked the panel to put on a different hat, and actually, we asked them to put on the NRC hat, but we're not treating this as a representation of the NRC. Okay, it's not. We were asking them to put on a hat of another agency, how they would look at this, simply as a way to scale the sensitivity analysis. Any good analyst always does a bunch of sensitivity analyses, and for all these things you'll see I did it on scientific confidence, I did it on the features, this, that, and the other thing. When you get to weights, the one problem with MUA is as long as the strategies are such that none dominates the other--in fact, oddly enough, none of the eight test strategies dominate another one. By dominate, I mean scores better on every dimension, scores better or at least as good.

So given that, you could come up with any--you could come up with an importance weight, which would give you any ordering you would want that would at least rank any one of the eight on top that you would want. So a sensitivity analysis for importance weights, you have to do somewhat differently than just saying, well, let's try this, that, and the other thing. We had to scale it.

So we scaled that sensitivity analysis by collecting a set of weights representing some other agency, and quite legitimately, you always have to give your respondents a dual job. You have to say, well, imagine yourselves if you were, for instance, the NRC. But again, I want to emphasize, this is not meant to represent the NRC. This is just something they had in their heads when they gave the answer, and you see it did have a higher weight for residual risk.

In fact, the other rankings stayed the same. Still, the delay was higher than cost, cost was higher than phasing potential, so all you basically did was take residual risk from the bottom up to just below confidence. Then I took the same data and just looked at, what if you just looked at confidence and residual risk, and came up with those weights, and they're different from what you'd expect from that ratio because of the way I treated the errors, and took the errors out.

DR. NORTH: What happened to F?

DR. LATHROP: F had to leave for a very worthy cause. I forget which cause it was, Cub Scouts or something, and I think seeing this, he probably wished he'd gotten somebody else to go. He, you know, I'm sorry. You know, he left at that crucial time at the end of the day when we got the final judgments there, and so these are averaged over six respondents, and these are averaged over seven.

So, you say, God, he's talked for so long and he hasn't given us the bottom line. Finally, here's the bottom line. It didn't come out on the printer very well. Here it is. 2,5 scores better than the others, not by very much, but scores better. Now, actually, this little dot here represents where it would score if we added in those outside features, but it does score somewhat better and so, of course, the next question is: How much better? What does it mean? And that's what we'll look at, but there it is. 2,5 scores best, 1 scores second, 7 scores third, and so forth on the DOE perspective.

Looking at the rankings, DOE perspective, 2 scores highest for five of the seven respondents--see, on the DOE, we have all seven. He hadn't left yet--scores highest on five of the seven, and the other two thought it should score second or third, relative to one, and then seven is the one that pops in there, and relatively good agreement on eight and four. They all agreed on ranking them fourth and seventh, and they flipped around between Strategies 3 and 6, but who cares? They're down in the fifth and sixth ranking. So the ones that are at the top of the rank, it's not perfect agreement at all, but on the average and five of the seven voted 2,5 at the top.

Putting in those other weights I talked about, Test Strategies 2,5 falls below 1, 7, and 8, in fact. So this ranking is sensitive to the relative weight, specifically; most easily seen as it's sensitive to the relative weight given the risk. I'll get into the contrasts in the sensitivity.

Just before I leave the basic results, the nice thing about the algorithm is you can run it any way you want. I re-ran it to evaluate the relative contribution to scientific confidence of the twelve features, and I came up with, here it is. And I said, oh, I've done it again. Dr. Lathrop you blew it. That can't possibly be true. The 12 features cannot possibly be so close to each other in the relative contributions to scientific confidence. You know, we're all Bayesians, right? We all look at this and say, which is more likely, that I screwed up, or that the real world is really such that the 12 features score about the same? And of course, my being modest, I said, well, I screwed up.

I went through it all, but really, you know, as best as I could tell, it really is true that each of the 12 features scores within 30 per cent of each other, and it falls out of this general orientation that we don't have a good idea about how we'll learn from each of these features. And so, generally speaking, just getting to a feature and letting me sample and test, and so forth and so on, is worth most of what getting to that feature is, and the differences because of the type of the feature and what you're actually doing and where it is, are small relative to getting to the feature at all, and that's why they only vary over 30 per cent.

I especially expected the outside repository ones, D and L, to score appreciably lower, and no, they're right in the middle of the pack, right in the middle of the pack. So there they are, so we'll be talking about that a little later. But basically, it basically says, design your test strategies to get to the features. The more features you get to, the better off you are. Simple.

Now, contrasting the top-ranked two, Test Strategies 2,5 versus 1--I do this for others. It's sort of interesting. The total column height is proportional to the importance weight for that dimension, and the shaded part is the--proportional to the degree of that sub-objective that you've attained. It's a little funny, because you have a--this means this has--2,5 has a lower cost than that one, so it scores higher on the minimizing cost one. The differences are too close to call here. It just gives you the feel that, yeah, they're fairly close to each other, and 2,5 gains more in cost--in less cost--than it loses in phasing potential and risk, and it gains more in reducing delay than it loses in scientific confidence. Let's look at the numbers. That's clearer.

The nice thing, again, with the additive function is you can look at the numbers, you can break them down; 2, 5, and 1, here's the data. This is right out of that data table, that 40-number 8 x 5 table. And we simply take the data and we transform them to the utilities for each of those measures. We take the difference in the utility, we multiply that by the importance weight. You know, simple, simple stuff. Here it is.

Here, in fact, is the difference in utility between the two. Don't pay attention to the magnitude. .02 sounds small, but we'll show what it is in the later part of the analysis. The important thing is that, how is 2,5 better than 1? Well, it's worse than 1 on confidence. It's worse than 1 on risk. But it's enough better than 1 on delay and cost that it scores better overall. You know, that's not at all the result that I expected, and I'm not sure I'm all that happy with it. I mean, if you like 2,5, you'd like it to be better than the next closest one on what you'd think was a fundamental part of this, was confidence and risk. Well, this is delay and cost. These are operational and procedural sort of things, but, okay, they do matter. They do matter.

Another way to look at it is, going from Strategy 2,5 to 1, you come out behind. What you gain in risk and confidence, you lose more in cost and potential delay. Well, the discomfort I have with this is it does depend on that relative weight given to risk, and remember, I did all these sensitivity analyses, and it's robust on the uncertainty in scientific confidence, and it's robust on a couple other things. It's not robust on weight.

A plausible relative weight for risk, somewhat higher than what was assessed from the DOE panel, would change the ordering, and that is a difficult value tradeoff, risk versus confidence, risk versus cost or delay; obviously, a difficult value tradeoff. Well, the nice thing about the methodology is we can use it to finesse that tradeoff by transforming it to a cost per life saved sort of evaluation, and I did that by basically taking the multiattribute utility function, taking the four non-risk attributes--and I won't go into details here, but I basically--well, you know, it's sort of interesting--took the four, generated a hypothetical set of two strategies--1-prime and 2-prime, 5-prime--where they're the same on three of the four non-risk dimensions, and only differ on cost to see what cost difference would give you the same difference in utility that you have between the actual two, and the bottom line said $61 million.

So basically, on the non-risk attributes, using the utility function as a transform, I find that, effectively, on the non-risk attributes, Test Strategy 1 costs $61 million more than 2,5. In fact, on the raw data, it costs $58.9

million. They fall about the same.

So now we can say, well, this is interesting, because now look how they differ on residual risk. They differed by, you know, .0153. Okay. They differ by the sixtieth in--a sixtieth of an expected fatality, or whatever that is. So what you can say is, we can use this to transform the question to, if you were on 2,5 and you tell it to change over to 1, that's equivalent to spending $61 million to reduce the expected fatalities by .015, which amounts to about $4 billion per life saved. Now, that's a lot.

I didn't--this is so much, I didn't go back to do the research. I think we can all remember there are statements in various regs and reg guides to the effect that one million per life saved and ten million per life saved, something about reactor containment vessel work. You might remember. There is something about that, and it was equivalent to $10 million per life saved. Whatever it is, it's a lot less than $4 billion per life saved.

So by the arcane logic of dollars per life saved, you say that, yes, I see now. Oh, okay, 2,5 is superior to 1 looked at in this way. In fact, because $4 billion per life saved is too much to spend for the benefit of the lower risk, it's such a tiny decrease in risk which you get; in fact, what's comfortable about this is, gee, even if we underestimate that risk by a factor of 100, that sounds like a lot, but actually, the risk assessment can do--that can happen. But even if you underestimated the risk by a factor of 100, you'd still be in a position where it would not be attractive to switch from 2,5 to 1, because that's still $40 million per life saved.

Now, I haven't been able to persuade my minister of the logic of this. He keeps saying, "John, that doesn't strike me as being a comfortable sort of logic," and we argue a lot and we wind up saying, "Well, you know, Jim, that's why you're a minister and I'm a decision analyst." He says, "Yes, you've got that right." We don't argue much, because we comprise two-thirds of the tenor section of the choir. If we got really upset and one of us quit, the choir would be ruined. I digress; I'm sorry.

Suffice it to say, with this logic, we could see that 2,5 is superior to 1. So, finally, bottom line time--I'm sorry, there will be some overlap with what Dave said--2,5 is the most desirable of the eight strategies considered, but it is not much more desirable than Strategy 1. In fact, I should have said, I did some other equivalence measures and using the utility function to transform it, what if they only differed on cost, they differed by $17 million on cost? I mean, they're right next to each other, okay? But more generally, it is--you can then extrapolate that the extensive excavation in the Calico Hills provides a net benefit, compared to minimum excavation there, considering these five dimensions.

And the robustness of that ranking is increased quite a bit by adding those outside accesses, which we've talked about, and this was from a previous talk. I think that if we did a more refined elicitation of that risk measure, we might wind up, actually, with a lower risk. Because of the format I used in doing this, it tended to over-emphasize the relative differences and not looking at the magnitude.

I have to qualify those findings with that other differential risk perspective, differential risk-averse perspective, and say, well, how would another agency look at this? That ranked 2,5 below 7, 1, and 8, ranked it fourth. So clearly, these results are not robust with respect to that, but adding features to Strategy 2,5 does increase the robustness with which it is ranked over Strategy 1. It doesn't help this, though. Even with adding those features, it still ranks below 7, 1, and 8 on that other perspective.

That's why I went into the cost per life saved. Well, let's get another handle on what the appropriate tradeoff might be, define that, and with that, in fact, you do have a robust rating of 2,5 over 1. Ranking results are robust with respect to uncertainty in scientific confidence, at least as that uncertainty is represented by differences among the respondents.

More generally, we found that access to each of the features provides similar increments in scientific confidence, so the more features you get to, the better. Obviously, the more features you get to sooner, the better. The relative weight given to residual risk as elicited here is critical to that ranking. A more refined elicitation would reduce that sensitivity, and I should have added here that in a cost per life saved rationale, you have a firm basis for ranking 2,5 over 1.

Delay and cost considerations that we find in this comparison can be just as significant as residual risk and scientific confidence, given the particular differences of the alternatives we were looking at. So, as this one did, any future evaluation should consider at least those attributes.

You may have noticed, I dropped phasing potential out. As it happened, phasing potential didn't happen to matter. So it seems clear to me that 2,5 is the superior sort of alternative. You can argue, but if you want proof, I'll give you proof in these specific areas.

I'm sorry I went through in double speed. I was conscious of the schedule, and I would entertain questions.

DR. NORTH: John, I must say, I confess to being a little dazzled by all these numbers and I, more than other Board members, have seen this before. I think we will need to spend some time studying the documentation.

With my background, having seen a number of applications of decision analysis in this area, I always find that my concerns are much more with what might have been left out or what wasn't covered in sufficient depth, rather than trying to make sense of all the material you've presented us.

I'd like to go back to my question to Dave Dobson early on, about the issue of the impact of exploring within the repository block on the integrity of the repository, and that effect on performance. I know that's in your analysis. It's an issue under review from NRC, and you haven't heard back from them. That's an area where I'd like to have much more insight into how that issue might distinguish between Strategy 1, where, as I understand it, most of your exploration is outside the block, from Strategy 2 and 5, where it's inside the block.

I'm not sure your analysis has given us as much insight on that issue as maybe it might. I think that may be potentially much more important than, for example, a dollars versus life saved tradeoff. What really might be concluded about the effect of the 2,5-type of exploration versus the 1-type of exploration and its effect on repository integrity?

DR. LATHROP: That actually, as you saw from this analysis, was numbers we frankly just took from the VOI study, so I'm not in a particularly good position to comment on that risk assessment, except to say that it does seem that it can be off by a significant amount and retain these results by that cost per life saved logic, but in terms of the questions to be raised, that's--

DR. DEERE: Maybe I will address this one to Dave. Isn't it true that 1, 2, and 5 might all be out of date, as we'll find tomorrow?

DR. DOBSON: Pardon me? Well, I mean, I don't think they're out of date. I mean, the recommendations went to the ESF group and they are being basically incorporated. There will be some modification. I mean, it's--they will likely not look precisely like they look now, and you don't see any accesses or anything on ours, whereas, obviously, in the ESF study you had a, you know, we have what are likely now, in our current configuration, to be ramp accesses to the Calico Hills.

I might note that as we go through the design process and we redo things like the waste isolation impact calculations, they will change when you consider ramps instead of shafts, for example.

To respond sort of briefly to Warner's question, we have in, you know, in some of the past briefings gone into a fair amount of detail describing what we thought the magnitude of the impacts was going to be, and--Ernie, if you feel like jumping in, please do--I guess I would say that we've tried to characterize what the likely maximum magnitudes of impacts as a result of the excavations could be, and that, of course, is a function, in part, about how the rest of the site is behaving. And so we've tried to bound the range of, you know, in a worst case, if you will, how much water could you move through those openings you've created, and what other kinds of site behavior need to be happening in order for that to happen, and I guess our, you know, the conclusions that we got were that there was a low probability of getting scenarios where you were releasing amounts of radionuclides that got anywhere near the standard, but even when those kinds of scenarios were happening--in other words, like concentrated flow that was collecting a lot of waste--your relative contribution of the engineered barriers or the engineered openings was small.

Matter of fact, the relative contribution, as you increase the total flux through the system, goes down. And so you're looking at a relatively lower impact. The maximum percentage impact that you see on the site is when you have very small amounts of water moving through most of the site, but you construct a model which concentrates a relatively large amount in the openings. And in that case, you can get significant per cent differences like you saw, the 13 per cent difference that we modeled between the--and, you know, kind of like John said, I would be wary of the number of significant figures there--but a relatively significant difference in the different strategies in terms of a relative contribution, but all of those are at numbers which are way below the standard, because you've got virtually no water moving through most of the site in that scenario.

So I guess my feeling is that the numbers that we've put up there were pretty conservative. In fact, when we start going back to redo the waste isolation calculations in the design process, especially when you consider that the majority of the contribution of the impact in our models came from having a direct connection, a shaft, between the main test level in the Calico Hills, when you start modeling the impact that you get from the ramps, with no direct connection, I suspect that you're going to be seeing the relative impacts from Strategy 1 move very close to the relative impacts from Strategies 2 and 5, but we haven't done all of those, you know. We did--we kind of had to call a halt to this analysis, and we felt like we had documented at least our view of the relative impacts adequately to support our recommendation, and so we hope we have, and we're waiting to hear comments from the NRC on that issue, obviously.

But again, you know, we also want to make it clear that this is not the end-all in waste isolation impact evaluations, nor should it be, and, you know, matter of fact, the best write-up you'll see on the subject will be in the license application, I hope--if we get that far--and, you know, assuming that the site otherwise proved to be acceptable, the license application requirements basically would require us to do an extensive job of documenting what the relative impacts were from what we had already built and what we intended to build if we went ahead with the repository.

DR. CORDING: The flip side of that, of course, is looking at what is the additional benefit, and obviously, you've been looking at that, of going into the repository block. I'm wondering if we've--when you compare being outside the block to being within the block, whether that benefit of actually being in and looking at the specific conditions that you encounter in the block is just--to what extent that has fully been factored in here, because it seems to me that, for example, just in looking at unknown conditions, faults are--we've got an idea of where the general faults are, but they're never exactly where you think they are, and they're never quite the same character. And you'll probably find some faults that are not anticipated at this point.

And, for example, it might be that if you go down in the Calico Hills in a different crossing, if you find a fault zone, that'd give you some information that says, we should avoid that area, or it'd be better if we avoided that area up at the repository level. It might give you some information that you couldn't find outside the block. So that benefit is another big part of this.

MR. BROCOUM: This is Steve Brocoum. One additional comment. It builds on yours, but I was also thinking about it before you started, and that's the NRC had a lot of comments on our SCP and SPC about representatives of data when we were just thinking of drifting in the northern portion of the block, and they made a lot of comments in terms of the southern portion of the block. And so part of the extensive drifting in within the block would address that comment, which was not explicitly addressed by this study, but it's another consideration I think you need to consider.

DR. DOBSON: Yeah, and I'd just like to support what Ed just said in that I don't know if you remember this table that John showed early on, that had check marks about how well each strategy provided certain kinds of information. Strategy 2 and 5 generally did better, and in some cases, significantly better than Strategy 1 for all the features inside the block, because it was there and it did exactly what you just did. I mean, you had the opportunity to look at it. But Strategy 2 and 5 got zero, as John noted, on two of the categories, and so when you added the total, when you summed everything, it came out slightly lower. But for those areas inside the block, you can look there. There are numerous cases where Strategy 2 got three checks versus two checks for Strategy 1, so...

DR. DEERE: Well, are you going to build inside the block or outside the block?

DR. DOBSON: Build the repository?

DR. DEERE: Yeah.

DR. DOBSON: Oh. I think that's pretty much a given.

DR. CORDING: I was involved in a court case where we didn't do quite as well as we should have, because I had to look at conditions around the area, not within the actual area. I knew the geology was the same, and I knew that I could extrapolate it, but I could not completely convince the Judge that was hearing the case, you know, and I think that's part of what you have to do.

But I think there are some very good technical reasons why things are going to be different than you expect, and if you find them in the block, that's where you're going to need--you really need to be looking in the block, in other words.

DR. DEERE: Well, if you're going to go outside the block, and you're into an area of a lot more faulting and closer faulting, which is not consistent with what you're going to be having within the block, you shouldn't expect to have information that you can use directly, I wouldn't think.

DR. LATHROP: This was somewhat of a surprise to me, that the scientific confidence measure didn't give as much credit as both of you are suggesting should be given being inside of the block, but those reflect the judgments of the technical panel.

It would be interesting to reconvene the technical panel. See, I hesitate--I always run these playing dumb, which is not hard for me to do. I don't, you know, I try to make sure I'm not leading the witness. So in the course of the technical panel sessions, I did not say, "Now, gee, you guys are giving all the same ratings to these outside ones as the inside ones. Are you sure you want to do that?" I didn't do that, and perhaps if I'd done that and we really sort of beat on them, they might have given a higher relative weight to the features that were inside the block, but they didn't.

DR. NORTH: Well, it's very valuable to have put this picture together, and we come up with a bunch of questions which would be very useful to go back into a second iteration and ask, "Do you really mean it? Can you justify this part of your story a little better?"

DR. LATHROP: Yes, right. All this should be done twice.

DR. NORTH: Right down to the reasoning.

DR. LATHROP: Yeah. I agree.

DR. REITER: Dave, I have a question, and perhaps you will answer it in the lessons learned. And that is, the difference between the VOI and the MUA, it seems to me there's two differences. The VOI, one, the way you constructed it dealt primarily with technical kinds of concerns; and second of all, it gave you an absolute answer. Should I or shouldn't--is the value of information, is it worth it or not?

The MUA allowed you to incorporate non-technical considerations, but it gave you a relative ranking; in other words, you didn't have--couldn't it have been possible to either construct a VOI so you included the value of non-technical information, or similarly, couldn't you have constructed the MUA so that you also had an option which said, compare the value of no testing?

DR. LATHROP: Yes. In fact, Hollis and I are attempting to write an article now, and in the course--I say we're attempting. I wrote my half of it, and it's 19 pages. It's going to get a little long. We'll have to talk about that, and Hollis probably has generated 25. But that's specifically what we've talked about. Yes, there's nothing intrinsic about the two approaches that would have prevented a good combination.

In fact, the VOI could certainly have been done with the MUA as an evaluation, and in fact, just a couple of days ago I happened to say, gee, I never actually compared the known testing at all. Test Strategy 6 is pretty close to that. Test Strategy 6 is basically the surface-based testing, and most of the strategies scored better than that. So again, on that sort of absolute measure, we do have an absolute positive benefit for testing, because they all ranked higher than--so many of them ranked higher than six. Therefore, they are, you know, testing is better than no testing. Testing down is better than surface-based testing.

I even, with some speculation, threw in an absolute zero testing, although on some of the scales you have to take some guesses, and there, all eight of them scored better than that, but that's sort of a hypothetical sort of conduct. What's the phasing potential in a zero test?

DR. DOBSON: Yeah. I guess I did want to also say with respect to the VOI model, as I mentioned in several of the various talks, we recognized early on that we could have expanded the VOI model, and it was recommended by some people. We chose not to, and the reason that we chose to do the MUA, when we finished the first phase of the VOI, instead of modifying the VOI, was--I'm not sure I could resurrect all of the reasons, but we had a lot of discussion about what the appropriate next step was and whether we ought to revise the VOI model or go with the MUA, and we chose this way, and we think--one of the reasons was that it gave us two completely different approaches to the problem, and we thought there was possibly some value in that in terms of what we were going to learn.

DR. DEERE: But isn't your next iteration going to have to be pretty well tied in with the accesses that--

DR. DOBSON: Oh, absolutely. I think, you know, the critical thing is what we do in terms of analyzing as we go through design. I mean, we now have a recommendation for a configuration, and we've put it together with the ESF alternatives group, and we're--now we need to start analyzing what the configurations that we're actually considering are going to do and are going to look like, and, you know, we've already learned, I think, a tremendous amount from the sensitivity studies that we've done here and in the ESF study, and you'll hear some about that tomorrow. So we've learned what some of the important factors are, and what affects your decisions, and hopefully, we won't forget what we've learned as we go through the process of putting the design together, because the idea, obviously, is to collect all of the most relevant information and come up with a final design that's the best one we can get.

DR. DEERE: Well, we think this has been a very useful part of the study.

DR. DOBSON: I think so.

DR. DEERE: But it's ready to incorporate, now, with the results of your alternative shaft study.

DR. DOBSON: Well, that's right, and from what Leon said, when I talk about lessons learned, one of them's going to say: Where do we go from here? Let's make sure that we keep it together.

DR. DEERE: Well, let's--should we move on, then, to the third? Thank you very much. I know there probably would be other questions, but we'd like to let John get started.

DR. DOMENICO: Excuse me, Don. Would it be--in view of the hour, quarter to twelve, would it be best to take--come back 15 minutes early so we can have a continued presentation here?

DR. DEERE: It might allow us to get in and get our lunch a little faster, because we'd beat the crowd if we go right now, but I'd leave it up to you.

DR. DOBSON: Well, it's your option. I mean, Jack's set for basically 20-25 minutes worth of presentation, with some questions, and whichever way you--I mean, if you want to have --if we run longer, then we may encroach on the stuff this afternoon, but we're here until you're satisfied, so...

MR. ROBERTSON: Fine with me either way.

DR. DEERE: Well, you know, last time we had a meeting we broke just a little early and we beat the crowd for lunch and it was no trouble whatsoever, and sometimes when we run over 10 or 15 minutes, we get caught up in getting served. So I would suggest, if it's okay with everybody, we come back at a quarter to one and then we'll let you--and we'll have plenty of time with you, if that's okay.

Thank you.

