Official Transcript of Proceedings ACRST-3244

NUCLEAR REGULATORY COMMISSION

ORIGINAL

Title: Advisory Committee on Reactor Safeguards Reliability and PRA and Plant Operations Subcommittees

Docket Number: (not applicable)

Location: Rockville, Maryland

Date: Tuesday, July 8, 2003

Work Order No.: NRC-990 Pages 1-150

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

TR04

UNITED STATES OF AMERICA

NUCLEAR REGULATORY COMMISSION

+ + + + +

ADVISORY COMMITTEE ON REACTOR SAFEGUARDS

RELIABILITY AND PRA AND PLANT OPERATIONS

SUBCOMMITTEES

MITIGATING SYSTEMS PERFORMANCE INDEX

+ + + + +

TUESDAY,

JULY 8, 2003

+ + + + +

ROCKVILLE, MARYLAND

+ + + + +

The Reliability and PRA and Plant Operations Subcommittees met at the Nuclear Regulatory Commission, Two White Flint North, Room T-2B1, 11545 Rockville Pike, at 1:00 p.m., Mario V. Bonaca, Acting Chairman, presiding.

SUBCOMMITTEE MEMBERS:

GEORGE APOSTALAKIS, Subcommittee Co-Chairman

JOHN D. SIEBER, Subcommittee Co-Chairman

MARIO V. BONACA, Acting Chairman

GRAHAM M. LEITCH, Member

STEPHEN L. ROSEN, Member

WILLIAM J. SHACK, Member

ACRS STAFF PRESENT:

MAGGALEAN W. WESTON, Staff Engineer

A-G-E-N-D-A

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

1:00 p.m.

MR. BONACA: This meeting will now come to order. This is the meeting of the Reliability and PRA Subcommittee. I'm Mario Bonaca acting as Chair of the Reliability and PRA Subcommittee for George Apostalakis who has been delayed.

Jack Sieber, Chair of the Plant Operations Subcommittee is the Co-Chair. He is not here but will be back I'm sure on time. ACRS members also in attendance are Graham Leitch, Stephen Rosen, who will come up, too, I guess. I didn't see his name listed there. And William Shack. Hopefully George Apostalakis will be here within the hour.

The purpose of this meeting is to discuss the progress of the mitigating systems performance index and to respond to questions raised in the main ACRS subcommittee briefing.

The subcommittee will gather information, analyze relevant issues and facts, and formulate proposed positions and actions as appropriate for deliberation by the full committee. Maggalean Weston is the staff engineer for this meeting.

The rules of prospective participation in today's meeting have been announced in the <u>Federal Register</u> on June 25, 2003. A transcript of the meeting is being kept and will be made available as stated in the <u>Federal Register</u> notice.

It is requested the speakers use one of the microphones available, identify themselves, and speak with sufficient clarity and volume so that you may be readily heard.

We have received no written comments from members of the public regarding today's meeting. We will now proceed with the meeting. Pat Baranowsky of the Office of Nuclear Research will begin.

MR. BARANOWSKY: Thank you. I'm the Chief of the Operating Experience Risk Analysis Branch and with me is Senior Risk and Reliability Analyst in the branch, Don Dube, and we're going to make a presentation today.

I would like to thank the subcommittee for giving us this opportunity to present the progress on this project. We found that airing the technical issues and getting input from the subcommittee has been quite valuable in the past and we would like to continue to do so.

Let me go to the first viewgraph here which pretty much states the purpose and objective that we had for coming to this meeting. First of all, as you had mentioned, we want to update you on the progress we've made on the mitigating systems performance index which is a performance indicator set that we've been working on for the past year or so.

We think we've addressed and will discuss how we have addressed the ACRS comments. Then ultimately, not after this meeting but perhaps a future meeting, we would be looking toward getting an ACRS letter on this particular developmental activity. Today --

MR. LEITCH: As I looked at the White Paper I guess I had not particularly focused on the difference between the word indicator and index. It seems to me there's a pretty significant difference there.

MR. BARANOWSKY: I'll explain why we chose that terminology when I get to my overview.

MR. LEITCH: Okay.

MR. BARANOWSKY: That's coming right up. The first thing I'm going to cover is some background on the MSPI. Then we'll identify what we

pulled out from the transcript as ACR's comments from our briefing about a year ago.

I'll cover the White Paper that we sent to you previously and then give a briefing on the status of the pilot program that we have been conducting and is coming to a close now.

Then really some of the meat of this presentation is to go over key technical issues that evolved as a result of comments received and the pilot program. Then summarize and get to an implementation time line that we are working toward.

Just for some background, the mitigating systems performance index, that approach evolved from a feasibility study that we did a couple of years ago on risk based performance indicators.

Basically it's a highly risk informed simplification to the risk based performance indicators. It was designed to address some recognized issues with the current performance indicators which are somewhat risk informed, simplified, generic, and so forth.

In particular, the MSPI addresses treatment of demand failures and fault exposure time which is causing problems in the implementation of the current set of performance indicators.

We addressed issues associated with the definition of availability and, in particular, spent a fair amount of time early on discussing inconsistencies with maintenance rule applications of unavailability and availability.

For the most part I think we have made as much progress as we can. We are pretty consistent now in terms of the way we define unavailability for at-power conditions, safety systems at-power conditions.

The other issue that was raised that was causing some problems was the lack of plant specific risk informed performance thresholds. In fact, the ACRS had brought that up quite some time ago in reviewing the current set of performance indicators.

There had been some problems with respect to the cascade failure treatment of cooling water systems where one cooling water system failure could cascade its impact as a dependent type system on to other front line systems and produce multiple hits on performance indicators in a way that they weren't designed to have multiple hits.

Now, the MSPI monitors risk impact of changes in performance for selected systems. That's why we call it the MSPI. That identifies a segment

of risk impact and it is not a risk indicator per se so it's like a conditional risk indicator with certain limitations.

It doesn't address shutdown, it doesn't address external events, and it doesn't address certain relatively rare events that don't have an occurrence interval -- recurrence interval that allows us to get a statistically valid analysis of performance implications.

We coined the phrase "index" even though we relate the indicator to CDF in trying to make it risk informed and plant specific. It's called an index to reflect the fact that it has a limited scope that it's trying to look at performance issues on.

MR. LEITCH: And that scope is basically at power?

MR. BARANOWSKY: That's basically the at power, on demand reliability and availability of the specified set of safety systems.

MR. LEITCH: It is by definition then plant specific?

MR. BARANOWSKY: It's plant specific, yes. It incorporates the plant specific -- well, that's sort of my next bullet there. It calls for

plant specific design, configuration, and plant specific data to assess the performance with respect to those six systems per plant.

The scope of the PIs -- I think this is an important point, too -- is consistent with the current PIs. It's meant to be a replacement for the current PIs. It's not meant to come up with new optimal ways of treating the whole oversight process scheme of PIs. It's meant to specifically address the mitigating system performance indicators for which there had been some problems identified by both the ACRS industry and NRR folks.

It does cover unavailability and unreliability and is consistent with PRA modeling which is why it's highly risk informed. The process uses a detailed definition of the scope and calculation specifics for the PIs in order to get consistency, reproducibility if you will, of the PI calculations.

The threshold bases are consistent with the current PI thresholds. Even more so, in fact, then when we first came here as you'll hear as we go through this. We've moved toward some performance thresholds which are consistent with PIs at the so-called green/white interface and more risk informed

thresholds at other threshold interfaces in the reactor oversight process threshold scheme.

MR. LEITCH: So what are the units of MSPIs, delta CDF?

MR. BARANOWSKY: Fundamentally delta CDF, yeah. Just to summarize some of the points --

MR. SIEBER: Maybe I could. In the list on the preceding slide, you don't talk about the treatment of common cause failure. Hopefully that will be discussed as we go.

MR. BARANOWSKY: Okay.

MR. SIEBER: That's already been dealt with.

MR. BARANOWSKY: I think I'll mention exactly how we handle that here and then there will be some additional information that Don will present.

Common cause failure is pretty much handled in two parts. One is an actual common cause failure incident which is quite rare and has significant risk impact on the plant is not something that we believe this indicator is capable of trending, if you will. Therefore, our proposal is that one would use a risk significance process like the significance determination process to look

at any real common cause failures.

At the same time the importance of common cause failure in the risk significance of the systems that we're monitoring is captured through the performance indicator. I don't know if that's clear enough.

MR. SIEBER: Well, let me ask a question. If you've got a non-green mitigating system performance index, you would get into the SDP as part of the ROP process. You wouldn't just go with the indicator.

MR. BARANOWSKY: What we would do is we have a scope split where we think the indicator can provide valid indication.

MR. SIEBER: Regardless of SDP?

MR. BARANOWSKY: Yes, because this indicator is designed to measure accumulated performance, if you will. Changes in performance over some period of time.

MR. DUBE: Three years.

MR. SIEBER: Three years.

MR. BARANOWSKY: Three years and then accumulation of data basically. Whereas the SDP is a one time, one episode incident.

MR. DUBE: Exactly. We have a slide,

too.

MR. BARANOWSKY: We'll cover that. So what we've tried to do is identify where we think this PI works best and where we think a risk determination type of activity works best. Either one of those can feed into the matrix.

MR. SIEBER: The question is let's say you've got a non-green index and say you went to the licensee and you said, "You are the ROP. You get some special attention."

MR. DUBE: It would be no different than we have now.

MR. SIEBER: Yeah. On the other hand, he would say, "Well, I don't think this is risk significant." They would then pick out the instruments that drove them over the edge. Perhaps it would be a green and now you have a conflict. I think if you use this, you have to clarify what takes precedence and why there's a difference because there will be instances where there will be differences.

MR. DUBE: That's a good point.

MR. BARANOWSKY: That's an important point and we also will cover that in a little more detail. Then if we don't satisfy you, I'm sure

you'll ask us more questions.

MR. SIEBER: Well, yeah. I think it's more understanding and writing down what these things mean as opposed to an argument as to whether it's valid or invalid.

MR. BARANOWSKY: Yeah, and we're trying to detail in the guidance documentation where one uses the mitigating system performance index and where one uses the significance determination process. We try to address, at least to some extent in the White Paper, some points as to why one might be preferable to the other in general.

