Official Transcript of Proceedings

NUCLEAR REGULATORY COMMISSION

Title: Advisory Committee on Reactor Safeguards

Docket Number: (not applicable)

Location: Rockville, Maryland

Date: Thursday, November 2, 2006

Work Order No.: NRC-1299 Pages 1-106

NEAL R. GROSS AND CO., INC.
Court Reporters and Transcribers
1323 Rhode Island Avenue, N.W.
Washington, D.C. 20005
(202) 234-4433

UNITED STATES OF AMERICA

NUCLEAR REGULATORY COMMISSION

+ + + + +

ADVISORY COMMITTEE ON REACTOR SAFEGUARDS (ACRS)

537th MEETING

+ + + + +

THURSDAY, NOVEMBER 2, 2006

+ + + + +

ROCKVILLE, MARYLAND

+ + + + +

The meeting was convened in Room T-2B3 of Two White Flint North, 11545 Rockville Pike, Rockville, Maryland, at 8:30 a.m., Graham B. Wallis, Chairman, presiding.

COMMITTEE MEMBERS PRESENT:

GRAHAM B. WALLIS	Chairman
WILLIAM J. SHACK	Vice Chairman
GEORGE E. APOSTOLAKIS	Member
J. SAM ARMIJO	Member
MARIO V. BONACA	Member
MICHAEL CORRADINI	Member
THOMAS S. KRESS	Member
OTTO L. MAYNARD	Member
DANA A. POWERS	Member
JOHN D. SIEBER	Member-At-Large

ALSO PRESENT:

SANJOY BANERJEE

SUSAN COOPER

JOHN FORESTER

JEFF JULIUS

ALAN KOLOKZCOWSKI

ERASMIA LOIS

JOHN MONNINGER

ERIC THORNSBERRY

I-N-D-E-X

P-R-O-C-E-E-D-I-N-G-S

(8:32 a.m.)

CHAIRMAN WALLIS: Good morning. The meeting will now come to order.

This is the second day of the 537th meeting of the Advisory Committee on Reactor Safeguards. During today's meeting the committee will consider the following: a status report on human reliability analysis research program. Further ACRS activities -- report of the Planning and Procedures Subcommittee, reconciliation of ACRS comments and recommendations, and the preparation of ACRS reports.

This meeting is being conducted in accordance with the provisions of the Federal Advisory Committee Act. Mr. Sam Duraiswamy is the Designated Federal Official for the initial portion of the meeting.

We have received no written comments or requests for time to make oral statements from members of the public regarding today's sessions.

A transcript of a portion of the meeting is being kept, and it is requested that the speakers use one of the microphones, identify themselves, and speak with sufficient clarity and volume so that they can be readily heard.

I'd now like to proceed with the meeting. I call upon George Apostolakis to get us started on the first item.

MEMBER APOSTOLAKIS: Thank you, Mr. Chairman. This first session deals with human reliability analysis models. In the last year or so, during various interactions with the staff, especially the subcommittee meetings, the latest -- the last one being last June, we realized that the agency has three -- that we know of -- models for handling human performance.

One is ATHEANA, which, of course we have reviewed in the past. The other is SPAR-H. That is used primarily for the significance determination process and other regulatory activities. And then, there was a NUREG that was discussed here that -- in the context of manual actions in response to fire, which also deals with human performance, but in a different way. It does not attempt to reduce any probabilities, but it works with margins.

Essentially, it says if you have a certain available time before you reach an undesirable state, then you have to demonstrate that the sum of diagnosis time and implementation action is less than this available time, and that margin has to satisfy certain

criteria. So we end up with three models, essentially.

Then, in the power uprates, we have seem that in most cases the most significant impact of the uprate is the shortening of the available time for action for the operators. So they might have under current power levels 25 minutes, and in the new -- at the new level they might have 20 minutes. So the question is now: what is the probability of a human error because of this shortening of time?

And, typically, we get numbers from the licensees which we understand are produced using a fourth model, which is part of the EPRI HRA calculator. It's a software package that allows you to use a number of models, and, in particularly, the so-called human cognitive reliability operator, reliability evaluation, which focuses on time. So they claim that if time changes, the available time for action changes, then they are able to produce probabilities for this new interval.

It is my understanding that the staff here has never reviewed that model, which bothers me. I really think that whenever we review something that the licensees submit the staff should have reviewed the model that the licensees are using.

So the result is that we have now three models here and one -- at least one in the industry, which if you look at the assumptions, I mean, they share a lot of assumptions, but also the focus is different, some other assumptions are different, and so on.

So the idea of having today's session is maybe to see where the staff is with respect to human reliability analysis and the idea that was proposed at the subcommittee meeting was first for the staff to see whether the three NRC models can be merged and have one model which may have different versions perhaps to satisfy different needs, but essentially would be one model with -- based on a common set of assumptions, and then explore also the possibility of bringing in the EPRI model into this. Now, of course, we cannot demand that EPRI collaborate with the staff, but the least we can do is to demand that we review it.

So with that in mind, I will turn it over to Mr. Monninger.

MR. MONNINGER: Good morning. I'm John Monninger. I'm the Deputy Director for Probabilistic Risk and Applications in the NRC's Office of Nuclear Regulatory Research.

The staff -- we are very pleased to be here this morning to talk to the ACRS about the NRC's research programs into HRA methods and applications. Looking back I guess over the past two years, we have had, you know, quite a few different meetings with the ACRS, approximately 10 different meetings, with the full committee and subcommittee.

You know, at a high level the NRC's research program in HRA it's -- we're going after various different areas. We have research ongoing in the use of HRA methods for materials applications, in addition to all the work that you've been hearing about the first two years for reactor applications.

We're also looking -- you know, with regards to the reactor applications we're looking at approaches for operating reactors in addition to preparing the agency for the use of HRA methods for advanced reactors.

We're also looking at the use of the HRA methods to solve ongoing regulatory technical issues, such as PTS, or as you mentioned ALFIRE. You know, our approach for HRA or, really, where we see it playing in predominantly right now is within the NRC's -- what we call the phased approach to PRA quality.

Last month the staff met with the ACRS to

go through Reg. Guide 1.200, which essentially establishes the quality standards for PRA and endorses various standards developed by ASME and ANS. Our efforts here on developing additional HRA guidance is in concert with that. The focus is to develop more detailed technical documents to -- that would ultimately be incorporated by reference into Reg. Guide 1.200.

You know, I think we have made considerable progress, and you will hear this morning I guess a discussion by Dr. Erasmia Lois on our future plans. One of the ones that we are most looking forward to very much so is a program that we are pursuing with Halden to benchmark various HRA methods. We think our planned efforts in terms of that HRA methods benchmarking project will go a long way to addressing many of the questions and issues that the ACRS has.

With that, we just look forward to a very interactive meeting, and we thank you very much for your questions and comments.

MEMBER APOSTOLAKIS: Okay.

MS. LOIS: Thank you. I'm Erasmia Lois. I work for the Office of Research. I believe that also on the -- through the telephone Jeff Julius of

EPRI is also participating.

MEMBER APOSTOLAKIS: Jeff, are you on the line? John, are you online?

MR. FORESTER: John Forester is here.

MEMBER APOSTOLAKIS: Okay. And who is the other guy?

MS. LOIS: Alan Kolokzcowski.

MEMBER APOSTOLAKIS: Alan?

MR. KOLOKZCOWSKI: Yes, I'm here.

MEMBER APOSTOLAKIS: So we don't have the industry.

MS. LOIS: You don't have the industry. Probably they will -- they will dial in while we are talking.

As Mr. Monninger mentioned, we are here more to listen to the ACRS today, but we thought that we would provide a brief overview of our activities up to now and what we have planned. And also, we plan to address some of the questions posed by the ACRS, so that we have a more productive interaction.

Okay. So what we'll do today is we'll summarize the HRA activities, focusing on reactor applications, although the activities for NMSS also are going to be mentioned. As I said, outline the plans for the next four years, and also discuss with

the ACRS the issues that have been raised.

Overall, the HRA research program objectives are to support risk-informed regulatory activities, and we do those by having activities that support improvement of the HRA quality, supporting specific regulatory issues that come -- are raised, and address new needs, new applications, and we tried to obtain that -- to achieve that through collaborative efforts with domestic and international organizations for efficiency and effectiveness.

The HRA quality has I guess three different perspectives -- developing guidance for performing tests for human reliabilities and raising outstanding technical issues, and also we have an activity which we call perform technology transfer. And I'm going to talk to each one of those very quickly, what we have done up 'til now, and what we'll do next.

In terms of HRA quality, recognition that the -- one of the biggest problems in human reliability is the lack of consistent applications among practitioners. Although the hardware performance aspect of it, the practices on how you model equipment, etcetera, pretty much are set and people are doing it consistently, that we recognize

that that was not the case in human reliability. That was one of the biggest insights we got from reviewing the IPs.

And, therefore, the first activity to address human reliability quality issues was to develop what we call good practices. We briefed the ACRS several times on those, and with the good practices I guess they -- they established the framework for consistency and also quality in terms of ensuring that the PRA model will include the important human actions that needed to be included, and also the model itself will be accurate in the sense that dependencies are going to build, etcetera.

The next step was now to evaluate the methods with respect against these good practices, and the result of that NUREG is -- was that different methods have different capabilities, and, therefore, should be applied for regulatory applications as needed, as they match.

So the idea that I'm choosing my HRA method and I'm trying to fit it with respectively to my regulatory application proved to be wrong, and that is the main thrust for -- of NUREG-1842 is to choose the right tool, the right method for your application.

In terms of issues that are outstanding

for human reliability --

MEMBER APOSTOLAKIS: Who is doing this?

MEMBER ARMIJO: They're on the phone. They have to put their phone on mute. Whoever is on the phone, just put it on mute.

MEMBER APOSTOLAKIS: Can you hear us, guys?

MR. FORESTER: Yes, I can hear you fine.

MEMBER APOSTOLAKIS: Can you mute your phone?

MR. FORESTER: Yes.

MEMBER APOSTOLAKIS: And don't speak unless spoken to.

(Laughter.)

MR. FORESTER: That's okay.

MEMBER APOSTOLAKIS: I'm kidding you, John.

MS. LOIS: One of the biggest problems with human reliability in terms of both testing the HRA methods and underlying assumptions is lack of data. So we have undertaken the activity of developing a repository of human events, and I guess we briefed the ACRS a couple of times on this activity.

We published NUREG/CR-6903 that describes

-- provides the overview of how we do this collection of data and loading into a database, and we are loading data right now based on LERs -- LER -- yes, LERs. And at the same time, we are developing what we call quantification tools that would allow the use of this data in human reliability.

MEMBER POWERS: Erasmia, I know that we have seen the information on bid practices. Have you sent us the methods evaluation against bid practices?

MS. LOIS: I'm sorry. I didn't get the question.

MEMBER POWERS: Have you sent us the methods evaluation against bid practices?

MEMBER APOSTOLAKIS: 1842. Do we have that?

MS. LOIS: Yes, it's going to be -- it's being -- in print right now.

MEMBER POWERS: Oh, okay.

MS. LOIS: But the ACRS saw it in a draft form before public comment, and I guess after public comment as well. So as we speak, if you look it up on the web, you'll find it. It's there.

In terms of addressing specific regulatory issues, we have done a human reliability for the PTS PRAs. We supported the screening analysis. We have

developed screening analysis for human reliability for the --

CHAIRMAN WALLIS: Would you tell us how you did this? I mean, George started off saying there are three different models. Which model did you use when you did these things?

MEMBER APOSTOLAKIS: And why?

MR. BANERJEE: And what are the models? I'm completely -- if you had started -- give us a brief introduction, like what is this all about? That will help. What are these models? How do they work?