(Whereupon, a lunch recess was taken.)

A F T E R N O O N   S E S S I O N

DR. DEERE: May we reconvene, please.

Okay, Dave, I guess we are ready for you.

DR. DOBSON: Okay. Jack Robertson is the next speaker, and he is going to talk about some of the saturated zone models that we used during the Calico Hills study.

MR. ROBERTSON: Well, Dave is right, I'll go into an area that didn't get the center of focus, of course, because our group was focused on the Calico Hills unsaturated zone.

But the reason the saturated zone took some attention and became a prominent sub-piece of our effort, really relates to the multiple barrier ideas that Dave got into earlier. And that was really a fundamental guiding factor in the Calico Hills Group, was not to consider the Calico Hills unit out of context with the entire system, but to look at its role as much as we could in the entire system. And that meant some consideration of the barriers above the Calico Hills, as well as, the performance, expected performance or the expected role of the saturated zone.

This meant some consideration of the stated knowledge of the saturated zone. There hasn't been a lot of characterization effort on the saturated zone in detail, although over the years there has been a number of good efforts done. There is some good data that's quite useful that I'll get into a bit.

Recognizing that there wasn't a lot of intensive detail quantified information on it, we wanted to use realistic best estimates of what the saturated zone pathway is like, how we as a group might expect it to perform with some degree of the uncertainty involved. We wanted to make sure that all the significant factors might be considered such as sorption and matrix diffusion and that we would approach its role from a radionuclide release perspective as we had in the other parts of the system, rather than travel time or some surrogate like that. We felt that it was really the amount of material released over 10,000 years was the issue here--using the measurement we were using for performance, although travel time plays a factor in that, that wasn't our primary role or factor that we were using.

We evaluated the expected performance of the saturated zone in a very crude, semi-quantified manner at different levels of confidence among the group to 99, 90, 50 and the 10 percent levels.

We began with considering an influence diagram, just to make sure we were considering the major factors that really related to transporter releases through the saturated zone. This came down to two major categories, the retardation factors and the flow factors. This says velocity distribution, but it's really more than that. It is really flux and velocity combined, so don't think of this as strictly velocity or travel time, it really incorporates both factors as you can see here. The flow distance, the effective porosity is a major factor here and the ground water flux which is pretty crucial. And then up in the retardation factors that influence this process are sorption, matrix diffusion effects and perhaps other chemical retardation measures.

We considered several aspects of this and kind of agreed on some general characteristics that we had some consensus of opinion on at least, and that is, one, that there is probably going to be a pretty flat gradient based on the available data and the SCP and other documents over the pathway, the expected arrange of pathways. The gradient is quite flat and in fact it is so flat it can't be measured, precisely. That there is probably an upward gradient from the carbonate aquifer into the overlying volcanic rocks, this might have some influence on keeping the expected flow paths in the upper volcanic rocks. That there would be some--that most of the pathway would be through the Calico Hills nonwelded zeolitized unit, although that is not very well known. I'll show you a cross-section that demonstrates that that might be expected, but the detailed stratigraphy along the pathways is not well characterized yet.

Fractured permeability was recognized as a significant role in controlling the flow parameters and the distribution frequency of fractures, we recognize is a major factor although that is not well characterized on a large scale yet either. The hydraulic conductivity probably declines with depth. There is some indication of that from field work and as a general expectation in these types of environments. That might play some role in the process.

The volcanic units have a high expected bulk porosity, particularly the Calico Hills unit. If it does occupy most of the full path its bulk porosity is up around the 20 to 40 percent range.

There's probably high ion exchange capacity in the zeolitized Calico Hills, which is expected to be a major part of the flow path. And that much of the porosity, not only the fractures, but much of the bulk porosity might participate in the flow, either both through hydraulics and through matrix diffusion. Both water and solutes diffuse in and out of the matrix even if it isn't heavily participating in the gradient driven flow.

And that there would be expected to be small effects from climatic change, although they could be larger than we expected. Climatic change could certainly raise the water table, change the gradient somewhat. It would not change retardation processes significantly, we didn't think. And, we didn't expect that it will change the gradient largely.

Just a cartoon to illustrate some of the concepts we were trying to incorporate in our model. And we used a model here in this title very loosely. It is not any kind of a numerical model we used. It is more of a qualitative professional judgment model and we tried to use the more realistic information we could.

But in these fractured volcanics, the highest part of the hydraulic conductivity is generally through the fractures, but there is a lot of porosity and water stored and moving in the matrix also at a much lower hydraulic conductivity. And there is exchange through matrix diffusion and hydraulic gradients between the fractures and the bulk matrix. We can't quantify that process for the site yet, but we know that at other sites it can be demonstrated this is very effective in relatively slow moving ground water systems and has a very large influence on how solutes move. And even if you are only looking at ground-water, at the movement of water itself one can say, well water is moving much faster through the fracture than through the matrix which is true at any one moment, but if you were somehow able to label the water molecules and paint the ones in the fracture red and the ones in the matrix blue, you'd find that they were interchanging also through diffusional processes. So if you were able to track molecules of water some of the molecules are moving faster in the fractures and some are moving slower in the matrix. There is a dynamic exchange.

It's very difficult to characterize--that's what makes the concept of ground-water travel time such a problem to all of us that are having to work with it. It's really difficult to pin that down in terms of a meaningful number in terms of the regulatory requirements. So we certainly didn't want to get bogged down in that in this process.

DR. DEERE: Jack, let me comment on something. I like that diagram. Could you put it back on again?

MR. ROBERTSON: Sure.

DR. DEER: I think it is just exactly what we see. In a lot of hydroelectric projects and almost every one of them, the abutments, which hold up the dam are not saturated. So every time and there have been thousands of these and many of them in tuff, every time we raise a reservoir, we are getting flow through an unsaturated medium. And it's been of interest and concern and they have been monitored for years and years. But the amount of water that really gets away and takes very much with it is usually very, very small, unless we have fractures in it, and then there can be some very important flows bypassed around the abutments or outside the grout curtain or the grout curtain gets dissolved at a later and we don't get a lot of flow through it. But, the interesting thing is, in this area which is slowly becoming saturated because we have peisometers in a great number of the areas, and we find that the regional water table including in the matrix is becoming saturated with the peisometric levels rising several feet to several meters per year, that the lead is always taken by fractures. In other words, the fractures may be carrying the water several hundred yards out ahead of the matrix, so it really appears that the lead is the fractures and the fractures have the access to the reservoir so they are bringing in fresh water.

Now in some of the desert environments or at least semi-arid environments, we have lots of salts in the matrix of the rock. And so the water that's getting into the matrix and moving through it, if we can take the matrix and grind it down, squeeze out the water, which we have done in some cases, we have found extremely high salt content. And yet when we sample the water that's coming out of the fractures, we have a very low salt. And when I say salt I mean all kinds of salts.

So the majority of the material that's being carried out of the environment, the concentration is so much higher in the matrix water, but the amount that really gets out and moves anyplace is the dilute suspension which is in fracture. In other words, when we first saw the salt content of some of the fractures we said, oh, gee we are not losing very much. But when we took the flow rate and multiplied it by the very low concentration, we still were getting much more material moving downstream through the fractures than we were out of the matrix.

And it is not uncommon that when we put drain holes in downstream trying to control this, that if the intersected fracture, we really can drain water and lower the peisometric, but if we put it into the saturated matrix, if anything comes out it so slowly evaporates before it drips out of the pipe.

MR. ROBERTSON: I think it's good to hear some practical perspective. And that's one of the things we are wrestling with here. We don't have any really good quantifying information on the site-specific behavior of these rocks. We know that matrix diffusion is quite dynamic in some rocks and in some rocks we've seen in that experience where the matrix water is much saltier than the fractured water. It depends a lot on many, many factors, but there are skin effects along the fracture lining that may be a factor, the spacing of the fractures is important, the contrast in permeability between the matrix and the fracture system is important. All of those come into play and there have been some numerical games played at where you set these problems up hypothetically on a model and assign various of these contrasts and play some games and show under what conditions you can get a very significant process.

If the bulk flow overall in this material is fairly slow, and we've as a group thought it was over the 5 kilometer distance, the slower the flow the more chance there is for interaction between the fractures and the matrix, even though there is a large contrast in permeability. If you are pushing water under a high head, say from a dam through fractures, then there is much less time, contact time, for the slope process of diffusion to work. So you'll see much less pronounced effect there, the higher the flux for the system is.

I'll expand on that a little bit as we go, but those are good points, I think.

We do have some concept of what the flow path looks like. Let me just show this one first. This is a figure out of the SCP that gives you an idea of the gradient. We see a so-called steep gradient zone up here. The flow path of the repository's southeast boundary here is generally along in this direction to the southeast. You can see there is no ground-water peisometric contours here. The gradient is quite flat. If you look at USW H-4 in J-13, there is about a meter and a half of head difference along that distance of approximately 5 kilometers. If you look at another well here, WT-2 here and WT-3, I think there is about a half a meter difference there.

These are down in the uncertainty measurement ranges of measuring these water levels. When you are raising the water level 1,000 meters deep, it's hard to get it plus or minus a meter. So we are in the high uncertainty levels of determining these gradients, but they are flat relatively across that zone.

The next cross-section I'll show you is this AA cross-section--

DR. DOMENICO: Jack, that is, what do you call it, a composite peisometric surface, correct?

MR. ROBERTSON: Right. That's another problem.

DR. DOMENICO: Are most of the measurements making up that composite come from the carbonates, or from any specific volcanic unit? Do you know that at all?

MR. ROBERTSON: I couldn't answer that question very knowledgeably, but there are probably people in this room that can.

DR. DOBSON: There is only one that's in the carbonates, and that's UE-15 p#1. That is an open hole all the way to the bottom, so it's a composite head.

DR. DOMENICO: So all the rest are in the volcanics, however many there are, four or five or six or whatever?

DR. ROBERTSON: Yes. Most of them are composite in the volcanic, they are not discrete interval is my understanding.

DR. DOMENICO: Yes.

MR. ROBERTSON: If we look at the cross-section along that AA prime, this is the Yucca Mountain area, the proposed repository area outlined in red. That well at J-13 is over on this end.

Looking down gradient, the water table is represented here as this kind of dotted, double dotted line going across there (indicating). This grayish brown unit is the Calico Hills that is in the right position to occupy most of the upper flow path along the saturated zone. So, one can estimate that most of the flow path along the saturated zone might be in the Calico Hills zeolitite. Although it is not certain what degree the zeolitization is along that whole pathway.

I notice John Czarnecki in the room. He'll recognize this, I'm sure. This is the--the USGS did a simulation of the regional flow pattern in the saturated zone with a numerical code and determined this type of flow pattern from their modeling efforts to be likely. It's a composite model, so it doesn't have discrete layers in it, but it tends to represent the general regional flow pattern.

As you can see from the south and eastern side of the repository, the flow path curves around. I've sketched in roughly the 5 kilometer accessible environment line. That's approximately where it might be. There is some flexibility in that. So you can see the type of pathway that might be expected in the saturated zone. Actually the length of that pathway, if this were the true pathway, is longer than 5 kilometers because it doesn't follow the most direct route to the accessible environment.

There is a number of things we considered. We did some back-of-the-envelope calculations and so on, discussed at length the known properties or expected properties of the Calico Hills and the related volcanics above and below it, and came out with an elicitation of probability levels on the expected release reduction that the unit might provide from the group. We had six members, six of the technical members. And it came that overall that we thought that the saturated zone will provide--has the potential of providing a pretty significant barrier if the little bit we know about it turns out to be representative of the system as a whole.

Now, these are, as was pointed out earlier, these were not conservative estimates. We tried to use our best guess of what we thought the system was like.

DR. DOMENICO: Well, reducing what, Jack? Reduction factor--what are you reducing?

MR. ROBERTSON: Okay, this is the degree of reduction of what is leaving the Calico Hills and entering the saturated zone, how much that is further reduced in its pathway over 10,000 years to the accessible environment.

DR. DOMENICO: How did you get those numbers? Is this based on model calculations, or--

MR. ROBERTSON: No.

DR. DOMENICO: No.

MR. ROBERTSON: Like we got all the other numbers in this process, the group of technical people sit around and guess with a lot of informative discussion about--and there were some calculations involved, but no sophisticated modeling done. We looked at--and I'll get into that a little bit, sort of the background that goes behind these numbers.

But it's basically looking at some semi-quantitative information, trying to factor that into our expert judgment and then soliciting from each person what he thinks is going to happen and why. And, then, there was a pretty good diversity, as you can see. There is a fairly big difference between geometrical and arithmetic averages which means that there is a pretty good spread in the numbers between people. But overall, they still felt that the 99 percent confidence level, everybody thought, you know, the aggregate felt that you are going to get a factor of 20 reduction provided by the saturated zone, and if you are getting down into, well enlightened, essentially nothing is going to get out of the saturated zone.

DR. NORTH: Could you give us a little bit more idea of the spread, the diversity within the panelists and the story, essentially. What kinds of things might lead to the 99 percent probability level? I assume looking at those two numbers you had one or two individuals with very high, very little reduction. And then others in the group that felt that you were going to get in the order of a factor of a 100.

MR. ROBERTSON: I don't have the table of numbers with me: I don't know if anybody here does; of the actual individuals; and I can't recall the degree; but you can kind of guess if you know the difference between geometric averages of the exponents these are the regular arithmetic averages. It's not--everybody thought it was going to perform well. When you get down in these levels, here's an order--five orders of magnitude between the geometric mean and the average, that means that there were probably a couple of individuals that said, it is never going to be better than 1,000, factor of a 1,000 reduction, and there was some that said essentially nothing is going to get out 10^{-12} or something. So we get down in these zones, and there is a lot of spread up in these zones and there wasn't so much spread, is my recollection. And I'm sorry I don't have those details off the top of my head, but they are available in the record, I know that.

DR. NORTH: What I'm interested in is what is the conceptual model that would lead to a very low reduction factor? I think about an underground river, is it that simple?

MR. ROBERTSON: Okay, just to--some members tend to be more conservative than others, that is one of the factors. The SCP is a very conservative travel time calculation. They assume all the flow is only going to be in fractures and fractures only occupy 4/10,000ths of the volume. So you are pushing the entire flux which is probably better known than anything else in this, and that is pretty vague too, through a very, very small percentage of the volume. That is really zipping it through a small channel, essentially, through fractures, that bulk matrix participates--has no participation in the flow and that there is no retardation mechanisms involved.

There is a group that felt that is really way overly conservative and not realistic in terms of how the transporter radionuclides are going to occur in this environment. Although, there is a lot of uncertainty regarding the true, effective porosity, the percentage of the porosity that is really transporting these things and carrying the flux of water, that's one of the biggest uncertainties. There is a lot of uncertainty on regional scale hydraulic conductivity. There are a couple of measurements of hydraulic conductivity at Calico Hills and underlying units, but they are pretty far between aerially, and it is hard to draw firm conclusions from those.

There are some tectonic features that might be present in the sub-surface that provide more fractured, permeable channeling zones through the system, than other zones. We assumed that the pathway pattern is going to be fairly uniform, but it wouldn't matter too much which pathway you have. This one is not going to be 1,000 times faster than this one. Well, that may not be true if you assume a little more uniform--

DR. NORTH: Well, I'm interested more in what are the scenarios for the low-reduction factors as opposed to what is called the expected 50 percentile scenario?

MR. ROBERTSON: Okay.

DR. DOBSON: Can I make just one suggestion? I think it might be worthwhile for Jack to go through the rest of the view graphs which explain some of the technical assumptions that I think you are asking about. And I think maybe if we kind of come back to this at the end and we could talk specifically about things like, you know, the most conservative and the least conservative scenarios and things like that, it might be better to do it in that sequence.

MR. ROBERTSON: I think it is true and I promise to get back to your question as we go or at the end. If I haven't, please get on my case here and I'll get to it.

Okay, how do we get to those numbers and what causes some of the divergence among the members is reflected partly in the uncertainty of the information on the parameters. The velocity distribution is controlled by hydraulic conductivity, effected porosity, in gradient. We have some measure of the gradients. We have some measures of the total porosity because there has been quite a few reports collected and their total porosity measures. But the effective porosity, that portion of the porosity that is really acting to conduct the fluids and the solutes is a real iffy parameter that's very hard to quantify without some field tests. And even then it's got difficulties.

But here are some ranges in measured porosities for instance. In the Calico Hills zeolitic, on cores the measurements run in generally the 20-40 percent range. And the vitric portion of the Calico Hills is a little more porous, around 40 percent plus or minus. There's been some estimates of saturated effective porosity and small scale effective porosity and there are some references in Scott Sinnock's report and some others that draw some estimates of this number, but it still not an appropriate number, not necessarily a representative number on a 5 kilometer scale that we are dealing with here. And that is where a lot of our uncertainty comes. If you use the most conservative assumption like was done in the SCP and say only a very small portion of the permeable fractures are carrying the flow, then the porosity number comes out four orders of magnitude lower than this. If you assume that some of the bulk porosities participating in the flow, then you can change it several orders of magnitude. And that was our position that--

DR. DOMENICO: Before you leave that, one could argue that if your concept is fracture flow, none of these values are worth anything. I mean, they are all too high.

MR. ROBERTSON: None of these are worth anything?

DR. DOMENICO: Either ones. Take either one, the so-called saturated effective or, if it is truly fracture flow, we are dealing with a very, very small effected porosity in large velocities.

MR. ROBERTSON: Well, we know we have fracture flow, but we also know that the matrix is saturated water too. And we know you can set a piece of rock in a beaker in the lab and get the stuff to diffuse into the matrix, so we know that happens. You can't rule that porosity out entirely because that is one of the factors that controls matrix diffusion effects.

DR. DOMENICO: Yeah, but what I am saying is, if I was on your panel and you gave me those choices of effected porosity, I would come up with certain thoughts about this unit as a transporting medium. But if you permitted me to consider 10^{-3} which might be more appropriate for fractures I would change my thoughts immediately and say this medium has no saving grace in terms of hydraulic characteristics, in terms of speed and movement.

MR. ROBERTSON: Okay. Well, keep in mind though that even in a fracture of rock, most of the flow was flowing through fractures, the water molecules are still going in and out--most of the water is in the matrix. And the water is still exchanging between the matrix even though there is a higher flux in the fracture. So, if you are trying to track the velocity of molecules, you can't just judge it by the velocity of the water in the fracture. We did consider those very low numbers of fracture porosity. We looked at the SCP numbers and we looked at other estimates. That hasn't been measured in the site either with any degree of confidence.

Another degree of uncertainty is in the hydraulic conductivity which is one of the direct parameters affecting flux and velocity. There have been several measurements on matrix, on cores of course, which don't tell us about much about large-scale, 5 kilometer scale effective hydraulic conductivities, but they tell us something about the matrix, hopefully.

DR. DOMENICO: Well, what's the units?

MR. ROBERTSON: These are centimeters per second. I'm sorry these units aren't on there.

On a bulk basis, there's been a few measurements in well tests even in the Calico Hills. And they tend to be falling in the range of 10^{-3}, 10^{-4} roughly, but a few measurements. Not enough to be representative of the whole 5 kilometer range. These are also consistent with some of the other conclusions drawn by the Czarnecki-Waddell study, and others that have looked at kind of gross conductivities or transmissivities that would explain the expected flux distribution on the water balance basis.

So one of the points here that is coming out of this, of course, is there's a contrast in even a small-scaled bulk hydraulic conductivity and a matrix hydraulic conductivity of a few orders of magnitude. That's what would be expected in these rocks. So that is still--that means that the reason the bulk conductivity is higher is because you've got some fractures in the bulk that are raising the hydraulic conductivity.

That still doesn't mean that all the water is moving from the fractures, that just mean that's where it's moving the fastest and it's still exchanging with the matrix. And solutes are still exchanging with the matrix, particularly in a slow-moving, long pathway environment.

Now we did some back-of-the-envelope travel time calculations and I'll get to those in a minute to show you again some of the uncertainties on this. Looking at travel time alone, which of course is a regulatory issue here, that has been looked at before, as I mentioned in the SCP for the 5 kilometer distance, we are using the most conservative number we felt you could justify for fracture flow only, you come out with a 170 year travel time through a fracture system occupying only 4,000th's of a percent of the volume.

If you--Czarnecki and Waddell said, well we don't know what the effected porosity is either, but it might be a range of this to that, whatever those numbers were, and they said they could get a travel time in the range of 100 to

20,000 years depending on what kind of assumptions you want to make about effected porosity. That's just travel time. That's an average travel time.

DR. DOMENICO: What was the panel's estimate?

MR. ROBERTSON: The panel did not come out--this is the panel estimate of travel time. But we did some back-of-the-envelope calculations and this is the kind of numbers we could come up with, but we are not saying these are the numbers. Say well, if you use some of the best guesses and some of these concepts of travel time including participation of some of the bulk matrix porosity in the flow through hydraulics and matrix diffusion, you come out with these numbers. Now if the gradients are in this range based on available numbers, you can scale those off the map. How good they are, I don't know, but at least there is some data accumulated to support those numbers. The hydraulic conductivities tend to fall in this range, .1 to 1 meters per day. And, if you assume that--if we had a total porosity of 30 percent and we allowed half of the total porosity to participate in the flow somehow, let's take this typical value here, typical value of K=.2 meters per day, that's in the range, the typical value of gradient, 2 x 10^{-4}, a typical effected porosity, this is sort of a surrogate way to account for some matrix diffusion, saying half the porosity is participating in the flow and retardation had some affect.

We come out with a typical number of 50,000 years.

And we are not saying that's the travel time. We are saying you can come to a reasonable calculation that shows that that is--that you can get a travel time like that using reasonable numbers.

DR. DEERE: That travel time is for any water that got into a fracture and it has to go to a fracture and then through the matrix and then back to the fracture and into a matrix and back again?

MR. ROBERTSON: This travel time would be an average travel time--that would be an average travel time of water going from here to here (indicating), not looking at the fastest pathway.

The travel time would be a very distributed function. And we are looking at sort of the central value of the travel time. And that is one of the things that is wrong with--that's difficult about travel time is the regulatory concept because it is a distributed parameter and some of the molecules are going to get out there, and even if the average is 50,000 years, some of the water is going to get there in 10,000 years or 5,000 years.

DR. DEERE: Or 170.

MR. ROBERTSON: Yes. And some molecules probably will get there at 170. Is that important if a few molecules get out there in 170 years?

DR. DEERE: That's the question.

MR. ROBERTSON: Especially if they don't have any of the radio isotopes with them.

DR. DEERE: That's the question.

MR. ROBERTSON: So, what we are saying is, yes, some of the water--the average travel time may be in the thousands of years or tens of thousands of years, some of the water is going to get there much faster. If it is carrying some dissolved contaminates with it or radionuclides, the cationic species are going to be heavily retarded because of the exchange minerals present along that pathway. The anionic species like iron or technetium perhaps, will enjoy some kind of retardation through matrix diffusion. We don't know what that quantity is. We believe it would be significant. That's all we could say. Somebody is going to have to do some pretty aggressive testing to look at the range of significance of matrix diffusion in retarding solutes in this process. But, we felt it would be significantly effective in causing more reduction of the releases to the environment. Over a fairly slow moving hydraulic system, which doesn't have a lot of flux in it, over a long pathway of 5 kilometers, that is an opportunity for a lot of matrix diffusion. Whether or to what degree is will be effective, we don't know, but we felt that it would be significant. And that's the factors that drove our numbers.