MR. DUBE: Your point is well taken. Addressing of technical issues is kind of leading in those kinds of implementation issues but they are very important.

MR. SIEBER: You've run a couple of workshops with the industry. In fact, you rely on an NEI document for part of the development of this. I would presume that during those workshops -- I didn't go to the workshops and I haven't read about them but other than the fact that they occurred you accomplished something.

I would presume that part of those workshops the industry understands what it is doing

here and probably will need in the process of implementation something in writing that says here's the policy and how we're going to employ it.

MR. BARANOWSKY: I think we actually have that.

MR. SIEBER: Okay.

MR. DUBE: You want to say anything, Mark or John?

MR. SATORIUS: What was the question?

MR. BARANOWSKY: Well, whether or not we have documented for the purpose of what we're doing with the MSPI pilot, for instance, the use of SDP versus the PI so it's clear for everybody. I think we've done that.

MR. HOUGHTON: Tom Houghton, NEI. Yes, we have. In the draft guidance document, I think right up in the very front of it, we list about five instances when you would use the SDP as opposed to the MSPI itself. We will be looking as we go through them with your advice if there are any others that we need to. Basically it's things that the indicator can't really measure very well or aren't included in the indicator's capability.

MR. SIEBER: This is in the NEI document?

MR. HOUGHTON: It is. And the inspection guidance, I think, would follow that.

MR. THOMPSON: This is John Thompson, Inspection Program Branch. Tom is exactly right, but the point that you were making earlier is an important point that where the indicator is valid and gets ahead and crosses a threshold, we have said in the working group meetings that will suffice as the input into the action matrix and we will not also do an SDP on it even if there is a performance issue.

MR. SIEBER: Well, I can see where you would get two different answers. You need to avoid that conflict by saying this is the one we will use.

MR. THOMPSON: So the challenge for us as NOR is to assure ourselves that this is at least as good an indicator of risk as it is what we have now where we currently do an SDP along with a PI and then take the higher color input into the occupation maker.

MR. SIEBER: Thank you.

MR. BARANOWSKY: Slide 5 identifies the points, comments, and questions that were identified from the May 2002 ACRS subcommittee briefing. Just to mention these bullets, the subcommittee did

indicate that we were moving in the right direction to solve many of the problems with the current mitigation system performance indicators.

They did want to know what we had learned from any pilot activities. At that time we were only formulating them but we said we would get back to you on that.

There was a question raised about should the PI that we are developing deal with risk in terms of thresholds. An issue was raised regarding some of the large numbers of SCRAMS that are needed to cross certain thresholds in the reactor oversight program.

We have looked at this and made some adjustments based on dealing with issues of validity of indicators where we have either two few hits that causes an indicator to cross a threshold, or so many hits that it's not really indicating anything. Don Dube will describe that a little later.

MR. ROSEN: You would be requiring so many hits that you would never get there?

MR. DUBE: Correct.

MR. BARANOWSKY: Basically. So-called ever-green indicator.

A question was raised and we think we

answered it but we put it here anyhow about whether we should be using a plant's own historical performance in a baseline or some industry performance. I think we discussed it at the last meeting but we have also concluded since then that we would like to use historical industry performance since.

If we were to use a plant specific performance for the baseline for plants that had -- then they would be rewarded by allowing to have a delta that goes even more in the core direction. The plants that have had a very good performance would be highly penalized. It seems to be more reasonable in light of what we are trying to achieve. Sort of a pragmatic as opposed to --

MR. ROSEN: This is the difference from the typical PRA approach where you would update the performance. It seems perfectly appropriate because you would tend to use this for a different reason. I think that is why I'm comfortable with that.

When you get done with talking about all these points, are you going to tell us -- is the industry, for instance, going to tell us what they have learned from the pilot and how they feel about it?

MR. BARANOWSKY: We're going to talk about the pilot. Tom Houghton is here from NEI and we would be more than happy to have him step up and say what he thinks.

The last point was there is sufficient data in EPIX. Even though the data currently isn't sufficient, there have been a number of interactions with INPO to get EPIX design and capable of handling this information.

It seems to be on track with respect to the time frame that we are talking about potentially implementing this indicator so that we would be able through INPO and their own so-called consolidated data entry system which is meant to be an efficient way of collecting various types of data to get the data that one needs in order to perform the calculations.

Okay. The next chart identifies some points regarding the White Paper that we sent, I believe, over a month ago. That's the one, dated April 28th. Let's make sure we understand what the White Paper is.

It's meant to provide the fundamental concepts and some related issues that give us a belief that we should pursue the development of the

mitigating system performance indicator. It's not an analysis of all the possible technical and implementation issues. It's pretty much a understanding that this looks like it has some merit.

It also gives the fundamental concept of how we would make some simplications in doing these risk informed calculations to keep the analytical part as simple as possible but no simpler than need be in order to get a reasonable indication.

We provide the mathematical formulation with the importance measure relationships, which I don't plan on going through anymore. We show how we treat unreliability, unavailability in such a way that we can combine them together and looking at both at the same time get an indication of the impact on the risk index.

This is about the simplest calculation, pretty straightforward algebra. It requires some bookkeeping but the equations are not really too complex. There's a parameter here, parameter there.

A lot of them are given by like the Bayesian update parameters, for instance, As, Bs, and things from analyses that we have done of industry data to come up with prior distributions

and that we have explored to understand how those parameters impact the calculations in the mitigating system performance index.

MR. ROSEN: The paper says that although the calculations in the paper can get complicated, the simplifications that you have proposed, that are being proposed, don't affect the results greatly. They are simplifications that have a limited impact except in some unusual cases.

MR. BARANOWSKY: We think we have identified just about all the little places where things can be unusual. The basis for some of the things that we're doing required some complicated analyses and Don is going to cover that. But then we believe we are able to boil it down into relatively simple and straightforward sets of parameters with these algebraic equations.

MR. ROSEN: That can be handled without a CRAY computer.

MR. BARANOWSKY: Yeah. This is just spreadsheet work.

MR. DUBE: This is not on a spreadsheet now.

MR. BARANOWSKY: If you can't track things that keep you honest with your tech specs,

then you wouldn't be able to do this. I would say vice versa is also true.

The benefits, of course, are identified in terms of some of the issues that we mentioned earlier; properly accounting for demand reliability and including plant specific designing data.

The limitations are called out to some extent in the paper. I think the interface is with where the significance determination process is proposed to be the appropriate methodology for evaluating the significance of performance issues is more detailed out in the NEI guidance document.

One thing that specifically needs to be recognized, that there are a lot of conditions that get discovered either by design reviews or by special tests that are not done routinely.

Those kinds of issues are also outside of the scope of this PI because they, in essence, are the discovery of conditions in which the plant would have been in a potentially significant risk state for a long period of time while all the indications can't possibly detect this because the data that forms the basis of the indicators is not being collective in those areas. That also goes into significance determination.

MR. SIEBER: But that does not include what you refer to as type two test results which is the 18 months as opposed to a correct test. So even though you may not discover something until you run the at-refueling 18-month test, that period where the deficiency is assumed to occur could be nine months under the old SDP process.

MR. BARANOWSKY: Actually, I think the main thing that would be discovered on those 18-month tests are running reliability issues. We have a way of dealing with those. The answer is yes and no.

If there was an issue that was identified that was, say, starting reliability on a diesel generator, for instance, that could not be detected for some reason during the normal monthly or quarterly test. We would have to take that into account. We couldn't just assume that the monthly or quarterly test provided valid numbers of demands. We haven't seen anything like that, by the way.

MR. LEITCH: For example, the recent -- at least a year ago or so red finding at Point Beach where they have daily mops. Under some conditions recirculation were correctly availed and the pumps would be run again and shut off. No amount of

testing would have revealed that after three months or 18 months or anything. I mean, it was just a recognition of the problem.

MR. BARANOWSKY: That's hopefully a rare event but it's a longstanding one that is not amenable to this type of indicator. There might be other types but not this one.

We mentioned issues that are related to differences between the mitigating system of performance index and a significant determination process which we have looked at and are continuing to look at.

There was a lot of discussion related to false negatives and false positives which we believe we have pretty good solutions for and validation issues which we are also addressing and have a pretty good handle on. Those are just highlighting the paper. I don't think we have solutions identified today. We can talk to you about some of the solutions that are in progress.

Of course, also since the last time we talked to you we did have an actual pilot program in which some 10 sites with 20 nuclear plants participated in testing out the guidance for identifying the scope of equipment within the

mitigating system including boundary and component identification. A data collection was done and computation using the original formulation that was put together about a year ago.

Another element of doing that pilot was to go through various validation and verification issues as we went along. Some of these involve special so-called table top studies with actually a significant amount of the pilot activity, for us at least, and that is the bulk of many of the things that we are going to talk about here today.

They included issues related to our own SPAR comparisons, SPAR being the standardized plant analysis risk models which the NRC used for our own risk analysis. We'll talk more about that. We wanted to look at a number of issues regarding differences between what the mitigating system performance index got and the significance determination process.

The other thing I want to point out is that the regions performed their temporary inspections per guidance and we got quite a bit of feedback on issues regarding burden and problems in following the guidance and that kind of stuff needs to be feed back into the updated guidance in order

to be more efficient if there is to be an implementation.

MR. LEITCH: One of the things that I don't quite understand is what the industry gives you versus what the industry does themself. In other words, in these pilots do they just provide -- let's say we're talking about the diesel generators. Do they just provide reliability or, I should say, unreliability and unavailability data? Then the expectation is that the NRC does the number scrunching to come up with the index?

MR. BARANOWSKY: Actually, that's a really good point. The original idea was that we would do a 100 percent parallel analysis of the data even though the licensees are responsible for it. They would use their PRAs and we would use ours. But that we would make sure that we did some PRA benchmarking so that we didn't have things like our models including designer operational features that were faulty based on understanding of incomplete information.

So the idea was to benchmark the SPAR models and then go off and do our own calculations because, as you will see when Don shows you, a number of technical issues came up for which we

needed to have a consistent set of models to look at these things across numerous plants. If we couldn't do that, I don't see anyway we could have done this project.

We would be working on this for years. Now we've got basically a set of, I guess, simulation models set up so we can look at a number of issues separately or together and look at the impact after we have benchmarked them against the licensee's models.