MEMBER CORRADINI: And does it only deal with pipes and valves? For the new members.

MR. BANERJEE: In a nutshell, one paragraph. What is an HRA model?

MS. LOIS: What is an HRA model?

MR. BANERJEE: Right.

MS. LOIS: I believe that an HRA model is a framework which you use in order to identify human actions that you would like to use or to take credit for in your PRA, and then once you identify the actions identify what would be the potential drivers for not performing the actual -- the action successfully.

And with that, you develop an algorithm or

a framework that would help you to come up with a probability, and there are several models. For example, the arm that is the latest one, ATHEANA, in order to come up with a probability estimate you use expert judgment.

Some other models, SPAR-H, the one that Dr. Apostolakis mentioned as being a different one, is -- it has guidance, starts out with what we call a generic human error probability, and then guide you through -- through different ways of, if you assume that, for example, stress is the most important factor, multiply your human -- your generic human error probability by a factor of 10 or 50 or whatever it is, if workload is another factor, multiply it by a factor of three. If now things are very good, reduce it with generic error probability. So it's a -- kind of a lookup table and cookbook, if you wish, guidance on how to come up with this -- with a probability.

MR. BANERJEE: Does a model try to predict the probability of success in performing a particular action?

MS. LOIS: If you put it in the positive -- I'm not sure --

MR. BANERJEE: Negative way --

MS. LOIS: That's right.

MR. BANERJEE: -- not performing.

MS. LOIS: Probability of failure.

MEMBER APOSTOLAKIS: Also, the factors that would affect that action. That's a major piece of these models. What is it -- as Erasmia said, first of all, the accident sequence context. Then, various what they call performance-shaping factors, like the stress level, whatever, all these -- the psychological factors.

MS. LOIS: Quality of training.

MEMBER APOSTOLAKIS: ATHEANA calls it the context. And the context is defined both by the control room context, the indications that they receive, and so on, plus the psychological factors. The SPAR model is more procedural, and it has levels of stress, levels -- I forget all the other factors. Give me a few.

MS. LOIS: Training procedures --

MEMBER APOSTOLAKIS: Training procedures. Okay. They say if -- this is the level of training, this is a factor that you multiply the basic human error probability, and so on. So they share a lot of assumptions, but also they differ in many ways. And then, you have EPRI that has curves over time that

give you the probability of an error, given the available time, with variations now. They refine it using the decision trees, and so on, so they focus on time.

This is one of my major problems here. ATHEANA and SPAR-H treat time as one of the performance-shaping factors. So you have stress, you have time, you have training, and so on. EPRI focuses on time and says, "How much time do they have?" They have five minutes. Now, what are the other performance-shaping factors that affect their performance? But the focus is always on the five minutes and what's the probability they will do it right or wrong. Okay?

So this is a major difference between models, and that's why EPRI -- and then, they use simulation exercises to claim that, you know, here is a curve applicable to these conditions. Here is another curve applicable to other conditions.

So, and this is a major difference between the --

MR. BANERJEE: What is the purpose of this presentation? Is it to tell us what's in these models, and how --

MEMBER APOSTOLAKIS: No. The complaint --

MR. BANERJEE: -- to rationalize them, or what?

MEMBER APOSTOLAKIS: The complaint from my party is, first of all, why should this agency have three different models?

MR. BANERJEE: Right.

MEMBER APOSTOLAKIS: Second, why haven't we reviewed the EPRI model, which is used in licensee submittals? And, third, why can't the community, the HRA community, develop -- start moving towards the development of a single model? So we don't have different assumptions, different --

MEMBER BONACA: You may want to mention the benchmark --

MEMBER APOSTOLAKIS: What?

MEMBER BONACA: You may want to mention the benchmark exercise that we always talk about.

MEMBER APOSTOLAKIS: Yes. And there was a benchmark -- there was a benchmark exercise conducted at ISPA more than 20 years now ago, where they had groups -- it was a European Commission at the time exercise. They had groups from the -- at that time there were I think 10 or 11 nations part of the Union, plus an American team.

They gave them an accident sequence in a

German reactor, explained it very well, and so on, and then they let them loose -- you go home and come back and give us probabilities for these human actions. And, of course, they were all over the place.

The same model at the time, which admittedly the models were not as sophisticated as today, the same model used by different teams led to widely different results. The same team, using different models, produced widely different results. In essence, it was a mess.

And somehow the community has ignored this, and we keep bringing it up, and, you know, nobody is willing to --

MR. BANERJEE: Are there experiments like simulators and --

MEMBER APOSTOLAKIS: Well, the staff --

MR. BANERJEE: -- virtual reality, or whatever?

MEMBER APOSTOLAKIS: The staff is sponsoring the simulator exercises at the Halden in Norway.

MS. LOIS: So I guess I should go to --

MEMBER APOSTOLAKIS: Right.

MS. LOIS: -- slide 11.

MEMBER APOSTOLAKIS: I think so.

MS. LOIS: Okay.

MEMBER APOSTOLAKIS: Now, the other thing is, for example, you started talking about the PTS and ATHEANA. In the spirit of my questioning today, I would ask, why ATHEANA? Why didn't you use SPAR-H for that? Why didn't you use the EPRI model for that? See, this is the comparative evaluation that we want to see eventually.

And, again, I'm asking these questions not to blame people for not doing this or that. I mean, whenever you have a new field, especially in the soft, so to speak, sciences of human reliability, it's natural that different teams around the world develop their own model. But it has been now more than 20, 25 years. Don't we need to start converging somewhere, especially as an agency? Why are we using SPAR-H for actual regulatory decisions and ATHEANA for research primarily?

MEMBER CORRADINI: What's the third one?

MEMBER APOSTOLAKIS: The third one is claimed -- claims to be deterministic. That was developed in the context of fires. So there is a fire, and somebody calculates that there is 18 minutes before there is core uncovery, for example, if you do nothing.

So now the analyst says, "Ah, okay. They will figure out there is a fire at that location within six minutes. They will develop a strategy, and then they will put it out or control it in seven minutes, so six plus seven is 13, you had 15, you have four-minute margin, you are okay." Nothing to do with probabilities.

These performance-shaping factors do not appear anywhere. They are a completely different approach developed by the same agency.

MS. LOIS: But could you -- could you clarify, what do you mean the performance-shaping factors do not --

MEMBER APOSTOLAKIS: Well, there is an appendix --

MS. LOIS: -- do not appear?

MEMBER APOSTOLAKIS: They appear in some sense, but it's not -- it's different from ATHEANA. It's different from SPAR-H. It just deals with time and takes the difference.

MS. LOIS: No. To the extent that we would have done an ATHENA analysis, all of the performance-shaping factors, if you will, are taken into consideration in NUREG-1852, we would have considered it, which is, do you have the staff? Are

they trained? All of these are part of the considerations. Does it mater what methodology you use -- SPAR-H or ATHEANA?

The only thing we did in the fire manual actions, because it was a deterministic approach, we believed that the acceptance criteria with respect to the adequacy of procedures, training, etcetera, are kind of the most basic criteria in order to ensure efficiency. And then, what -- what you ensure reliability?

Well, we thought that given that all of these things may not happen the way we anticipate, let's take the step to require a margin of time.

MEMBER APOSTOLAKIS: Well, yes, and I agree. And, obviously, in a short, brief summary, I cannot go into the details.

MS. LOIS: Yes.

MEMBER APOSTOLAKIS: But this is exactly what I would like to -- this is exactly -- I would like to see the three models next to each other, and the assumptions, and comparison, and so on, because if I were a licensee right now I would go with that. It's an easy way out. Calculate the times, give an estimate -- why bother about --

CHAIRMAN WALLIS: Does the word

"verification" appear in the vocabulary here? The word "verification" doesn't appear on any slides. Does it appear in the vocabulary of HRA?

MEMBER APOSTOLAKIS: No.

CHAIRMAN WALLIS: No? There's no verification of any model?

MEMBER CORRADINI: This is actually -- can I just -- I think this is what Sanjoy was eventually getting to, which I was wondering, which is, if I'm looking for some sort of experiment -- some sort of comparison, whether it be model to model, or something, because that's where I'm trying to judge something.

CHAIRMAN WALLIS: If you could start off with verification procedure or something --

MS. LOIS: Yes, that's what we plan for benchmarking. But we haven't been so far, but -- we haven't done any --

DR. COOPER: Yes, and I guess -- Dr. Susan Cooper, Research. Verification, in the sense that you can go out and do an experiment and exactly duplicate what operators can do, is not going to happen in HRA, nor are we going to be able to collect data. But it doesn't mean that there isn't anything behind these models.

The second generation HRA models, of which ATHEANA is one, and there are others like REMROS for media, have a lot of psychological, behavioral science, and cognitive science behind them, in addition to the fact that they've looked at a lot of operational experience from the nuclear power industry as well as others.

CHAIRMAN WALLIS: You answer questions such as, what's the uncertainty in your estimate? And you come up with a number of .1 that the person will make the right decision -- probability. How do you assess how good that is?

DR. COOPER: Well, that's a different question, and most HRA --

CHAIRMAN WALLIS: But isn't that the sort of question we are asking, should be asked?

DR. COOPER: I don't know that I've heard anyone ask, really, what uncertainty is.

CHAIRMAN WALLIS: Well, I see -- I mean, I see numbers. It's either .1 or it's --

DR. COOPER: No, I think that's a quite different --

CHAIRMAN WALLIS: -- 01 or it's --

DR. COOPER: -- question, to be real honest. I think uncertainty has to --

CHAIRMAN WALLIS: When someone gives it to me, I say, "Well, where does it come from, and how sure is it?" And there doesn't seem to be an answer.

MS. LOIS: Okay. So, then, I would appreciate if the committee lets me go through the slides. I just jumped to slide 11. This is the treatment of the difference between the fire manual actions and --

MEMBER APOSTOLAKIS: The committee hasn't seen my questions, I think. Have they seen my questions, Eric?

MR. THORNSBERRY: In the status report they are discussed.

MEMBER APOSTOLAKIS: All right. If you have that slide. Okay. So they know what it is.

MS. LOIS: So these are the questions.

MEMBER APOSTOLAKIS: Okay.

MS. LOIS: All right? And we'll try -- and we believe that we have some answers to these questions, and we believe that the benchmarking using simulator data activity will help us address some additional questions.

MEMBER APOSTOLAKIS: Okay.

MR. BANERJEE: In a simulator, what is actually measured? Like suppose the simulator has to

shut down something, that's the activity the human has to undertake. What do you measure, and how do you measure this? How does the simulator --

MS. LOIS: The coding facilities where we plan to have this -- what we call benchmarking exercise, they have -- their facilities is to perform -- are set to perform experiments for collecting human performance data. So, in actuality, what happens is they have operator crews, real crews, mostly from European countries, although there is a very good possibility that U.S. countries are going to be used as well.

And there are various analyses, like in training, when the operators are trained in the simulators they -- they have to deal with LOCA events or loss of offsite power events.

So there are very well pre-set scenario set out, and there is a very detailed data collection in terms that are video cameras that are observing, there are experts that are observing human performance, as well as there are debriefing protocols that help identify, if operators did a mistake, why they did it. And that's a very crucial aspect for human reliability perspective.

MEMBER APOSTOLAKIS: But what they measure

-- what they measure is the response time. That's what they measure.

MR. BANERJEE: That's the key -- time.

MEMBER APOSTOLAKIS: The time.

MR. BANERJEE: Okay.

MEMBER APOSTOLAKIS: So what they do in some of the experiments, they change the conditions, for example.