DR. DOMENICO: But Jack, I'm still curious, based on the panel and their thinking, do they feel that the saturated zone is more of an invected barrier or a geochemical barrier? By that I mean if you threw out all the geochemical aspects, retardation, matrix diffusion et cetera, and you took that range of travel times that you look at there, does this saturated zone do anything for us in terms of releases or do you need that geochemical aspect?

MR. ROBERTSON: Well if you are calling matrix diffusion geochemical--

DR. DOMENICO: I do.

MR. ROBERTSON: I would say you need it, but I would also so I don't think you could deny it.

DR. DOMENICO: I didn't say whether you could deny it. I'm just trying to assess its worth in terms of the importance of geochemical barriers in this project in the saturated zone or the lack of such importance.

MR. ROBERTSON: I view--you know, geochemical barriers is like geochemical reactions and ion exchanged--I have used matrix diffusion as a physical process. That's a movement of molecules.

DR. DOMENICO: Well I don't believe that. I believe it is a chemical driven by concentration gradients, but how about the interaction of retardation by zeolites? I mean that does the same thing as matrix diffusion. Is that incorporated in here?

MR. ROBERTSON: Yes. Retardation was factored into our analysis.

DR. DOMENICO: Retardation specifically by this--

MR. ROBERTSON: Qualitatively, in the analysis, the group was asked to consider whatever you think is important, and generally there was a feeling that because of the availability of ion exchanged minerals in this pathway that that would be an important retardation factor for the sorbable isotopes.

DR. DOMENICO: But, you did say the geochemical aspects are important to your conclusion that you may have some significant--the saturated zone can contribute to this problem, as far as geochemistry?

MR. ROBERTSON: Right. Yes, that's a fundamental part of it, yes.

DR. DOMENICO: Okay.

MR. ROBERTSON: I believe that--I think I speak for the whole group on that, but that is certainly what I believe.

DR. DOMENICO: I only ask because I've only been doing this for ten years now, and I've heard people say at certain times that we've taken no credit for the geochemical aspects, so I just want to ask and see how the program is changing.

DR. DOBSON: I just want to clarify what Jack said or to add one other perspective, and that is that as he noted for the sorbable species, the cationic species primarily, mineral distribution and things like where the zeolites are may play a very large role, but for the soluble species, like the anionic species like technesium, I think Jack is right. Probably the way in the group we talked about it, we didn't really think of matrix diffusion as a geochemical process, but in the way that you describe it, if you include that in what you would call-- and concentration gradients could be argued to be a geochemical process, so to that extent it certainly is, but if you allow that difference in our definitions, yeah, we think that the process of matrix diffusion would certainly be important to our conclusions with respect to retarding soluble species.

DR. DOMENICO: Technetium who theoretically has the distribution coefficient of zero theoretically, it has been noticed, but technetium should partake in matrix diffusion.

DR. DOBSON: Right. Exactly.

DR. DOMENICO: Which is just another retardation phenomenon.

DR. DOBSON: Exactly. I think our views are concurrent.

DR. DOMENICO: However, if the velocity gets too large, matrix diffusion becomes tremendously ineffective, so you have a competition between rates always going on.

DR. DOBSON: Yes.

MR. ROBERTSON: And that's, getting back to Warner's question of regarding what were the scenarios that spread these numbers. Those were the kinds of discussions we had.

Somebody didn't believe that, well maybe fracture coatings would really inhibit matrix diffusion for some of these species. And that's possible that in another scenario that this is not a uniform hydrogeologic environment. There are some linear structural trends in that part of the world, and maybe there are channeling zones within the general pathway. There may be a non-obvious fracture zone or path fault zone going down through part of this pathway that would carry 20 percent of the water, and the other 80 percent would be moving very slowing, 20 percent would be moving ten times faster.

DR. DOMENICO: And the zeolites can be embedded in the matrix and with the flow taking place in the fractures would never come in contact with those zeolites and so you could--

DR. DOBSON: Well, we do have enough data to know that the zeolites are both in the fractures and the matrix.

DR. DOMENICO: The cores do show that.

MR. ROBERTSON: So, zeolites are not the only sort of mineral.

DR. NORTH: It would seem useful to look at some analog areas, N-tunnel, for example. I'm not sure how good an analog it is, but a number of us have been in there and it's clear that there are gallons per minute coming through some of those zones and in nearby areas, the rock appears to be completely dry.

MR. ROBERTSON: I think that's a good point to bring out.

The problem--and I've been in N-tunnel, too, and I think there is some great value in looking at those analogs stated in writing, but I think the problem in using that analog for this problem is the flow regime is so much different. The gradient is primarily what is different--is one of the things. You've got practically a one-to-one gradient in Rainier Mesa downward. And here we are talking about a gradient on the order of 10^{-4}, so we are getting four orders of magnitude different in gradient. That has a great effect on the role of matrix versus fractures, the general rate of flux to the system which may be driven by the gradient. We could do some experiments and that--

DR. NORTH: That's a story that would be very interesting. What you are saying is that essentially even if the fractures are there, they don't make any difference because there is no gradient to drive the moisture--drive the water through those fractures. It will go into the matrix in the absence of the gradient.

MR. ROBERTSON: Well, we know we do have a gradient. We know that we have flux to this system, because this system is a dynamic system. Water comes out at the bottom end, so we know water is moving through this system. We just don't know much about the details of how it is moving through on that scale. It is looking at the scale of N-tunnel in Ranier Mesa could tell us some things about that, particularly if we could do some controlled tests where we could do a matrix diffusion test for instance on the scale of meters in one of the tunnels and that is one of the kinds of tests that I would endorse doing if feasible, there or somewhere, where we could force some tracer fluid in a controlled system and run it through the higher permeated fractures and see how much matrix diffusion effects we are getting.

We still won't be able to assimilate probably the slow speed of this system, because we'll have to do the accelerated tests on this as we only do in these kinds of things, but I think there are things to be learned at Ranier and Mesa and other analog sites, particularly if we can do some tests in there of looking at--it gets back to the question of matrix versus fracture regime in the flow system, which was one of the driving factors that we don't understand about the Calico Hills and that is one big reason for our spread of numbers there. We don't understand how the fracture versus matrix system is going to work, particularly in the unsaturated zone there, which is more complex than the saturated zone.

DR. NORTH: Well, from the performance assessment aspect, I'll give a very simple value of information calculation. If we could assure ourselves with some data gathering activities, that the kinds of scenarios that led the members of your group to come up with the 1 in 20 geometrically characterization of the 99 percentile can be ruled out, because we don't have those situations now that we've had the opportunity to get some data. Then you potentially can draw some rather strong conclusions with respect to the protection offered by the saturated zone. And many of us on the Board have taken the position, maybe we ought to be putting more emphasis on the saturated zone as opposed to Calico Hills, and the same issue about value of information. If you take the, I'll call it multi-attribute utility perspective, that there may be a lot of value in the scientific confidence area, perhaps there are some opportunities here which would be very valuable.

MR. ROBERTSON: I think we agree with you. I think we came to that conclusion too that the saturated zone offers the potential of a very significant barrier, that's probably not been given its appropriate level of attention in most of the process--so much focus has been promoted on the unsaturated barrier concept, which is good. That's good too, but the saturated zone pathway has sort of been left as the ugly step-sister and not been given too much attention.

And, it may turn out that yeah, there is so much conservity there we can never be confident that it is going to provide these 10^{-4} additional level of protection, but I think it is worth--it is in some ways easier to do some testing in a saturated zone. You can do pumping tests and alteration of tests and things like that and you don't have to do them in the block. You can do them away from the block. And there are lots of things you might do that relatively speaking can tell you a lot about the saturation in the system. And there are some planned, some good planned tests in the saturated zone. But, it might focus, maybe there might be worth re-visiting the testing plans of the saturated zone to see if it is getting its fair share of attention. We are not saying it is or it isn't, but it certainly deserves some serious attention.

DR. DOBSON: Yeah, I wanted to sort of reiterate the last point that Jack made, which is that we may--the saturated zone may have gotten something of a short shrift in terms of publicity in terms of its capability to be an effective barrier, but we haven't ignored it from a testing perspective. And we can talk about whether we might want to do more tests, but we haven't gone into this context talking about testing program that we've already planned, but we do plan things like tracer tests and pump tests in the saturated zone. And I think we've always believed that it was important to do that. And it may be that we may wish to consider at some point expanding that program, I'm not sure. But, I would just hasten to say that it is not our position that we shouldn't rely on or test the saturated zone. We've planned on doing that all along.

I just want to add one other thing because you didn't put up that one view graph. I had to throw in the geologist's perspective a little bit on the estimates of ground-water travel time. Not only are the model results that Jack talked about, I think defensible in kind of the general ballpark estimates that he mentioned, but there is another way to take a look at the problem, and that's in trying to get a handle on how long the ground-water that's out there now has been setting there and particularly there is a summary of C-14 ages which to me are actually rather remarkable in that they are very consistent with the estimates that were made by John Czarnecki six or seven years ago in terms of total residence times for ground-waters in the saturated zone. And it doesn't prove anything in any ultimate sense of the word, but the fact that these numbers in the ranges of a few thousand to perhaps a couple--up to a range of 20,000 years are consistent with the hydrologic estimates, I think that adds some level of confidence to your feeling about the--

DR. LANGMUIR: Dave, I've been reminded to remind that those are apparent ages.

DR. DOBSON: Apparent ages, that's correct. And I am not trying to prove to anybody--I am reminded. I am reminded.

DR. NORTH: How well could we do at the 20,000 year limit there given C-14. I mean could it be 50,000 to 100,000 years?

DR. DOBSON: Well, I don't even know--I'm not sure if we've got any Los Alamos people here today. I do know that

Ted Norris at one point had done some estimates based on Chlorine-36 that were much older than that, but I don't know how to interpret those numbers either. So, the range is from zero to several hundred thousand years, and picking out the expected value out of that range is a little difficult, but--

DR. LANGMUIR: What you've got to do here is assume that the lowest number, could be, zero, could be present. So you are looking at a range from zero in each case.

MR. ROBERTSON: Yeah, there is--you know, as most of you know there--you have to use a lot of caution in saying much about what these numbers really mean quantitatively. But at least they are not inconsistent with some of the things we were looking at in terms of travel times. And these numbers probably represent blended water for one thing, a mixture of old water with some modern water mixing, so you get some kind of a funny blend of age out of that that isn't a true age, plus there's some geochemical factors going on and you are stretching the limits of the methods and so on.

At least these numbers are indicative that there may be some pretty long travel times. These are not--don't confuse these with travel times from the repository to the accessible environment. These just happen to be some apparent ages on water that is collected at different spots in the area. They represent a complex history of that water that we don't know.

DR. DOMENICO: I don't think you can lose with the saturated zones even if the flows are ten times more than you assume they are, they you can always invoke dilution to help dilute that small voiced stream that is going to be dribbling through the unsaturated zone. So, you really can't lose. Do you agree?

MR. ROBERTSON: Yeah. I think it gets you to some good points. And it certainly is worth getting serious attention.

Just to summarize and I'll get down, we think the low gradient, the relatively low permeabilities plus the high porosity causes slow travel time in general on an average through the expected pathway zone. We think that matrix diffusion and ion exchange can cause high retardation effects; can be expected to cause high retardation effects, in a relative sense.

We think the release factors can be in the order of several orders of magnitude if some of these things are representative for the entire pathway.

The saturated zone is a very significant potential barrier. And finally, that the effect of having a saturated zone, if it truly is up in that level as a potential barrier, diminishes the relative importance of the unsaturated zone in the Calico Hills unit so that not all the eggs have to be placed in one basket in other words. Again, in the multiple barrier concept, the redundancy prevails.

DR. DEERE: When you are speaking of the saturated zone here, are you talking about all of the volcanics down below or just the saturated zone within the lower part of Calico Hills?

MR. ROBERTSON: We are talking about the way we've talked about this is on the scale of this drawing, the Calico Hills is basically the unit right in here. Once you hit the water table which is this dotted line, that's the new zone, that's the saturated zone which we switch over from one to another. And from there on the flow is--

DR. DEERE: Still, we're in the Calico Hills?

MR. ROBERTSON: Some of it, yes, it stays in the Calico Hills. It probably gets into other units at different places. This diagram indicates that most of that, at least--it's saturated all the way down here. These are the carbonate--well I don't know whether the carbonates are on here, but these are older tuffs down here. The carbonates are maybe deeper than this. But, we don't know the exact pathway--if a release were to occur from the repository, because there isn't a lot of recharge occuring in this system, a lot of downward gradient, in fact the gradient tends to be upward, it would reach the saturated zone and pretty much stay lateral in the upper part of the volcanic rock.

DR. DEERE: Because there is an upward hydraulic gradient at the contact of the carbonates.

MR. ROBERTSON: At the few places it has been measured, it's been upward in this area. I may not be familiar with all the data, but that is my understanding.

DR. DOBSON: Well, UE-25 p#1 encountered dolomites I think at about 1,000 meters, is that right, oh, and there was an upward gradient. It probably went to a faulted contact and there was a positive gradient from the carbonates into the tuffs. But that really is the only measurement that we have in this vicinity. We don't know if everywhere along the carbonate tuff contact, there was an upward gradient.

MR. ROBERTSON: In other words we don't feel that there is any evidence that the flow path is going to be down deep into say a faster more permeable, faster flow path and deeper carbonates come zipping out down here at a faster rate. The indication is the path rate probably will stay in the upper part. That's another one of the uncertainties here. Another scenario was for something to get down in a faster flowing dolomitic or something and make a faster track. Scott?

MR. SINNOCK: Scott Sinnock of Sandia, excuse me, just something quick to help maybe for scale, just that unlabeled line before UE-25 P#1, represents a symbol for where the carbonates are.

MR. ROBERTSON: This one?

MR. SINNOCK: No. Come over to the right, under UE-25, P#1, down at the very bottom.

MR. ROBERTSON: There?

MR. SINNOCK: Right there.

MR. ROBERTSON: Oh, okay. There is a good carbonate aquifer deep in this system.

DR. DOMENICO: As long as there is lateral permeability, there is no reason for that stuff to go any deeper than the unit that it finds its lateral permeability. Water is no fool.

MR. ROBERTSON: Okay. I'm glad to hear that. That was our thought too.

DR. DEERE: I wonder what controls the flat gradient?

MR. ROBERTSON: Everybody wonders about that.

DR. DEERE: I wonder if it is the carbonate itself?

MR. ROBERTSON: It's a complex--we know the basic principles that control it, but we don't know the details of the site of course. And I know that in the Czarnecki and Waddell model we know that there is a steep grading up in here and that's--I think they assimilated a lower transmissivity zone, practically a very low transmissivity zone, in this one and that's why these flow lines are going around it. And then you get the steep gradient up in this area describing this large flux through the system. Then we have the higher, relatively higher transmissivity or permeability in this zone relative to this zone, so we sort of get this back water effect. We have the amount of flux moving through this system is--can easily be handled by the hydraulic conductivity in this zone without a steep gradient. It's a low flux, and a moderate hydraulic conductivity is basically what probably controls it.

DR. DEERE: And one of the reasons the Swedes have one of their facilities located offshore or near shore, is because they have such low hydraulic gradients, they feel this is a very positive situation for reducing flux.

MR. ROBERTSON: Well, often in hydrogeology, we find low hydraulic gradient areas usually indicative of finding hydraulic conductivity.

DR. DEERE: Particularly in carbonates.

MR. ROBERTSON: But, that is not always true. It depends, you have to understand how much water is moving through the system here. And in an humid eastern environment, if I see a flat gradient and then a fairly permeable rock, you know, the kind of rock I can expect to have some permeability, I am pretty sure it is going to be highly permeable rock. In this case it is partly controlled by the modern permeability of the rock, as well as, the relatively low amount of flux in this environment. I don't know whether John or anybody else wants to comment on that, they have looked a little harder at this question.

MR. CZARNECKI: I'm John Czarnecki. Could you put the vector diagram up again?

MR. ROBERTSON: Yes.

MR. CZARNECKI: One uncertainty that we have in a model like this is whether or not water actually does come in from the north. In fact I gave a paper to many of you last fall on that topic that if you examine the chemistry of the water from Paiute Mesa and compare that with what we see at Yucca Mountain, the only way that you could account for Paiute Mesa water making it to Yucca Mountain was to mix it with 40 Mile Wash Water. And we certainly have little data. And we have very little data to say that indeed water does make it from the north as suggested in this model.

So, if that indeed is the case, if water is not coming in from Paiute Mesa, one has to ask the question, where that water is coming from. One possibility is that what we are looking at is a draining system that was established sometime during higher recharge rates and the system is responding in a transient mode. And that would suggest then that the flow in the flat gradient area is a lot smaller than is represented here. That's just something to keep in mind.

MR. ROBERTSON: That would give us more effects--more barrier effect.

DR. WILLIAMS: The open file reports I've read on the saturated zone seem to indicate that maybe 10 to 20 percent of the whole of any given drill hole, like the H-wells, produce during pumping. Do you think you can design a drill and test program to answer the question about matrix flow with that kind of a section?

MR. ROBERTSON: I think a test program can be designed to learn a lot more about matrix flow, but whether or not it will be sufficient to build enough confidence on characterizing the 5 kilometer pathway is another question. We've still got dealing with scale problems even if we want to get up to well scales. And there are multiple well tests in the plan where we inject tracer in one well and pump it out of another one, well interference tests and so on for the characterization program. And I think those will be--we'll learn a lot from those, but it may not be enough to build the level of confidence we would like to have in knowing particularly if it is going to perform well, but there will be problems in designing and interpreting tests and knowing whether they are representative of a much larger scale or not. I don't know the answer, until we do the tests and see what they look like and do more than one and see how they change from test to test. Anybody else?

DR. DEERE: Thank you. Just a question. Is this included in the report, or not, the comments of the saturation zones presentations in the report that you turned in?

DR. DOBSON: Well there is a description in the report of what we did.

MR. ROBERTSON: These figures we dealt with today are not in that report. Some of the numbers are and there is a little more text in there, but the figures are not. It's a pretty brief description of what we did and what the bottom line was.

DR. NORTH: Is there a transcription of the panel session that provides a lot of the background in this area?

DR. DOBSON: There are minutes in that section of the report of the models, but there is no transcript for that part.

MR. ROBERTSON: There wasn't a court reporter, but we took notes.

DR. DOBSON: Okay. Hopefully we'll wrap this up so that the rest of the group can get up. There was a request when we were negotiating the agenda for this meeting for some perspective on what we have learned in the process of conducting the Calico Hills analysis as well as some of the other analyses that we did. So, we put together a few view graphs that address that general topic. Most of these are statements, that you've probably heard before, kind of gather here and some of them verge on the philosophical. But, I would kind of like to go through them, to sort of see if they may provide a basis for discussion, or if they might be actually something to think about as we go through some of these other task force results, because I think that although I am making these statements with respect primarily to the Calico Hills analysis, many of them apply to several of the things that we have done.

In the value of information model, we went through a lot of effort to try and determine what was the basis for our testing program and why we were doing things. And one of the things that, we think, we learned as I noted on the first bullet here is, that, in many cases the testing program, as we have it written down, is not, and I have this in quotes, "Performance Based". And that is--what I mean by that, is what we talked about--is that the expectation that decisions will change, as a result of your changed understanding of the performance of the site, is small. In other words, we don't anticipate that's going to happen.

Now, that doesn't mean, as we've said many times, that the testing has no value. If you have a prior expectation that performance is going to be at a certain level and you do a test program, and your posterior expectation is still at the same level; does that mean that the test had no value because it didn't change your decision basis? Or does that mean that the test had great value because it confirmed your previous expectation? So there is value that we've tried capture here (indicating) in MUA here (indicating): in the confirmation aspects of these testing programs; and, certainly we think, in the confidence building aspects of the program.

Most people here will note, that the kind of program that we described in the Calico Hills from a testing perspective, and really the program that we now have defined in the ESF, is really fundamentally exploratory. We have a certain set of tests that are going to be done at the main test level which are specific and process oriented. But, most of the testing plans that we now have, for the underground facility at Yucca Mountain, are basically exploratory. And that is why all the rationale that John just went through this morning becomes so important. If what you need to know is what is the chance that you are going to be surprised by something that you learned, then the ability to go and do that exploration and to demonstrate that you have tried to capture the range of how the site is going to perform becomes a potentially pretty large value.

And so to the extent that these are basically exploratory programs, we think they have great value in completing our understanding of how the site is going to behave and in terms: of they have a great performance confirmation value, if you will; and the way you can almost look at it as the first phase of what is already regulatorially (sic) mandated--if I can get that adverb right. But, it is already required in the regulations that we conduct a performance confirmation program. In a certain way you can look at some of these as confirmation tests during the characterization period because you think you know how the site will behave and you think you know what you might see when you come to a fault zone or a fracture zone and the ability to go there and verify that you were correct could be of some value, especially, we think in the kind of environment that we are likely be in in the licensing hearing, which is a very regulated one.

DR. DEERE: Well, even before that time, Dave, I think it has a lot of value in allowing the Board members here to have confidence that you have the information and it is not inference but that you actually have had a chance, because our reports become public and if we have skepticism and criticism in them, I don't think it is going to help public acceptance any.

DR. DOBSON: Well, that's exactly right. I mean high demonstration of what you assert is very valuable. You can--I can stand up here or Jack can stand up here all day and say this is the way the world is, but having the ability to walk underground and show other people that it is not based just on my best guess of how the world is--it is potentially of great value.

A second conclusion is, that, even when you consider things like human intrusion and gaseous release, which we did not do as you will recall in the Calico Hills study, but some of the other tasks did, and Bruce Judd may wish to address this again later today, it would appear that the value of information, as defined, is still low in that both the prior estimates of performance and the posterior estimates of performance are still well below the EPA standard, and well below the levels that we would expect you would need to start changing decisions about the program.

And finally the last one here, which we've, I think, have probably discussed adequately, is that we do think that the saturated zone should not be ignored. And it is important to keep that in mind to the extent that we have suggested it was not a potentially an effected barrier, then we should correct that notion, because, we think that it is.

DR. NORTH: The conclusion from putting several of these points together is maybe that it is worth considering some underground exploration that would give us more information on the saturated zone here. Again, it may be for public acceptance as opposed to performance based values simply assuring that there is not an unpleasant surprise in the saturated zone and the reasoning that the panel went through is accurate.

DR. DOBSON: You may be right. I guess I don't think that I could support that with any of the analyses I've done so far because we didn't consider the change in the program--

DR. NORTH: Well, because your analysis hasn't addressed that issue.

DR. DOBSON: That's right.

DR. NORTH: But, it may be worth addressing an iteration too.

DR. DOBSON: I might add, too, that we are in the process of putting out the study plans for the saturated zone program, as well as, for the unsaturated zone program. And I know at least a couple of them are out. I don't remember if John's--John's I think is approved. And I think we also have a regional one, so we are trying to describe the plans we currently have in the saturated zone and we may wish to consider whether--

DR. NORTH: Well, in particular, if you take Strategy 1, as we were considering it, where there is extensive access outside the repository block into the area down gradient from the block, going all the way down into the saturated zone, might allow you to get a lot of information without necessarily compromising repository integrity without an exploration.

DR. DOBSON: Well, I--

DR. NORTH: I submit it for as a candidate for iteration, too.

DR. DOBSON: Okay. Anybody taking notes?

DR. CORDING: Dave, just before you go off of that, I still find that statement under testing program--to me does not represent what we are talking about, what the need for testing is. It is not a matter--saying public acceptance put's it at--certainly this has to be, this whole site has to be approved and it has to go through a hearing process and that is public in that sense, but the arguments are going to be technical arguments. And you are not going to make the technical arguments unless you get down there and look at it. And, I think it's a technical question not just characterizing at a low level of a public acceptance; or even thinking of it as a political sort of thing. So, I mean that I think that the way that is stated, I can't quite agree with that type of characterization of the need for testing.