MR. DUBE: But, in answer to your question, the licensees submit historical performance, the number of demands for a particular quarter, all the importance measures, totaling the spreadsheet and automatically calculated what the equivalent delta CDF, core damage frequency, and what color designation is projected. It's been done for about six months. It was done monthly but if it were implemented the data would be submitted only every three months, every quarter.

MR. LEITCH: For example, in the diesels there's many, many different configurations so the impact of diesel unreliability and unavailability is factored into the model for that particular plant.

MR. DUBE: Exactly.

MR. LEITCH: Be it the smaller models or --

MR. DUBE: The importance measure. If you would find, for example, that a particular plant was, let's say, just two diesel generators where loss of off-site power or station blackout was a dominant sequence, the importance measures for those would be reflected in the high importance measures for that particular component for that plant.

Whereas another plant that had more diesel generators and station blackout or loss of off-site power was not an important contributor to core damage frequency might have importance measures that were lower for that particular plant.

MR. ROSEN: this was ACRS' specific point, that the new system had to account for these site specific differences in order to be fully robust.

MR. DUBE: Exactly, and that's what it does.

MR. ROSEN: And that's what it does.

MR. BARANOWSKY: So it's expected even if we implement this the licensees will make the calculations but they will make the data available for us once we have enough confidence in the

calculations and whatever because this is very different from the current set of indicators and does use plant specific PRAs so it's a step up in terms of what we've been doing in the past.

MR. SHACK: And he'll be calculating the importance measure with his PRA rather than you supplying him an importance measure.

MR. BARANOWSKY: That's right.

MR. DUBE: I'll talk a little bit about that when we get there.

MR. BARANOWSKY: Let me talk a little bit about the status of the pilot program and then if others have some points they want to mention, that might be a good time to.

We did hold a workshop in July of 2002 in which we went over the draft proposed guidance and we made some changes as a result of that workshop. Then finally we issued guidelines for the pilot as modification NEI 9902 in September.

Then from September through February the licensees collected and submitted the data. We performed the temporary instruction at the pilot plants basically from -- I don't know exactly. I think it was September because didn't we start in September?

MR. DUBE: Yeah.

MR. BARANOWSKY: It went at least all the way through March. We had another workshop in January for sort of mid-course assessment. Then we identified a number of technical issues regarding temporary instructions and details of calculations and anomalies and results and things like that. We redirected our efforts to look at the issues that Don is going to talk about shortly.

One of the things that we found that we had to spend a fair amount of time on in order to do all this was to bring the SPAR models up to a state where they could be used to give a pretty good reproduction of the licensee's risk down to a fairly low level.

Normally when we use the SPAR models we use them for the absence sequence precursor program and it pictured generic issues. We try to get our total risk of core damage pretty close, say a factor two or three on the total core damage frequency.

We think if we get that close and most of the top 10 or 20 dominant contributors are in there, we're happy because we're going to work with this on a case-by-case basis if it's a special issue or an accident sequence precursor.

MR. ROSEN: You mean close to the plant CDF?

MR. BARANOWSKY: Right. Or, as a minimum if we don't get it that close, we're going to say it's not there because we don't believe the plant CDF. That's a possibility, too. I might as well be fair about it.

In this case we had to understand differences that took us into the second and third decimal place because we are measuring delta CDF impacts on the order of 10 to the -6 or less and the total CDF at the plant is about five times 10 to the minus five. That's a pretty important thing to keep in mind.

We've got uncertainty on these core damage frequency estimates that might be a factor of three to 10 on the first significant figure. We're going now into the second and maybe third significant figure. That's a pretty significant calculational activity.

Here are the set of key technical issues that I was mentioning and Don's going to go over them. I think I'll just leave this with you and turn it over at this point to Don unless there are other questions, or would someone like to talk about

implementation before we move into the issues? Other implementation factors.

MR. SATORIUS: Mark Satorius from the Inspection Program Branch, NRR. Now or maybe after you touch on the key technical issues. One thing I just wanted to say is we have a process that we go through when we pilot these new performance indicators or any part of the ROP.

We went through that same process, you may recall, when we looked at SCRAMS. We would count manual SCRAMS or not count manual SCRAMS back in the beginning of the ROP. That process is in inspection manual chapter 608.

Notwithstanding the technical issues which Don is going to go over right now, there are what I call nontechnical or program type issues or success criteria. These are the things like having the ability to have license report to requested data without problems, whether the new PI will continue to maintain safety and meet some of the other criteria that the ROP has in front of us.

We are still analyzing those nontechnical aspects of the success criteria and notwithstanding all the good work that research has been doing on the technical issues we are still

looking at those nontechnical issues.

MR. ROSEN: Before you get to the next slide, the last bullet on your prior slide on the status of the pilot programs implies that having made a major effort to reconcile differences with the pilots, that you will have a similar major effort with all the other plants left. Is that not true?

MR. BARANOWSKY: We don't necessarily plan on having every single model capable to this degree unless there is some issue that causes us to believe we have to. We are looking at a scheme in which we use the SPAR models as a audit tool.

Based on our understanding of the normal SPAR model, QA process, and differences that are identified during that versus this much more enhanced activity, we can determine where we think we would like to spend the effort to bring SPAR models up to this level and then do an audit of licensee calculations.

Ultimately we might get there for all the plants, and actually I believe we have that budgeted but it doesn't necessarily need to be done immediately.

MR. DUBE: This is a different level

than in the i, 3.0(i)?

MR. BARANOWSKY: Yeah.

MR. DUBE: This is a notch up from that?

MR. BARANOWSKY: Yeah.

MR. ROSEN: Definitely.

MR. BARANOWSKY: It's enhanced models. It includes additional detail on support systems, recovery actions, and other things that were found to be important.

MR. ROSEN: But all plants will use the new indicator if we go to a new indicator. All plants will be using their PRAs to give you the data to be manipulated to find the importance measures.

You'll be taking -- if you don't do this level of effort on all the other plants, those that were not in the pilot, then to a degree you will be relying on those licensee models more than you did rely entirely on the pilot plant's models. Right?

MR. BARANOWSKY: That's true but, at the same time, we are identifying insights that we have obtained from both the normal SPAR QA work and this enhanced activity which we will take a look at and determine if that needs to be fed into this process such that the models have some level of consistency in that regard.

A couple of issues that Don is going to cover like support system initiators and things like that. Those came out of our reviews and we have different ways of dealing with that if they are not, for instance, included in the PRA that a licensee has.

MR. SHACK: Can you identify in the licensee PRA elements that must be of a certain NEI quality standard that you would feel comfortable with, the results from them?

MR. BARANOWSKY: I think we've got sort of a list of things, a tentative list that we put together already. We need to look at it and we need to ask ourselves what do we gain by spending effort making anybody do these things? Is it the third significant figure? Does that change what the outcome would be in using this PI because there are several aspects about the way we have looked at the so-called invalid and -- oh, what was the other indicator?

MR. DUBE: Insensitive.

MR. BARANOWSKY: Insensitive indicators which make some of this a little bit moot, actually, which is good. It doesn't have to be so twitchy, so to speak.

MR. SIEBER: This is also in terms of delta CDF so you could have some fundamental error and still have the delta come out of it.

MR. BARANOWSKY: Well, I'm not sure about that. I'll be honest with you, I think we learned that when you're working with delta CDFs of 10 to the -6 or smaller, it doesn't take much to get factors of two differences. If you've got eight times 10 to the minus seven here and 1.6 times 10 to the -6 there, there is not a lot of difference that gets you that.

That's a small delta CDF. Yet, that's the level at which 174 is being applied. It's a level above where risk informed tech specs are being applied. They are even going down into the 10 to the minus seven range. So whatever we've learned here certainly has some implications for other applications.

Nonetheless, I think we can identify how we can address concerns about how accurate one needs to be rather than calling it quality. Quality sometimes means documentation and does it look right and everything. For our purposes we just want to calculate things consistently, sort of robustly if you will.

MR. LEITCH: When the twenty plants in the pilot were selected, was the intention to cover the gamut of designs from very robust to --

MR. BARANOWSKY: That wasn't the intention. I don't know that we exactly did.

MR. LEITCH: Volunteers.

MR. BARANOWSKY: Will volunteers that cover the gamut step forward. They did all right.

MR. DUBE: And we have a mixture of Westinghouse low combustion engineering and preboiling water reactors. No BMWs but it's a reasonably representative of old and new plants.

MR. LEITCH: I was not thinking so much about the reactor manufacturers as diesel configurations.

MR. DUBE: Oh, we have from two diesels to four diesels, for example. From two aux feed pumps to four aux feed pumps.

MR. LEITCH: I guess another questions that comes into my mind is that it's plant specific indicator but is the green and white thresholds plant specific or is that one number?

MR. BARANOWSKY: That's actually a program threshold. How you calculate your plant's performance is plant specific. Everybody has to do

55 miles an hour or less but how you accelerate and break and whatever to do that, that's going to be a little different.

MR. LEITCH: So a plant with, let's say, more robust safety systems could be less conservative in the way he manages those safety systems and cannot cross the threshold.

MR. DUBE: It could tolerate more failures and more unavailability all other things being equal.

MR. BARANOWSKY: But we moved away from using a purely risk benchmark to a performance benchmark which doesn't allow such a wide spread. It allows some spread. It gives some credit. I think from what I'm hearing feedback wise it's about the right amount. That's a judgment call.

MR. LEITCH: One thing I was curious about and I kind of got lost a little bit in the White Paper was the merger of unavailability and unreliability because they say things are very much inter-related.

In other words, if one tries to drive the unavailability to zero, you could, and likely would, raise the unreliability. If you don't take the outages to do your preventive maintenance, your

unreliabilities kind of go up.

There's an inter-relationship between these two numbers. I'm just wondering how this indicator deals with that. I tried to figure my way through the math so I understand the inter-relationship but I couldn't quite see how that factored in here.

MR. DUBE: I can handle that. I think one of the reasons why this is an improvement is because the current indicator deals only with unavailability. You can find situations and industry representatives who will admit that they will manage to the indicator.

If there is a threshold here and their unavailability is going up, they will manage the indicator and perhaps in the long run to the detriment of reliability. Why I think this is an improvement it properly balances unavailability and unreliability so that in theory once you find that optimum, hopefully it's a broad optimum where the right preventive maintenance will give you an optimum.