Okay. There is a LOCA in the simulator, things are going as expected, and they will see how they respond. Then, they create a diversion that the staff now -- the crew has to take care of something else while the LOCA is occurring. What is the new response time? Okay? And they do things like that. I mean, they are very well thought out experiments, and they produce --

MS. LOIS: But we have helped them to get away a little bit away from the -- that's what they were doing in the past for -- called human factors applications. But we -- for human reliability, we believe that you should allow the crews to take the time and see actually what happens if they have a little bit more time.

And, therefore, there is -- those scenarios are set out to -- for an hour, an hour and

a half, so it's not just if you didn't perform within 10 minutes or you failed, if you didn't perform -- I mean, we have observations that actually crews meets to perform the action, and probably started doing something which was entirely different than what they were supposed to do.

MEMBER APOSTOLAKIS: Well, yes, there were variations. There are variations, and so on.

CHAIRMAN WALLIS: Well, that must be a variable. Did they do the right thing? That must be --

MEMBER APOSTOLAKIS: Yes.

CHAIRMAN WALLIS: The time is one of the drivers.

MS. LOIS: It isn't a response time related -- only that. I mean, that's one aspect of it.

MEMBER APOSTOLAKIS: But that's a key aspect, though, is that not? It is a key aspect, because the thermal hydraulics controls that. You have certain time for response.

DR. COOPER: Unless it's an error of commission.

MEMBER APOSTOLAKIS: Unless it's an error of commission.

MEMBER BONACA: Unless you have an error. I mean, Halden was quite interesting. I mean, what they showed here. And, again, it was complicated, too, in the sense that some crews were very effective in one way, and then the were ineffective in other ways.

MEMBER APOSTOLAKIS: Yes. And the EPRI exercises of 15 years ago or something. There were also simulator exercises. In fact, they produced curves using the results from the simulators, and they had certain hypotheses that they tested, you know, that the operator response time behaves this way, they got the test of it, sometimes it failed, sometimes it worked. So there were significant efforts.

MR. BANERJEE: Maybe that's why EPRI focuses on time.

MEMBER APOSTOLAKIS: They focus on time. There is also a point of view. I mean, they really want to develop a method. That's why they have this calculator, which is software based. They claim that we should have a method that a good engineer can use without being an expert on HRA. I mean, he should -- he or she should know something about it. I mean, it's not like you go blindly or apply it, but they try to proceduralize it as much as they can.

MS. LOIS: Which we object.

DR. COOPER: Yes, we do.

MEMBER APOSTOLAKIS: Okay. Well --

MS. LOIS: This is our fundamental objection with this methodology, the fact that you have a capability. My thermal hydraulics tells me I have half an hour. I put in half an hour. Here is my probability. It's -- you can become blind as to what are your potential drivers of human error, and why half an hour was not -- was not enough, etcetera. So that mechanistic approach to human reliability is the --

CHAIRMAN WALLIS: Something is missing from this. The output of this must be a probability you put in a PRA or something.

MEMBER APOSTOLAKIS: That's correct.

DR. COOPER: That's the result.

MEMBER APOSTOLAKIS: An output. I mean, the question is: how do you arrive at that output? And time presumably is a means to an end. You've got to do something with that time when you get it.

CHAIRMAN WALLIS: Right.

MEMBER APOSTOLAKIS: So I haven't heard that mentioned yet.

MEMBER ARMIJO: I have a question. This

has been going on for --

MEMBER APOSTOLAKIS: A long time.

MEMBER ARMIJO: -- 20, 30 years. I know in Halden, NRC has been a member forever. I believe EPRI has been a member forever. It's surprising it hasn't converged to a point where there is maybe two major competing models, and the issues are well defined, and now you're going to set up some sort of method to resolve those issues. Is that where we are?

MS. LOIS: Halden has been involved in human reliability-related research the last three years or four years.

MEMBER ARMIJO: Before that it was called human factors.

MS. LOIS: It was human factors.

MEMBER ARMIJO: Not the same thing?

MS. LOIS: It is not the same thing.

MEMBER ARMIJO: Okay.

MEMBER BONACA: But, you know, I mean, the one big issue -- question is always, how representative is this of what takes place in the powerplant? Because the powerplant has certainly a huge edge in the sense that they have the procedures. The operators are trained continuously on those procedures. The simulator puts in front of them

problems which have to do with their own specific console, controls, procedures.

So they operate now -- that may set also a stage for some trap for them, because they are so used to certain circumstances. Something can happen that throws them off. But this is different from what takes place at Halden, for example, where you have these people going in, and I don't know how trained they are in specific procedures for the powerplant.

MS. LOIS: They are actual --

MEMBER APOSTOLAKIS: They are.

MEMBER BONACA: They are. But, you know, I don't know how it compares to crews that live at the plant for years and years.

MEMBER APOSTOLAKIS: Let's pursue this a little bit, what Erasmia mentioned. I said that the EPRI guys tried to proceduralize, produce curves, and so on. On the other side, ATHEANA -- both methods have a very detailed evaluation of the context. EPRI doesn't call it context, but essentially they are also looking at the performance-shaping factors, what is the accident sequence, and so on. This is the shop framework. So there is a commonality there.

Then, when it comes to producing probabilities, ATHEANA says essentially that you

should assemble a group of experts who will evaluate all these factors and the context, and so on, and will develop a probability distribution for the human error.

The other side is the EPRI guys, which Erasmia said they were objecting, that has proceduralized that. It has curves, and so on, and there is some flexibility, but essentially you have to follow what they are telling you.

And one question, for example, that I think we should try to address as an agency is: can we merge these two? Is it possible to bring some of the EPRI approach into ATHEANA and some of the ATHEANA approach into EPRI, and come up -- because, you know, when you tell people who are doing a PRA, for example, that they have to have a group of experts to do this, that's a very expensive proposition.

So there are advantages and disadvantages. It probably is a more thoughtful and detailed evaluation of the context if you have experts that are doing it, and perhaps you should do that for one or two or three events or human errors that are of extreme importance to the plant. But should you be doing it for all of them?

CHAIRMAN WALLIS: George, can I ask you

something? When anything matures as an engineering discipline, people can teach it. And if I were to teach students this, I'd want to teach them how to do it. I'd want to have some way of testing whether they did it well. All those things that we do in every other discipline -- this doesn't exist in this area at all?

MEMBER APOSTOLAKIS: Well, I think, Graham, you can only take the analogy from the sciences, the hard sciences so far here. I mean, yes, it would be nice to have experiments that would validate and --

CHAIRMAN WALLIS: Oh, right. If I had to teach --

MEMBER APOSTOLAKIS: -- how you --

CHAIRMAN WALLIS: -- what would I say?

MS. LOIS: We are going to do it, to the degree that we can.

MEMBER APOSTOLAKIS: What do you mean? You will present these models and tell them how to do that.

CHAIRMAN WALLIS: By saying the way you do it is hire a group of experts?

MEMBER APOSTOLAKIS: Well --

CHAIRMAN WALLIS: You said homework on

that basis.

DR. COOPER: No. But for this particular case you need experts. At least, in the ATHEANA case, it would be trainers from nuclear powerplants. And so we present to them the context, which could include time and does often include time as a factor. And have them evaluate for their own plant and their crews, what their judgment is so far as whether or not those crews would fail in that particular situation.

So, and the purpose of having the experts together and not just one person is so that, in fact, you can have discussion about that context, because one of the things that was found in the old, old benchmarking study is that many of the differences had to do with the fact that people were studying the different problem.

They were thinking -- they had to make assumptions. In other words, it wasn't a completely defined problem, and they had to make assumptions about what the context was. And so, in fact, they were actually analyzing different things.

So if you get people in the same room and they -- you know, you describe the context in a certain way, which may include time, and ask them to -- okay, so this is the situation, what is your

opinion, which is an expert opinion, because they've been watching their crews going through the simulators, and probably they've even done some work for this particular analysis, what's your opinion/judgment on whether or not they'd be successful in this particular case?

And you might find out in the discussion that somebody says, okay, I think -- I think it's this, and I think this is a really important factor and we haven't considered it yet. And then, maybe you, you know, change the context or you define two different events, and, you know, the idea is to reach some consensus but also a common understanding of what that context is.

Now, it's our opinion that, although we do believe that time is important, but for a long time now, even back in the '90s, there was discussion in the HRA community that to be too focused on time can get you in trouble.

I mean, you plug in 30 minutes for this event, you plug in 30 minutes for this particular human failure event, and the conditions are very different. In some cases, it could be very close. You may not be successful. In another it's plenty of time. But you've got the same time.

CHAIRMAN WALLIS: But the question I have is: how does it work? I mean, the agency has to evaluate something submitted by some utility or something. They say, "We hired a group of experts, and the experts said that the probability of doing this right is point one." Now, how does the agency evaluate whether that's a reasonable number or not?

DR. COOPER: Maybe we ought to let Gareth --

CHAIRMAN WALLIS: They must have some way.

MR. PERRY: Yes, this is Gareth Perry from NRR. I've been trying to sit in my seat without jumping up too much, because I've heard a lot of things I disagree with here.

CHAIRMAN WALLIS: What, with the questions or the answers?

(Laughter.)

MR. PERRY: Everything.

CHAIRMAN WALLIS: The questions are hard to disagree with.

(Laughter.)

MR. PERRY: One of the problems that I -- I'm not sure that you're tackling the right problem. I don't think we need a one size fits all method for every application. For example -- let me give you --

we use, for the SDP, the significance determination program, we use the SPAR-H model. The question we should be asking is: is that good enough for that purpose? Because certainly any proposal to use ATHEANA for SDP is like using a sledgehammer to crack a nut.

So I think the things we have to understand is where the applicability of these models is, and that I think is what the comparison of the good practices -- comparison of the methods against the good practices provides a good basis for this document.

Now, another thing, I want to raise an objection to Erasmia's objection that you can't have a method that you could proceduralize that could be used by non-experts. I think that was one of the bases of many of the EPRI methods, particularly the CBDT method, which is the decision tree approach.

And the idea behind that is is you bring in the knowledge that you have about the aspects of human performance, put that in the model, and then train the person to recognize which of those factors are relevant for the particular sequence that he's dealing with.

CHAIRMAN WALLIS: Well, you're teaching

someone to do something.

MR. PERRY: Well, that's the idea, to try and -- at least to embed it in the structure of the model, so that you can do that. And personally, I think that's a very useful thing for people in NRR who are reviewing licensee's applications.

MEMBER APOSTOLAKIS: But let me disagree with you now. It's one thing to say we need a model for this particular application that's appropriate for it, and quite another to actually look at what the model does.

Let me give you an example of reactor physics -- obey the Boltzmann question, period. That's how they move in a reactor. It's also -- you can solve it that way.

CHAIRMAN WALLIS: Wait, wait.

MEMBER APOSTOLAKIS: You can solve it. You can solve it. You can solve the equation. There are methods for different applications. In a time-dependent situation, the simplest one is the point kinetics. For certain application, it's hard. For other applications you go to multi-group --

CHAIRMAN WALLIS: You can test it, George.

MEMBER APOSTOLAKIS: -- you go to -- well, wait a minute. You go to multi-group diffusion

equations, and you solve those using sophisticated methods. Okay? Sledgehammer and all of that.

But all of these methods are produced from the Boltzmann equation, making approximations. I've done it many times, Graham.

CHAIRMAN WALLIS: But you don't have to do it that way.

MEMBER APOSTOLAKIS: But we -- my point is that they are all based on the same -- on the same physical processes, and then you make approximations. SPAR-H used different assumptions from ATHEANA. That's my problem. I don't mind having a simple way of handling routine regulatory applications, but it should not really be different --

CHAIRMAN WALLIS: But, George --

MEMBER APOSTOLAKIS: -- it's not entirely different, but it --

CHAIRMAN WALLIS: But, George, it's hopeless, because you say, first, I believe the Boltzmann equation, and then I deduce everything. In this area there's nothing you can believe as the fundamental equation, deduce things.