DR. DOBSON: Well, okay--I did not mean to suggest that the reasons for testing were non-technical. As I said, I think that we've gone through the--if you value a testing on the basis of the expectation that it changes your CCDF, you tend to get low values. That does not imply to me that the reasons for testing are non-technical. So I mean I don't think I disagree with anything you said.

DR. CORDING: We've discussed it before. I think we do tend to agree, but that statement I think really perhaps doesn't give it the importance that I think it deserves, that's all.

DR. DOMENICO: There is value there. And if you have got your low release it is because you also tied into an unsaturated zone that you wrote off eight years ago. And let's keep in mind that the data base for the saturated zone has not changed in eight years, only your interpretation of it. That's the only thing that's changed. Where once it was written off, today it is valuable and that feeds right back into discussing the value of testing the Calico Hills. I still feel that Calico Hills is a main barrier and that's to be demonstrated to me that that is not true. I think it is the main primary barrier, in my heart of hearts.

DR. DOBSON: In my heart of hearts I agree with you. I guess just getting into a discussion about what primary and what backup and everything, and in the real world when we assess the total performance of a system, we will take credit to the extent we think it is appropriate to take credit for each of the barriers. So, I think maybe I'm agreeing with you. I don't know if it is all that productive in exercise to start ranking the performance of the various barriers at this point in time. We are just trying to point out that there is potential performance in the other barriers that we have not perhaps taken credit for.

And I didn't mean to suggest that this was in anyway any comprehensive list of what the benefits are. It was just an example of something.

DR. CORDING: I think the term public is one that can mean different things. That could mean media, for example.

DR. LANGMUIR: I'll try to keep you on that past overhead, Dave?

DR. DOBSON: Yes.

DR. LANGMUIR: Just the comment that even if you considered gaseous releases which left me with a question of where have they have been most recently considered? Is there a document you can site for us where discussions of gaseous releases are covered?

DR. DOBSON: You bet. The Testing Prioritization Task Force, Volume I. And you will hear about that as soon as you get rid of me here.

DR. NORTH: But we have no yet seen it?

DR. DOBSON: You have not yet seen it. It's soon to be available. Yeah, he's pointing at it back there.

Some of what we learned from the multiattribute analysis and I think this is important and probably intuitive, but we wanted to sort of reiterate it and that is that in the view of this group of scientists and regulatory people who did this analysis, you don't have to choose the biggest, most expensive strategy every time. In other words, a maximum strong look is not necessarily--all the maximum strong looks, as we define them, are not necessarily required in order to have enough information to get a license. It's a very hard thing drawing that adequacy line of how much testing is enough. And I don't think that we really solved that problem, but we did point out that we think that you can acquire adequate information with less than all the possible information that you can get. And I think that is from a planning perspective and from where we go now, it's very important for us to keep that in mind. That, you know, it's

--an important thing for us to do is start defining when we can bound the problem adequately and when we can say that from a performance perspective we have enough information to address a regulatory issue.

The second aspect of that first bullet is that the you can get useful information from outside the block. In the view of this study, when you compare the benefit of the information inside the block with the benefit of the information outside of the block, we felt the information inside the block was better. I think basically for the same reasons that Ed cited them. When you put them in a regulatory frame work and the question is, "well, I want you to show me what it is there you are talking about." Then you have some extrapolation problems and representative problems. But that doesn't mean that that information from outside the block is not potentially useful. I think that's one thing that we believe very strongly in. And that is why Strategy 7 which was entirely outside the block still turned up pretty well in terms of the amount information it provides, because it was a comprehensive testing program.

Finally the--or thirdly, I guess, no secondly, the UTF which is being renamed now, we are currently in a naming exercise for what our new underground facility is going to be called, the last I heard we were considering: exploratory laboratory facility; underground laboratory facility; and at the point we did this view graph, it was underground testing facility; should be designed, we believe this very strongly, to be capable of drifting to any part of the repository block. And one of the main reasons for doing that is, that, if you don't have that capability, then there is a possibility for delay during, that, essentially during the licensing process or during that docketing process, because you may not be able to demonstrate that you know enough about an important question.

So, where do we go from here? Well, in our view there are a lot of things--we think that this first cut of the study is basically completed and it is in the report, but there are a lot of things that we need to carry along with us in terms of what we learn. And, some of them are listed on this view graph. We need to continue to pay attention to how we manage this program in terms of defining the emphasis and the scope of the test program throughout the design process, and throughout the site characterization process. And there are several components that we just wanted you to know that we have, kind of high up, in our minds of things that we know we need to keep an eye on. But, I don't know if it is to reassure you but to let you know that we don't intend to forget about any of these things. We need to take the next step in terms of defining the test designs and locations.

So the Calico Hills laid out, in general: the test program; and the scope; and about where we need to go in more detail to define what really precisely we are going to do; and where; to the extent that we can.

We need to evaluate the sequence of exploration versus specific tests. Should we go back and start in the north or start in the south, or does that matter? We need to continuously, as we've talked about several times today, re-evaluate the impacts and characterization and during the design process I can assure you that the current configuration will be evaluated. What you see in the Calico Hills Risk/ Benefits analysis is an evaluation of an assumed configuration for the purposes of this analysis. And whatever we end up with for a final configuration will be evaluated and that will show up in the design reports that we do as we go through the design process.

We also need to reassess our performance estimates based on the new information that we get. I think that does provide some sort of measure of how we are doing on the testing program. I think we documented in the VOI model, that although we may not change the performance estimates to a level where they are likely to affect a decision, certain of the testing strategies do provide better updated estimates of performance than other ones. So, we need to keep track of how well we are doing in effect. And, also in connection with that, we use that information to determine when we think we might be getting done with the testing program.

We do need, I think, and I think there is recognition that we need to have that--the independent expert oversight of the program. Of course, the Nuclear Waste Technical Review Board is one significant component to that independent oversight, but we may wish to use other independent reviewers as well. And I presume within the next few months, we'll probably be briefing you on some examples of that. For example, we did recently complete a peer review of the unsaturated zone hydrology program and that kind of review of our program we think is important to building some consensus in the total technical community that what we are doing is credible.

And, finally, I keep coming back to it, but assessing the adequacy of information and our ability to use it to close regulatory issues--well, I guess I said it all.

The last two are view graphs that I presented a few weeks ago to the Nuclear Regulatory Commission and, of course, (indicating) they are at least in part what the NRC is currently working on evaluating the Calico Hill analysis. We think that the CHRBA was adequate to meet the commitment that we originally made in response to the CDSCP. We have considered the benefits of testing as measured by several techniques, versus the risks that the performance was measured primarily in terms of adverse impacts to waste isolation. And we think that the analysis that we've done and the input and the exchanges we have had with the ESF group will provide us with an adequate basis for moving forward with design for the current design study that we are doing or we've integrated with the recommendation from the ESF alternative study and then following completion of that resumption of the Title 2 design.

We want to caution, as I have said a number of times, that neither the precise configuration or the treatment of waste isolation impacts are regarded by us, or should be regarded by anybody else as the final word. It will be redone continuously between now and when we write a license application, should we determine the site to be suitable.

And, also, and this is important because: we have promised; and we are obligated; to continue to consult with the NRC. We will do that through: the design process; and the design review; and continued meetings with the NRC; as well as, I am sure with the TRB.

And this is just kind of an intro to a talk when I first put the view graph package together I thought was going to be next, but we've change the sequence, so the ESF alternatives is not following now. But, I did want to make a point that we think that the integration between the task forces have been effective and that the current top ranked options from the ESF alternative study support very well the recommendation of the Calico Hills study. So there has been a very much confluence of the recommendations of the two task forces in that regard.

We think that access to the Calico Hills via ramps from the east, which are the recommended means of access in the ESF study, will provide excellent site characterization data which will supplement and improve the site characterization program. And they will probably, also, improve the performance; the waste isolation impact. Because, as I have noted on the second bullet there--that (indicating) we will eliminate, given a configuration like that, any direct vertical pathways between the main test level of excavations in the Calico Hills excavations.

And finally the last statement that was made earlier by myself and John, that the robustness of the recommendation for strategy 2 and 5 would be increased by the ramp accesses that would be the simple term everybody is using

MR. BLANCHARD: Dave, your last--your previous view graph indicated that the report had been given to the Nuclear Regulatory Commission for review, and since we have some staff from the NRC here who were involved in that meeting and who might understand or are also involved in conducting the review, would any of them care to indicate what the current status of the review of the Calico Hills Risk Benefit Analysis?

MR. STABLEIN: I would.

DR. DOBSON: By the way, before King starts, I should note that I guess there must be a letter hung up on our system, because we certainly intend to transmit copies of the Calico Hills report with the Board and if you haven't received them, I hope they will be on their way shortly. I know that Russ has a copy. But we will get copies for everybody on the Board.

MR. STABLEIN: My name is King Stablein. I'm with the NRC. And when we reviewed the CHRBA, I guess I'll be heading up that review, so I would like to take about one minute to tell you the status of that review, since it's been referred to a couple of times today.

We met with the DOE January 29, 30 and 31st in Washington in a technical meeting where DOE laid out in a lot of detail various aspects of the CHRBA as well as ESF alternative study. We received the CHRBA at that time or perhaps a day or so before, at any rate we received the report coincident with that meeting.

Subsequent to the meeting, some of the headquarters personnel for NRC had a conference call with three representatives of DOE headquarters at which time it was requested that DOE provide us with certain information so that we could start the review of the CHRBA. The information included a crosswalk that DOE said is being developed between

CDSCP objection 2, that was our objection, and the CHRBA. It also included further supporting material on the VOI, we as some of the TRB members have expressed had some difficulty understanding the ramifications of the VOI both by itself as to what it meant, what the results meant, and then in connection with the MUA. So we requested certain supporting information in that regard, and we had a couple of other requests with regard to the CHRBA that we asked be put in that letter from DOE to NRC.

We haven't received that material yet. When we receive that material, depending on what it looks like, I'll be able to devise a review schedule, which we would like to of course be as timely as possible, and we will of course make that available to DOE and anyone else who is interested in knowing about the progress of the review.

Right now the staff is of course looking at the document informally, becoming familiar with it. We also have feedback to give DOE via the meeting notes for that January 29th meeting, but those two are just a little bit hung up at the headquarters level, DOE and NRC headquarters level.

If there are any questions on this, I'd be happy to answer them.

MR. BLANCHARD: King, I'm not sure that we are aware of the list of items that you've asked for except for the first one where it was clear in the exchange that we had there in

Bethesda, Maryland, you were looking for a crosswalk. And we sent that crosswalk out. We've checked on it and they have a letter number for it, and so, we know that it went out.

MR. STABLEIN: Just a second.

MR. BLANCHARD: But we don't know that it's been sent to you.

MR. STABLEIN: Oh, you don't know it's been sent to me. Where might it have been headed?

MR. BLANCHARD: To our focal point in Washington who distributes all information to you. But, I don't think we've gotten anything from OSC to our office that identifies these other items so far as I know. Do you Dave?

DR. DOBSON: Not that I know of.

MR. STABLEIN: Well, it wasn't my purpose to point any fingers at anybody at any rate for these not having arrived to us. I'm merely mentioning the phone call that took place on February 11 between John Linehan of the NRC and myself and Steve Brocum, Linda DeSalle and Dwight Shelor at DOE, during which we made verbal requests only. The cross walk, of course, was mentioned, as well, in the meeting. The other items weren't specifically mentioned.

MR. BLANCHARD: Well, we should be able to get the cross-walk to you immediately, because I know it's out. The other things I'm not sure what we'll have to do to finish those off. First, we'll have to find out what they were.

MR. STABLEIN: Fine.

MR. BLANCHARD: Thanks for the information.

MR. STABLEIN: Okay. Are there any questions, Dr. Deere?

DR. DEERE: Thank you.

DR. DOBSON: Okay. Well I guess we are done with the Calico Hills.

DR. DEERE: Thank you very much, Dave, Max, and all of the presenters.

We would like to move into the second topic and we see the man arriving.

I think I have to apologize for us being behind schedule like this, but we did have, as you know quite a number of questions about the presentations and particularly the last item on the saturated Calico Hills because we hadn't had the chance to discuss this very much in the past. So, sorry we have taken the time away, but we will give it to you on the end.

I used to tell my students when I arrived about 15 minutes late for my lecture, don't worry, we'll make it up on the end.

DR. DYER: I'm Russ Dyer from the Department of Energy in Las Vegas with the Regulatory Evaluation Division. What we've just passed out is a report that DOE just received last Friday. And this is the first report of the test prioritization task force. Please note that this is not a final report. This is just the starting point for this effort. And, I don't think that we've talked with the Board since last July in Atlanta. And I'm going to spend a little bit of time talking about the--if I can get the sequence of events right here.

We have two presenters for you today, myself, I'll be the lead off and I'll be given you an introduction and bringing you up-to-date and what's happened with what used to be called the Surface Base Testing Prioritization Task. The name has now changed of course with a new acronym, TPT. The objectives of the task have changed somewhat since you were last briefed on it. We went through a phase approach, and I'll talk a little about the scope of the phase approach. Then Bruce Judd will follow, and Bruce's presentation is split into two parts. One of which is the analysis and methodology that underlies this first report and some of the future activities of Phase II and the coordination with the site-suitability effort that we are looking at right now.

This slide is not in your briefing package, but it is one that I thought I would pull out just to remind you of what originally drove this effort. And that was a decision by the Secretary of Energy back in 1989 to focus or refocus the near-term site testing program. And specifically we were directed to focus the near-term scientific investigation specifically at evaluating whether the site has any feature that would indicate that it is not suitable as a potential repository site. So, the specific focus was to look at unsuitability as an issue early on in the testing program. And our charter was try to identify those tests that we needed to bring on-line to address this issue.

The Test Prioritization Task has two objectives, one of which is to develop a method or methodology which will allow us prioritize tests and to assist in this early detection of any unsuitable conditions. And the second objective here in the icon here, the results of the application of this method, would be a rank ordering of tests, one, two, three, four five.

The second objective is to recommend methods to re-prioritize the testing at any point during site character-ization, based on updated information, understandings that developed during the testing program. And this must explicitly include a method for deciding when to stop testing, when do you have enough testing? And in the icon here we show that test 1 and test 2 have been completed. You have gained some information, some understanding based on the completion of these first two tests. Based on this information that you've acquired, you have changed the rank order in your following tests. The tests that you were going to do third, trades place with the test that you were going to do fourth and based on this information that you have from the completion of test 1 and 2, there is no need to even do test 5. So that was the second objective of the test prioritization task.

And, one thing that we would like to point out is that the method should provide a management tool for test prioritization, but it must be consistent with the site-suitability evaluation methods, and you'll hear more about the interaction between this effort and site-suitability as we go along with it.

Just a brief recap of history here. I alluded to the directive we received from Admiral Watkins back in 1989. In January of 1990, we initiated this program. We had the actual plan approved under a QA program. And the original charter of this group included the responsibility for making recommendation of possible site-suitability methods, methodology. And, we were proceeding along that path in the summer when it became obvious that site-suitability was a very large endeavor in and of itself. And we could either do one or the other since this group was well along at looking at the prioritization we constituted another group to look at site-suitability.

This August 1990 letter, there is a critical word missing in that little line. It should say "Letter report on site-suitability methodology". There was no recommendation made on site-suitability at that time.

In October 1990, we officially modified the scope of this task force, split out the site-suitability effort as a separate effort, directed that the test prioritization task do a couple of things. We both expanded and decreased the scope of the study. We expanded the scope of the study in that the effort was originally targeted to look only at the surface base prioritization, surface base testing. And we expanded the scope to look at the entire testing program, hence we changed the name to the Test Prioritization Task.

The scope was decreased in that the charter for looking at site-suitability was moved into another group. And also the other thing that was established was decision to go into a two phased approach, an initial effort to provide some quick information that we could use and then a follow on long-term effort. What you are going to hear today is the report on the initial effort which is based on a simple spreadsheet model, which Bruce will tell you about.

And I guess I'm a little ahead of myself here. Here is a little description of the spreadsheet, the two different phases of this effort. The spreadsheet model was due 1 March 1991, it was delivered on time. We not only developed a spreadsheet model, I use the "we" in the vicarious sense, I oversaw the hard work of Bruce Judd and Steve Mattson, Dwight Hoxie, Scott Sinnock, who were core team members for this effort. A method was not only developed, but it was also applied and you will hear a little about both the development, the logic behind the method and also the application of the method to date.

Phase II is a model that we perceive as being the next generation of this tool, and it would be based on a simple total system performance model, with using that in lieu of some of the expert assessments of performance, and also incorporating a larger sampling of experts in the assessments that are made.

Just a little background before Bruce launches in here on his talk. The Phase I approach is a five step approach. The first step consisted of compiling a list of potential concerns. And what we used as a source for concerns were: the potentially adverse conditions of 10 CFR 60 Part 122; the potentially adverse conditions; potential concerns; disqualifying conditions of 10 CFR 960. We also on February 8, 1990 held an elicitation meeting at the project office, to the tune of about 50 scientists, where we elicited for any other concerns that may have arisen within the technical community that were not captured either in 10 CFR 60 or 10 CFR 960. We came up with a list of over 100 different concerns at that time. And, we went through a process, which is step 2, of assessing and ranking the importance of each of the potential concerns to waste isolation. So there was a measure of importance that was created and attached to each potential concern, and from this we started out with a long list of potential concerns, by the time we assigned or affiliated an importance ranking with each potential concern, then we could rank order the potential concerns. Step 3 consisted of identifying tests which relate to determining information about these potential concerns and that was done here in Phase 3. And we did not look at all tests. We had to screen in order to make this a dual project, we had to screen the potential concerns. I think we only looked at about the top dozen to fifteen, made the cut-off there, looked at the tests that were associated with our topped-ranked potential concerns.

The 4th part of this task is to assess and rank the tests that address the important potential concern. That is assign--we have a rank ordering of potential concerns, how can we associate a rank ordering of the testing program. So that was step 4. Step 5 was then to merge these and to evaluate the testing priorities. That takes us through Phase I.

In a slightly different format, this slide captures the essence of what I just went through verbally. Our inputs were potential concerns of 10 CFR 960 and 10 CFR 60. Those provided us with potentially adverse conditions, disqualifiers of the other input that we had here was elicitation of the technical community. We had the testing programs outlined in the SCP. One of the things we wanted to look for was to see if we could identify any holes in the testing program. Were there things that needed to be done that weren't identified in the testing program? But we used the testing program as outlined in the SCP for our starting point.

The criteria that we used for prioritization, that is to give us our measure was radionuclide and release limits based on the performance standards out of 40 CFR 191. And finally, we relied very heavily on expert judgments in order to combine all of these individual inputs.

I said we did some screening. Well, this just gives you an idea of how--what we had to do to make this a workable task. We started out with over 100 potential concerns in the beginning. By combining some things, certainly some parts of 10 CFR 60 and some parts of 10 CFR 960 are different ways of saying the same thing. So, we could capture the essence of a particular concern, perhaps in one statement, instead of having to treat it as two or three concerns. So, by combining some of the things, we boiled it down to 32 potential concerns. And, these were ranked. These were assessed. There was a measure of importance assigned to each of the potential concerns and we ranked them.

Then we got it down to--well, it says here, the ten most important potential concerns and we associated tests that were related to those potential concerns. We used very heavily the PARATRAC data base in order to identify relations between the testing program and the regulatory issues here. And we had 15 test packages for ten of these concerns. Of course some test packages addressed more than one concern.

Just a quick word here, test package doesn't mean an individual experiment at a particular locality. It's a generally a set of tests that's looking at some particular issue. We'll talk a little bit more about that in a little while.

DR. DEERE: Were some of those other than those that have been presented in the SCP?

DR. DYER: We identified, well maybe--let me jump ahead here and give the bottom line for what Bruce is going to say.

We identified a gap in the testing program. We know that there are some things that we need to identify tests for and those tests are--we need to develop tests to address particular issues. Those tests have not yet been developed.

The next step was to assess and rank the test, as much as we did the concerns, and then to evaluate the test priorities. Once the test priorities came out, as you will see, they fell into three natural groupings based on what--I hesitate to use the term value of information, but how much good the test provides to us.

Now, having followed the Calico Hills, let me state from the very outset here that we recognize that this measure we are using is based on performance. There are many other reasons for doing tests besides a performance based rationale. And these are some of them that we recognize exist. These were not considered in this initial analysis. Evaluating pre-closure health and safety for design input, providing other information required for licensing, facilitating other tests, it maybe that your test sequencing requires that you do a test before you do a follow-on test. It's actually the follow-on test that is most important, but you can't do the follow-on test until you do a preceding test. Building scientific consensus. You heard quite a bit of discussion about that in the preceding talks.

And this is the group that steered us through this effort. Steve Mattson was the team lead out of SAIC; Scott Sinnock from Sandia was the performance assessment representative; Bruce Judd, Decision Analysis Company, provided the decision analysis insight, and Dwight Hoxie of the USGS was the last USGS representative on the team. We went through--we seemed to burn up USGS representatives on this effort. We started with Tim Barber and then went to Bill Wilson and finally ended up with Dwight, so congratulations Dwight, for making it through.

Let me now turn over the talk to Bruce Judd who is going to tell you about the Phase I analysis and results which is what you have in front of you in the subtle orange covers.

DR. JUDD: While I'm getting wired up, let me put this on. It came out of the New York Times. I realize I'm getting close to that.

What earlier tests showed was critical, sometimes between three and four, so I hope you will excuse occasional interjections of either attempts at humor or variety or something, I'm not sure what you want to call it, but it's just a consideration for the time of the day.

The other thing I would like to say before I start out that the names that Russ showed you including Dwight, and Scott and Steve and Russ, all are integral to this presentation as well as the analysis behind it. And every now and then you'll ask a question that I am positive I can't answer as well as either Russ or one of those others, and if it is okay with you, I would like to be able to defer a couple of those.

DR. DEERE: Sure.

DR. JUDD: I'm going to go follow the order of this five step process that Russ explained to you and I'll get started with compiling the lists of the 32 potential concerns that we considered in our Phase I analysis. And Russ has mentioned the sources of these and they were Parts 60 and 960 including the PACs, as well as, disqualifying conditions. Now there are several issues discussed in those regulations, including qualifying conditions, for example, and those qualifying conditions were not included explicitly here. We were focusing on tests, and this phrase you'll hear me say over and over again: tests that could help detect early in site characterization, potentially unsuitable site conditions, because of that directive from the Secretary, that Russ showed. That's why we took that focus. So, we are including the adverse conditions and the disqualifying conditions because it was that subset test that we were focusing on.

Here's an example of some of the potential considerates that we looked at related to gas flow, ground water travel time, et cetera. It's a familiar list to many of you and I'll get more explicit as we go on about the particular list. And Russ mentioned that we use the regulations as a source as well as other concerns raised by EPRI and EEI, the Regulatory Commission the State of Nevada and the participants in early meetings we held in February, just a little over a year ago. This produced a list of over a hundred potential concerns which our core team then boiled down to a list of 32.

In order to do our quantitative analysis, we needed to define or specify measures, quantitative measures, for these potential concerns and, as well as, some thresholds and I'll show you how the thresholds were used in a minute. And rather than giving you more words on this, let me give you an example. Let's take the disqualifying condition in 10 CFR Part 960, ground-water travel time and measure for that. In other words a way to quantify it is listed actually in that document. Along with Part 60, there is a discussion of this as a PAC. The measure is expected ground-water travel time. Now you have to specify that a little more carefully and the specification that we used was ground-water travel time along the fastest paths has significant, and likely, radionuclide transport. And so we had to specify carefully what the measure was in order to any quantitative analysis.