I won't say zero but optimum unreliability and that's what the theory always tells you. And it weights unavailability and

unreliability by the importance measures. Vessel over UR which is kind of like a risk achievement minus one but it's an importance measure. It appropriately weights unavailability and unreliability in the appropriate amount and that's why I don't believe it is an improvement in that sense.

You are exactly right. In theory, if you're doing the right maintenance the sum of unavailability contribution to CDF and unreliability contribution to CDF should be a minimum if you're doing it just right.

MR. BARANOWSKY: I think the maintenance rule also pushes one in the direction of balancing unavailability and unreliability so that was another area where we were trying to be consistent. We could have taken these separately which is, by the way, what we did with the risk based performance indicators when we had a lot more indicators.

That becomes problematic with lots of indicators and not doing this tradeoff in one program in NRC whereas another one allows the tradeoff and you get inconsistencies and all of a sudden you've got two different requirements and it's not working.

MR. APOSTALAKIS: Is unavailability still defined in terms of maintenance? Unavailability is the ratio or what?

MR. DUBE: Yes. Planned maintenance and unplanned maintenance. There should also be some contribution picked up. If there is a failure and is corrected that should find its way in, too.

MR. APOSTALAKIS: But isn't that failure part of the evaluation? That's how you find it?

MR. BARANOWSKY: No. That failure goes into the unreliability, but what he means is if you take a component down to perform corrective maintenance, then that goes into unavailability. The so-called fault exposure time is captured by the unreliability term. There's no fault exposure time.

MR. APOSTALAKIS: There is a fault. They say you have to start. That goes to the unreliability. So what is it that goes to the unreliability?

MR. DUBE: If it was down three days for repair so that's an unplanned maintenance, that would find its way in the unavailability.

MR. APOSTALAKIS: Oh, okay. I see now. So the unreliability contribution then is just modified and there's no time.

MR. BARANOWSKY: Right. Failures per demand.

MR. ROSEN: And there's no assumption about how long it was unreliable prior to being discovered.

MR. BARANOWSKY: No, but it has to be failure that is detectable by the routine testing. You can't have something that was so unique that they went and did special test time and we've seen this. That's like an accident sequence. That gets special treatment.

MR. ROSEN: And that's what Graham referred to earlier was the point situation. That would be handled by the SDP, right?

MR. BARANOWSKY: Exactly.

MR. DUBE: The design deficiency of that nature.

MR. APOSTALAKIS: So if you find a phase to start on the 1st of February, you're not going to speculate how long it will be?

MR. DUBE: Exactly. That's correct.

MR. BARANOWSKY: We're just going to count up the number of demands over the period which we are measuring and the number of failures and we are going to do a calculation. Just the usual type

of PRA type calculation.

MR. SIEBER: That's unreliable.

MR. BARANOWSKY: It's demand reliability.

MR. SIEBER: Yes, it's demand.

MR. APOSTALAKIS: And then, of course, you have a separate reservation and you start to phrase for 45 minutes.

MR. BARANOWSKY: That's the usual run for reliability.

MR. APOSTALAKIS: So why did you decide not to speculate on how long it had been down? Isn't the average time usually one half?

MR. BARANOWSKY: That works if you have a long period for which you are going to collect data for that so-called speculative unavailability. For very short periods of time it gives you spikes and nothing. Spikes and nothing. What we are trying to do is over a period of three years taking demands and failures.

You also can't update that one, or at least we don't know how very well, using Bayesian statistics. It's consistent with the way we do PRAs. It's consistent with the way people do maintenance rule. That's the reason it shows it.

MR. APOSTALAKIS: In a PRA if you have the other test, you are averaging over time.

MR. DUBE: Right.

MR. APOSTALAKIS: For a single component if there is a failure it's not very large which is usually the amount. The average unavailability over that period is one half. That means this is the average probability. The average fraction of time or the interval is down.

MR. BARANOWSKY: Yeah.

MR. APOSTALAKIS: Now you're finding your failure on the test and you decide not to go that way but you're saying this is not a demand unavailability or failure.

MR. BARANOWSKY: I think maybe I can explain it. That's the constant failure rate assumption. As you ingrate over time T goes to infinity the probability of failure on demand equals exactly one half lambda for the constant failure assumption. So they are the same exact values. When you get into trouble is when you do it over short periods of time. Then your statistics get out of whack.

MR. DUBE: That's why we use a free year interval, too, to average things out.

MR. APOSTALAKIS: That's another thing. Why do you use the years? Can you use the years? Is this going to be used by the ROP?

MR. BARANOWSKY: Yes.

MR. APOSTALAKIS: And the ROP doesn't go by date.

MR. BARANOWSKY: It's a rolling three-year indicator. Rolling three years.

MR. APOSTALAKIS: I thought it was three-quarters.

MR. BARANOWSKY: No. That would be -- the statistics would be so poor for three-quarters you couldn't really use these kinds of performance indicators. It's not clear that you would be chasing noise or real performance changes if you look at things over quarters.

MR. DUBE: There was a study of the risk based, NUREG 17 I believe it is, where we looked at varying intervals. That's a whole separate report but it was found that three years was about as optimum as what it could get. Too little and it's too sensitive. Any more and you're not really seeing the trend. Three years seemed all right.

MR. BARANOWSKY: That just happens to be the interval that is currently used with the

performance indicators.

MR. DUBE: Well, on to the key technical issues. I' Donald Dube. I came to the Commission in October and pretty much took over for Hussain Hamzehee so you've got a new face here.

When I took it over I thought this is going to be pretty easy, but it didn't take too long, two months into the project, to realize there were a number of key technical issues.

Certainly during the pilot program or workshop in January a large number of technical issues, as well as some implementation issues, came to the surface. I'm going to touch upon a lot of the major issues that came about over the next few hours or so.

I do want to say that there is no way I could have come on board in such a short time and tackled these issues without the assistance of the primary contractors, ISL and Idaho, and also Corey Atwood. I want to give them acknowledgement.

MR. APOSTALAKIS: You just left me out.

MR. BARANOWSKY: So is that an endorsement?

MR. APOSTALAKIS: Maybe in the future when you send those reports you can identify those

because I had no idea.

MS. WESTON: Right.

MR. DUBE: I've listed them here on the overview. Independent verification. One of the first things we found were the significant differences between the SPAR model and the plant PRA. Pat Baranowsky mentioned this earlier.

We thought when we were going into this that the importance measures, let's say for a particular component, might vary from the SPAR model to the plant PRA, let's say, for a diesel generator by maybe tens of percent or 50 percent or maybe a factor of two kinds of numbers.

Lo and behold we found significant differences. In many cases one order of magnitude difference and in other cases two orders of magnitude difference. It really begged a lot of questions.

As we were doing this and they started rolling in we asked ourselves what is going on here and what are the differences. We undertook a major effort to reconcile the differences and I'll be talking about that. A lot of it had to do with the fact that the SPAR's original intent, the S stands for standardized. One started by having more or

less a template for different series or classes of plants.

As years have gone by they have become somewhat more customized. But there were still significant differences specifically in many of the balance of plant systems and the cooling water support systems such as service water and cooling water we later found out.

So in parallel I will be addressing a lot of these other technical issues. A major effort was undertaken at Idaho to understand these differences and explain them. I'll show some examples.

We also early on had come to an issue that we called an invalid indicator. What that basically means is that if one component failure resulted in the system indication turning to white, one failure does not make a trend.

While one failure may result in exceeding delta CDF of 10 to the -6 on paper, does that necessarily mean that performance is degraded to the point that an indication should be actually white? It has come to be called invalid indicator. It has certain connotations so it's really a bad name for it by definition but it's come to be kind

of associated with that.

MR. APOSTALAKIS: So the first would be green, white, and green is still CDF based?

MR. BARANOWSKY: It's changing to performance based.

MR. APOSTALAKIS: It was always performance based.

MR. BARANOWSKY: We were delta CDF based the last time we talked. Now we are changing back to performance based.

MR. APOSTALAKIS: Just green?

MR. DUBE: It's a mixture.

MR. BARANOWSKY: Green and white.

MR. DUBE: No, it's for other, too.

MR. BARANOWSKY: It's for cases where we have to deal with these invalid and then insensitive indicators. I would just like to add one more thing about the invalid indicator.

The reason why we called it invalid, I think, was because we are collecting data over such a short period of time that the number of demands that one can get are too small to get a good statistical indication of what the demand failure rate is. One could get false/positive indications very easily and you need more information if you are

dealing with that period of time.

MR. DUBE: So what we're saying then is the process is implemented then and runs along for years those invalid indicators would be washed out. They would not --

MR. BARANOWSKY: We have a different way of treating them.

MR. LEITCH: If you're looking at three-month quarterly interval when you do a test once and it fails, is that an example of the kind of thing that would be an invalid indicator?

MR. DUBE: No. An invalid indicator would be an example where if there were no failures and the indicator was less than 10 to the -6 or green, one should not have a situation where just one failure of a particular component would turn it white above 10 to the -6. In that circumstance we are going to have an alternate formulation that will not call that white.

MR. APOSTALAKIS: I don't know about that. If it's 10 to the -6 and you've got one component, that is pretty significant. If you expect it is 10 to the -6 and you get one, yeah, you should worry.

MR. BARANOWSKY: Suppose you do this?

You look at data over a 10-year period of time and you have one failure. Take a look at five. You still have one failure. Then you look at three, then you look at two, then you look at one. Well, I take that same exact data and when I look at 10 years it's not risk significant. When I decided to make the one-year look, it was down in risk contributor. What does that mean?

MR. APOSTALAKIS: But you have already decided to go back to years. What was the rationale?

MR. BARANOWSKY: Going to three years was we got enough data so we didn't have that kind of situation occurring. In the meantime between failures was such that we could collect several.

MR. APOSTALAKIS: I'm curious. You're still keeping the white, yellow, yellow, red based on delta CDF?

MR. DUBE: Right. Ten to the -6, 10 to -5, and 10 to -4.

MR. APOSTALAKIS: You know this Committee has criticized that.

MR. DUBE: Doesn't like it.

MR. BARANOWSKY: We're also looking at --

MR. APOSTALAKIS: Performance guys, why didn't you ask the experts in the field to tell you what the yellow should be? There are so many people who have long experience. Two of them are here. When would you worry?