MEMBER APOSTOLAKIS: But give me the benefit of the --

MEMBER POWERS: Graham, come on.

MEMBER APOSTOLAKIS: I mean, if ATHEANA believes that certain assumptions are very important, and these assumptions are not in SPAR-H, you're going to have a problem.

MR. BANERJEE: Well, the Boltzmann equation is a model, albeit not very exact, for some types of behavior. So is there an equivalent, however approximate, model for human behavior? If there is not, you don't have the equivalent to it.

MEMBER APOSTOLAKIS: Well, wait, wait, wait. I think my point is that even in cases where you have the fundamental equation, you have to develop models like Gareth says that are applicable to different situations and have different degrees of flexibility and accuracy, point kinetics being the crudest.

But all these models have these fundamental -- this fundamental process under them. Okay? They are approximations and can show how you produce them. Here you don't have that. You don't have that. But that doesn't mean that you can make any kind of assumption you want to develop your own model.

At some point you have to compare them. You have to compare them and say, "When SPAR-H gives

me six or seven levels of stress and puts a factor of ten here and a factor of five there, how does that compare with something that ATHEANA does in a similar situation?"

MR. BANERJEE: Sounds like biology.

MEMBER APOSTOLAKIS: Well, it could be.

MS. LOIS: But the history of human reliability, I don't think we can take it back. It was -- this is the evolution of these methodologies and we do believe --

MEMBER APOSTOLAKIS: And nobody disagrees with that.

MS. LOIS: -- that the -- through these benchmarking exercises we will be able to address exactly those questions, in the sense that we are going to test the underlying assumptions of SPAR-H an ATHENA and THERP -- that has been still used -- and through that exercise we'll be able to compare, to see the differences, and then also determine the applicability of the method or how we can improve the method.

MEMBER APOSTOLAKIS: Well, then you are on the way of doing what I want. But what I would like to see first is a comparison on a table with columns. Before you do any benchmark exercises, you say, okay,

this guy does this, this other guy does that, how do these two compare?

MS. LOIS: But we've done that in 1842.

MEMBER APOSTOLAKIS: No.

MS. LOIS: We have tables where we compared -- I mean, we do not compare it, but we identify the basic assumptions in these methods.

MEMBER APOSTOLAKIS: No, you haven't' done it the way I want it. I want you to go to SPAR-H when they have any questions for the dependency and beat the hell out of it, and say, "Why is this true? How do other models handle this?" There are some equations there that come out of the sky and you're a -- I'm scratching my head to why this is true and nobody questions it.

MS. LOIS: That can be done only through collection of data and -- because we are not going -- in the benchmarking exercise, we are not going to compare methods. What we are going to do is we're going to evaluate, if you wish, every individual method in its merit.

So the plan is, and probably I have that in my backup slides --

CHAIRMAN WALLIS: I'm wondering, what do we expect the committee to do with this? I'm sort of

struggling with this, and --

MEMBER APOSTOLAKIS: Well, the first question, is the committee happy with having three NRC models and one industry model?

CHAIRMAN WALLIS: Well, three codes for thermal hydraulics.

MEMBER POWERS: George, I think there is no inherent reason that you wouldn't have three models. Now, to have them on a different philosophical and technical bases is a little more distressing. But there is nothing inherently wrong with having three models.

MEMBER APOSTOLAKIS: Oh, no. No, I think it's --

MEMBER POWERS: And I don't think you think that either.

MEMBER APOSTOLAKIS: That's not what I mean, no.

MEMBER POWERS: But I don't -- I don't quibble any with your objectives here. I a little bit quibble with educating the members at the table. I think it would be useful to go through the presentation. Paging through it, it looked like it was a useful exposition on what the research program is.

I would like to explore further what's meant by benchmarking, and I think they'll get to it. And, of course, I'm very interested in how they use maybe flight distributions.

MEMBER BONACA: The only place where I would have a major problem with the three different models would be if for the same scenario people not familiar necessarily with all of them will come up with very significant differences. That would be troublesome, because then, how do I judge that, you know, for the SDP it -- is it proper to use SPAR-H? I mean, do we know that?

MEMBER APOSTOLAKIS: I really think it is inappropriate -- is for us to accept results from the EPRI calculator without a review of the model.

MEMBER MAYNARD: I believe that it's worthwhile to continue to try to come closer together on these things, but I'm not sure you're ever going to get to one method. And I know that from experience when the industry has an issue or a model, the NRC will use their models, and where there's a difference then they get together.

And it's up to the industry to then prove that -- you know, if the industry is coming up with better numbers or so to speak, that -- you know, the

NRC number prevails unless the industry can convince them that, you know, their model doesn't account for everything.

And the industry -- there's a lot of information out there. There's a lot of things that are being done in the simulators. There are exercises going on all the time that toss in a lot of just things that distract you and different things like that. So I think we're getting more and more data that -- it's not just an opinion by somebody as to what or may not happen. There's a lot of data to back up the performance.

MEMBER BONACA: But again, I mean -- but again, however, you should have some consistency. What I mean is that -- take a critical scenario that everybody is taking credit for in PWRs -- bleed and feed. Now that's a fundamental scenario for some type of plants. For example, the C plants or the early design, there is a very narrow window for being successful.

If you do it too late, you're not going to succeed. It will be interesting to know, given the scenario with some complications or whatever they may be, if you get very different results that says with this model you are never going to make it, with this

model you'll make it with success, and then I would like to understand, you know -- that's really what I would have --

MS. LOIS: Well, exactly that's what we are going to try and hold it, to set up scenarios that probably pertain to some --

MEMBER BONACA: Okay. Good.

MS. LOIS: -- very important human actions.

MEMBER BONACA: Yes.

MS. LOIS: And the analysts, the experts, ahead of time they will do their predictions. Given that scenario, that specific plan characteristic, you know, get all the collection and collect all the information you would have when you -- in HRA by yourself. And you would do your predictions.

And then, afterwards, we are going to see what --

MEMBER BONACA: Yes. No, I understand.

MS. LOIS: -- how well and why if you didn't predict well, why -- and if you did predict well, why. And we are going to compare all of that.

MEMBER BONACA: Okay. So you'd use three different methods of -- all different ones.

CHAIRMAN WALLIS: The thing is, if I read

-- I mean, I read the Halden report, the previous report. What I would like to see in the future, for example, from a similar report is to say, okay, we have this scenario. If I take SPAR-H, this is what I'm supposed to do to calculate some probability. And this is what the simulation will be to test whether these guys are doing the right thing.

Now, ATHEANA will do something else. So this is how we're going to test ATHEANA. Rather than have those guys run their scenarios and do whatever --

MS. LOIS: Exactly. That's what we tried to do. Here is -- define the measure --

MEMBER APOSTOLAKIS: Which is sort of --

CHAIRMAN WALLIS: We're supposed to write a letter on this research program, and I don't have a clue what it is yet. So how can I write a letter on it? I mean --

MEMBER APOSTOLAKIS: Well, no, we are not writing a letter on the research program.

CHAIRMAN WALLIS: We seem to be going into all sorts of stuff, which is very interesting, but what is the program we're reviewing?

MEMBER APOSTOLAKIS: This is not a letter on the research program. We have already done that.

CHAIRMAN WALLIS: It says it's a

presentation on the HRA research program.

MEMBER APOSTOLAKIS: Well, no, it was on this -- the fundamental question in my mind that triggered this meeting is, is it appropriate for the NRC to have three different models based on three different assumptions? Not completely different -- I mean, they share a lot.

If, for example, SPAR-H was presented and developed as an approximation to ATHEANA, then I wouldn't have any problem, but that's not how it was developed.

CHAIRMAN WALLIS: How can you approximate an expert elicitation? I mean --

MEMBER APOSTOLAKIS: If it was presented that way, there would be no problem. The other thing that bothers me is that in the regulatory arena we are accepting results from a model that there are answers that have not been reviewed officially. I would like that --

VICE CHAIRMAN SHACK: But they reviewed the results, George.

MR. PERRY: I'm not sure that we necessarily accept those results. We just don't find them unacceptable, but it's not that we're endorsing those results. And, no, that's -- there's a

difference in that statement. Personally, I don't believe that for most of these power uprates that you need to use these time reliability curves, because I don't think -- I don't believe, from my understanding of the procedures, that the shift in time from 25 minutes to 20 minutes makes much of an impact on the way the operators are --

MEMBER APOSTOLAKIS: There was a case of eight to six. There was a case of eight minutes to six minutes. Any time we use a model, we have to make sure we review it and we understand it. And that's not in this case. I don't know. Is this a different field where we don't apply these --

MS. LOIS: This activity is going to give us that opportunity.

MEMBER APOSTOLAKIS: Okay. Great. If it does, it does.

MS. LOIS: And I'm sure that Jeff Julius is on the telephone. But the assumption is that every --

MEMBER APOSTOLAKIS: So tell us what -- I mean, we interrupted you. What are you planning to do? Maybe that's what's missing from this discussion.

MS. LOIS: Okay.

MR. JULIUS: I'm online.

MS. LOIS: Hello, Jeff.

MR. JULIUS: Hi.

MS. LOIS: Do you want to speak out for --

MEMBER APOSTOLAKIS: It's very early for him, by the way. He's on the west coast.

MS. LOIS: Yes. For your participation in the benchmarking exercise.

MR. JULIUS: Yes. I guess the statement I wanted to make was that where we are converging is on these performance-shaping factors. If you look at the basis for the SPAR-H performance-shaping factors, and the SPAR-H -- and the EPRI HR calculator to perform the shaping factors, and this is the same performance-shaping factors I believe are used in the ATHEANA -- the baseline quantification or as part of the ATHEANA process to look for deviations from the baseline quantification.

We have converged on those, and those are what are published and being collected in the NUREG on HERA. And then, the question is now, are we looking at the -- how these are wired up or what -- the impacts of these performance-shaping factors. So we have I guess reached beyond the methods, and agreed at least upon a baseline set of performance-shaping factors that we're looking at.

MS. LOIS: But, however, in terms of participating in the benchmarking exercise, that would allow us to compare notes in a much more detailed manner. It appears that it's a guess, right?

MEMBER APOSTOLAKIS: Is EPRI participating?

MS. LOIS: I assume so. Jeff?

MR. JULIUS: Yes. Yes, EPRI is.

MEMBER APOSTOLAKIS: The thing is this: if you look at the methods, they share a lot of common elements. They do. It's not that they are completely in different directions, but they also have differences. And the simple question I'm asking is: has anybody sat down, looked at them critically, and said, "This is where they really differ, this is where they are doing the same thing," and perhaps by doing so start creating the basis of a more unified approach. It's a very simple question.

Because I don't -- I repeat: the purpose of this is not to blame anybody. I agree with Erasmia that historically that's how methods evolve in a new field. Okay? People develop what they believe is the appropriate way to approach it, but at some point --

CHAIRMAN WALLIS: Let's see the plan for evolution, then. I mean --

MEMBER APOSTOLAKIS: That's exactly what the letter is going to say. That's what the letter says.

CHAIRMAN WALLIS: Well, I was hoping I was going to see. It was sort of --

MEMBER APOSTOLAKIS: That's what the letter says.

CHAIRMAN WALLIS: I understand people are converging on performance-shaping factors. That's a step forward. Now, how do they shape performance? Is the next question perhaps, the performance-shaping factors and how they're being addressed.