And then we picked a single point on that dimension, that measure. The single point we referred to as an assessment threshold, and in this particular case our job was made easy, because there is a number of 1,000 years given in the guidelines in Part 960. So we picked 1,000 years as our assessment threshold. What was an assessment threshold? It was a point such that if the measure is above that point, where above means more severe towards waste isolation, more detrimental to waste isolation, or above that point the potential concern is present. If we are below that point that we pick as a threshold, the potential concern is not present. And the icon that you see here is intended to look like one of those speedometers that didn't last very long on cars. They were a little red bar that rotated and the line came across. And you can imagine if we had good knowledge of what the ground-water travel times actually were, and this meter indicated that accurately, it would point somewhere here and any value above the assessment threshold, we would say, that concern is present.

Now, what does that mean, that concern is present? What does that mean? Well, we have to define a little more carefully then what it means for this potential concern to be present. What are the consequences for the detriment to waste isolation if this is present. So, that was one of the key factors in our analysis that we had to analyze.

Now we quantified and analyzed that in Step 2 of our methodology. Step 2 was conducted by a panel including the core team. A total of nine experts on performance assessment and the site. These were experts from the project, the Yucca Mountain Project, and they were the ones that we felt were the most knowledgeable, and those who could contribute on a time frame that was necessitated by our own schedule. So we had a set of experts provide that information and we felt that they were the most knowledgeable at the time.

Now I'll show the kind of information that these experts provided us, the kind of judgments. And I'll do that as you might expect, since the word decision analyst was in my introduction. I'll do that with a probability chain to show you the type of assessment that gets at, how important--how important is this potential concern.

Let's take an example. Expected ground-water travel time less than 1,000 years along the fastest paths of significant likely radionuclide transport. Several concerns are raised here or several issues that had to be assessed. One was, does the concern--is the speedometer on the last view graph--is the speedometer above or below 1,000? Is the ground-water travel time less than 1,000 years, in which case the potential concern is present, or is it greater? So one of the first assessments that we made was that of probability. There are nine people in the room, each contributing judgments to this, we would have a lot of discussion of it applying some of the same probability assessment procedures that you've heard discussed back in July and earlier today.

This number here represented consensus from the group in a sense that it was a geometric mean of the wide-range of assessments provided by the group members. In this case it was .002, is the likelihood that the ground-water travel time along this fastest path was less than 1,000 years.

The next is a conditional probability assessment meaning that if, indeed, the concern is present today and exceeds the threshold, then what is the likelihood that this concern will be present sometime during the next 10,000 years? In this case, it was judged to be very likely. And finally, the question was asked, the third question was asked, was what is the likelihood of this concern, if it is present during the next 10,000 years, affecting waste isolation? Again, this number that represented the judgments of the group was a high number, .6.

These are probabilities that lead us along a scenario, if you will, to this concern having affect on waste isolation. The product of those probabilities is a relatively small number, which means it's unlikely that we get out to this end point here, but it is certainly possible. The probability would be the product of those three numbers.

And the last question that we asked this group in assessing importance, and I'll use this term quite a bit, what is the consequence, if you will, of this potential concern affecting waste isolation? If it affects waste isolation, by how much? The way we quantified that was to say, what is the increase in radionuclide releases because this PC or this potential concern is present? So what's the increase that this concern causes in radionuclide releases? And that has to be relative to something. This is incremental over something. The something was the baseline judgment that was the judgment on the radionuclide releases assuming that the potential concern is not present. In fact none of the potential concerns is present.

We normalize that to the EPA limits so that if this had a value of 1, that would say that releases are exactly what EPA sets as a limit in Part 191.13. If this number then is .002, this is saying that if we get along this chain of events then .002 percent of the EPA limits is what the incremental releases due to, or what are the incremental releases are due to this concern existing?

It's a judgment by this group of the relative importance of this potential concern. And we need it to do that kind of assessment then for each of the 32 potential concerns in order to get an idea of which were the most important from the perspective of contributing to potential unsuitability of site or contributing to the loss of waste isolation. We are trying to get incremental contributions. The importance number that, you'll see me use on subsequent view graphs, is the expected contribution of normalized curies or normalized releases which is the product of these three then times that number and notice that this is a very small number. And the other numbers that we will be assessing are also small numbers, and what is important here is their relative ranking of those importance numbers, and I'll show you that in the next view graph.

Okay, so these are the types of probability assessment questions that we use as input. I will on subsequent view graphs define it or use a slightly different aggregation of these numbers. I'll call this the probability that the concern is present, which is this number right here, (indicating), that the concern exceeds the threshold. That's this number. And I'll take these three numbers and roll them together and refer to them roughly as consequences if the PC or the potential concern is present. This is its likelihood of occurring and that's the consequence if it occurs. Probability and consequence, you could think of this as a risk due to this potential concern.

Now, I will show you the results at several points throughout the discussion today, and this is the first set of results and it is a ranking of importance, and the dots that you see down the graph are the relative importance of these potential concerns. But across the bottom is something that I need to add a little bit of explanation to, so if you will allow me to turn this thing on its side, and then for those of you fortunate enough to have a copy of the orange report in front of you, it's a little bit like watching a Broadway musical and you need to have the play bill open on your lap, the page in Executive Summary ES-1, has a table, that actually continues onto page ES-2, a table that gives the short definitions of these potential concerns across the bottom. I'll mention a couple of them now verbally, but I suggest that if you have a copy of this report, turn it open to that page, and let me just introduce you to some of the lead players in the musical here.

This is the transport of the gas radionuclide from the repository primarily upward through the unsaturated zone. CG stands for complex geology. This was a measure of the degree to which we might make a mistake or incorrectly predict the releases of radioactivity, because there are some unexpectedly complicated or complex site features, or some difficulties in modeling that weren't anticipated. So if this potential concern is present, it is a set of features that cause the modeling or the analysis of the site to be very difficult, and furthermore, cause us to underestimate significantly, the releases at the site. The amount of the estimation was 10 percent of the EPA release limits.

This was human intrusion of the waste package, a direct penetration of the waste package, either due to some systematic drilling program or search for water, a search for natural resources, etc. This was the ground-water travel time potential concern that I mentioned before.

This one I was giving a briefing a couple of weeks ago and said what is "Eh" and I turned to the person and said, "Eh?" It's "Eh", the measure of oxidizing, the oxidation potential of the ground-water proximity to the waste package. So this is a very shorthand abbreviation for that oxidation potential.

This was climate affects on the unsaturated zone hydrology. This is human intrusion that also affects the geohydrology, or geohydrologic system, human intrusion due to natural resources, perched water, this one is relating, and I'm going to introduce only three more to you, I won't go down this whole list. This one, UO2 Solubility, and Total

Dissolved Solids were both measures of the reactive ground-water chemistry in the host rock. And volcanism was a direct igneous intrusion into the repository itself. We had another volcanism down here which is an igneous intrusion into the controlled area.

So I've listed the top 15 or so for you and these are the characters that we will come back to as we go through this discussion.

Now that you have some of the players, let me introduce the axes. The vertical axis here for measuring the importance as we've defined it in this study--again I'll turn this around so you can read this a little bit better, this is the increase in radionuclide releases due to each of these potential concerns. And that's normalized by the EPA limit, which means if you had a value of 1 on that scale, in other words, we are up at the top, then a particular concern, one of these potential concerns would be contributing to releases an amount exactly equal to the EPA limit. Notice that each of these at least individually are contributing less than the EPA limits.

Now there is one more feature of the view graph or the chart that I need to explain, and that is the difference between dots and the vertical bars. The dots are the importance numbers and here importance for ground-water travel time was 2×10^{-6} which was the product of a probability, .002

and a consequence. That's what the dots are.

The vertical bars are just that, consequence measure, and so for instance, let's take ground-water, complex geology aqueous releases, the consequences here were higher than the dot if the potential concern is present, there are some significant consequences associated with that, but that's not a 100 percent sure that the concern is present. And because the probability is less than 1 that it is present, this dot, which is the product of a probability and this height here, the dot is lower.

I said this is an intermediate result, time for around 3:00 in the afternoon when drowsiness is supposedly at its maximum, so let me say, what do these results say? Well first off to a decision-maker who is setting some priorities in testing, we now have a ranking here, a relative ranking of the potential concerns and there are 32 of them and they correspond in one way or another to the concerns raised in the PACs and disqualifiers. So we have something there that makes a relative comparison.

What else do we see? Well these three, those associated with gas, complex geology that results in gaseous releases and complex geology that results in aqueous releases, those three have an importance here that's at least a factor of 200 greater than all the others. If I compare those up there to these down here the relative ranking is many, many orders of magnitude, 12 or 14 orders of magnitude. But compared to some of these others in here there is at least a factor of 200 difference. You will see that throughout the discussion of our results. Those three keep popping up on top.

Then there is another set here that have lower contributions, lower importance, lower potential contributions to the releases from the site and they come out here and then we--there is sort of a plateau there, and then we head downhill at a slope that gets us all the way down to 10^{-14} and 10^{-16} contribution releases.

And what we did with this was divide it right here between where this thing plateaus and starts heading off again. And we took these 14 as Russ mentioned and did more analysis of the testing related to those. So we are ranking these potential concerns. We are now going to eliminate about half of them and focus our attention on these upper ones up here.

Now when you do that, you always run the risk of either leaving something out or mis-quantifying because of something that happened in the assessment process, and so, I've shown a sensitivity analysis here, that recognizes the fact that we had many experts. In fact, the dot represents the geometric mean or the consensus number from the experts. These extremes, I'm not tall enough to show you the top of that one, maybe I could stretch up to this one and show you the high and the low importance number for the nine individuals that we had in the room.

Well you can see that there is quite a bit of difference of opinion, often spanning three or four or five orders of magnitude in the importance assessed by these individuals. That may be a worry. There seems to be quite a bit of opinion, range of opinion being expressed here. But let's look again at our top three and notice that almost without exception and this is the one exception, but almost without exception, all of the others, or I should say it the other way, none of the others reach up into that zone with those three. And that we feel was a significant conclusion, that everyone in that group agreed that these were the most important.

Now we haven't answered the question yet about the testing associated with those and how accurate that testing is, but at least just in terms of sheer importance, those are the most important.

What we are going to do as I mentioned before is take this, divide it roughly in half and continue to work with these up here. For the ones that were just starting, in no cases do they ever jump--for the ones that are being set aside at this point, discard may be a bit strong, but the ones being set aside, none of them jump into that upper region up there.

They certainly jump up to the region with these, but they don't jump up into the other region, and, our conclusion through this study, or the analysis throughout the study, supports the conclusion that these are the ones that you ought to go after first if looking for potentially unsuitable site conditions.

One other thing I should mention about this chart is that we can relate this importance number to testing. This importance number, the way we defined it and the way we calculated it, sets an upper bound on the value of information, VOI as it has been called today, the value of information provided by the tests for each of these PC's. In other words, if you could know perfectly that this natural resource, potential concern indeed does exist, if you could know that perfectly, then you could take some actions regarding this site, either walking away from the site or somehow mitigating that concern, you could take some action to eliminate those incremental radionuclide releases. That would be a benefit of the testing. It has told you something that led you to an action that eliminated some of the potential radionuclide releases. That's our measure of benefit and this set of dots here sets an upper bound on the value of the information provided.

Now, why am I mincing words here on this upper bound? Why is that important? Well, our tests are not perfect. Tests can miss a potential concern in which case the benefit of the test will drop a little bit relative to this. But, this sets an upper bound on the testing.

I'll conclude this view graph with the following caveat, this is an upper bound on the value of testing when testing is being done. Why? To detect potentially unsuitable site conditions. There are many other reasons for testing as Russ mentioned.

This Step 2 generated we think, some significant insights about the importance of these potential concerns. There is certainly substantial variation in the relative importance. That is obvious to anyone who has gotten involved in the problem over the years. However, gas flow and complex geology were agreed to by this group as being more important by the others, by at least a factor of 200, so those are the ones to keep in mind.

And, third, the screening that we did in Step 2 identified a set of 14 potential concerns that were carried forward into the subsequent parts of the analysis. Four of those were later either eliminated or combined with others, but these were the top ones for further consideration.

That's sort of a natural point, it's a little less than 30 minutes. Are there any questions to this point?

DR. DEERE: Any questions up to this point?

Let's take a break for about ten to fifteen minutes.

(Whereupon, a break was taken off the record.)

DR. DEERE: Okay, let's get started again.

DR. JUDD: They are keeping the time clock on us for about 45 minutes into the total presentation of Russ and myself, and we have about an hour to go.

So we've identified a set of potential concerns or PCs, not personal computers, that have relatively high importance compared to some of the others. So the next step then was to compile a list of the studies and tests associated with that.

As Russ mentioned, the site characterization plan provided a set of these. Now, we need to make it clear that we did not look at everything in the site characterization plan. We only took those tests that could be done early in characterization and were focused on the intent of the test, or the output of the test could be used to detect potentially unsuitable site conditions. That is only a portion of what is in the SCP. We used PARATRAC as he mentioned to correlate the potential concern PACs the disqualifiers to actual tests and also analysts themselves had some contributions to what the tests might be.

When I say tests, what we are referring to is actually packages of activities that include SCP investigation studies, activities and sub activities. So, they are packages, suites of tests, and within any particular package, we did not set priorities within that package. We set the priority comparing one package for a particular PC to let's say test for a different potential concern found within the packages, and this method that we've developed, could be applied in a similar fashion to set priorities within the tests.

The third thing that we did was, that on some of our potential concerns and it is just on four of them, four of the 14 remaining, we created levels of testing that were progressive in the sense that they involved more and more testing activities as we got to higher level numbers. For example, the testing for ground-water travel time, we constructed three levels. Now for some of them we constructed two--some of the other four potential concerns we constructed two levels, some of them--I think one other one we did three levels, these levels were the first and the simplest package was no new drilling. You use available data and non-surface disturbing work.

Level 2 took that body of information provided by testing, and added to it some surface based drilling. So, Level 2 was the sum of that surface base drilling and Level 1. And then Level 3 added the--Dave, what is it these days? Here it is referred to the exploratory shaft underground test facility, et cetera, data from getting underground to the data available from the bore holes and from the non-surface disturbing work.

Here's an example related to the tests for that thing that we've called complex geology, which as you will recall from the definition was complexities on the site that are unexpected but contribute to difficulties in modeling and estimating releases that causes us to underestimate significantly.

Here is a list of 10 SCP activities and the simple way to interpret this is, these are tests that were going after three specific features: the Solitario Canyon fault; Ghost Dance fault; and the steep hydraulic gradient north of the repository. Those three site features plus a baseline set of data from the systematic drilling program, that involved one to three, feature independent bore holes. Those were the activities associated with one of our levels of testing for complex geology. This level involved two to six total bore holes split between these feature investigating bore holes and feature independent--

So that provides us with these lists. We took the 14 potential concerns, knocked out four of them for reasons that are described in our report, or combined them with others, that got us down to ten, and then we expanded the multiple levels for a couple of these which got us up to a total of 15 test packages which were then assessed and ranked by another series of expert panels. We did this in a series of workshops, and in some case, there was some overlap with the panels. In other cases there was no overlap. In other words we got some new people involved in that phase of our investigation or our analysis.

To show you again the type of information that we assessed I'll use another probability tree. What we are trying to get at in the workshops related to testing was really one main issue and that was the accuracy of the tests. How did we quantify accuracy? We used conditional probabilities which I'll show you down here, conditional probabilities related to true positive results and false positive results. And what do I mean by positive results?

I had a case last week, not another case, but a need last week to dig into Dr. Spock because my daughter had the chicken pox, and I couldn't resist this picture which shows one form of testing that he was suggesting for chicken pox. In other words, the temperature that you would take from the child is an indication of whether the kid has got chicken pox or not. It's not the key indicator I later learned, but it was an indicator. So let me use this picture out of Dr. Spock to talk about what I mean by a positive result.

In my case I defined the positive result to be temperature greater than 101👌 Fahrenheit. That's not necessarily positive. That's not necessarily good. In fact that's bad. And so when I use the word positive, I don't necessarily mean that it is a good result. I am saying it means that the potential concern is present and the test indicates that it is present. I'll define that a little more closely in a minute.

Since tests can be accurate or inaccurate, we might ask the question, what is the probability that the child's true temperature is greater than 101 given that my thermometer, my test in other words says, it's greater than 101. What is that probability? Well that is one of our measures of the accuracy of the test. And of course, my daughter who is not an infant, but knows how to manipulate the thermometer with a cup of soup or something like that has learned how to give a false positive result which would be the probability that a false positive now says that the temperature indicated by taking the temperature was greater than 101, even though in fact it wasn't.

Now let me be a little more careful with my use of probabilities here and explain that. But does everybody get the idea that when I'm talking about a positive result, I don't necessarily mean a good one. And when we talk about testing in general, we have to recognize that there can be false positives. Let's follow this probability tree.

I have two cases. The concern is either present or it's not. And this is in some sense truth. It really is present. For instance the case of ground-water travel time, let's say the expected ground-water travel time along the path the significant and likely rate of ground-water travel, is indeed less than 1,000 years. If I conduct my test, the test result might "find the potential concern". In other words, it finds the test result is--the travel time is less than 1,000. Since it truly is, and since our test has accurately predicted that it is, I'll call this a true positive outcome and the probability of getting a true positive outcome is P2. This is a conditional probability. It's conditional that indeed the travel time is less than 1,000, that's the way it really is down in the repository block or below it or where ever we are measuring it. That's what is there and we find that it is there. This was a good test. It ferreted it out. It detected a concern that was truly present.

The other case that we worry about down here is that it is not present. In fact ground-water travel time is greater than 1,000 years. But our test, because it is inaccurate indicates that the concern exists, i.e., were less than 1,000. So it's truly greater than 1,000 and we find that it is less, this probability is the probability of a false positive. We call it P3 on the view graph and we called it P3 in our assessment, but just think of it as the probability of a false positive or slang or shorthand, the probability of a false alarm. This tells us there is a problem when indeed there isn't. Those two numbers are the numbers that I'm going to use to quantify the accuracy of the test.

And here are the results of that assessment for the 15 combinations of potential concerns and the tests associated with them. I'm plotting two axes. One is that the good dimension here, the conditional probability of detecting the concern given that exists, that's the true positive. It's the accuracy in the test in finding something that is truly a problem and that is plotted up along this axis. The best test would be one that gave you the 100 percent.

Here's the other consideration in testing. There may be a false alarm. Where is that woman that speaks through that speaker up there and says the beeping was a false alarm. I think that is how we interpret that set. Well that's this other axis and here we quantify the accuracy on that dimension as the conditional probability of a false alarm. We want it to be zero, that means the right-hand into this scale is an inaccurate test, one that has a high probability of a false alarm. And by the way, if we are way down here, if this point were down here, it would also be a test that has very low probability of a true positive.

So tests where the dots fall over here are inaccurate. Tests that fall up here are highly accurate.

Notice, for example, if we take these three ground-water travel time tests, this was Level 1, Level 2, Level 3, no surface disturbing work, bore holes in ESF. Notice that the tests get more accurate as we move to additional testing, in other words, adding the bore holes, adding the ESF. Things get better in that direction, because our accuracy is going up. Both our false alarm rate is dropping and probability our true positive is increasing. That's good.

What about the tests for gas and complex geology? Well complex geology is right here, fairly close to those ground-water travel time tests, but gas had a high probability of a false alarm. Also it had a high probability of giving us a true positive result. So, how do we trade those off? Which is more important? The negative associated with having a high false alarms or the positive associated with gas.

DR. DOMENICO: Why does gas give you a high probability of a false alarm when gas releases virtually is inevitable in all unsaturated zones?

DR. JUDD: Gas release is--

DR. DOMENICO: You are talking about release, is that what you mean by gas? The gaseous releases of let's say Carbon-14?

DR. JUDD: Carbon-14 is exactly what we are talking about here, being released and then moving up through the unsaturated zone above the repository and here, can I call one of my technical experts who knows the answer?

DR. DOMENICO: Yeah, I'd like to hear the answer.

DR. JUDD: Scott Sinnock.

MR. SINNOCK: Dr. Domenico, just the reason that we think it is inevitable it can occur. Remember this is an assessment. If the gas flows is not a problem. What's the likelihood that we will find it is the problem. I think part of the reason we will find that it is a problem is because our expectation is so high that it is. We are likely to interpret that data in a way that would confirm our belief.

So the question was asked if it is not a problem, if the gas flow time is more than 10,000 years, what's the probability that we will conclude after testing we know it is less than 10,000. And so I think part of that high false negative is based on an expectation as you said, that it is less than 10,000 or a high expectation of that. So I think this factors into it that we would be likely to interpret the test result to tell us that the gas flow is fairly rapid even if indeed the gas flow was quite lengthy.

MR. BLANCHARD: Scott, did your group look into the inherent uncertainty in gas phase measurement techniques and the range of methods that are used that may give you conflicting results?

MR. SINNOCK: If you look at this and you read the report we used a surrogate of permeability of the host rock, and we assessed the accuracy of the permeability measures if I remember correctly as a surrogate for the gaseous releases. I don't know if that answers your question. But at the level that we assessed, I'm pretty sure that we just assessed the accuracy of permeability testings.

MR. MATTSON: I'm Steve Mattson with SAIC. And another consideration here is that when made this estimate, we were only looking at the tests that were presently described in the SCP. And I think as we go onto the view graphs you'll see that we recommend a strategy to re-visit this issue right here.

DR. REITER: Bruce, as a point of clarification of since accuracy includes both false positives and true negatives, an inaccuracy is false positives and false negatives, what exactly is plotted on each one of those axes?

DR. JUDD: The true positive probability is this one right here (indicating), given that it is present, the ground-water travel time in this case is less 10,000 or a 1,000 years, what's the probability that the test will indicate that it is. That is what is plotted here.

DR. REITER: And what's on the other axis?

DR. JUDD: And on the other axis is that given that we are okay, in other words the concern is not present, given that we are okay and it is not present, we are okay from a waste isolation standpoint, but the test falsely concludes that we are not okay, i.e., the potential concern exists. That probability, that conditional probability is what is plotted on this axis.

DR. WILLIAMS: Bruce, can I ask you one more question?

DR. JUDD: Yes.

DR. WILLIAMS: I don't understand why you are dealing with accuracy rather than something else, like validity. Because, it is pretty difficult for me to see how the tests that can be performed that are available to be performed using ground-water travel time can be inaccurate. They may be invalid, but it seems to me that accuracy is not what you are after.

DR. JUDD: Well think of this as a--well the way I think of it is for instance an instrument that has an inherent either random--either mis-calibration or random error associated with the measurement, and so even though truth is right here at .006, the instrument registers .004 because of the problem with an experimental design or with the instrument itself--

DR. WILLIAMS: I'm not sure that is what is most likely to happen to you. It's more likely that you will be using the wrong test and therefore getting an invalid result, not an inaccurate result. You can think about that.

DR. JUDD: That's a good point. Thank you.

Now one other thing I need to say about this chart and that is a very strong point that this chart, while it does talk about the relative--these two probabilities, it is not a basis for setting priorities on tests. Why not?

Well two reasons. One of them is listed down here, but let me give another one first. You can manipulate where the dot falls very easily. Either by getting the experiment designer to change what I call decision point. Think of the metal detector you walk through in order to get here. Those metal detectors as you all know can be turned way up in terms of sensitivity or turned way down. What happens if you turn them way up in terms of sensitivity? Probability of a correct alarm goes up. Somebody is walking in with concealed metal, the probability of detection goes way up.