MR. BARANOWSKY: The fact of the matter is once you get into the white zone you address the issue.

MR. APOSTALAKIS: Exactly. That's actually true.

MR. DUBE: Well, that's a good point but, in essence --

MR. APOSTALAKIS: There are four, not two.

MR. DUBE: The best way to address the plant specific variations is why using some constant measure like a 10 to the -6 threshold and let the plant PRA manifest itself through the importance measures and the performance data into how many failures does that equate to to turn white or yellow or red as opposed to just picking some numbers out of the air that may be a one size fits all. Well, we can talk about it.

The third issue is the large number of failures to turn the system to white which is called

an insensitive indicator. In the sense that if something has a relatively high risk achievement or importance measure, it may take one failure to exceed the delta CDF of 10 to the -6.

At the other end there may be certain components that have such low contribution to CDF, have low importance measures that it may take theoretically a large number of failures before it turns to white. When I say large, I'm talking many, many dozens, for example. That's not indicative of a good measure.

MR. APOSTALAKIS: But you said white is performance based so it shouldn't take that many. Only when you have delta CDF phase thresholds you get that problem. Because if the expert tells you yeah, it's not very significant but if it should go about two failures over a certain period of time, I would worry.

If it should go from white to yellow, then your argument is valid because now it's rigid calculation. That's the advantage of using expert opinion. Anyway, isn't this the issue, though? Both of these colors are statistically minded.

If it has its own process and I want to establish a quality control program, isn't it the

issue of what is the number of failures that I should worry about and if I see more, I have a problem? That's really the issue we're facing here. The peculiarities of the rate is so low.

Corey, you want to say something?

MR. ATWOOD: Corey Atwood, Scott Wood Consulting. White was based on delta CDF.

MR. APOSTALAKIS: I understand that.

MR. ATWOOD: But if you're concerned in performance, then you would say how many do we expect, how many do we really not expect.

MR. APOSTALAKIS: Forget about the practice of the NRC. You want to establish a quality control program for its own process. I mean, the first quantity you're looking at is lambda, the average number you expect to see over a period of observation. What makes this complicated is that lambda is very low so you are trying to make it reasonable by going to three years.

In other cases it's going to be unreasonable. So I think fundamentally that's what we're facing. It was a side remarks. There was no question. There was no praise either. There should be, though.

MR. DUBE: Okay. Well, I appreciate

that. I stepped into the program here and we were developing a risk-based performance indicator and that's basically what it is. There are -- we have as a result of bullets No. 2 and 3 realized that relying on a strict algorithm that estimates the delta CDF and translates into number of failures.

Ruling on that can result in kind of ridiculously low numbers. On the one hand we call that invalid and ridiculously high. On the other end we call that insensitive. We will be proposing -- are proposing limits on both ends to avoid that situation. It will be fundamentally risk-based but with performance based limits at the upper end and the lower end to an event.

MR. APOSTALAKIS: Ultimately all of the thresholds will be performance based, right? It's about four or five years.

MR. DUBE: Yeah, four or five years.

MR. LEITCH: Do both of these programs greatly impact at the present? In other words, once the program runs for the full three years, will be issues disappear?

MR. DUBE: No. I mean we will resolve these issues but the fundamental reason of why one failure might result in delta CDF more than 10 to

the -6 is because it's a finite time frame of three years.

Certain components like a steam driven or steam pump have such a high importance measure, risk achievement where all it takes is one failure to give you a delta CDF in that three-year time frame or the 10 to the -6. Averaged over many years it's probably a wash because it's a fundamental issue that 10 to the -6 is kind of a low threshold. It's a very sensitive threshold.

MR. BONACA: You said you will propose leads so you have to have some criteria on what is reasonable.

MR. DUBE: And I'll talk a little bit about that. The fourth one is identification of system boundaries. This is more of a mechanistic thing having to do with bookkeeping and realizing, for example, that if there's a service water system providing cooling for a diesel generator and there's a valve and that valve's function is only to isolate or open flow to the diesel generator, then the way we are considering it is that valve is part of the diesel generator boundary as we define it because it only serves the function to that diesel generator and not as part of the service water.

I bring that up because there were a number of issues along these lines as the pilot program and a lot more issues than we thought. I'll show you how we addressed those.

Data collection burden. Many licensees did say that at the time it has been a burden to collect this data. Certainly the first time and there is a lot of data collection that has to be done up front because we are going back three years of historical performance data, demands and failures so there is quite a bit of effort there.

Then to some extent maintaining it but I will address how we are planning to integrate this with INPO, WANO, consolidated data entry system so that it would minimize that burden we think.

MR. APOSTALAKIS: There will be additional burden in the issue of SPAR versus plant specific model, right? Even though the plant works with you and makes sure that these discrepancies are resolved. At some point an indicator like this will require every licensee.

MR. DUBE: That's a good point. We have said going into the program for the purpose of this pilot that we don't expect any of the pilot plans to make any changes to their PRA. It's a voluntary

process. We just want to exercise the method, collect the data, and see how it turns out.

But as part of the SPAR enhancement process and reconciliation, if there are significant differences between the plant PRA and the SPAR model, and we truly believe that the SPAR model is correct and the plant PRA model has an absolute error, it is expected that error has to be corrected or certainly addressed in one way, shape, or form or another.

MR. BONACA: But you found many instances where, in fact, the plant specific model had an error?

MR. DUBE: A number. Not a lot. I would say not a lot. A few, and these were the region inspectors that found these in the process.

MR. BONACA: I'm surprised. That's an interesting thing because the assumption is always the problem is going to be with SPAR and you're telling me that instead you found PRAs out there with the errors.

MR. DUBE: In some cases there were omissions. A particular valve that is needed for recirculation flow of a pump was not modeled. Things along those lines.

MR. BARANOWSKY: How about that Point Beach PRA? That didn't have anything about the requirement of instrument error to make the auxiliary feed water pumps work. That's the kind of thing we're talking about.

MR. DUBE: I think when I get to the next couple slides you'll have some eye openers.

MR. APOSTALAKIS: Judging from the examples that Don and Pat gave us, that was the motivation for asking in the last letter that people look at the operating experience much more carefully because I'm not sure that the word error would apply if somebody didn't analyze a particular failure mode. Unless everybody else in the world is doing it and it's a well-known fact, why would you call that an error? The way you learn is by looking at operating experience. I mean, I would, and that was the motivation.

MR. BARANOWSKY: Well, the second way, and maybe Tom can comment on this, was I think we all learned a lot when we tried to compare the two PRAs, the SPAR model versus the licensee's model. The bulk of the times we had to change the SPAR models. We have to fair about the situation.

MR. ROSEN: Well, that's what I would

expect. You would expect to have to change the SPAR model because of the PRA. Site specific PRA is more detailed and, in fact, says that SPAR model gives you an answer that is not conservative. The case that Mario raises is the one that is more troublesome and more surprising.

MR. DUBE: I can count on one hand those number of examples, but there were a number. As I said, it went both ways. If SPAR was out of sync with the plant PRA, they most likely would have been issued of not modeling cross connections between unit one and unit two or from train one to train two or some kind of things like that. It's pretty much not part of the standard.

MR. ROSEN: But they, in fact, affect a plant specific PRA in a very substantive way.

MR. DUBE: Oh, yes. Definitely.

MR. BARANOWSKY: Usually though only a factor of two or three on the total core damage frequency.

MR. ROSEN: Only.

MR. BARANOWSKY: Well, I say only because let's be honest about what the uncertainty is in these calculations.

MR. ROSEN: Ah, yes. When you reflect

it against uncertainty, I agree.

MR. BARANOWSKY: That's the big, big errors. Most of the changes we saw weren't even touching the first significant figure but they do impact the Fussell-Vesely importance measures. If you want to get the pecking order right, let's say right is the correct term. What is the most important thing and work your way down. Then you have to go beyond the first significant figure.

MR. DUBE: Okay. The next two issues I'll get into more detail but it basically relates to how does one treat the common cause failure contribution to Fussell-Vesely. That has to do with the fact that many models, plant PRAs, take into account the fact that if the independent failure rate or single failure rate changes, then there is some connection and some coupling to the common cause failure rate. We need to address that and I will talk a little bit about that.

Another item is support system contribution to Fussell-Vesely in that we are dealing with mitigating systems such as service water and component cooling water which are often sometimes called support systems. But those support systems can also be initiated as loss of service

water, loss of component cooling water.

A particular component that is an initiator will have a Fussell-Vesely associated with that initiator. Many PRAs use a single point, a point estimate for the initiating event frequency so it will get properly captured for a particular support system. We propose a --

MR. APOSTALAKIS: It's described in terms of frequency, right? It's a support system, it's unavailability and unreliability.

MR. DUBE: Yeah, but it would have also Fussell-Vesely associated with it, particularly like a service water pump is part of the loss of service water initiator there would be a contribution on the pump to the loss of service water frequency.

MR. APOSTALAKIS: A different quantity.

MR. DUBE: Correct.

MR. APOSTALAKIS: Are you going to come to that?

MR. DUBE: Yes.

MR. APOSTALAKIS: I noticed you are avoiding RAW. Will you explain to the subcommittee why? I couldn't find it anywhere.

MR. DUBE: Well, because the formulation is delta CDF as opposed to RAW is given a base what

is the factor by which a CDF increases so we use Fussell-Vesely over UR and that's approximately the risk achievement -1.

MR. APOSTALAKIS: Why do you bring into this the burn bomb? Are you going to talk about these things? The burn bomb measure is described but why I couldn't figure out.

MR. DUBE: Sometimes it's more convenient to use burn bomb.

MR. APOSTALAKIS: But what are these times? Are you ever using it? More convenient on Fussell-Vesely?

MR. BARANOWSKY: I think it's the way that things were originally thought out was in terms of burn bomb importance measure being the proportionality constant, if you will.

MR. APOSTALAKIS: Yeah, but it wasn't clear why it was discussed in the report.

MR. BARANOWSKY: The reason we went to Fussell-Vesely, this happened before Don was on board, was because everyone calculates Fussell-Vesely importance measures and not everyone calculates burn bomb importance measures. We just said, well, let's take this burn bomb and burn it into a Fussell-Vesely divided by a parameter. It's

the same thing. It's all proportionality.