MEMBER BONACA: I would like to say, I mean, on behalf of what George is trying to do, I mean, the issue of human reliability is very important. When the IPEs were submitted originally, or at least -- you know, the estimations from plant to plant, they were all over the place. I mean, they were wild. There were order of magnitude estimation differences between different plants, etcetera.

So how can you believe the results of PRAs that we, you know, base our judgment so much when you have embedded in those these wide variations?

Now, if something has been done, but still now there are big variations between the reliability

of human action in different PRAs, and that skews the results. That's the fundamental reason why I've always believed the SPAR program is so fundamental for the agency, because it's one model and hopefully also in the HRA is going to be some consistency there, so that you have some consistent approach.

So if you have the type of plant, you know, there are five or six of those plants, you will have certain expectations if the -- this proper training is also -- so that's a very important issue, because --

MEMBER APOSTOLAKIS: And the --

CHAIRMAN WALLIS: George, I'm going to assert some authority here. I mean, it seems to me we're asked to decide to comment on what the staff is doing, and we have to know what it is. And we just keep talking around this thing. Can we sort of agree that they have 20 minutes or something to tell us what they're doing?

MEMBER APOSTOLAKIS: Yes.

CHAIRMAN WALLIS: Can we agree that? Because we just -- they never get going on anything here. Can we agree that?

MS. LOIS: I guess at this time we have kind of exhausted our presentation. The only thing I

can -- the only discussion, only topic, and we would like to really --

VICE CHAIRMAN SHACK: Let me ask a question about the benchmarking studies.

MEMBER APOSTOLAKIS: That's why we don't do that.

VICE CHAIRMAN SHACK: When you have something that has a relatively low air probability, just how are these experiments done? I mean, I presume that these people don't fail all that often. You can't run the same experiment over and over again. How is it actually done?

MS. LOIS: It's what Dr. Apostolakis explained before. You start out with a -- it's been called basic scenario, which is a well-trained scenario, and you have the capability to observe --

CHAIRMAN WALLIS: Can I establish something? Erasmia, you said you didn't want to follow my process. You don't have enough to present. You'd rather have a conversation with the committee, is that okay?

MS. LOIS: Yes.

CHAIRMAN WALLIS: That's okay. All right.

MS. LOIS: I would like to finish with --

VICE CHAIRMAN SHACK: I have to understand

what it is they're doing.

MEMBER APOSTOLAKIS: Well, let her finish first, and then -- why don't you finish, Erasmia, and then we'll --

MS. LOIS: Okay. I will answer your question. The only thing I would like to add here is that we are not quite sure how we are going to determine success in the benchmarking exercise. It's a very early process. I would like to, you know, personally express appreciation for what the committee is doing. It helped -- you are helping us on that. It's not -- I think we are in full agreement here.

It's not that we know how to do it, but -- and we both recognize that the variability in the bottom line number among methods is an issue that we have to address. We believe that this benchmarking exercise -- I don't know if the right word "benchmarking" -- but this exercise, by observing simulator crews to perform, and then collecting the data, having the experts have -- ahead of time to have predicted -- predict what are the potential failures and to what degree --

VICE CHAIRMAN SHACK: But suppose he predicts an error rate of .01, how do you -- how do you measure that in the experiment?

MEMBER APOSTOLAKIS: You don't.

VICE CHAIRMAN SHACK: You don't. Okay.

MEMBER APOSTOLAKIS: That's a straight answer. You don't. The simulator exercises will not produce Monte Carlo simulations where you calculate the probability. They are evaluating assumptions in performance-shaping factors. Like, you know, one of the results, as I remember, they tested four crews or five. Four of them did something within five and a half to six minutes. One of them was 11 minutes. And then, they asked, what? What happened? What was the factor that affected them? This is the kind of fundamental insight you are going to get from this.

MR. BANERJEE: And what was the factor? I mean, can we have something concrete to -- as examples?

MEMBER APOSTOLAKIS: It's in the report. I don't remember.

MS. LOIS: The main factor in that specific case was communications among the crews.

MEMBER APOSTOLAKIS: Yes.

MS. LOIS: How the SDA was not -- the way they were doing their work, people were totally not communicating about what they had to do. So we really find some very important things. And to answer your

question here, you do not have the capability to run a thousand experiments of the same, but you do have the capability in the simulator to have -- to make scenarios a little bit more difficult. And then, you are observing some failures, and you understand why the --

VICE CHAIRMAN SHACK: But you'll still have to make a judgment, then, to get your probabilities.

MEMBER APOSTOLAKIS: Exactly. Exactly.

VICE CHAIRMAN SHACK: It will just be a more informed judgment.

MS. LOIS: That's the method, the judgment --

MEMBER APOSTOLAKIS: Yes.

MEMBER CORRADINI: Can I repeat what I've heard, since we're -- you have -- there is a few -- I heard you agree to a few ground rules, which surprised me but it's great, which is the presentation is kind of over, conversation is okay. So as part of the conversation, I want to repeat some things so I get it right.

One is, there is three NRC models, that you do agree with what George's hypothesis -- or thesis was, and Dana restated it, but it -- it struck

me as interesting, which is they do have fundamentally different assumptions. They're not --

MEMBER APOSTOLAKIS: Some.

MEMBER CORRADINI: Some. They're not like, you know, Model X, and then Model Y is just a more detailed Model X, and Model Z, which is just a deterministic, less detailed, or a different branch of Model X. There are literally three models with potentially three different sets of some -- some of the assumptions were fundamentally different. I heard you kind of agree to that. Is that true?

MS. LOIS: That is true.

MEMBER CORRADINI: Okay. All right. So that's one.

Two, that the Halden exercise, as you've been trying to explain it -- I'm still not sure exactly what it is, but it's the equivalent of I went to Kewaunee, and I watched them run essentially a small break LOCA in their simulator, watched the crews respond to it, except that you run it with five or six different crews, and then you threw curves at them in terms of what should be the standard operating procedure to address a small break and try to give them deviations and things that will try to knock them off and see if they either succeed or don't succeed.

Did I hear that right?

MEMBER APOSTOLAKIS: That's also correct.

MS. LOIS: Yes. But ahead of time, you are -- now your expert -- your HRA expert would know that Kewaunee has a -- had a small break --

MEMBER CORRADINI: Yes.

MS. LOIS: -- and these the correct --

MEMBER CORRADINI: So the three models of the four models were predicted.

MS. LOIS: Use your numbers --

MEMBER CORRADINI: Okay.

MS. LOIS: -- for those situations, and then you observe what happens.

MEMBER CORRADINI: But then you said something that really got me, which is after you did that you said you're still not sure how to evaluate the results of the experiment relative to the predictions. Did I mishear that?

MS. LOIS: No, no. I said it will give us the capability to evaluate.

MEMBER CORRADINI: And so now, like an experiment that I do in my lab, or somebody does for me in my lab, since I don't do that anymore, is so do you have the attributes and the procedure to do the comparison, or are you going to do the comparison

procedures on the fly?

MS. LOIS: Okay. So then, we are determining the -- we are determining the -- what we call "experimental design." We have to -- to define what we call "success," at what point we would be able to say that, yes, SPAR-H successfully predicted, to come up with the number it's -- you know, given the small number of experiments, it's very unlikely.

But if, for example, we see that workload is a big issue in this specific example, and SPAR-H identified workload as the driver of the human failure, we believe that this is really good.

MEMBER APOSTOLAKIS: In the experimental design, though, you will bring all four models.

MS. LOIS: Yes.

MEMBER APOSTOLAKIS: Okay.

MS. LOIS: We are going to bring every -- I believe -- correct me if I'm wrong -- ACRO and CBDT.

MEMBER APOSTOLAKIS: Okay.

MS. LOIS: And the NRC and I guess -- NRC and EPRI is going to benchmark or to test PIRT, and then we have ATHEANA and SPAR-H.

MEMBER APOSTOLAKIS: That's very good. That's very good. Now -- I forgot what I was going to say.

MR. BANERJEE: What's even more important, something that you said, was you tried to understand the reasons why these models don't agree. And I don't know if such a small set of experiments can actually shed light on what is right or wrong, but it's worthwhile finding out at least what the reasons are, you know?

MEMBER CORRADINI: So that actually leads me to another question, which INPO, for their training and their reaccreditation of all the plants, part of the observations are always these crew observations. Is there just a disconnect on -- or is it inappropriate to understand from all of the simulator training and all of the various events at all the plants, to try to extract something that you can use as a comparison to these models? It just seems to me they are doing this again and again and again at all the plants, at least when I was at Kewaunee watching this

MEMBER BONACA: Yes, but they don't -- they don't put in the monkey wrench that they do at --

MEMBER CORRADINI: Well, yes, they do, and the simulator training is --

MEMBER BONACA: Yes, you do some variations, but I'm saying that when you are looking

at the center-oriented procedures to see if they -- they seem to fall back on the center oriented. But this is something -- something different.

MEMBER APOSTOLAKIS: It is different.

MEMBER CORRADINI: So it's a more controlled environment.

MEMBER APOSTOLAKIS: I would like -- IU mean, this is great. But what I also would like to see is, before that or in parallel with this, to see a critical evaluation of the details of the models by people like you. What do I mean by that?

You mentioned the two reports, the NUREGs that looked at model's best -- good practices, and so on. They are -- as I recall the good practices report, it said you had a number of steps that you thought were the good -- a good thing to do, and then you searched to see whether each model -- how each model addressed these steps.

And it was at a certain level that said, yes, this model does do this. Okay? This model does it peripherally, but not in detail. What I'm saying is -- what I mean by "critical evaluation" is to go a couple of levels down and say, yes, this model does it. They account for dependencies, they account this way, and we think that's not right for such-and-such

a reason.

In other words, evaluate the way they do it, not just the fact that they do it. And I think that will be a great thing to have in addition to the experiments, and then I think everybody will be -- will begin to -- first of all, another thing that I have noticed, and maybe you disagree, but as you know I had an opportunity to look at the EPRI model and the ATHEANA model in more detail, and I must say I was surprised by how much -- how similar they are in many respects.

I thought that, you know, the EPRI model they -- and I hope Jeff will forgive me. I thought it was on a much shakier ground than it turned out to be. And we had access to somebody from a utility who is actually using the model, and his response to a lot of questions -- of our questions were very reasonable. In fact, they were doing many of the things that ATHEANA does.

So my -- the thought in my mind is: why, then, not try to blend them?

MS. LOIS: Okay. I think there are two things.

MEMBER MAYNARD: At least approve it. You would approve --

MS. LOIS: Two different things. One is what we call good practices. And this is the SHARP-1 framework that EPRI created at the beginning of HRA, and then the good practices that we lately documented. Those guidance documents tell you how to create your HRA model as interaction with the PRA.

I believe that the calculator has been improved tremendously after the good practices computation. And, therefore, a lot of the fundamentals that -- if this was derived from the ATHEANA development and from reviewing the IPEs and really developing an experience of what's going wrong, you know, why results in human reliability are so different, have addressed through those guidance documents.

And we believe that the calculators probably will be a very good tool to do --

CHAIRMAN WALLIS: Erasmia, can --

MS. LOIS: We do not object to that.

CHAIRMAN WALLIS: -- can I now begin the conversation? I see you've got a slide up here. I'd like to address that. I mean, you have a plan. Now, when you have a plan, I first like to see what's the objective. And the objective, I gather, is to assess the validity of several models. If it's not, then

tell me something else.

MS. LOIS: Yes.

CHAIRMAN WALLIS: Then, you're going to run some experiments and collect data. So where we might be able to help would be if you could tell us why these particular experiments measure the key things which enable you to evaluate the models, and what kind of data you need to collect, you know, maybe how many you need to collect.

You know, all those kind of -- are you just exploring the kinds of things which might influence behavior, or are you actually assessing and validating some models, which seems to be part of the discussion? In that case, the plan has to address that in some specific way.