If the person is not walking in with concealed metal, but maybe something else that is dense, because you have turned the sensitivity up so high, you have also created a high probability of a false alarm and so the point plots way up there. As you turn the sensitivity down, the probability of detecting the metal goes way down, and the probability of the false alarm goes down. Eventually, you turn the thing off and it gets down to zero. So the experimental designer can vary where the dot falls on this plot moving this way. So these points aren't by themselves a good indicator of what the test priority ought to be.

Secondly, these dots are just talking about that one term-accuracy. We haven't considered at all, on this chart, the probability that the concern is actually present. In other words, are we looking for a needle in a haystack or are we looking for something that is very likely? We also haven't considered the consequences of the particular PCs or the potential concerns. And as you saw some of the consequences associated with this up here, those consequences are greater than this one up here, certainly greater than this one down here or greater than these over here. So, consequences, as well as, likelihood are important. And so this while it plots the assessments that we got, it's not a basis for making decisions. It should not be viewed as a basis for making decisions.

And that takes us to the last stage of our analysis where we combine the information that we had on test accuracy, the test themselves, with the information that we had on the relative importance probability and consequences, put those all together to evaluate testing priorities.

Now there are three view graphs I need to show you here and let's see how we are doing on the progress through time. It is now a 3:45, which I think takes us up to the top of the scale, and unfortunately, the time right here is when I am going to show three view graphs that have only words on them, no pictures. It's the hardest kind to stay with. So I will implore you, if you will, to follow some of words here because they are important to understanding the results of our study, but I realize that it is coming at a bad time.

First we will define the benefits of conducting the tests, the value of conducting the tests. Not the down side associated with false alarms, but the good side associated with what I call the benefits of the test, and these are quantified, of making some assumption about actions that would be taken if you detect the concern. If the report comes back and DOE interprets it as the concern is present, will some action be taken? We are assuming, yes. We are making the following assumptions. We are assuming that possible actions will be taken and yet we are not specifying or analyzing specifically what actions those are. They might range from mitigating the consequences of the concern to abandoning the site. There is a wide spectrum in there. But we are assuming that some action will be taken, although these were not analyzed specifically.

Secondly, you take the same action, whether it is a false alarm or not. When the fire alarm rang, some people stood up and got ready to go. We didn't know whether it was a false alarm or not. You assigned a high probability that it was, so not everybody jumped. But a few people did and they were taking an action that would be the same action if it were a true alarm or false alarm.

We are assuming that the action prevents, completely eliminates, and this is a strong assumption, it eliminates any of the potential radionuclide releases associated with this potential concern. In other words, because you've detected it you can take strong mitigation measures or you could walk away from the site so that you are preventing any of the releases associated with this potential concern. That admittedly is a strong assumption and it sets an upper bound on the value of the test. It is as beneficial as you can be to that test, because you are giving it the maximum possible value. You ar crediting that test with saving a lot of incremental curies; the maximum number of incremental curies.

Finally we are assuming that the amount of savings are proportional to the expected increase in curies released, in other words, if we didn't catch this thing, some releases would have occurred. We are assuming the benefits were proportional to eliminating all of those releases, and that makes it proportional to sort of the magnitude of that concern. That's one set of general thoughts on how we approach the benefit question.

Now let me be specific and define a new term which I will underline in green because it is the good side of this. I'll underline the bad side of testing with red. Detection benefits measure the maximum value of a test. The greatest value would contribute in terms of detecting potentially unsuitable site condition. It is assumed to be proportion to the avoided releases, which is what I just said about the last view graph, and finally for those of you who are following the equations of how this analysis was conducted, I know Leon for example will do this, so let me mention specifically, the detection benefit is the product of three terms, the likelihood of the concern is this--if you are trying to have a test that finds something, you have to be concerned with what is the likelihood that something exists.

Next question, will the test detect it if it exists? That's this term here. If it exists will we catch it?

And the third term is, how much better off are we if we catch it than if we don't? Well, recall my assumption that we can eliminate the radionuclide releases associated with this concern if we catch it, so the third term of the equation is the expected releases that we can avoid. So it is a probability it exists times a probability we will catch it, times how much better off we are because we caught it. That's our measure of the benefit of the test.

The units are going to be expressed in radionuclide releases avoided divided by the EPA limits, so if it were equal to one, we are just exactly saving something that would have equated to the EPA limits.

That's a lot of words and no pictures. So here's the results of quantifying that. So this is again another set of results that are going to show the relative benefits of tests for each of these. This is the same set of concerns you saw before cut in half, gas, complex geology, etc. I have now added to each of these the level of tests that was being done.

So this is complex geology tests being done in bore holes, complex geology tests being done in the ESF. Ground-water travel time, you were told there were three.

The vertical axis is the ability to avoid releases because we've detected the potentially unsuitable site condition. As we get up to the top, we went all the way up to a value of 1 or 10^0, we would be at a benefit equated to avoiding the releases associated with just meeting the EPA limits. We are below these on all of them. The green dots are the detection benefits that I just defined for you. And notice that they vary by many orders of magnitude and again we see our grouping of complex geology and gas, and then another set here and then a third set down here.

DR. LANGMUIR: What happened to human intrusion? It disappeared.

DR. JUDD: Good question. Out of the 15 things--

DR. LANGMUIR: It sounds like you can't measure it.

DR. JUDD: Human intrusion was the one where you intercept the canister directly. And, once we started assessing the tests for that, the workshop group judged that that is really a consequence of natural resources. Exploration for natural resources has as one possible consequence intercepting the canister. It also has as a possible consequence disturbing the hydraulic system, the geohydraulic system.

But we combined human intrusion related to the direct intercept of the canister as a consequence of this natural resources, and so it is part of the benefit here.

DR. DOMENICO: Possibly the majority of the benefit too?

DR. JUDD: Yes, dominate as I recall But that has the greatest expected contribution to releases, or at least a large one relative to the other consequences of that potential concern.

Okay, now I have sneaked in something new on you. This vertical axis is similar to what you saw before. It is measuring the benefits of the test according to what it can save us in the way of potential releases, because we are avoiding those releases because we've detected the test. But what have I done over here? I keep saying what I've done, I'm trying to personalize this--it's the core team that is doing all that and I do that all the time and I apologize. It is certainly all of us doing the analysis that have done this. And that is we have translated using EPA's number straight from Part 191.13 to translate avoided releases into avoided excess cancer deaths over a period of 10,000 years. If you avoid one times the EPA standard which would be one decade above this chart here, you would avoid something like 1,000, in the order of a 1,000 which was the number given earlier today, cancer deaths. Now in fact the 1,000 I think was for 100,000 metric tons. We have 70,000 metric tons of heavy metal in this repository so we actually used the number 700 equivalent to 1.0 times EPA releases.

So if you are concerned then with these levels of tests down here, what's the maximum benefit provided by those tests? Well it can avoid 10^{-3} cancer deaths over 10,000 years or that's about 1 cancer death every 10,000,000 years I think was the number.

I'm trying to put this in a perspective that relates to cancers, but of course that involves accepting EPA's numbers here and there are some people who don't accept those. So you can either think of the benefits of the tests on this vertical axis or that vertical axis.

One more slide with no pictures on it to define something that is very important and that's false alarms. Is it possible that the test will tell us there is a problem when there isn't? And if we have one conclusion that kind of jumps out of this study, it is that you can't ignore these false alarms. Thet potentially can occur, especially when you are in some cases looking for a needle in a haystack you may come up with something that looks like a rusted needle, and in fact it is not.

So let me go through carefully how we analyzed the false alarms. These are what we call false alarm costs. They are associated with a test and they are costs of actions that might be taken to mitigate or abandon the site or mitigate the consequences that were taken in error. These actions were taken unnecessarily because in fact the concern is not there, but there was a detection of it, and so in the context of a fire alarm we evacuated the building unnecessarily. In this case we are taking some other action to mitigate the consequences.

Why is that a problem? Well the consequences might be costly, or abandoning a perfectly good site can be very costly. So there is a potential cost there.

We are assuming that the consequences of that false alarm are proportional to the importance of the potential concern. In other words you've got a potential concern that looks like it has a very large number of radionuclide releases associated with it. We are saying that assuming you take greater action in that case than if it is a small potential concern in terms of its contribution to radionuclide releases.

Finally, here's the definition. It is very similar to the one before. It's the probability that the PC isn't there times the probability that we catch it even though it is not there times the amount of curie releases or radionuclide releases that have been avoided unnecessarily.

It's very possible in this scheme of things that when you weigh the benefits and the costs of a test, you find that the false alarm costs are greater than the detection benefits. These costs might exceed any benefit because you are looking for a needle in a haystack, you are very likely to come up with a false alarm, very apt to take a potentially costly action in response to that alarm, and it is all being done unnecessarily. So this is an important conclusion from this study.

So I am now going to plot some of this same information we've had before and add to it the dimension of the false alarms. I'll try and do this with color and I am going to do it in sequence making the diagram a little more complicated in each step. But I'll start out with a fairly simply diagram that on this axis, and this axis, plots exactly what you saw on the last chart, it's the benefit associated with the detection: either measured in releases, that have been avoided because you fortunately detected the concern; or, over here, cancer deaths avoided, because you've accepted that number in Part 191.13.

Down along this axis we have introduced something new which is the false alarm cost as I just defined them. If the false alarm costs are high associated with a particular investigation, it plots over here to the right. If they are low it plots on the left. As you look at the chart, what do you see? Well, just as the gas concern had a high probability of a false alarm, therefore it falls to the right of the chart. So if we compare what we saw previously, we are seeing a similar pattern in some respect.   Let me show you that one chart here.

If you recall, gas was up to the right. Gas is up to the right again. High benefits, but high false alarm costs. Complex geology was way over here on the left on this diagram, but it has now migrated to the right side of this diagram. Why? Because we took into account a couple of more factors. The probability of occurrence and the consequences, by taking those things into account, shifts this to the right of these diagrams. And so the complex geology are right up there with gas in terms of both the false alarm costs and detection benefits.

We see a group of tests in the middle primarily related to ground-water travel time. There is a set down here, then, related to perched water and volcanism and a couple of others in both of those categories. This is the result of our study plotted where we showed test benefits and test costs. The evaluation was based on all the workshops that we've held.

What do we see in this? Well let's draw some more lines on the chart. I am now going to draw some boundary lines in between that group and this group in the middle and a boundary line here (indicating), separating the lower priority groups.

Russ mentioned in his introduction that we had identified three test priority categories and here they are and I've drawn somewhat arbitrarily that line in between the two. It could go anywhere from here up to about five orders or magnitude higher.

These lines are based just on the detection benefit and they ignore false alarms, but this is an important exercise. It's important to draw a line here and line there because there is certainly a distinction between the benefits of those tests and the benefits of these tests and analogously another distinction down here. We call this a value judgment where you draw that line.

And we say in our report that this is a value judgment that needs to be made by those who are setting priorities on tests because you draw the line here and say well that clearly distinguishes these two, or if you draw the line down here, the next step might be to say, well then, which test shall I conduct? These clearly have higher benefits for detecting pontentially disqualifying site conditions than do these, or I might go down to this third category and include these, even though they have the lowest on the chart.

How many deaths over 10,000 years are we preventing by doing the tests? In other words by detecting these unsuitable site conditions and being able to mitigate them. Let's see these tests here are about 10,000,000, one death every 10,000,000 years. This is a death in some number of billions, one billion years or so. These numbers are very small.

DR. LANGMUIR: Bruce, what this suggests to me is the need for public education, because clearly the public is way down in those items on priority 3 in terms of their concerns in many cases. So educating them as to what the real risks are is a major part of proceeding with this.

DR. JUDD: I agree with that, but we don't want to lose sight of the fact that there are other reasons for testing besides detecting potentially unsuitable site conditions.

DR. LANGMUIR: Yes. Yes.

DR. JUDD: And those are other motivations for doing some of these tests down here and a strong motivation in some cases. But, yes there is a prioritization here.

DR. LATHROP: Why did you decide to bring the categorization independent of the false alarm costs?

DR. JUDD: Because I'm doing this sequentially, and I am going to get to that on the next slide. And John was actually paid off to ask that. That was John Lathrop.

DR. DOMENICO: Bruce, one more question on that.

DR. JUDD: Sure.

DR. DOMENICO: How do you translate complex geology into specific tests? That is probably not a question for you to address.

DR. JUDD: Well, we had the list. He's getting a clue as to what I know and don't know. But this was that first level of tests related to the complex geology. And there were those things going--trying to understand three features that could potentially cause us serious mis-estimation of the releases.

MR. BLANCHARD: Bruce, let me help. If you go to Volume II, Appendix D, around page 124 in that general section, you will find a more detailed, itemized list of the specific SCP tests that go with the category of complex geology and there is some text there and some other information that associates that.

DR. JUDD: Thank you, Max. Appendix D has each of the assessments and the technical discussion of each one of them. All right let me move fairly quickly then to the point that John brought up.

The conclusion from this was there are decreasing benefits associated with these tests as we move down the chart. On the other hand as I said, as I gave the caveat, there may be other reasons for conducting the test besides detecting potentially unsuitable site conditions. Nevertheless, we have to watch out for false alarm costs and this chart here which then gets us--this is as complicated as it gets, now has something drawn on it related to the false alarm costs. It starts from--this extra graphics that you see on this slide, start from a diagonal line. I'll refer to that as the diagonal line. The diagonal line corresponds to a judgment and this is a judgment. A judgment as to the relative weight one wants to give false alarms relative to the weight one wants to give detection benefits.

So, for example, if I said the consequences to me as a decision-maker, if I am a decision-maker, the consequences of accurately detecting a concern are equal to the consequences of inaccurately detecting a concern although there is minus sign in one of them, in other words the consequences of a true detection are equivalent to the consequences of an inaccurate detection, I don't weight one any more than the other, then I would draw the purple line, and what that would say on that purple line is some of these tests, in particular the complex geology test up at the top is borderline. Some of the tests are borderline. Their false alarm costs are roughly equivalent to their detection benefits and we should very carefully consider whether we want to conduct those or not, because they are right on the line.

Any of those tests to the right of that purple line are ones for whom the false alarms are more likely or more consequential than the benefits of the tests. It is like looking for a needle in a haystack. And so, the detection benefits are outweighed by the false alarm costs. If they are to the right of that diagram, the decision-maker wouldn't conduct the test for the purposes of detecting potentially unsuitable site conditions. The ones that are to the left of the line are the ones that would be conducted because their benefits are greater than their costs. And this line simply represents a point of equality between benefits and costs.

If a decision-maker instead says, no, I think the benefits of accurately detecting a concern outweigh the false alarm costs by a factor of 10, it's ten times more valuable to me to detect a concern than it is costly to detect it inaccurately, then you draw the line here. And now notice that these complex geology tests are both worthwhile in this sense. Some of these down here related to ground-water are worthwhile. Others are still on the not worthwhile side and some of these are to the left of that line.

So there really are two value judgments that are required here. One is how low do you test in going down the diagram, and the second is how do you react to false alarms and how do you weight those?

DR. LANGMUIR: How accurately are the points known in their locations? If you are fiddling around with the point itself, you can move it around.

DR. JUDD: Exactly. And if you used a different set of individuals, you might find a different--it might fall on the other side of the line. But what that says to me is that this judgment about where I draw the line is an important judgment because some of these close to that line. And I might want to do some reassessment of the numbers that lead to it being close and I also want to think very carefully about where I draw that line.

Here's an important conclusion of our study. Anything that is to the right of the line has a potential false alarm cost associated with it, and for whatever we are conducting that test, because: we have to; or because it's pre-closure related; or because it provides scientific confidence; or whatever. Whatever reason we conduct the test for, we have to be prepared for a false alarm. And we have to have a strategy for dealing with tests that are conducted for other reasons, reasons other than detecting unsuitable site conditions, but can have false alarms. We need a strategy for dealing with those tests.

DR. ALLEN: But of course that is based, I presume, on the judgment of these nine individuals--on the basis of what they know about the test that projects the tests. We could learn something about a new kind of test in the next six months that could radically change some of those.

DR. JUDD: That is true, with one caveat. And, that is that the term that drives this result in many cases is its location; this way which is determined by the importance of the PC itself not by the accuracy of the test. Where we fall this way is determined by the accuracy of the test. And what this might say to me is let's look at those ground-water travel time tests and let's see if we could do something so we reduce our likelihood of a false alarm, because false alarms apparently are our real concern.

DR. REITER: Bruce, a quick question.

DR. JUDD: Yes.

DR. REITER: Why did you use unnecessarily avoided release as a cost? Couldn't you translate that, I mean, couldn't you use some sort of monetary value, the cost of abandonment? The cost of an unnecessary mitigation?

DR. JUDD: Yes. We think the logical next step for getting at this important trade-off is to quantify what these costs are. And we got to the point in our study where that was clearly important based on these results, but we hadn't taken that next step yet. So we recommend taking that next step to develop a logical structure for assessing this trade-off and it would involve identifying what some of those costs are.

DR. REITER: That could rearrange the order of the false alarms. If one assumed, because all you--if I understand, unnecessarily avoid releases, just the exact same number you got from avoided releases, right?

DR. JUDD: Yes.

DR. REITER: That element is the same. However, the cost which may be monetary may not scale directly with the avoided releases.

DR. JUDD: That's exactly right and yet these weights reflect the judgment as to what that scaling is. Drawing this line here or here, using one-to-one or ten-to-one weights reflect the judgment of that scaling. It's how do you scale false alarms versus detection benefits. And we are recommended that you should take the next step and develop that scaling algorithm a little more carefully or a little more in detail than has been so far. Here it is simply treated as a weighting.

DR. DOMENICO: Bruce, that complex geology contains nine or ten items that have been lumped together as complex geology. Is it possible that each of them by themselves if you were looking at them would give a different shift in that point there? I'm speaking in particular to a thing like the net flux of the unsaturated zone which is very, very important. And we have great expectations that it is going to be small. Does that mean that there is a large probability that we are going to get a false alarm because our expectations are--no, it does not necessarily mean that?

DR. JUDD: This was aggregate for the whole package.

DR. DOMENICO: Aggregate for the package.

DR. JUDD: So, there are a few in there that are excellent tests for particular things, they are being, and have low probability of false alarm, they are being penalized on this diagram because they are lumped with others.

DR. DOMENICO: Okay.

DR. JUDD: On the other hand, if they will be done as a package, then it is appropriate to lump them because that package as a whole will give this high probability a false alarm.

DR. DOMENICO: But you do say that each one of these if you did them independently would have their own point on that curve and some are being penalized because of others.

DR. JUDD: That's correct.

DR. DOMENICO: Which ones?

DR. JUDD: Now we did one other thing that illustrates another point, and that is, this ground-water travel time is in some sense a special case along the EVS performance and a couple of other concerns. Because there is a performance objective, an NRC performance objective that says a 1,000 years has to be independent of the EPA releases, and so what happens if we elevate the importance of ground-water travel time? From an analysis as you saw on the last view graph, where consequences of ground-water travel time are measured in terms of radionuclide releases, just as are the consequences or everything on this chart, what happens if we elevate ground-water travel time to say, it has its own NRC performance objective, it's a disqualifier in 960, let's elevate its importance and see where it falls on the chart.

Now how did we elevate the performance? We equated, we equated missing the ground-water travel time objective. In other words having less than 1,000 years. We equated that in terms of consequence over here to missing the EPA performance limits. What that does is increase this by a factor of 600, and does two interesting things. It drives us above into priority category 1 above that arbitrarily drawn line there; of course, if the border line had been drawn up there we would still be in priority category 2. But it drives us up and then the second thing is it also magnifies the false alarm consequences, and therefore shifts us straight over that way, straight up this way, straight up this way (indicating), because we are just magnifying the spread among these three tests.

What do we see? Well, again we see that we are on border line on one of these. The other two are to the right. And this produces an insight related to the way that test results are revealed. If all three of these tests were going to be conducted anyway, these three levels of testing, this cautions against making conclusions sequentially as opposed to waiting until, if you are going to do the ESF, waiting until the ESF results are in. Because here, because there is such a high probability of false alarm we are way over here to the right, whereas if we had the more accurate tests that get into the ESF type of testing, we would be a little farther to the left. So you can use the analysis to derive insight such as this.

Okay, that's the conclusion to the graphs. Let me summarize now what we've learned from the Phase 1 of this study.

We've produced a list of the potential concerns and their relative importance, identified some priority tests for those important potential concerns. That in itself we think is valuable. We've also identified an area where there is a current program strategy with respect to Carbon-14. That may need to be re-evaluated because that gas-radionuclide-Carbon-14 comes out so high on the chart. We may need to re-evaluate that strategy.

We've shown that the priorities are not absolute. They are dependent, sensitive to, if you will, some of these value judgments. We've gotten some insight into the potential significance of the false alarms. And let me emphasize again, that if tests are conducted for any reason there is that possibility of a false alarm. So, we need a strategy for decision-making that takes into account that potential for false alarms.

We did the sensitivity for ground-water travel time, and finally we provided a tool that has been applied once, could be reapplied with much less effort as we go through the testing program site characterization, for the purpose of revising priorities as Russ mentioned in his introduction. I feel those are valuable products.

Let's repeat though the caveats about what we've done and what we haven't done. The potential concerns were related in PACs and disqualifiers. They did not include the qualifying conditions and in particular they did not look at pre-closure conditions, all post-closure PACs disqualifiers affecting waste isolation. We've taken just a subset of the SCP test, those related to "early" detection, then we've used just one performance measure and it is the EPA standard, although we have done sensitivity to the NRC performance objectives, and finally, our information source as Russ mentioned with subjected judgments.

Let's not forget the other reasons for testing even though we focused on early detection of potential unsuitable site conditions, pre-closure issues, design and construction issues, information required for licensing, all of those being strong motivations for doing these other tests, initiating long-term performance confirmation tests, facilitating tests, building scientific consensus. When you roll those into the analysis, priorities may get revised.

Two view graphs if you will on Phase II of this study. The task force recommends the following actions based on Phase I. Those judgments as Leon pointed out, those judgments are important and we need more structure. In particular related to that trade-off given false alarm costs and detection benefits. We need more structure there. That's the recommendation.

The second recommendation is use some of these results in the process of setting priorities. You need to look carefully at Carbon-14 and, depending on the trade-off one makes between false alarm costs, we need to look carefully at complex geology. There may be other criteria for test prioritization such as the pre-closure costs of the tests, etc. Those things were not included in this, and they could be wrapped into the analysis.

Phase II of this study I will explain on my final view graph and we recommend completing that, and I'll show you what that is in a minute. Jean Younker is going to be talking about, if I ever get out of here, she's left. That's what happens. That's not the first time. She's given up on me.

She will provide results and insights to the effort on site suitability, and finally as information comes along, not only the method here, but the numbers, the assessments that have been made can be used to re-prioritize tests without a whole lot of additional assessment. So, we recommend that.

Phase II just to give you an idea of the ways that this scope could be expanded in Phase II, two basic ways, one is expanding the criteria and the other is expanding the assessment analysis. The criteria that we use did not include the costs of tests or other reasons for testing and those could be folded in. The expanded assessment and analysis could go in the direction of a broader range of experts. Something that we have found to be very important is these assessments are difficult at best. They often had to be assessed up at this level, what is the ground-water travel time?

By developing a more detailed model and making assessments such as the effective porosity, the flux, the distance for ground-water flow, assessing down here and computing that, the assessments would be a lot easier. And, you may recall that we reported to you last July 24th and 25th, that we were in the process of doing that and had found those to be much easier to assess for the experts. That effort has been continuing in parallel with the effort that we described today, but it is not complete yet and that is why we call it part of Phase II.