MR. APOSTALAKIS: I think the fact that everybody calculate Fussell-Vesely and RAW becomes the driving force behind the analysis and I don't like that. The convenience of getting these things is making us do mental acrobatics to justify what we get and treat everything to Fussell-Vesely and RAW.

MR. DUBE: Yeah. One could very well use burn bomb and maybe we should have started with burn bomb.

MR. APOSTALAKIS: Yeah, but I mean more for analytical convenience but, for heavens sake, it shouldn't really drive what you do.

MR. BARANOWSKY: It really isn't driving anything. I mean --

MR. APOSTALAKIS: You dare tell the world that you don't like Fussell-Vesely, you see, because everybody gets it from the quotes.

MR. BARANOWSKY: We chose it because it's not going to make any difference whether we use burn bomb or Fussell-Vesely divided by unreliability. We're going to get the same exact values. Since everybody has it, it's a burden issue. If it was going to affect the way we did the calculation, then we would have said something about

it but it's really not going to change the calculation and the burden.

MR. DUBE: Okay. I mentioned support system contribution and then touched upon a relationship of SDP and PI and we'll talk about that again a little bit later.

MR. ROSEN: Are you going to go through each one of those things? Why don't we just summarize it. We ought to speed it up is what I'm trying to say.

MR. BONACA: I think we're getting some good results.

MR. DUBE: Okay. Let me talk about the independent verification. The original intent was to replicate the MSPI submittals from the licensee using the SPAR model. I mentioned before we expected them to be in pretty good agreement but in many cases we found significant differences in the importance measures, Fussell-Vesely over UR.

Just because there is high-level agreement doesn't mean that 10 to the -6 and lower level there is agreement. In many cases the importance measures weren't just off by factors of two and three but by one to two orders of magnitude, especially cooling water support systems.

We had to reconcile the differences at the lower level. We had to go one level deeper in this SPAR model, we're calling it the SPAR enhancement, and either change the SPAR or recommend to the plant the PRA change was justified or both in some instances. We undertook this effort to do it for 11 distinct SPAR models for 20 nuclear units.

Let me give you an example for Bravewood. The PRA internal events CDF is 3 E to -5 per year. The SPAR before looking at it was 7 5 E to -5 per year. I'll skip the third yellow bullet for now. On average the Fussell-Vesely over UR was too low in the old SPAR model by about a factor of 10.

I mean, there was some factors of 30, 40, some factors of 2. Sometimes it would be close. Far to great of a difference for this particular application. I show you in this bottom table some typical components, RHR pump, aux feed pump, diesel generator, service water pump, volume control isolation valve.

What this is is the Fussell-Vesely over UR ratio. That is the fundamental importance measure in the MSPI. It's the Fussell-Vesely divided by the unreliability. The middle column is

the ratio of the ratio. It's the old SPAR model Fussell-Vesely over UR ratio to the plant PRA ratio.

If the plant PRA and the old SPAR model were in perfect agreement, these factors would all be one, would be normalized to one. You can see that they are pretty much all over the universe.

On average, geometric average if you will, the old SPAR model was too low for a factor of 10 so if the purpose of trying to replicate the MSPI results were to use the SPAR model and importance measures, we are already far off to begin with so we have to understand the differences.

At least understand the differences, reconcile the differences, and then back off. An effort was made to enhance -- I use the word enhance. It modified the SPAR models for Bravewood as well as all the other pilot plans to understand the differences and change the model.

With the enhancements to the SPAR model on average the agreement is within a factor of 2 high or low. That's the last column so the new SPAR models for the plant PRA model. There's numbers high and there's numbers lower than 1 but take a geometric average it's within a factor of 2.

MR. APOSTALAKIS: I'm confused. The White Paper says on page 4 the MSPI is formulated as the sum of changes related to unavailability and unreliability so it's the sum of the change.

MR. DUBE: Right.

MR. APOSTALAKIS: Why are you focusing on Fussell-Vesely divided by unreliability?

MR. DUBE: Because in the formulation the change in CDF is a factor of FV/UR. If that factor doesn't agree between the plant PRA and SPAR model, then everything else thereafter might as well not even continue.

We're saying that importance measures, which is a reflection of how much does this particular component contribute to the overall core damage frequency differed by factors of 10 and we're biased low in the SPAR and that says how can we even continue with the independent verification if we can't even understand where those differences are.

MR. APOSTALAKIS: So what were you verifying?

MR. BARANOWSKY: Verifying that we can calculate basically the same MSPI value by paying a little bit of attention to enhancing details of the SPAR models so that we understand the accuracy of

can't even understand where those differences are.

MEMBER APOSTOLAKIS: So what were you verifying?

MR. BARANOWSKY: Verifying that we can calculate basically the same MSPI value by paying a little bit of attention to enhancing details of the SPAR models so that we understand the accuracy of the licensee's calculation.

MEMBER APOSTOLAKIS: So you are working with the infrastructure.

MR. BARANOWSKY: Yes. Otherwise people just calculate things.

MEMBER ROSEN: And they have to buy it.

MR. BARANOWSKY: That's it.

MEMBER ROSEN: This way the SPAR model is tuned up, you could say, and you could go in and independently judge what you get.

MR. DUBE: That's exactly right.

MR. BARANOWSKY: If you want a risk-weight thing, you've got to be able to do this.

MEMBER ROSEN: You don't have to do this every time and you wouldn't.

MR. DUBE: I mean, to me this is the ultimate, quote, quality check, in that you are taking one PRA with all its models and assumptions,

and success criteria, and data, and bringing it to a whole different PRA developed for a separate purpose and separate applications, and trying to understand the differences.

And changing the one, or in some cases, the both, to get at least reasonable agreement typically within a factor of two on importance measures.

I don't think that we could ever do better than a factor of two. In some cases we do much better than a factor of two, and in some of the other plants, we just can't come to two.

MEMBER ROSEN: Well, what do we get? It's four. This is not about the search for ultimate --

MR. DUBE: No.

MEMBER ROSEN: This is about trying to decide what to do in the action matrix based upon inspection result in PRAs. So it has a very pragmatic reason. So if it didn't have that pragmatic reason, you might want to keep on working it until you get near perfect agreement.

But that is not the objective and we are only using a pragmatic reason to get into the action matrix and get it right.

MEMBER APOSTOLAKIS: And the action matrix already has intervals, and so the same accuracy is not really --

MR. DUBE: That's correct.

MEMBER LEITCH: I would like to think there is a backfit issue here, and suppose you run into a licensee who says that is my PRA and I am sticking to it.

MR. DUBE: Well, for the purposes of the pilot, I said that we were not going to make it, but in my opinion if there is -- an outright error has to be corrected, and if there is a difference of opinion, then there needs to be some reconciliation.

For example, and I will bring this up. I have a parentheses here, assume same success criteria for PORV. This is important, and given this assumption, that last column is what -- you know, the comparison, and then the third yellow bullet, used by Model 31E to the minus 5, is almost in perfect agreement with the plant PRA.

But that is an important assumption, and having supervised on one of my previous jobs about a dozen feed and bleed calculations on a plant specific basis, and realizing how sensitive the results are in terms of timing of operator action,

and the number of pressurizer pores, and the high pressure injection pumps and so forth, the success criteria is so sensitive to a number of assumptions that it is important that we have those analyses done with a high degree of accuracy.

The spy model right now has a success criteria of 2 out of 2 porches. The Braywood plant PRA is 1 out of 2 porches. Now, I an not saying that it is not impossible to have (inaudible) one out of two porch, but my own experience has been that Westinghouse plants of this vintage and the amounts that I am familiar with indicates that it may be closer to two porch than one porch.

But to show you how sensitive these core damage frequency results are to this one success criteria, because it is a two ox feed water plant, and a motor driven pump and a diesel driven pump, if the success criteria was changed from 1 out of 2 to 2 out of 2, the core damage frequency goes up by a factor of three.

And it is a most sensitive -- I won't use the word assumption -- success criteria that I have probably ever seen in my career. So it is important to understand where these differences are and reconcile them before moving on.

And that is why there were examples not as dramatic as this, but there have been a number of them

MEMBER APOSTOLAKIS: Are you familiar with any PRA or aware of any PRAs where this uncertainty was explicitly stated?

MR. DUBE: Not in uncertainties.

MEMBER APOSTOLAKIS: Well, you are not sure it is 2 out of 2 either.

MR. DUBE: Right.

MEMBER APOSTOLAKIS: Are you aware of any PRAs that acknowledge this explicitly and do something about it?

MR. DUBE: No.

MEMBER APOSTOLAKIS: It is a factor of three, right?

MR. DUBE: Yes.

MEMBER APOSTOLAKIS: Okay. But you know about PRAs, right?

MR. DUBE: My point is that it is an open -- you know --

MEMBER APOSTOLAKIS: I am just trying to make a point.

MR. BARANOWSKY: I think the other thing is that there were some issues raised, well, what if

these plants meet the ASME standard for quality or whatever.

I don't know whether they meet it or not, but we can come up with a list of things that if you want to talk about payoff in terms of implications on the quantitative results, and the pecking order of what is important, we pretty much know what they are.

You can go and talk about whether their documentation is good, and if they have got all this other stuff. If you want to get the so-called right answer, these are the things that you are going to have to look at.

I think this is the heart if you will, the kernel of PRA quality issues, and you pass down the quantification values, and the pecking order of what is important.

MEMBER APOSTOLAKIS: When you say this, what are you referring to?

MR. BARANOWSKY: The items that we are able to find by doing this work.

MEMBER APOSTOLAKIS: Do you have a list of those?

MR. BARANOWSKY: We have a tentative list of those insights that we have gained by not

only looking at these 20 plants, but by doing a little more simplified look at the other eight.

MEMBER APOSTOLAKIS: And you can give that to the committee?

MR. BARANOWSKY: We are going to make that available.

MEMBER APOSTOLAKIS: We appreciate it.

ACTING CHAIRMAN BONACA: For this kind of work it is a fundamental level of the cooling, and in almost every scenario and therefore you have a measure mode of cooling with this kind of sensitivity, and it is not recognized as a sensitivity position, but yet in these other items, it is not stated or documented.

But yet it is not surprising that you would have a sensitivity to it, and whether or not you need 1 or 2, you know, it is a key element.