MS. LOIS: So we'll be happy to come back in January and address that.

MR. BANERJEE: Is this a major facet of your verification and validation program for these models, these experiments?

MS. LOIS: I believe it is.

MR. BANERJEE: So, then, it would be nice to see the experimental plan and how they're addressing each issue with regard --

MEMBER APOSTOLAKIS: Yes.

MR. BANERJEE: What are the issues with regard to the models? How are these experiments addressing that? What do you expect to get out of them at the end of the day?

MEMBER APOSTOLAKIS: Well, let me ask you another question, because we keep talking about agreements, and so on. Where do you think, Erasmia, that there is a disagreement around the table? Is there a disagreement anywhere, or are we just violently agreeing?

MS. LOIS: I think we agree. I believe we agree.

MEMBER APOSTOLAKIS: You are focusing -- let me, then, see if I -- if we -- do we -- I mean, the experimental design, I agree with what Professor Banerjee just said, and it will be great. We can meet with the subcommittee if you'd like to discuss it.

But I still think that before we jump into it, maybe in parallel or a little bit ahead, you should produce a document like the good practices that goes with a different name, goes deeper into the models, and evaluate -- and say the fundamental premise of this model is this -- the fundamental premises. I mean, there are a number. And then, start comparing them, and then of course you will need

to collaborate with Jeff or other representatives of EPRI to make sure that you get the right perspective from their side.

MS. LOIS: Can I ask Alan Kolokzcowski to answer this question?

MEMBER APOSTOLAKIS: Oh, absolutely.

CHAIRMAN WALLIS: Well, I think what you're saying, George, is what we're saying, too -- find out the differences between the models, find out the way to run the experiment to tell which one is right or how -- if they're both wrong or something, how good they are --

MEMBER APOSTOLAKIS: But I don't want to do only -- yes, there needs to be something about the dependence between human errors and the models do it. I wanted to go down to how they do it and whether the analysts agree. And, again, the objective here is not to blame anybody.

The objective is not to say, "You are bad and I'm good." The objective is, you know, after 20, 25 years of working in this field it's time to listen to the other guy, and it's time to try to see where we agree and where we disagree at the detailed level. That's my objective here.

MS. LOIS: We totally agree with you, but

we believe -- and that's why I would like to have either John Forester or Alan Kolokzcowski do -- answer that. We believe that in the methods evaluation, with respect to good practices, although that's -- we naturally -- we went beneath that and we identified the characteristics and the basic assumptions of these models, and we tabulated it.

And we characterize them in terms of, you know, goodness or a lack of goodness, whatever that is. We didn't do it in a -- the level of detail that you probably asked to do.

MEMBER APOSTOLAKIS: Well, then it will be very easy for you to do what I want.

MS. LOIS: But I don't know if it's possible. Is it? We can do these things?

MEMBER APOSTOLAKIS: It's possible.

DR. COOPER: It's very context-specific, and that's one of the reasons why there was first a good practices, and then a methods evaluation with respect to the good practices. As Erasmia explained, good practices addresses really how you do an HRA, the various steps of an HRA process overall, whereas many of them -- the HRA methods in fact only address quantification.

So many of those process steps are really

sort of outside of a specific quantification method, and they really rely on how they do things. So those good practices were laid out, and then we -- the methods evaluation, then, we're supposed to look at, then not just -- those methods, how they matched up to good practices, but also something with respect to implementation.

But the challenge all along is: how do you make that generic? Because for one application -- let's say for at power -- it's different than it is for shutdown or for fire, because you have -- you need different capabilities. So what may be good for at power may not be -- it may not be good enough for fire.

You know, the set of performance-shaping factors that you want for at power may be different for fire. Maybe they should be. So that's the -- to be able to do something, you know, very detailed about, you know, evaluating some aspect of their model, really has to be within the context of what specific application you're trying to use HRA for.

MEMBER BONACA: When -- go ahead.

MEMBER SIEBER: I don't think you can successfully benchmark these models either. There is a lot of aleatory uncertainty involved in human

performance, and it's not like you're measuring some engineering property and writing an algorithm that will predict some engineering performance.

I think it's much more difficult in the human performance area, and probably the best you can do is evaluate these qualitatively to decide whether the right factors are there and properly treated, rather than put in your mind in advance that you want to reduce the number of models that you have. I think that's --

MEMBER BONACA: That's why for major actions in PRA the licensees depend heavily on the simulator observation.

MEMBER APOSTOLAKIS: If you look at SPAR-H, it says somewhere there, here are the various levels of stress. And if the stress level is at this level -- if the stress is at this level, multiply the human error probability, the nominal probability, by eight. And I'm sitting there and I'm saying, "Why?" Is that consistent with what ATHEANA says you should do? Why 8 and not 15?

See, this is the question that I think somebody has to address.

MS. LOIS: You have to have data, and we are creating the HERA database, which will help us

to --

MEMBER APOSTOLAKIS: No.

MS. LOIS: -- potentially address some of these questions, and we are going to do simulator --

MEMBER APOSTOLAKIS: My point is that first you have to ask the question, and I'd like to see a document that says, "Here is what they do," and it's maybe open to question whether these are the appropriate levels, maybe the factors are up in the air and they have to be validated. That's what I don't see.

MR. FORESTER: George, this is John Forester. I'd just like to comment on that. In terms of the HRA reviews, there is a discussion section in each of the method reviews where we do address the underlying assumptions of the method and try and address, you know, what are the problems with the matters, what are the advantages and disadvantages, and really what are the weaknesses and strengths in terms of their assumptions, and so forth.

It is a bit buried in there, but I do believe there is a fairly sound discussion that gets at the strengths and weaknesses of the methods. Now, that could be extracted out, but I think there is some fairly good information in there on that.

CHAIRMAN WALLIS: Guys, when you design something, you have a customer, and you are producing these methods for someone to use. And, really, the only question is: when you tell the customer something, how good is it? How -- what are you going to tell the customer this is good for? That's what matters eventually.

MEMBER APOSTOLAKIS: That's right.

MEMBER KRESS: So why not think very hard about how can one determine the uncertainty in the models.

MEMBER APOSTOLAKIS: Yes. I mean --

MEMBER KRESS: That looks like a tough chore to me. You can't just look at the model and do a Monte Carlo uncertainty.

MEMBER APOSTOLAKIS: No.

MEMBER KRESS: You just don't have the information. So it seems like you need to think about how to conduct the benchmark tests to arrive at some uncertainty in the predictions.

MS. LOIS: And we could potentially have to go back and do exactly what Dr. Apostolakis says, in the sense in order to ask the right questions for each one of the models, we'll have to go deeper as to what the models are assuming.

MEMBER APOSTOLAKIS: That's what I'm saying.

MS. LOIS: Yes.

MEMBER APOSTOLAKIS: And you have done already a lot of it. I'm not denying that.

MS. LOIS: Gareth, do you want to --

MR. PERRY: This is Gareth Perry again. I do think that it would be useful to do some of these simulator exercises at Halden, but I think we have to be realistic. Really, if you think about it, I can't remember how many PSFs SPAR-H has. Eight? Okay. ATHEANA has 60 or something. CBDT has several.

There's no way that you're going to be able to conduct experiments that will enable you to calculate the impact of changing PSFs on human error probabilities for sure. So, for example, asking whether the stress changes by a factor of five, because it increased the -- sorry, a high stress increases the failure probability by five is not something we're going to be able to answer with these issues.

MEMBER APOSTOLAKIS: That's why I want this evaluation separate.

CHAIRMAN WALLIS: That's the kind of information I like to hear. I mean, that's useless to

me. If you've got 60 parameters in the model, and you're going to try and run experiments, that's an awful lot.

MEMBER POWERS: Well, that's a question I think that -- I don't know if this -- these benchmarking --

CHAIRMAN WALLIS: Well, I would say right up front, that's a useless model. If it has 60 parameters, which you're going to adjust to get an answer, that's absolutely useless.

MEMBER ARMIJO: Yes. You know, I'm not in -- from this area, but I would -- I would look for something -- a ranking of what you currently believe are the most important --

CHAIRMAN WALLIS: Yes.

MEMBER ARMIJO: -- shaping factors, and then separate effects tests in some way, maybe in the laboratory environment like Halden, to really see if these stress or operator fatigue or some other factor really does have that much of an effect.

MEMBER SIEBER: But you aren't going to be able to tell.

MEMBER ARMIJO: Well, I'm just saying, I don't know how to do this sort of stuff, but it seems to me when you have that many variables operating

simultaneously, I don't know how you --

MEMBER APOSTOLAKIS: It's precisely for that reason that I don't think we should rely only on the experiments. That's why I think we need the experts in the field to create this comparative evaluation and raise questions. Maybe you don't want to say that this is wrong, but at least ask the question, because it's true what Gareth said. You cannot test all these things here.

MR. BANERJEE: No. But what they were saying, if I understood it correctly, was that they would be pre-predictions of this --

MEMBER APOSTOLAKIS: Of importance.

MR. BANERJEE: Yes, of --

MEMBER APOSTOLAKIS: Importance of the factors.

MR. BANERJEE: -- of these, let's say, benchmarking exercises. And these pre-predictions may or may not be right. But if they were not right, they would try to understand why.

MEMBER APOSTOLAKIS: That's right.

MR. BANERJEE: I mean, I think this field is going to be open to qualitative attacks for a long, long time. I mean, you're not going to have quantitative --

MEMBER APOSTOLAKIS: Yes.

MR. BANERJEE: But even if you look at, say, the so-called predictive methods for good mechanics of something, these codes have hundreds of parameters in them which are adjustable. And the number of experiments that can be done are very limited.

So the situation isn't all that different. I mean, we don't test every parameter. We may choose five or six which we think are really important, and that's our judgment call. In some, you know, exercise of the models we find out where the main uncertainties lie. I'm sure you guys do the same thing in some way. You try to figure out, what are the most important factors in these models? And see how they are affected in an exercise like this.

But I think we should encourage this and get back to really seeing what the results are. It would be very interesting.

MEMBER APOSTOLAKIS: Yes, we are not discouraging it.

MR. BANERJEE: And I'm just really wondering whether there is more data around from even the day-to-day simulator training exercises and all, which must be somehow put into these models, right?

I mean, you are extracting and inducing data every day, I would think, from all these training exercises.

MS. LOIS: So we have this activity on collecting data, which is very resourceful. Actually, it requires a lot of resources to take an event and evaluate it from a human reliability perspective, and then put it in the database.

And the activity that starts out with LERs, more are looking at events that are -- have had some kind of precursor analysis, etcetera. So it's one activity that we're pursuing, and we hope two or three years from now to be able to do -- to use this data as objective measures of the -- to test the underlying hypothesis or predictability of the HRA methods.

That's a long-term activity, which we have, but this one -- the Halden experiments give us a controlled environment to do experiments, which I -- we believe that will help us to expedite our process for understanding the methods.

MR. BANERJEE: Well, I like the pre-prediction --

MEMBER APOSTOLAKIS: To summarize now -- we have to summarize. To summarize, it seems to me if there is a disagreement it is -- it is the degree to

which you rely on the experiments. I would like to see -- and I'm willing to say, fine, you've done some of it already in the good practices, but I'd like to see a critical evaluation of these models and their assumptions in parallel with this activity, which I believe is very important, but I wouldn't rely only on this activity.

And I think if you have a group of experts who are familiar with these methods, have used them, you certainly have people that have used SPAR-H, ATHEANA obviously, but maybe get Jeff or somebody who is experienced with the EPRI model who can go a little deeper than what we have done. I think that would be extremely valuable to everybody, and then we'll see what happens. Then we'll see what happens, but it --

CHAIRMAN WALLIS: George, how do you critically evaluate without an experiment? Unless you -- I can evaluate whether it's mathematically consistent or something, but that's not the question here. The question is: are these hypotheses valid? Isn't that the thing?