And finally, using a total-system-performance to compute performance of the system from these lower level assessments would be a significant, in our opinion on the task force, significant improvement in our results.

So, that concludes our presentation.

DR. ALLEN: Bruce, can I ask a question? Perhaps the results of this analysis based on approval of the task force comes up with some concerns, the priorities of the public perception of picking the stated tests , is there any way we can test whether or not, and for some reasons citizens of

Nevada and perhaps the nation will not believe it because of the fact that all of these people are associated with the project, all of them technical people, is there any way we can test whether--if we went out to a group of nine technical people, science and engineers in Las Vegas or Nevada or the country as a whole and did the same thing, whether the results would any different and whether we could use that to reinforce this as something that should be believed, that this kind of study should be believed by the public?

DR. JUDD: Good question. You are saying can we test it, can we test with actually doing it. I mean one way to test it is to do it. And that is not that hard if we could identify the right set of people.

DR. ALLEN: Except that that involves an awful lot of their time to come up to speed on all these technical issues. That's the real problem I see in it.

DR. JUDD: The other way that is in a small way a test of that, and we did this for instance on volcanism, we had Bruce Crowe as one of our experts and we asked Bruce not only to give us judgments, but to give his judgments on what the others would say if they were in the room, other people that weren't among the small group of experts that we had together. So, that provided some sort of sensitivity analysis, because we had that person's judgments of what the other people are saying. We did that and it didn't affect volcanism very much, but I think that takes us a little bit along the way towards trying to assess what would be the effect if we had the others in the room.

DR. ALLEN: Well, if this kind of analysis is correct, and I have no reason to think that it is not valid, it would certainly help us impact if we could somehow demonstrate to the public that indeed it was representative and didn't represent somehow a biased viewpoint of people associated with the project. I don't know how to do that.

MR. BLANCHARD: Clarence, let me help get Bruce off the hot spot so to speak. We have been quite aware of your all's encouragement for us to incorporate additional outside experts, people outside the program and involved in these kinds of activities. And as a consequence, the next speaker will describe our early site suitability evaluation process. And in that process I think you'll find two things. One, important aspects of what Bruce has just described in our test prioritization will be incorporated in that process. And, the plan calls for a complete peer review with outside people, people not associated with the program. Experts in each one of the disciplines that are selected.

Now, we are not taking the whole suite of tests that come out of this test prioritization, because, you can't necessarily link those to the early site suitability evaluation in that direct way. But some of the tests, especially the ones that are easily seen or recognized that there is a leakage, that will occur in that process and there will be some outside peer review. Now, how large that group gets is up for discussion at this point, but clearly if the group gets too large it becomes unmanageable to get the product done in a reasonable time frame.

DR. ALLEN: Well I am well conscious of the fact that when you start bringing outside people in it takes two years to learn the acronyms much less the what they mean.

MR. BROCUM: In a subject like this, when you are kind of addressing, when is enough enough, when is it not, you could actually structure something with outside people in an area where you have lots of time. And in the when's enough enough issue, we have a lot of time, because you know, we haven't gone in the mountain yet. So, you could in that area and we have had a little bit of informal discussion on this of following up something like this on filing something like this of perhaps using outside groups, okay. So, your comment was an excellent comment. But it would be nice to find a subject like this one where you are not on a very short time fuse so we can keep the program itself going.

MR. BLANCHARD: Also, there have been ongoing studies where people have been looking at suitability of Yucca Mountain for evolving their own independent view and as you are aware, you have probably listen to Bob Shaw discuss the

EPRI effort which did not use the staff from our project. And he did that for the utility industry. And so I think the utility industry was interested in their own independent appraisal with respect to what do they think of the site and its capabilities to perform in waste isolation. And we have been very anxious to watch that product evolve, and have been very pleased to see that it evolved in a way that was not inconsistent with our early understanding of the site's capabilities too.

One other thing if you look at the view graph number 41 which Bruce discussed in terms of the recommendations, the first three steps, we are very interested in doing relevantly soon because in order to determine whether or not the department should change it's technical baseline, in other words, go to our change control board and make a formal modification of our baseline, we need those three inputs, plus some additional ones to support the defense of a change in the baseline. And so we will be interested in pursuing those recommendations in an attempt to determine how should we change that baseline, what studies plans should we change or should we create some new ones?

DR. LANGMUIR: You've already gotten the EPRI team in place at the point. I gather that they are not working for EPRI, but that team of folks would be an obvious group to look at this issue and this approach and evaluate without having to get up to steam. They are at steam right now. That would be an independent group that would be suited.

MR. BLANCHARD: Don, that's an interesting observation. Both you and Clarence know that it takes a good bit of homework in order to get to a point where you are reasonably convinced, you could move out and make some intelligent application and interpretations on such a complex subject. And I certainly agree with you that it would be a big benefit to have people that have gone through that process on their own independently to factor them into a peer review process because the time involved would be much shorter.

DR. JUDD: Dave Dobson mentioned about the public concern and that is that we've used a single criterion, EPA performance limits. And if we used a different criterion it might reflect more the public concern that might change the results, hence the recommendation here.

DR. DEERE: Well, at our meeting in June, isn't it Russ, on the testing, will we be getting into some of these tests themselves and the relations that you might be considering? This was sort of a selection of tests and priorities. We really haven't got into--

DR. DOBSON: I don't believe we've started negotiating the June agenda yet, so I can't say what we are going to talk about in June. We are certainly willing to talk about aspects like that.

DR. DEERE: Well, you see we've come almost two years in our activities and we really haven't heard a test yet. We really haven't discussed-- a lot of them we are really not in agreement with.

DR. DOBSON: I'm not quite sure--

DR. DEERE: Not necessarily.

DR. DOBSON: I'm not sure what--like I say, we are just beginning the process of trying to figure out what it is that you would like to hear in the testing meeting.

DR. DOMENICO: We have an agenda set for that.

DR. DOBSON: Okay.

DR. DOMENICO: We are very clear what we want to year.

DR. DOBSON: Good. That's okay. I'm not sure--did you say you were not in agreement with some tests? Did you make that statement or did I hear something?

DR. DEERE: It could well be.

DR. DOBSON: Oh, it could well be.

DR. ALLEN: The way we understand it now, the answer is probably yes. We are not in agreement.

DR. DOBSON: We'd be perfectly happy to talk about that and any aspect of the testing program in the June meeting.

DR. ALLEN: But I think we would have to make a very specific aspect and go into some detail.

DR. DOBSON: I agree. When I looked at the title of the meeting, testing, I was a little at a loss as to where I was going to go. You know, that's a big topic.

DR. DEERE: Well, you know it does have some specifics in it.

DR. DOBSON: I'm sure it does.

DR. DOMENICO: Actually, I think we are thinking of this as a follow-up to the December 18, 1989 meeting that we had where we listened to people talk about isotopes and we listened to all the measurements in the unsaturated zone and all of those sorts of things. We know the actors at that time and those are the kinds of--I believe those are the kinds of things that we will be requesting.

DR. DOBSON: Now, that sounds like something that is potentially very useful. And I think you are aware, Pat, too, that we have done a review of at least the unsaturated zone testing program since then, and so, we have given it some more thought.

DR. DOMENICO: We'd like to come to some ideas of closure on something of that sort.

DR. DEERE: Some of the concerns that we may have also may not fit into this early--into the prior possession and maybe in some length but not necessarily involved in these sections.

MR. BROCUM: Are we ready to move on?

DR. DEERE: Yes.

MR. BROCUM: The last topic today, I guess, site suitability and it is also a new topic I think basically for us and for the Board.

I'm going to give some of the background and status of what the core team has done to date. When the Secretary reported to Congress he committed to an early focus on the evaluation of site suitability. I think Russ actually put the words up a little earlier on the substantive words the Secretary used.

In order to comply with the Waste Policy Act and its amendments and 960 and the Secretary's commitment we see that we need two kinds of evaluations. We need the early evaluations, interlude perhaps, the kind of evaluations where we focus on conditions that might make the site unsuitable. And that responds to the Secretary's admittance in November '89, and then later on more comprehensive evaluations that may lead up to decision to recommend a site for development of repository or perhaps recommend disqualification.

When we wrote the site characterization plan, we kind of envisioned the comprehensive site suitability evaluation and envisioned early error of that waste site suitability. At that time we were taking of a site characterization period of three to five years.

In October of 1990, John Bartlett assigned our office the lead for developing an approach to the evaluation of site suitability. Leading up to the October letter from

John Bartlett, we had a meeting among staff in Nevada. In Nevada we have a very small workshop that we try to line up the basis of what we want to do, and then Bartlett wrote his letter. And then we had a workshop in Albuquerque, there must have been about 100 people, two and a half day workshop, to review the status of site suitability. The DOE of course was there, EPRI and Golder made presentations on the status. NWTRB had observers, Leon and I think Russ were there. NRC had observers at this meeting. This was not an official public meeting, but it was an open meeting.

The objectives of that workshops were to obtain input from the attendees about site suitability concerns and methods and there was a lot of open discussion. And, to begin developing an approach for evaluating suitability or non-suitability of the potential probably waste repository site.

Many issues came up at that meeting and have been discussed subsequently by the core group and these are the kind of issues that are still in discussion to various degrees.

The role of our siting guidelines 10 CFR Part 960, what is their role? We are using those guidelines to which we are doing early suitability evaluation. The use of suitability criteria and considerations of residual uncertainty, this is a concern that Bartlett expressed, he expressed it at that meeting in Albuquerque and he expresses it over and over. He is very concerned about this, that we will ever be able to reduce residual certainty enough. I think that is one of his major concerns.

What assumptions should we make regarding engineered barriers in terms of early site suitability. When you are doing total system performance assessments, you obviously need to have source terms. When you are looking at a site early, what's the role? There was a lot of debate at that Albuquerque workshop and I would say there was some disagreement. That's a fair way to characterize it. Some people thought it was adequate and other people thought, no you didn't necessarily have to. And think these are under consideration by the core group.

What is the role of performance assessment in early site suitability evaluation. When you are looking at disqualifiers how much performance assessment do you need? How soon can we incorporate total system performance assessments? That's an incorrect word, refinement of test prioritization task group, how can we use their results in our early site suitability? What role should expert judgment have? This is not only a concern within our group, it is a concern at two levels. We have had quite a bit of experience now, in the last year and a half, I think we ought to learn from our experience.

What is the relationship between suitability and licensability? The main point here is that suitability is DOE's responsibility, licensability is NRC's responsibility. However, we don't want to be in the situation where we can find the suit suitable and then NRC is likely to find unlicensable. We always have to keep the licensability issues in the back of our minds.

After the Albuquerque workshop, we made a decision in December 1990, last December to this year conduct an early evaluation of suite suitability. We were directed by John Bartlett to implement a plan to do two things. One was to develop a general approach to the evaluation site suitability, and the second is to make an assessment of the suitability or non-suitability of Yucca Mountain.

One of the things that we discussed in the Albuquerque workshop is that the last time we have made an assessment in any kind of formal sense was in April of 1986 and that is five years ago, or will be five years ago soon. And we perhaps ought to look at the information we've got since then.

The last bullet that this evaluation will be one component and the decision-making process reflects the fact that site suitability transcends, the Yucca Mountain Project, transcends the Office of Geologic Development. It's really an Bartlett and a Secretary level issue. There are many other factors besides this study. There's public factors. There's interested parties, there's budget, there's schedule, there are many other factors that Bartlett or the Secretary have to consider. So this study will go to that and in part for that consideration.

At the end of January, DOE approved a management plan for conducting this study. That plan has been issued. It includes a scope of a study and a schedule of activities to be conducted. It has responsibilities. The responsibility for conducting that study was given to T&MSS and they were asked to produce a detailed implementation plan. The plan was produced. It was approved by DOE. It includes things that this QA--has a lot of details of exactly how the study will be conducted.

The schedule which I am kind of a little bit hesitant to show because we are under a very tight schedule, with all the other things going on, is to define what I call the general approach. That is the exact wording that John Bartlett used in the letter and that should be done by May of 1991. John Bartlett would like to go public and discuss this sometime in the summer. I think we have a milestone date down of about August 15th, to submit to management the results of the early site suitability evaluation on or about June 1991, to put that in peer review, that Max has mentioned, outside peer review and complete that on or about November of '91, and to submit this report to the Director of the OCRWM on or about early 1992. That is our current schedule.

I think Jean has perhaps a more detailed scheduled. So I think Jean now is going to talk of the activities to-date of the core team.

DR. DOMENICO: Who is going to do this work, Steve? Who is going to do this work?

MS. YOUNKER: I am going to tell you.

DR. DOMENICO: You are going to tell me.

MS. YOUNKER: Yes.

MR. BROCUM: The actual day-to-day work will be done by a core team which is ran by Jean Younker.

MR. BLANCHARD: Leon, perhaps I need to point out that the contract with the project office is one that includes Westinghouse, Harza, and SAIC. And so the complex there in Las Vegas in the Valley Bank Building, although it is mostly SAIC staff, there are some Harza and Westinghouse staff in support of that contract. And that is why it is called Technical Management Support Services contract.

MS. YOUNKER: Right. That was another acronym that got by you Leon.

All right, thank you, Steve.

Okay, what we will tell you about will be the general approach that we've laid out given the background that Steve's established, the status of the task, and as you saw on Steve's schedule, we are on a fairly short fuse here. And there is a really good reason for that I believe, but the way we have defined it, I think you will find what we have defined the scope to be is doable in this time frame. And I'll tell you the plans of the detailed activities through 1991.

If we step back and look at the general approach and for those of you who were at the Albuquerque meeting, you'll recall that John Bartlett used a view graph kind of similar to this one which we have found evolved by thinking a little bit, and added a few things that makes sense to us.

Steve told you the general guidelines from 10 CFR Part 960 are the basis for the frame work for this early site suitability evaluation. And I am going to be mostly talking today, even though every now and then I'll broaden out to the total comprehensive suitability evaluation. Everything that we are talking about except for the general approach that Steve mentioned is really now addressing this early site suitability evaluation product.

But the general picture that we have in our minds, in the minds of the core team and the DOE people that have helped us set this up are that in order for us to use 960 and apply it to Yucca Mountain, we believe we are going through kind of a thought process that we have defined here for you as developing a site specific technical approach.

As you recall, the last time we talked about this, in fact I think it was during the discussion on test prioritization, when test prioritization still had suitabilities as a component, we talked about the fact the guidelines are really in some cases quite general because they were meant initially for site comparison, although it is very clear when you read them they are also meant to be the basis for your final recommendation of a site for repository development.

So, when we look at the guidelines, guideline by guideline and I have a list I'll run you through in a few minutes or at least pick out some examples, you look at the qualifying conditions and the disqualifying conditions, you will find that in order to talk about them and decide what it is you will do with each one, you do something that we've captured as a site specific technical approach. And clearly part of that, as you'll hear me say is it kind of is thinking about a guideline like dissolution as an extreme example. You know dissolution is in 960 because of other type of media besides the one we now have given the act of Congress that amended the NWPA that chose Yucca Mountain as the site to be characterized.

So, when we look at one like dissolution, the site specific approach that we would take on dissolution is different than one such as hydrology, geohydrology, as you have talked about today. Geohydrology is clearly one of the ones that is a site specific technical concern for Yucca

Mountain. So, the attention we are going to pay and the way we are going to look at that one, the method will define for evaluation of early site suitability will be quite different. So that is what the thought process is that we are going on. We are using for each guideline, we think about what is it about this guideline, about this specific technical concern is really a potential concern and even a potential disqualifying factor for Yucca Mountain.

Okay, given that you have a background of information that is built up back in the EA and of course of the site characterization plan pulled a broader base of information together, you go into this phase that we are calling early site suitability evaluation, and from here on, as I talk through this, it isn't probably too different than the way you would visualize this general approach that we will produce as one product of the team effort. Because I suspect it will look a lot like this, but with some further thought.

You'll ask the question, is the site suitable and, of course, you are going to do that, given that you are using 960. You are going to do it over and over because 960 is set up so that: if you don't meet one of the qualifying conditions; or if you have one of the disqualifying conditions; then that is an out for the site. That is essentially it. It's a yes/no on any one of them. So you must be able to meet every qualifying condition either now or eventually, and the same thing for the disqualifying conditions, no one condition has it. And the same thing is true of course with the system, those who are not familiar with the system performance, both pre-closure and post-closure, is captured in qualifying conditions. So that term qualifying conditions applies to total system performance for both pre and post closure as well.

All right, you asked the question is the site suitable and clearly there are some different answers that can come out of that. One would be yes, you really right now have all the information you need. You can support a higher confidence finding on every qualifying and disqualifying condition then you go ahead--this team would recommend to the DOE that they have the basis to recommend the site.

Another outcome is that the answer you get when you ask that question is no, and in that case you then would ask the question, should additional data be acquired, because you don't know if the site is suitable, but you also don't have a definitive answer. So, you say, should additional data be acquired? Well, if in this case they are wrapped up in this decision as a whole of further thought because if the answer is no, then you have to ask the question--if the answer is the site suitable is no, then you have to ask the question is there additional data that I can get that will help me to answer that question in a more definitive way.

If the answer was no, ultimately then you are facing one of these determinability issues or the idea that you will never be able to gather enough information about this particular site and therefore an element of the disqualification or the abandonment of a site is because you don't think you will ever be able to gain the confidence that you need to make the positive findings.

DR. ALLEN: I don't quite understand the timing here. Is the site suitable? Well no one is in the position to say yes as of next January, are they? I mean haven't we already agreed that a characterization program is necessary?

MS. YOUNKER: Certainly there is a wide range of agreement that some kind of site characterization program of the site makes sense, but if you were to take the message that you've heard in some of these task forces that we are on the order of 10^5 or 10^6 or 10^7 better than the EPA standard, then you have to come back and ask yourself a question, why am I doing that characterization which is kind of what you heard Dave Dobson say in his management, I don't know what you called it, what we learned about the study. Basically, the answer is clearly we know we are going to do some site characterization. Clearly we don't think we don't think we are going to yet, but on some of the specific guidelines that I'm going to tell you about, it is conceivable that we don't need any further information about a particular area.

DR. ALLEN: Oh, on certain areas, yes, but is anyone prepared to say by next January that we are ready to go to the Congress in the country and say this site is being recommended by the DOE, beyond our wildest--

DR. DOBSON: I think there is a low probability of that outcome as a result of this study.

DR. JUDD: Small negative exponent.

DR. DOBSON: No, I mean I think there is substantial uncertainty in some technical areas, and certainly we've written a site characterization program which we intend to pursue but as Gene noted and as part of the reason is you can do this as a whole site or you can take this on an issue-by-issue basis. And we do think that there are some issues where we are rather close.

DR. ALLEN: Don't you think there is a 99 percent probability as of the moment and perhaps as of January, that we are going to come down to that second prime and we'll ask is additional information needed and we'll answer yes.

DR. DOBSON: Sure.

MS. YOUNKER: On the other hand--go ahead.

DR. DOBSON: Well, I was going to say we have been asked to reassess where we are now five years after the last time we made an assessment. And we have acquired some new data. As Pat pointed out, not a whole lot in terms of drill holes, but there has been substantial progress in some technical areas in isotope geochemistry and some of the volcanic age dating techniques and things. So we need to aggregate all the new data we require to figure out if we have kind of leapt that hurdle into higher level compliance on a few issues.

MR. BROCUM: Let me just say one thing. You know we just talked about the next year, but we've also had a lot of internal discussion about doing this in an iterative way, and we are trying to come up with the methodology that addresses the overall site suitability as Jean said. And if you do this iteratively or periodically, sooner or later, you might shoot off to the right there. Maybe five years, ten years, you may do it. Okay, so if you are talking about a comprehensive methodology, you need to have that box.

MS. YOUNKER: If you'll let me talk you through the actual method, the way we have the thing set, I think it will become clear why the representation is like it is, Clarence.

In terms of refining the testing program, and analyzing this question of should additional data be acquired, we felt very clearly that was one of the important cross-overs to the test prioritization methodology, because obviously what Dave set up is an approach for us to ask the question, how much is it going to buy us if we go after additional information. And so this is why we have--one reason for this dashed outline around the name of test prioritization is because as you see me go through this you'll see that that is clearly one of the areas where we've made a very strong proposal to DOE that we need to bring those two tasks together because the method that they have established is the right way to think about this, we believe.

In terms of what you do next, obviously, you go make some changes as Max talked about. Maybe we need a new study plan. Maybe we need at least a new activity within an existing study plan to go after some of the things that they've already highlighted as being potentially important from site suitability.

DR. ALLEN: Well we many not need to refine anything. We may just need to start the program as planned.

MS. YOUNKER: Right. Sure.

DR. DOBSON: Well, I think that is good--this program, I mean the diagram was originally done assuming that you were at the point where you had collected the data you already said you were going to collect. At the current point in the program you are absolutely right. Should additional data be acquired, it may not require refining any programs, it may just require conducting the ones that we planned.

MS. YOUNKER: But, you'll see in the logic that we have laid out that there are some cases where we might, right now, answer the question that we don't think in that specific area in question that there is a need for very much more information to move ahead with this site.

So what this, then, does is take you back around to go get some new data, and go back through the evaluation that Steve just mentioned.

And once again, just to express this idea of the close integration of what, we hope, will be an actual running of the two tests. The recommendation is that basically the site suitability decision that we are talking about, which is dependent on when you are making it. It has three branches of recommendations not a definitive outcome, so you go back and get some more information, or it is definitive but it is definitive that I should abandon this site because there is some information available now or at some point during testing that tells you that the site should not be taken forward.

The site testing decision that Bruce talked about in terms of continuing testing and stopping testing, clearly this is all part of this question of what do I do when I get a non-definitive answer to the question of do I have enough information or is the site suitable if I am asking the bottom question.

What do they give us? This is just another attempt to tell you that our view right now, and I hope that this is how it comes out, is that we take the two tasks and basically make them one and the same. We want to use their preliminary strategies for looking at the importance of post-closure suitability concerns as the general approach for looking at suitability concerns. It's a good way to think about it, to structure our thinking.

The value of information, value of additional testing from the value of information standpoint we certainly intend to use. And then in terms of test prioritization, once you figure out, we'll say that in our case, we'll go through this evaluation and we end up with a potential disqualifying factor that wasn't represented by one of the concerns that they have in their study right now. We clearly want that to be looked at, mapped in and looked at what the testing program can do.

DR. DOMENICO: Jean, can you hold that on there?

MS. YOUNKER: We'll do that, Pat.

DR. DOMENICO: The first bullet, preliminary strategy for evaluating importance of postclosure suitability concerns. How can you do that when the 10 CFR Part 960 did not incorporate or even think of or mention the EPA release standards?

MS. YOUNKER: Oh, it certainly did, Pat. The total system guideline for postclosure is the EPA release standard plus the two--

DR. DOMENICO: The 960 did? I thought you were basing this on the guidance given when you had nine sites and you were looking at a site suitability--

MS. YOUNKER: 960 required, at that time it actually required a total system performance calculation for each site that was being prepared for both 10,000--did we have to do 100,000 years too? I think we did.

DR. DOMENICO: So this includes not only the concerns when we were in a site selection procedure as well as all of the documents that have since come along?

MS. YOUNKER: You bet.

DR. DOMENICO: Okay.