MR. BARANOWSKY: And we don't propose that this vehicle is the vehicle for going out and ensuring some, quote, level of quality with licensee PRAs. We are saying that we can provide insights.

Right now you are using PRAs to do all kinds of other things. This is a voluntary program, and we are just saying that these are the areas that we have learned can have significant quantitative

differences.

ACTING CHAIRMAN BONACA: You see, it opens up all kinds of questions and so let me ask it this way. For example, are these (inaudible), and I would say probably not.

So you have to assume that now how do you decide that if you need 2 out of 2 that you would be able to open both? How do you decide that you will have all this success with 1 out of 2; and you attempt to open both and you only open one?

I mean, you have so many issues that drive the issue of sensitivity. And again I don't want to raise too many questions on the source of the PRA.

MR. DUBE: Well, my whole point of bringing this was that there was a lot of lessons learned, and a lot of information that has gone actually both ways, in terms of making enhancements to this part, but identifying where SPAR models in the plant PRA had significant differences that still need to be reconciled.

The next example is Palo Verde, which is where I believe is the best example where the enhancements made to the plant PRA were extremely good, and we didn't find those kinds of gotchas if

you will in the particular PRA model.

You can see that there is a bunch of columns here, and I am showing them a little bit differently. I say at the top of the page that the Fussell-Vesely over the UR, and that is the importance measure, and on average within plus or minus 25 percent.

I mean, for the major components to be concerned in the MSVR. Previously I said that we can get it within a factor of two, and this is within 25 percent, which is even a closer agreement.

There is three columns; the plant PRA, and the SPAR enhanced, which is what we have done after we have made these efforts to reconcile the differences in the SPAR 3-i, which was before if you will. So that kind of flip-flopped there.

But it is important not only to get the overall core damage frequency, but to have agreement in terms of the contributors to the core damage frequency.

And while the first column, plant PRA in the SPAR 3-i, the core damage frequencies were like within 25 to 30 percent. The constituents that made it up in terms of contributions of transients and tube ruptures, and LOCAs, were not so close.

But as a result of the enhancement effort, not only did the overall core damage frequency come up reasonably close, but the contributions of the next level, which is contribution by percent to each particular initiator, has pretty darn good agreement.

And then even at the third level, which is the Fussell-Vesely over URs, we were able to get it to pretty good agreement. And in this particular example, we made a lot more changes in the proposal to the SPAR, and I believe there may have been a handful of recommendations that the plant PRA would take.

But again one of the important lessons learned if you will, and benefits of the overall record. Any questions on this?

(No response.)

MR. DUBE: Okay. I am going to talk about invalid indicators, and it has to do with the fact that components with high points measure one component failure and can result in a delta CDF of 10 to the minus 6.

I won't go through the math, but if there is a high Fussell-Vesely UR, which is like a high risk achievement work, if the change in

reliability or unreliability times a high number, can exceed 10 to the minus 6 for this one particular failure, in this program we have been calling it an invalid indicator.

MEMBER APOSTOLAKIS: Again, this is a non-issue, because you are not going without a threshold. You are not establishing a threshold between why it is based on the CDM.

MR. DUBE: We are using 10 to the minus 6 CDF as the primary means --

MEMBER APOSTOLAKIS: I thought you were switching to performance.

MR. DUBE: Well, we were using performance based at the lower end and the upper end. In this particular case, we would use a front stop here, which says that we are not going to allow one failure to become --

MR. BARANOWSKY: It is the false-positive fix. It is the fix for false-positive

MEMBER APOSTOLAKIS: And based on performance do you find that you can tolerate one phase, where does this come into play?

MR. BARANOWSKY: This is not based on performance.

MEMBER APOSTOLAKIS: Yes, but I thought

you said earlier that you are not using this anymore, and that you are switching back to performance. I understand what this is based on, but I am just questioning whether it is relevant anymore.

MR. BARANOWSKY: It is relevant for those indicators where if a single failure pushes you over the green-white interface from normal baesline to the one failure, and it takes you over the green-white interface, that is where this comes into play. And only for those cases.

MEMBER APOSTOLAKIS: Well, if you have a delta CDF criteria threshold.

MR. BARANOWSKY: Right. And if it turns out that one failure does take you over the delta CDF, then you go to this so-called frontstop approach, which allows more than one failure based on our analysis of concerns concerning false-positive indications.

MR. DUBE: And the frontstop would be the minimum number of failures within a system before the performance indicator turns white.

MEMBER APOSTOLAKIS: I still don't understand it.

MEMBER SIEBER: You may have missed one

of the flip-flops.

MR. THOMPSON: Let me see if I can clarify this. This is John Thompson from the Inspection Program Branch. We are implementing generic risk informed thresholds for every plant. If that plant determines that they have either the invalid or the insensitive issue, they will use the alternate means of determining what is the threshold.

But for purposes of the public, and they go on the webpage, they will see that 10 to the minus 6, and minus 5, and minus 4, and we have yet to work out the details.

But for those plants that have a system that might meet one of these two alternate approaches, there will be an asterisk, and then you will see what the new threshold is.

So thresholds are risk-informed, but it is just that for some systems at some plants the research is proposing to use the alternate. It is adding a degree of complexity that we in the program office have to deal with, and we are working with that.

MEMBER APOSTOLAKIS: Green to white was never risk based. Green to white was always

performance.

MR. BARANOWSKY: For the current performance indicators.

MEMBER APOSTOLAKIS: yes.

MR. BARANOWSKY: But the concept right from the beginning was risk-based, or risk-informed.

MEMBER APOSTOLAKIS: Okay. Now we are on record as opposing risk-based thresholds for all the (inaudible), and so in that sense what you are saying is interesting from the mathematical point of view.

But the committee does not accept your premise. Is that clear enough to everyone, or what is it that you are not understanding?

MR. BARANOWSKY: Well, let me also --

MEMBER SHACK: It's clear.

MR. BARANOWSKY: Let me also point out that presumably --

MEMBER APOSTOLAKIS: That does not mean what you are doing is wrong.

MR. BARANOWSKY: -- you may change your mind some day. What we are trying to say here is that you have some concerns about using risk thresholds because they give some results that just look ridiculous.

What we did is we said, well, why don't we try to get the best of both worlds. We will try to use risk as much as we can, because that is what the Commission told us to do.

But when it starts to look ridiculous, either on a false positive indication or false negative, we won't let things get way out of hand. We don't want it to be twitchy, and we don't want it to be so forgiving that it looks like anything goes.

So there is a vast number of systems and cases where we can use this thinking and get what looks like pretty reasonable results, and there are some that don't, and we take care of them with this.

MEMBER APOSTOLAKIS: On the other hand, you can say that this is a self-created problem? if it is one of 10 to the minus 6 for CDF, then that creates a problem.

MR. BARANOWSKY: But one would have to change the premise of the reactor oversight program's threshold evaluations from what was put in 99-007 to something else.

MEMBER APOSTOLAKIS: Well, 99-007 did it right for green and white.

MR. BARANOWSKY: Well, that was an expedient thing, and they said it was.

MEMBER APOSTOLAKIS: Which turned out to be right.

MR. BARANOWSKY: Since I wrote that section of 99-007, I will accept that compliment.

MEMBER ROSEN: It's better to be lucky than smart.

MEMBER APOSTOLAKIS: You see, that is my point though. It was mentioned earlier that the committed doesn't like. It's not what the committee doesn't like. The committee wrote an argument in the report on why one should not do that. So it is not a matter of liking.

Now all the problems that you are having here could go or would go away if you went performance based, because the experts then would have told you, look, this is unacceptable. If I see one failure, you know.

So most people tolerate two failures. So the whole thing goes away.

MEMBER SHACK: From a pragmatic point of view, you have solved the problem, George.

MR. BARANOWSKY: George, I think you have a logical inconsistency if you will excuse me. If you want this thing to be risk informed --

MR. DUBE: And plant specific.

MR. BARANOWSKY: -- and plant specific --

MR. DUBE: There is no other way.

MR. BARANOWSKY: -- it can't be purely performance based. You have got to bring risk into the picture somehow, and I don't see how you do it by just saying everybody can take two failures on this end, or six failures on that end.

ACTING CHAIRMAN BONACA: But we want to certify that --

MEMBER APOSTOLAKIS: But it is risk informed.

MR. BARANOWSKY: But it is not plant specific. This is an ACRS comment. You are going to have to go back and change that one, too. You set it to reflect configuration of plant specific data, and now you are telling me not really. So just change everything.

MEMBER APOSTOLAKIS: Listen, listen, the way --

MEMBER ROSEN: One member in ACRS doesn't make. You can have George's opinion on that, and maybe the whole committee would --

MEMBER APOSTOLAKIS: I have not expressed an opinion that is inconsistent with the

letter so far.

MEMBER SIEBER: So far.

MEMBER APOSTOLAKIS: Now, the risk-informed, I think what this is going ultimately is that what would really matter would be the results of the SDP, and not the performance indicators. Performance indicators are just an indication of how you are rating with respect to your colleagues, the peers.

What really matters is what you find in the inspection and the risk (inaudible), which I think should be calculated, because how many PRAs have you seen where you go to core melt because one thing is of high frequency? No. It is a combination of events. And usual combinations are there.

It is not that something happened too many times, but it is interesting to know whether it happened too many times. If it happens 10 times to my plant, and everybody else is below three, well, then we have to know about it and do something about it. and this went below the level.

ACTING CHAIRMAN BONACA: Yes, but this committee took exception on that because certain issues were where it didn't make sense, okay? So to

some degree, I think we are converging.

MR. BARANOWSKY: I think we tried to look at your whole argument and see what it was about, rather than just the one sentence, and we tried to address the whole argument.

MR. DUBE: For example, the second bullet on the resolution, the concept of front stop, and later we will talk about back stop, we adapted from the ERISA front tech spec initiative. There the effort is to identify and allow outage time that may result in a delta CDF of 5 to the minus 7.

Of am algorithm and calculation results in a number less than the existing tech spec, and let's say 72 hours, the front stop is that you use 72 hours.

If the algorithm comes out with an allowed outage time or completion time of more than 30 days, and let's say 80 days, the back stop is 30 days.

So the limit, the lower limit if you will, which is the existing allowed outage time, and upper limit, which is 30 days, and the plant specific variation, and the Fussell-Vesely's, and the importance measures, and the plant performance, allows some variation in between.