And then, if they are valid, how do you quantify them in some way?

MEMBER APOSTOLAKIS: But, first, I want to have an identification of these hypotheses. I want

them to raise the questions first.

CHAIRMAN WALLIS: So you want a logical evaluation of the hypotheses.

MEMBER APOSTOLAKIS: Yes.

MR. BANERJEE: But there was -- and somebody said --

MEMBER APOSTOLAKIS: And there is -- these people are very experienced, right?

MR. BANERJEE: -- that there was an evaluation done.

MEMBER APOSTOLAKIS: Sorry?

MR. BANERJEE: Somebody said on the phone that the evaluation is buried in --

MEMBER APOSTOLAKIS: No, to some extent it's done. And what I'm saying is, great, build on that and go a couple of levels deeper to actually look at how each model is doing certain things and raise questions, compare with what -- it's a comparative evaluation, really. ATHEANA does this in this area, this other model does that, and maybe there are some questions.

CHAIRMAN WALLIS: Now, can we ask Erasmia to summarize at this time?

MEMBER APOSTOLAKIS: I'm done. Do you have any summary responses?

CHAIRMAN WALLIS: What do you want us to take away from this?

MEMBER APOSTOLAKIS: What I'm really struggling for is: how do we add any value in a letter? I mean, we discussed all kinds of stuff, and I don't see there's any sort of real focus on what we need to say.

MR. BANERJEE: It's very important.

CHAIRMAN WALLIS: George, can you tell us what you want us to say, and then maybe we can say it -- we can see if that's appropriate? What would you like us to say? I had to ask Gareth to -- are you the customer, Gareth, for this work?

MR. PERRY: Sort of, yes.

CHAIRMAN WALLIS: Yes, I think it would be nice to hear from the customer, too. So can you both summarize your --

MEMBER APOSTOLAKIS: Can I also tell you what I would say?

CHAIRMAN WALLIS: No, no, George. You're not allowed to say anything.

(Laughter.)

MEMBER APOSTOLAKIS: No. Because I want them to react to it.

CHAIRMAN WALLIS: I want them to tell us

what would be most useful for us to say in a letter to help them. Now, can you tell us --

MS. LOIS: I would ask John Monninger to do the --

CHAIRMAN WALLIS: Your manager?

MS. LOIS: John Monninger.

CHAIRMAN WALLIS: Okay. Ask the manager, then. You have any opportunity now to tell us what you'd like to -- like us to say in our letter.

MEMBER APOSTOLAKIS: What is it that you would like?

MS. LOIS: John, do you want to make --

MR. MONNINGER: This is John Monninger from the Office of Research. I guess first off, you know, we weren't explicitly requesting a letter. But if a letter was to come, you know, from the ACRS, one thing we think is important to recognize, the advancements or the contributions, or the work that has been done to date in, you know, the establishment of the good practices and the evaluation of the methods against the good practices.

CHAIRMAN WALLIS: Didn't we send you a letter on that already?

MEMBER APOSTOLAKIS: Yes, we did.

MR. MONNINGER: Yes. Yes, you did. The

second thing with regard to the planned experiments with Halden, we are still, you know, in the planning phases with them. The kickoff meeting is later I guess in the beginning part of December.

I think, you know, something along the lines of, you know, a qualitative endorsement of the proposal to go forward with the program, but with, you know, some type of caveats to the extent that the committee would like to be informed of the objectives, the approach, etcetera. You know, further briefings, interactions on the program would be, you know, helpful.

You know, the motion with regards to the critical evaluation of the HRA methods, I guess one question comes to my mind, you know, do you do that, you know, if we were to do that, or, you know, if we had the resources to do that? Would you do that in parallel, or would you proceed first with the -- you know, the benchmarking exercises? And then, you know, see what the results of that are, and then, you know, go a step below into the critical evaluation of the methods and models. So --

MR. THORNSBERRY: Dr. Wallis, I'd like to also -- the staff is here on our request. This falls under the category of an ACRS initiative. So like

John said, they weren't specifically coming to us to ask for a letter. It was our initiative led by Dr. Apostolakis.

MEMBER APOSTOLAKIS: They'd rather be left alone.

MR. THORNSBERRY: And it fell out of our subcommittee meeting in the summer. So it's really kind of our initiative is why they're here, to set the stage and tell us what they've been doing. But the things that came out of the subcommittee led us to this, so that we could give some additional guidance beyond what they're already doing.

CHAIRMAN WALLIS: Do you have anything to say, Gareth?

MR. PERRY: Actually, I think a lot of what we need is probably being done in 1842, to the extent that there is a review of the models that explains what the models can and what they can't do. It probably would help to have maybe a little more confidence in some of the models that we use, such as SPAR-H. But I think certainly an attachment of whether it's good enough for the purposes for which we use it is -- certainly would be helpful.

Personally, I'm interested in the results of these experiments, particularly if they -- if they

-- what they can do is to highlight for us what are the most important performance-shaping factors. And I would prefer them to be performance-shaping factors that somehow you can measure as opposed to something like stress, which is something you have to -- you have to think about what it means.

But I think -- I think certainly I'm very happy with the 1842 document. I think that was really helpful to us.

CHAIRMAN WALLIS: This is the one that we reviewed before.

MR. PERRY: Yes, this is the one you reviewed before.

CHAIRMAN WALLIS: We've written a letter on that already.

MR. PERRY: Yes, yes.

MEMBER BONACA: I have a question for Gareth. Right now, since the SPAR-H is being used to model different plants out there for which there are already PRAs, I'm sure there are many instances the staff reviews HRA assumptions in this -- in this report against what you're predicting yourself. I mean, do you find some convergence there taking place? I mean, with respect to what you used in the past, or do you find wide differences still?

MR. PERRY: Actually, that's hard for me to say, since I don't get involved with the reviews myself.

MEMBER BONACA: I understand.

MR. PERRY: But let me tell you one of the things that we do have problems with, and particularly in things like the significance determination process. It's not the routine human error probabilities like failure to depressurize, for example. We don't -- there seems to be general agreement that it's within a certain band.

But it's the -- it's the unusual things like the recovery actions that our licensees claim that they can do in a certain time to demonstrate that this particular event was not a high-risk event. And I think a lot of those things are actually not even addressed by many of the models. The models just don't apply in those situations.

CHAIRMAN WALLIS: Okay. George, was it --

MEMBER APOSTOLAKIS: Well, let me tell you what I think.

MR. BANERJEE: Excuse me. Why don't the models apply in those regions?

MR. PERRY: Because the majority of the models have been developed to address control room

responses to procedures, I think, in the main. I mean, yes, you can adapt ATHEANA probably to go beyond that. But ATHEANA isn't the method that is widely used by industry, for example.

MR. BANERJEE: But wouldn't that be an important initiative on the part of Research, to try to collect this information that through the significance evaluation process is being really developed and used, and to gather an understanding of what is happening out there insofar as -- I mean, there's a wealth of information being generated there at the working level on -- in the field.

MR. PERRY: Yes.

MR. BANERJEE: And I think that that's something that could be mined.

MR. PERRY: Yes, and I think probably that could go into here, probably is where that -- that would be useful input to HERA I think.

MR. BANERJEE: But coming back to -- you said that one of the major areas where you have a need currently -- recovery actions or whatever 00 which are not necessarily control room oriented, how do you handle these right now?

MR. PERRY: Typically, we handle them through discussion with the licensee. They'll tell us

what they think is the case, and we will inquire -- typically, what we look at actually is the -- is what ATHEANA could call the context, and make a decision on that basis whether we think the action is feasible, and then we either reach agreement or disagreement with the licensee on whether we think it's a feasible action.

MR. BANERJEE: But would you like to have something a little more -- or is this a satisfactory situation?

MR. PERRY: You know, for what we're dealing with, this -- I'm not sure if you're familiar with the significance determination process, but it's really meant to be a quick evaluation to determine the extent of the additional inspection that we give to plants. I don't think we need a major new tool to do that. I think we can -- we should be able to deal with it.

But I think what we do need, though, is a little bit more basis perhaps on what are the important factors that decide whether an action is feasible or not.

MEMBER APOSTOLAKIS: Yes, let me -- I'd like to read the three lines that they have here for recommendations.

MR. BANERJEE: Well, just one thing I wanted to ask you, because this hasn't really -- I haven't got to the end of this. We hear about people having to switch from hot leg injection to cold leg injection, or cold leg injection to hot leg injection, a variety of stuff that people have to do in various accidents.

Is that covered by these models, or is that falling under what is --

MR. PERRY: No, that would be covered by the models.

MR. BANERJEE: That will be covered.

MR. PERRY: Because those are proceduralized actions.

MR. BANERJEE: Okay.

CHAIRMAN WALLIS: Okay. Are we --

MEMBER APOSTOLAKIS: So what -- I'd like to know in the next minute the reaction of the staff -- this is important -- if the recommendation was this. The staff should evaluate the agency's human reliability models and the models included in the EPRI HRA calculator and create a plan for the development of either a single model for the agency to use or an integrated suite of models to be used in specific circumstances. Would that be something that would

cause a heartburn?

CHAIRMAN WALLIS: Yes, you should create a model which can be used for the purposes of --

MS. LOIS: A model or a suite of models.

MEMBER APOSTOLAKIS: Yes, or a suite of models. But right now we're asking for a plan.

CHAIRMAN WALLIS: Aren't they doing that already? I mean, I hope they're doing that already.

MR. MONNINGER: It sounds very broad and open. I mean, it sounds very feasible. I mean, you're saying either a single one or a suite of models, that would be appropriate for the circumstances.

MEMBER APOSTOLAKIS: But I would like to see a plan that says by this time we are comparing these, we are hoping to get these conclusions. Then, we do these experiments. This is the objective. This is what we are getting by this time. Then, by that time we're going to do something else. Having in mind this ultimate goal of either one model or a suite of models that are not developed independently for specific applications. So, yes, it's broad.

CHAIRMAN WALLIS: I guess what we're saying, though, is that the plan should be more structured.

MEMBER APOSTOLAKIS: Right.

CHAIRMAN WALLIS: And there should be some logical threads which we can look at, and so on. That's --

MEMBER APOSTOLAKIS: So what would be something that --

CHAIRMAN WALLIS: But I don't think they asked them to develop a plan. They already have a plan, so we are commenting on it, aren't we? George?

MEMBER APOSTOLAKIS: Well, they have a plan, and I'm asking for a new plan. I don't know. A sub-plan. So the plan that they presented to us is much broader. It's a human factors, human -- this is specifically human reliability, a plan to achieve a specific goal, either a single model or a suite of models, an integrated --

CHAIRMAN WALLIS: Well, we don't quite know which is appropriate yet, until we look at the context of this, do we?

MEMBER APOSTOLAKIS: What?

CHAIRMAN WALLIS: It may be that the use requires several models. I don't know.

MEMBER APOSTOLAKIS: But that's what it says, or an integrated suite of models.

CHAIRMAN WALLIS: So a plan -- you're

looking for a logical plan is what you're --

MEMBER APOSTOLAKIS: Yes. And I don't hear any objection.

CHAIRMAN WALLIS: But that's what anybody would do, isn't it? I mean, it's almost like saying that they're having an illogical plan now. Therefore, they need a logical one. Is that --

MEMBER APOSTOLAKIS: Well, it's a matter of focus and direction, and I'm sure a lot of it they are already doing, but now --

CHAIRMAN WALLIS: Do we want to change --

MEMBER APOSTOLAKIS: -- it's going to be specific: this is where we want to go, and this is how we're going to get there.