MS. YOUNKER: And, in fact, the next two view graphs show you, just to refresh your memory about what is in 960, it is extremely comprehensive. It runs through in the system guideline--it's EPA releases, it's also subsystem releases. So you have to look at EBS release and containment as well. And then for each of the disciplines where I think anyone would think from the standpoint of performance of the repository system there might be a concern. We have a bin for that. And so if you go through them, ground-water travel time--some of the guidelines do not have disqualifying conditions, but they have qualifying conditions, which if not met are disqualifying conditions. So, for every one of these, there is a statement which says if you can't--if for example geochemistry is not compatible with waste containment and isolation, which means geochemistry of the site doesn't allow me to comply with the total system requirements, then my site is not suitable and my site is disqualified, in fact.

So you see mapped out here, geochemistry, rock characteristics, which includes kind of a lot of the complex geology type of concerns, climate changes, erosion, dissolution, tectonics and this is your postclosure tectonics, natural resources, which includes the human interference concerns, and then the site ownership and control is one that is both a preclosure and a postclosure guideline in 960.

DR. DOBSON: You've got one in L and H.

MS. YOUNKER: Oh, I'm sorry, I didn't mention what the rest of the table is about. Thanks. What you see tabulated over here is whether or not there is disqualifying condition present if there is an "X" in this column it tells you in the postclosure guidelines there are six disqualifying conditions and you can see where they are by the "Xs" in this column. And as I said before, there is a qualifying condition for every guideline.

The final column is either an L or an H, and what that means is that 960 has a philosophy that you have a lower confidence finding, which is called lower-level finding, that does basically look at all of the available information and decide if the site appears to be okay on that basis. That is the lower level or the L that I have up there. The H is look at the available information and determine if the site is suitable on the basis of that, and are you confident enough that you think it will remain suitable on the basis of any further information that you obtain. So, it's the higher confidence finding.

And what we tabulated for you here was to show you that in the case of dissolution, as I earlier used that as my example, at the time of the environmental assessment, we were confident enough that dissolution in the way that 960 intended it to be looked at, is not a potential disqualifying factor at this site. And so we did make higher level findings and both the qualifying and disqualifying condition for dissolution in the environmental assessment. And what that says is, there is nothing that we are going to find out--we were confident enough that there was nothing we are going to find out, about the site doing site characterization that would change our minds. That we are confident enough to go ahead an make that higher confidence finding now.

Okay, remember that there is a whole set of preclosure guidelines, as well, in 960. And, this covers essentially all the preclosure radiological safety concerns and at the top of the list up there. You see, this is the meeting the preclosure radiological safety criteria that are specified. It includes the things--the technical guidelines within that system of guidelines that you need to know about or that you might have concerns about and this is another example where you see a couple of H's.

The population density and distribution guideline is another one where it was basically, what is the population distribution and does it meet the criterion that was set in that guideline. So that is another one where, for this site, DOE was confident enough to make the finding on that, because we don't think the population distribution is going to change that drastically, that, this one would become a problem for this site. So that is another example of what that higher confidence finding really is.

In terms of the rest of these now, when you get down to, and this is an important point, going back to what Bruce Judd has talked about, when you get down below right here (indicating), you have some pre-closure geotechnical type of guidelines, where you get at the question of seismic hazards, you get at the question of preclosure hydrologic concerns from the standpoint of construction or if there is any problem of surface flooding, you get it either here or here. And rock characteristics brings in the question of, are there any health hazards relating to mining, either the actual minerals present or is there a reason to believe that you can have a safety problem, so it brings in all of the pre-closure type of potential factors that everyone would say you should look at when you compare sites or when you evaluate a site.

One thing, the reason I brought this up is because Bruce made the statement, you know, they only looked at post-closure. So from the standpoint of driving and testing the program, for what they've looked at right now, you would not find priority being placed on the kinds of tests that get a pre-closure seismic hazard analysis or any of the things related to pre-closure, the geoengineering, geomechanical part of pre-closure concerns, because, that wasn't included. But that is in 960 and we will look at it as a part of our study.

Okay, Steve already told you that our objectives are, basically, for the core team to define an approach to evaluate the suitability of Yucca Mountain within the frame work of 960. To look at 960 and--I'll show you on a guideline-by-guideline basis, we want to look at them to determine where the data may actually already be sufficient. We asked the question, is there already information such that, in this area, we really don't think there is a potential site suitability concern there. And then to conduct our earlier evaluations, and we don't mean necessarily site evaluations. We are talking about evaluations of each guideline.

Okay, we have a general DOE plan, that Steven mentioned was prepared through the T&MSS contractor which is just an acronym for the SAIC and the contractors that work with us in support of the Yucca Mountain project office, to prepare an implementation plan for their plan which basically described what we would do over the one year period of this task. And we are in the process of integrating with the test prioritization group. What that really amounts to is I'm trying to get Bruce Judd to come over and work with me, and the rest of the team. I didn't mean to leave you guys out, but I think we need Bruce to work. Some of his early thinking in site suitability underlies the test prioritization group and it is just perfect to bring that right in and evolve that.

We put the QA controls in place for this task. Since some of the information, although clearly not being written to go into a licensed application, some of it--it would be very nice to be able to use it as efficiently as possible. The kinds of controls we have are heavy documentation, do everything you can to make sure that you document every step along the way such that if someone wants to use the information later everything is there that they need. We have other controls, the obvious ones, but that is the one we are really attempting to be just as careful as we can.

We have an implementation plan that was written by SAIC to put this together. And, we selected a team, and coming up on the next page, just so you know, once again it is an in-house team in that it is the support contractor, National Labs, and USGS for the Yucca Mountain Project Office. But as you heard just a few minutes ago, and I'll get to it in just a minute, we are going to have a peer review as part of the process so we'll end up with, I think, some good external input into this.

The people that are listed here, I won't go through it with you except to say we have some of the people that have worked in the program for a long time like Bill Dudley from USGS who of course provides the geologic, hydrologic expertise for us, and taps into his organization in the areas where we need that kind of support.

The same thing with Art Ducharme from Sandia, tapping into the Sandia performance assessment and rock mechanics, geomechanics type of expertise for the project, bringing it to this team for us.

Some of the other people like Bill Andrews, or Greg Fasano may be two names that some of your panels will have heard from. But, these are the people that are in the environmental socioeconomic transportation side of the house. And they, of course, are needed on our team because we are going to look at the complete set of guidelines, not just the post-closure or pre-closure geotechnical.

And the next one just tells you that the way we are set up, this is a T&MSS directed task and so, as a result, the DOE people wind up being members of the team or observers. And they are there with us, to know what we are doing every step of the way, because it is clearly in our best interest not to have any surprises and to make sure that they know what the product will look like when we get there, but they are not members of the team. And we also have an observer, another decision analyst, Rex Brown, so we have a pretty diverse team to work with.

Okay, we are formally underway, we've done the things we have to do to be legitimate from a quality assurance standpoint. We have worked together. One of the first things we faced, which I think you all would probably guess that, was what do we mean by suitability? So we are working on a definition for our purposes. It may not be the definition that you all would necessarily have developed, but we have a definition and it will be the one that we will say, for our purposes, this is what we think suitability means.

And, what we've done is to do what, I was kind of thinking, was a pilot study, and then I didn't want to confuse you because we talked about a pilot study earlier in our thought process, I think back at the Albuquerque meeting we were talking about a pilot study. It was a different pilot study. So, we called this a preliminary evaluation. But think of it, with me, as a pilot study, because what we did was to go through every one of those technical guidelines that were listed on an earlier view graph and basically do a round table of what we think, what new information and analysis is available for each guideline since the last time we really focused in on this topic.

We then said what is the current status? Is there a real concern about this guideline? Is there a lot of information we don't have, or is it one where we basically feel that this is one where--let me give you an example like pre-closure surface characteristics. It gets at the potential for flooding that would require measures beyond reasonably available technology, or terrain that causes you to go beyond reasonably available technology: in your design considera- tions; or in your actual construction.

There are some, like that, where the team in our first evaluation, preliminary evaluation, which hasn't been approved by anybody--so I am just sharing with you kind of our developing ideas, where we suspect that from the standpoint of suitability, that is not a real concern at this site. We don't think that there is a disqualifying factor related to terrain and ability to design a service facility, an under- ground facility, using reasonably available technology we'll be able to resist the flooding potentials at this site. So that might be one where we would propose to DOE, we think there may be other reasons you need to get some information about terrain, may need some detailed maps to do your designs, but we, as a group, recommend that you consider a higher level of finding, a higher confidence finding on this particular guideline. It is not a suitability issue for this site. That's what I mean by that little box, develop the site specific approach.

And, of course the next step in that is that once you've gone through this little pilot study is well, I don't have the information now, what is it I'm going to need to get that higher confidence finding. And there you are mapping back into the site characterization plan, as Clarence pointed out, and furthermore, you are asking the question, am I really going to get what I think I'm going to get by going through a value of information type of thought process, because we may define something as a group that we think we need, we may get our value of information type of analysis and find out we are never going to get it.

So we will have to, then, think about, well, did we really need it? Was our thought process off? Or is this a real potential disqualifying condition for the site because you'll never be able to get the confidence about that particular potential disqualifying condition.

So, we are going to use this pilot study, the preliminary evaluations that we've completed as a basis for figuring out what we can really do with each guideline and by what we can do, I am meaning, do we recommend to DOE as part of this product that we believe that surface characteristics of pre-closure hydrology or pre-closure rock characteristics or perhaps erosion--that erosion is not a potential disqualifying factor at this site. Go ahead--from our viewpoint, we believe the information may support the higher level finding, a higher confidence finding, and begin kind of checking off that list.

Now in some cases we may come in and say, there is a couple--like geohydrology is a good example, ground-water travel time disqualifying condition, 960 gives you some help. It says, this is a guideline that is not intended to be evaluated until after site characterization. It tells you to be careful about when you apply certain guidelines, 960 being one specific example. So we get some guidance that we need to carefully think about, as a team, as we make a recommendation to DOE.

Now of course because we are only making recommendations, DOE can use that guidance themselves and we aren't going to make high level findings or propose high level findings on any of these guidelines. It doesn't make sense to us right now, but the value of what we are doing, I believe, is that we will lay out that information basis, it will be there and be available for other people to review.

DR. DOMENICO: Jean, I'm bothered--just one minute, Jean I'm looking at your post-closure guidelines. I don't--do you plan to do a preliminary performance assessment?

MS. YOUNKER: Yes. For the total system guideline--

DR. DOMENICO: I see nothing in these that relates to that--

MS. YOUNKER: That's what that is.

DR. DOMENICO: What is that system and s/system, what does that mean?

MS. YOUNKER: It is the total system guideline in 960, it is EPA.

DR. DOMENICO: The EPA, okay that's locked up in there, thank you.

MS. YOUNKER: And, just to follow up, Pat, we clearly--at this point in time we are not going to do a comprehensive, a CCDF of the kind we could do after site characterization. But we are going to make every attempt to get some good sensitivity studies done. And in this case we--in fact there was a parallel meeting going on this afternoon, with--where's Larry--Larry Rickertson, who is the subteam leader for the total system guideline evaluation, talking with the people who are the best able to provide some input and some sensitivity studies to support that guideline evaluation.

DR. DOMENICO: I might add that there is a disqualifying condition for that too that is not noted there. For example, not meeting the EPA requirements.

MS. YOUNKER: Well, remember now, a qualifying condition not met is a disqualifying condition. So everyone of these X's, every guideline has a disqualifying condition because it has a qualifying condition. If you can't meet it you are disqualified. It's an on-off switch.

In certain cases there is also a disqualifying condition. In general the disqualifying conditions are conditions: that you can evaluate earlier on the basis of less information; that is meant to be, kind of, used in the earlier phase of the site screening process if you read 960.

So there are more things, like, 200 meter over-burden, for example, is one. It's not true of ground-water travel time, however. 960 says don't evaluate that one early, or at least be careful if you do.

Steve mentioned there are two parts to what we are doing. We are doing these preliminary evaluations and developing this general method, which you can clearly see now, we've had--in taking the first step in parallel with developing the general method for comprehensive suitability. We are taking the first step in looking at this early site suitability evaluation or developing this.

I didn't have another view graph in here that told you about the peer review, but I started to say that we've been in the preliminary scoping phase for several months. Trying to get this thing defined, and get the scope of it defined, in a manner that, we felt, we could be successful in completing it in the one year that we have. This also included the pilot study that I just described to you. We basically have gone through a round robin, a couple of times, with the key people on the core team, and they have gone home and consulted with their support staff, so it isn't just the core team members making our first round of decisions about this.

What we intend to do in this evaluation package preparation is, for each guideline, to have available all of the information that goes into our conclusions and this should update the information base that was used in the EA and the SCP. In this case we are not totally confining ourselves to only published data, although we will make every attempt to get it published in a: letter; report; or in some form; so that it can be referenced. But we are not saying that it has to be published in a referenced journal, because we want to look at all information. We want this information package to be as complete as humanly possible in the time we have.

And, starting out in about August 1st, I think the date is August 1st or the 15th, we want to have an extra peer review of that package and that would be of the conclusions reached by the core team on each guideline. And that should allow us to, in that package, have all the information that anyone would want to look at, either available to the person, or--certainly there is a reference citation so that anybody else on that peer review team can look at: the package for ground-water travel time; for mineral resources; for any of the ones where there is contention about the suitability of the site. Look at that same package that we've looked at and see whether or not they can draw the same conclusion that we have.

You know, my personal view of the task is that it is a big step in this whole scientific consensus building process for us to bring a broader group of people up to speed on the package of information that is available in each of these potential, suitability or unsuitability areas.

DR. DOMENICO: Will the data be provided in the final package for the public when you publish this?

MS. YOUNKERS: Everything we use, everything we can document, you know, all decisions, the basis for all recommendations, any data, yes.

DR. DOMENICO: This will not be a small publication then?

MS. YOUNKERS: No.

DR. DEERE: Six volumes?

MS. YOUNKERS: Well, I guess the question is, what we actually--how much do we actually have to assemble and summarize versus how much of it is just in reference information. We tell you the page number where we got the information. I think that would be true to a certain extent. The time we have to do it in determines how much we are going to be able to actually pull in and summarize, versus send you to the right place to look for it, but I'm looking at this as a major kind of data acquisition task with an executive summary that is not too thick, and then this room full of references.

DR. DEERE: The end report will say, for each one of these features, we recommend the site, or we recommend abandoning the site, or conduct additional testing. Is this right for each one of the features?

MS. YOUNKERS: It will say based on this core team's analysis that we think the information is sufficient to go ahead and make the higher confidence finding, or that we have some real concerns about this particular potential suitability, unsuitability factor and we think you should either gather more information or, if the information basis was really there to support disqualification, then we would recommend disqualification on the basis of any one of these factors.

DR. DEERE: When you define the factors, will every one of them be defined, because if you have just one that says we recommend abandoning this site?

MS. YOUNKERS: 960 says that these are potential disqualifying conditions and every qualifying condition is the same thing, so any one of them, right. In this evaluation using 960, any one of them.

DR. ALLEN: Presumably, if you lose such an area, you would already recommend abandoning the site.

MR. BROCUM: We went through this once in EA, so for every one of these findings, we have a lower level finding and for some we have a higher level finding already.

MS. YOUNKERS: The lower level, yeah, meaning that on available information then you believed that you met the qualifying condition or it didn't have the disqualifying condition. But you know, there is a contention about that in some cases that having this information, all together in one place, having us look at it and then having our external peer review look at it to see whether or not the basis is really there for disqualification, or whether you need additional information to make that decision or to make the positive decision, it seems to me that the value of it is, kind of at, the process that we are going through.

DR. DOMENICO: Was this the format used in EA? I forget. The same format more or less, the item by-item-by-item?

MS. YOUNKERS: Oh, yes, you had to go through it item-by-item.

MR. BLANCHARD: I believe it is Chapter 6 isn't it in the EA?

MS. YOUNKERS: Uh-huh.

MR. BLANCHARD: Not Chapter 7.

MS. YOUNKERS: Right, Chapter 6.

MR. BLANCHARD: Because, Chapter 7 compared all of the sites where Chapter 6 ranked the attributes of that site with the guidelines.

DR. ALLEN: What the Secretary is asking for and what you are proposing to do is to give them a very thorough well organized project.

MS. YOUNKERS: That's a reasonable assessment. But to also--

MR. BROCUM: But the Nuclear Waste Policy Act requires us if there is something wrong with the site to notify Congress I think within six months or something like that. This is a method of doing that, you know of doing the progress report, if you like. The progress year-by-year or however we follow-up after this year, see what I'm saying.

So, I think it is prudent for us to look at the information as it comes in.

MR. MCFARLAND: Jean, have you done any thinking about how you will draw together for peer review considering the magnitude of undertaking?

MS. YOUNKERS: We haven't spent as much time thinking about that so far as we need Russ. But I guess my impression is because we have the pre-closure nongeotechnical and we have the pre-closure geotechnical and the post-closure geotechnical plus the risk assessment, performance assessment type of content, that we are probably going to have to have subpanels of a group. And I'm assuming that the people who do the pre-closure transportation and environmental socioeconomic type of review probably wouldn't even judge themselves qualified to participate in the other part of it. So we may have to structure the panels somehow so that they each look at their own part and then there is some kind of roll up of that.

But how we get people educated quickly enough; and the kind of people to use; and what it is we ask them to do; --I mean one of the things I'm very concerned about is, what is it we can really ask them to do in the kind of time frame given the mass of information that will be there? They are very difficult questions and I know that I am going to need to talk to some people, who have thought about this and who are experts in this area, to get advice as to how to proceed.

So the answer is I don't know.

DR. REITER: Jean, could you give us an idea, this sounds like a major project, that could be a high visibility item, is this something that is going to be the order of the ESF study in terms of effort like the Calico Hills?

MS. YOUNKERS: I'm not personally looking at it from the viewpoint that it will expand to anything like, probably even the Calico Hills, certainly not the ESF alternatives, partly because we have to deliver the product to a peer review by basically August 1st, and so that is why I said, when I said, what we will actually do in looking at a lot of it is assemble the information bases. And so a lot of it is basically figure out what it is I need, and then having people get it for me. Get that information together, and: document it; tabulate it; have it in a way that other people can access it in a reasonable fashion.

The executive summary of the group decisions, the consensus that is reached in the core team, as well as, after we incorporate the peer review results, to me is, probably, a fairly short executive summary.

DR. REITER: So the core team is going to make the decision on each one of those conditions?

MS. YOUNKERS: Right.

DR. REITER: That's the one. It's not separate. Okay.

MR. BROCUM: But the extent of the effort is clearly defined in the management implementation--FTE's and a separate document to Carl Gertz, the cost. The idea was not to let this become a very large effort.

DR. DOMENICO: Yes, but there is work involved other than looking at stuff and making value judgments. The performance assessment requires work work.

MS. YOUNKERS: That's exactly right. That's why they were, off, meeting today. They are figuring out and defining this--the work to be done is to establish what explicit sensitivity studies can be run given: available data; and available models; codes; and getting those defined well enough such that we can get any kind of results that we need in this time frame. And that's what the group was, off, doing today.

DR. DEERE: You see--in the future meetings we talked about we hadn't defined the scope yet of the testing coming up and some of the things that we feel that we haven't had a chance to discuss and to have any input or to understand some of the things in detail. We keep running into the task force and the new programs. So it makes us feel a little guilty in saying well this is a subject we want to go into when you are on a different task or a different study with a lot of priorities. You certainly have to coordinate it because we don't wish to interrupt a flow of studies that you have to go through. By the same token, we want to make sure that we are getting enough information that we can make worthwhile assessments.

MR. BROCUM: Right. But if we could agree on the methodology if you like, or the approach for evaluating site suitability will help a lot then when we discuss the detail studies to implement that, okay. I think that is an important factor.

DR. DEERE: One concern we have, I'll just mention it, I'd say we wouldn't be in complete agreement with some of the tests that you have. We are not sure they have been evaluated with the new layouts of ramps and accessibility potential with turnouts and things like this.

MR. BROCUM: And the testing people, I am well aware of that. They themselves have raised those issues internally. Yes, you are right. That's an important issue.

DR. DEERE: So, we should be discussing over the next two or three weeks with you, as to, what is the most efficient thing that we could all be doing on this to let us evaluate your testing?

Any other comments? Panel members? Ed? Russ?

DR. DOMENICO: I have a question for Bruce. Bruce, I'm still bothered. A lot of things bother me. The gas release, you know we go to a lot of meetings, and my memory is not so good sometimes, but now as I was thinking back, I recall talking to the engineered barrier people and hearing that Carbon-14 release in excess of EPA standards is inevitable, totally inevitable.

We got that information, I don't know a year ago and filed it away someplace. And then thinking about that we thought that was a problem for the waste package people. In other words it is problem, either of designing something in the waste package, or it is a regulatory problem in the sense that the people are even thinking about talking to the EPA about: maybe the limits on that particular constituent were set arbitrarily; and maybe not; reflect the real situation, so is accepted as an interval. But out of your studies come, today, it turns out to be, a very high priority testing operation; when to me it still seems to be a matter to be dealt with in the waste package or in a regulatory sense. Because what the experts told me is it is inevitable in unsaturated zones--there is no question.

Would you address that? I don't know if that is a question or a statement.

DR. JUDD: If I could point out a couple of things and I am sure others at the table will amplify and provide others. One thing is that, our recommendation was to re-evaluate the strategy. And I left the impression that that was to re-evaluate the testing strategy, and yet, strategy can include testing the site for the transport of gas, it can also include looking more carefully at the waste package design. Other alternatives like venting the waste package prior to emplacement, review of the regulatory requirements and that review of the strategy needs to be taken very broadly.

DR. DOMENICO: It doesn't necessarily mean just physical tests that we are talking about?

DR. JUDD: That's correct.

And the second is on the inevitability, the assessment by our group was a 62 percent chance that this concern is present. In other words the gas flow will be short enough to be a significant problem, relative to our assessment threshold. So, that's a very high probability. And the assessment threshold was; we will exceed 2 percent of the EPA limits during the next 10,000 years. Those probabilities, the 62 percent and the 2 percent of the EPA limits, those are very high numbers relative to almost everything else in our study.

So inevitably, no, it didn't say that. But the probability was quite high.

DR. DOMENICO: Well I was talking to deterministic people.

DR. JUDD: It is inevitable.

DR. DEERE: Just a second I'll try to check on the time.

It seems like we have two times listed for tomorrow in two different documents, an 8:00 starting time and an 8:30. And I just want to make sure that we select the one that doesn't foul up a lot of people that are going to be coming in at 8:30 and we are going for half an hour. We have made a change like that in the past and it was not well received.

In the same token, I am sure all of you who would come in at 8:00 ready to go and nobody shows up until 8:30 to make the presentations, your's said 8:30.

DR. DOBSON: We'll be here when you say you are going to be here.

DR. DOMENICO: 8:30 is much more civilized.

DR. DEERE: Everybody wants 8:00. 8:00 all right.

DR. DOMENICO: Which is the decision here?

DR. DEERE: Well, as you see we ran over about an hour or an hour and a half because we had questions we wanted to ask you.

DR. DOBSON: I think 8:00 is fine if you want to go at 8:00.

DR. DEERE: Since we are all here, I would say 8:00.

(Whereupon, the meeting was concluded at 5:30 p.m., March 6, 1991, to reconvene at 8:00 a.m., March 7, 1991.)

CERTIFICATE

This is to certify that the attached proceedings before:

UNITED STATES NUCLEAR WASTE TECHNICAL REVIEW BOARD

In the Matter of:

STRUCTURAL GEOLOGY & GEOENGINEERING

and

HYDROGEOLOGY & GEOCHEMISTRY

JOINT PANEL MEETING

Location: DENVER, COLORADO Date: MARCH 6, 1991

was held as herein appears, and that this is the original transcript thereof for the file of the Board.