CHAIRMAN WALLIS: Okay.

MEMBER APOSTOLAKIS: Okay. And with that happy thought, I'll turn to back to you, unless -- Dana?

MEMBER POWERS: I have a variety of questions to ask.

MEMBER APOSTOLAKIS: Do you want to ask them now or during the -- oh, okay.

MEMBER POWERS: I don't want to ask them during the break, no.

(Laughter.)

MEMBER APOSTOLAKIS: You don't. Okay. Go ahead, then.

MEMBER POWERS: If we could turn to page 3 on the viewgraphs. It indicates that you're collaborating international entities. I'm wondering what those entities were. Is that just Halden? I see later on you'd interact with IRSN.

MS. LOIS: In actuality, the plan -- and probably I should use my backup slides. Let's do that. To have -- to have a steering committee which would be two members from the U.S., and that would be EPRI and NRC, and then have representatives from other countries. So far, India has expressed an interest, and, of course, WG risk, IAEA, would be -- no, I'm sorry, the OECD facilities.

And, of course, the have expressed an interest to participate, so what we plan is to have a steering committee which would kind of do this thought process, come up with a plan, come up with an experimental design, and communicate that with those organizations signatory to Halden that would like to participate, and then hopefully have an agreed-upon plan and design for the experiments that everybody would agree, and then try to do the experiment.

Now, there are many intermediate steps for

that. For example, right now at Halden there are about 16 crews are running experiments, and we are going to have a crew. A few of us will go there to observe and see how the experiments are run, so that we understand what it takes to come up with a design.

It will be hopefully a meeting in January with prospectives and other signatory countries, and debate what it will take. Everybody has its own method they would like to test, its method or methods. So it's -- there is going to be many, many steps in between in order to come up with the plan and the design of the experiment.

MEMBER POWERS: Very helpful. Not apparent from the soliloquy that was conducted here.

MS. LOIS: I'm sorry. The presentation was not on the benchmarking, and probably I should have done that.

MEMBER POWERS: Will you turn to page 4? You indicate that you're putting operational data into HERA. Could you just give me a thumbnail sketch of what that data are?

MS. LOIS: They are -- right now it's LER data.

MEMBER POWERS: LER data. That's enough for me. Thank you.

If I could turn to page 8. You indicate you're working in the SOARCA program. I understand there is some debate within SOARCA on how much HRA help they need. Would you elaborate?

DR. COOPER: Susan Cooper. They are getting HRA help. I'm the HRA representative within NRC. We have Sandia and subcontractor SAIC. I don't know if John and Alan are still on the line, but they're helping out.

But in any case, we are getting HRA support. I would say that the principal uncertainties with regard to HRA support right now have to do with how the overall project is going to proceed. But it is the expectation of everyone involved at this point in time that HRA will be an important factor in how scenarios are refined and developed for developing MALCOR inputs. But, you know, this -- may of the specifics have not been decided, and we're very early on in the process.

MEMBER POWERS: Okay. So you think you're going to be looking at interfered accidents and not just hands-off accidents.

DR. COOPER: That's -- yes, we will be looking at accidents that involve operator actions.

MEMBER POWERS: Okay. Thank you.

On page 9, you indicate ATHEANA. And you say "trial applications" under there. And I thought we had gone through about a year of trial applications, so I was trying to understand what -- it doesn't say what trial applications, so I --

MS. LOIS: What I'm talking about here is the ATHEANA user's guide, which is an addendum to the existing ATHEANA NUREG, NUREG-1624 I believe. And the question is how good the user's guide would be. Could an HRA expert pick the user's guide and apply ATHEANA, given that he's an expert? So --

MEMBER POWERS: So it's --

MS. LOIS: -- we are going to do some of those, or we hope we will.

MEMBER POWERS: It's a novice HRA professional, not a novice ACRS member.

MS. LOIS: Correct.

MEMBER POWERS: Which we have several. Thank you.

Let's see, if we come down to page 10, it says, "HRA have implications on burnup credit for spent fuel pools." I found that surprising. I wondered what you meant by that.

DR. COOPER: This is a preliminary suggestion or indication of where we might help.

There is a user need in draft from NMSS, and we have been doing some development and demonstration of HRA capability for them in the area of spent fuel misloads and cask drops. And because of that, we're -- they see the potential usefulness of HRA in answering the questions, so far as allowing burnup credit.

MEMBER POWERS: So it's really a question of spent fuel pool operations and not so much about what the actual neutron count is going to be.

DR. COOPER: That's correct.

MEMBER POWERS: Okay. If we could come to page 12, and, boy, do I have trepidation here. You're talking here about benchmarking, and I'm a little unclear what you mean by "benchmarking." And I note that during your -- the free-form discussion that was held that you said it may not be quite the right word. I wonder if you could give me two sentences on what you mean by "benchmarking."

MS. LOIS: Well, in the engineering sciences, I believe that "benchmarking" has a very concrete definition. In here, in human reliability, given that the environment will not allow benchmarking exercise in a very consistent way, we are not quite sure if that's the right word.

However, what we tried to do there is to

-- to do a testing or a determination of each method capability to predict on its own. And then, given that all those methods will be tested out, then we'll have the capability to compare the methods.

MEMBER POWERS: I see. In other words, you're using "benchmark" much as it would be used for any deterministic code. How does that code work on this problem by itself? Thank you.

MS. LOIS: I'm happy to hear that.

MEMBER POWERS: Let's see if I can come back to page 10. You used the word "modality," and I'm not sure what you meant by that.

DR. COOPER: "Modality" is a term used by those folks involved in evaluating medical applications of radioactive material. That could be brachytherapy, it could be gamma knife, could be -- there are any number of different treatments --

MEMBER POWERS: So it's what's being done.

DR. COOPER: Exactly, what's being done and how it's -- you know, what vehicle by which it's being done. It's a -- the work is being focused on gamma knife right now, and so that's why I say "other modalities." It could be extended.

MEMBER POWERS: You will quickly learn in front of this committee the less you say, the less

likely you are to get in trouble.

(Laughter.)

Thank you.

Now, let me ask you about one question that the committee has wrestled with a couple of times, and that has been in the power uprates for boiling water reactors. Especially when we are working with BWR-4s, we have a short period of time for the operators to respond to indications of core instability.

And in that discussion with the applicants for the power uprates, they consistently came in and used THERP to estimate the reduction in operator reliability. But they indicated to us that this particular evolution is practiced regularly by each crew, each year, operates this.

And, for instance, one of our applicants indicated they had 50 data points with no failures, yet he took a -- there was an increase in operator and reliability of .01. Okay? How do you respond to that? And can you -- is there anything that's going to be able to help us on that question in the future?

MS. LOIS: I'll try -- I'll answer that question from a high level and then probably go to --

(Laughter.)

I believe because, as Dr. Apostolakis mentioned before, the Halden experience -- experiments can be time-driven. I believe that we could set -- and I'm not quite sure if this is the time, this -- this time around is going to be the time. but we could potentially set up experiments with various time intervals allowed for the operators and observe their capability on well-trained actions and observe their capability to do the action as reliably as before.

Now, I would like to note that even in those well-trained scenarios that we have observations so far, you do see a crew that was a little bit delayed to complete the action, even for a well-trained scenario. Now, that's an indication that not all crews may complete the action as reliably, but I guess it's within the variability of human performance that one expects.

MEMBER POWERS: Okay. Finally, I'd like to ask a question on -- in the course of looking at some NUREG guides on hazardous materials around nuclear powerplants, we several times have had licensees tell us that they equip their control rooms with self-contained breathing apparati, so that should some noxious material come into the control room that operators could stay on station and continue to work

there.

But when we've asked them, do they ever train on that, in the simulators, I have yet to have any of them say yes. Now, I cannot claim to have done a complete survey, but I've asked the question at every place I go. How would ATHEANA or any of these human reliability models handle operations under -- with self-contained breathing apparatus and any degradation in reliability that would come from that? And wouldn't the agency be interested in that kind of information?

DR. COOPER: I can't answer the last question. Maybe you folks can direct us to that. I can say that with respect to the first question that there are some methods that would at least identify that as being a potentially important aspect to be considered as an influence on human performance.

That's only one piece of it -- an important piece to know that you actually ought to address it. The other part is, so how does it influence human performance? And that question I can't answer. I mean, we have not tried to analyze something like that.

Right now, I don't know if there's any information out there right now. It's not -- that

kind of knowledge, if you will, is not contained in any HRA method that we have right now, just as there isn't any knowledge in any HRA method right now so far as the effect of smoke on human performance.

Now, whether or not there is information more broadly across the U.S. and other industries or in psychological data, I don't know. But if you were to try to address that, that would be my first step. But it's not in any HRA method right now.

MEMBER POWERS: I use it -- something is a stocking horse, because I've been interested in it. But we have had challenges in the agency with control room habitability issues, and there's quite a lot assumed and argued in connection with control room habitability.

And I might suggest that because of the central role of the control room that you might want to look at some of those FSARs to identify areas of human reliability that need to be explored as you develop those models. That completes my questions.

CHAIRMAN WALLIS: George, are you through?

MEMBER APOSTOLAKIS: I'm through.

CHAIRMAN WALLIS: So it's back to me?

MEMBER APOSTOLAKIS: Back to you.

CHAIRMAN WALLIS: Well, my first comment

was that the sort of questions that my colleague Dr. Power has asked are the sort of thing -- one of the sort of things I thought we were going to be doing here.

I thought we were going to be looking at your slides and your plan and your activities, and I was sort of pleased when your answer to the first of his questions, you actually showed us a plan. I mean, he said, "What's your plan for international work?" and it turned out you did have a plan, which we didn't know, you know?

So, you know, I think the -- when folks come before this committee, you have to have a plan. You have to say, "I want this committee to look at our plan or look at our list of activities," or something specific, and then we can respond to that. So that --

MEMBER APOSTOLAKIS: I think it's --

CHAIRMAN WALLIS: -- off track, I wasn't sure what you were asking us to do, and we got into this discussion and again it wasn't clear to me what you were asking us to do. So --

MEMBER APOSTOLAKIS: I think one of the things we have to do, Graham, is we -- we saw today what we saw also a little bit yesterday. We have too many new members.

CHAIRMAN WALLIS: That's not the problem, George.

MEMBER APOSTOLAKIS: It is a problem, because there were questions -- there were questions --

CHAIRMAN WALLIS: The question is: is the staff controlling their own presentation and being allowed to do so?

MEMBER APOSTOLAKIS: There were questions that the staff assumed had been answered in many meetings in the past, and it was true.

CHAIRMAN WALLIS: Well, how do you know what they assumed? George, I'm not going to get into this conversation. I'm going to stop this now.

MEMBER APOSTOLAKIS: They assume we know what ATHEANA is, for example.

CHAIRMAN WALLIS: I think we need --

MS. LOIS: One clarification is that we believe that this is more the committee's meeting, I suppose, and we had to prepare something --

CHAIRMAN WALLIS: But if there is a -- I think staff has to come to the committee saying, "This is what we're going to present to you, this is our -- this is our" -- you know, you, it's your presentation. These are the kind of issues where you think that we

can contribute. Of course, we'll jump into all kinds of things, but if we let it go everywhere it's going to go everywhere, and you have to bring it back again.

So on that note, I'd like to stop, if the committee is happy to stop now. And we'll take a break until quarter to 11:00, and then we will do the P&P and a few sort of administrative matters, and then we'll go to letter writing. We don't need the Court Reporter anymore. Thank you.

(Whereupon, at 10:31 a.m., the proceedings in the foregoing matter went off the record.)