And in many ways this is how this

project when I took it on had no front stop and no back stop. It was whatever the computer or spread sheet spit out is how many failures one would allow.

What we have done over the last few months is propose an adaption, which says that we are going to have a front stop with a lower limit, and a back stop, which is an upper limit number of failures, and the plant specific variation, the four-diesel configuration versus two diesel-configuration, will allow some variation in between.

MEMBER ROSEN: So you don't penalize people who have better, more robust, designs, by giving them the same text specs, or the same indicators that you give class or less robust designs.

MR. DUBE: Exactly.

MEMBER ROSEN: You get some credit for doing better.

MR. DUBE: Exactly, that is the fundamental purpose that we are proposing.

MR. BARANOWSKY: And we think that is a point that the ACRS made a few years ago, and we followed that --

MEMBER ROSEN: Over and over.

MR. BARANOWSKY: And so help me out.

That's why I say logically inconsistent. That's why we went back to the words and discussion on this issue, and not just the one sentence that said don't use risk.

MEMBER APOSTOLAKIS: You are right, Pat. The ACRS 3 or 4 years ago was not of the opinion that the performance indicators should be strictly performance based. You are absolutely right. We changed our mind on the way. Actually, we formulated an opinion on the way. So your confusion is justified.

MR. BARANOWSKY: We appreciate that.

MEMBER APOSTOLAKIS: And I just say that I, for example, at least am very pleased that you are actually paying attention to what we like.

MR. BARANOWSKY: Very astute.

MR. DUBE: So on that note --

MEMBER APOSTOLAKIS: We have to move on. I am the chairman, and --

MR. DUBE: The next slide shows you -- and I will go over quickly the preliminary results. This is without any changes to the methodology. These were the first results.

Where were the invalid indicators coming and is there a pattern, and lo and behold, one did

find HRS, which is heat removal system, which includes ox feed water for pressurized water reactors or RCIC, or steam driven HPCI.

For boiling water reactors, we found a pattern where one of a kind steam driven ox feed pumps, for example, tended to have or to be more, quote, invalid, than other particular systems, because they had high importance measures.

If there was a failure, you couldn't spread that failure over many like components for the failure rate, because failures are over the number of demands. If you have got two pumps or four diesels, a given failure or one failure over a number of demands, you could spread it out and the failure probability that resulted would be low.

But when there is one of a kind that has high importance measures, they tend to show itself out as an invalid indicator. So this is what we had coming in pretty much in January, and this was the challenge before us.

The insensitive indicator is the opposite. If something has a low importance measure, it is going to take a lot of failures calculationally to exceed 10 to the minus 6.

And it can be 10, and it can be 20, it

could be 80, it could be hundreds even. And it is a result of the fact that we originally came into the MSPI using a deterministic criteria. We wanted to have enough components within a particular system.

We did not want to exclude stuff, because in some cases if you exclude everything with a low importance measure, there would be nothing left in the system.

MEMBER ROSEN: Now you are getting the idea.

MR. DUBE: And it is a result of the fact that we have some low important systems in here, but that was --

MEMBER ROSEN: Well, one design philosophy might be that to build a plant that is so robust that no one component matters much, and tell me what is exactly wrong with that?

MR. DUBE: There is nothing wrong with that. That is a good idea. But going into the program, the program is that you will include emergency A/C power, ox feed, RHR, service water component cooling water.

But some particular plants have such robust cooling water systems, and service water systems, and so plants have like four pumps in unit

one, and four pumps on unit two, and you can cross-tie on train A an train B.

And you can cross-tie across units, and so the particular components aren't going to have low importance measures.

MEMBER ROSEN: Now, this is a good thing, except for the argument that therefore you can take all kinds of failures and the plant can summarily with a completely degraded maintenance program, because it is designed so robust, and that is the back stop.

MR. DUBE: That is exactly right, because otherwise we would have 80 or a hundred -- I mean, there was one calculation, and it was in many significant digits, number of failures to cross, and obviously that is not reasonable. So that's why the back stop comes in.

MEMBER LEITCH: Trying another input into that process is taking a corrective action program, and if you are having that many repetitive failures.

ACTING CHAIRMAN BONACA: But still I think the backstop puts some sense into the -0-

MR. DUBE: Exactly. So I mentioned the 30 days, which is the time of the back stop, and it

is called completion time in the proposed risk-informed tech specs.

The next slide shows you on a first cut, and this is plant specific for San Onofre what exactly I mean. And here is the system on the left, the particular component, the failure modes, the number of failures to reach white, and in the Fussell-Vesely there will be UR, which again is roughly risk achievement worth minus one.

It shows an inverse relationship. The lower Fussell-Vesely over the UR, the more failures to get white. I mean, it is just basic algebra, basic math. The higher Fussell-Vesely over the UR, the lower number of failures.

MEMBER APOSTOLAKIS: Are you saying that Fussell-Vesely and risk are deterministically related? You keep saying that FV over U minus one is wrong.

MR. DUBE: Fussell-Vesely over UR for a low UR is approximately risk achievement minus one algebraically.

MEMBER APOSTOLAKIS: So why aren't we looking at both Fussell-Vesely and UR?

MR. BARANOWSKY: We're not.

MEMBER APOSTOLAKIS: Not here, but the

special treatment requirements staff does, and the argument is that they are independent. I mean, the risk reduction work is related to Fussell-Vesely, but that is a separate thing and now you are saying no.

MR. DUBE: Algebraically you can approximate Fussell-Vesely over UR, is approximately risk achievement worth minus one or a low UR.

MR. YOUNBLOOD: This is Bob Youngblood. Dividing by UR is the critical element.

MEMBER APOSTOLAKIS: I agree.

MEMBER SHACK: For a passive component where the unreliability is zip, the numbers sort of become meaningless.

MEMBER ROSEN: And the risk achievement for those very high reliability components gets to be enormous.

MR. DUBE: So there is an inverse relationship here, and lo and behold, the next transparency, which on your sheets are black and white, and my is colored, and the overhead transparency is colored, shows you the red, which are those which -- well, this is a phenomena that we didn't know how to deal with when it first came out. I will be honest, okay?

We artificially said greater than 20 failures is insensitive. Well, one can even argue something lower, but we have to pick some number, because when we were trying to adjust this model to address invalid indicators and insensitive indicators, and do sensitivity studies, we had to start with something to fine tune it, and so that's what we called it.

But our backstop would not be minus 20, but this gives you an idea of where we were, and it was something like 11 percent of the systems are insensitive.

So we have a number of the systems that are invalid, and a number of systems that are insensitive.

ACTING CHAIRMAN BONACA: It is three o'clock, and why don't we take a break.

(Whereupon, at 3:00 p.m., the meeting was recessed and resumed at 3:21 p.m.)

ACTING CHAIRMAN BONACA: Let's get back to the meeting, and you were I believe at the identification system, page 18.

MEMBER SIEBER: Page 18.

MR. DUBE: Page 18. The next several issues are not maybe as profound as the issue in

value indicators and sensitive indicators, but they were major issues that came out of the workshop.

Something is -- you know, such as identifying the system boundaries, there is a definition in the guidance, but you find what is a train, and it is based on parallel heat exchanges, pumps, and flow path. But there is some different configurations out there that may not fit neatly into those definitions.

So the way that we are resolving this is that we have got a website where we pose and frequently ask questions, and we discuss them in the public meeting. We will revise -- or NEI will revise 99-02 with improved guidance.

And then before final implementation, assuming that this goes forward, there will be at least one, and probably several, lessons learned workshops where these experiences are shared and hopefully in an effort so that the plant implementing this won't have the same issues.

Data collection. For a number of plants, I have had an issue where they had a large number of components that needed to be monitored. On average, the number of components we found is about 50 per plant, which is not an unreasonable

number.

I mean, when you really think of it, the internal events, core damage frequency at the plant, at least for these six systems, can be represented by in large part by 50 components, which tells you something right there.

That much of the risk from active components falls on a small population relatively speaking. But because we had deterministic criteria way back when in the program for identifying whether a component needs to be in scope or not if you will, or monitored, there were some plants that had a large number of valves to monitor, like 35 or 40, or 45, and so there were some concerns with that.

It had been a burden, but the resolution as we are coordinating this with INPO consolidated data entry program, so that licensees will be able to report the data through this mechanism and not have to make a separate report for the MSPI. It will be uploaded and downloaded relatively easily.

MEMBER SIEBER: That is the EPIX.

MR. DUBE: Correct.

MR. BARANOWSKY: The EPIX is a subelement of that whole thing.

MEMBER SIEBER: Right.

MR. BARANOWSKY: It used to be MPRDS, but with this consolidated data entry is going to include several things, like the old monthly operating reports. It will include the actual PI values that are not -- are they coming through NEI? How does that work now? They come through NEI?

MR. SATORIUS: Yes, the come through NEI.

MR. BARANOWSKY: So this CDE would be the place that they would stream into.

MR. SATORIUS: Yes.

MR. DUBE: A second way to handle the number of values and to reduce the number that need to be monitored is kind of a risk-based approach, which I am proposing to use 10 to the minus 6 per year.

And I know that George is going to say, well, you are mixing up the Fussell-Vesely and Bromberg all over again, but it turns out that we looked at Fussell-Vesely over UR as a cut-off means, as well as Bernbaum, and I think Bernbaum is the best, because it has the core damage frequency already impacted into it, and I will show you some slides in a second.

And the third item that I think we need

before we move forward, and this is my recommendation to the industry, I will proffer some software and some interface for data entry, because we did find a number of data entry problems, or I should say the inspectors when they did the inspection found this.

This is a number of valves that need to be monitored, and on the X-axis is the Bernbaum cutoff. In other words, if we set anything with an importance measure below this or not, we are going to exclude, and so the number of valves per plant on the Y-axis, and I have showed you for the 20 pilot plants, red is the plant with the highest number of valves to be monitored.

Blue is average, and I guess black is lowest, and it gives you an idea of how many valves have to be -- and you can see that there is a quick drop for very low Bernbaum, and then it kind of levels off. So this is the benefit if you will of having a cutoff value on Bernbaum for the number of valves to be monitored.

And keep that in the back of your mind, and then look at the next graph, which is what I call the unaccounted for URI, the unaccounted for delta CDF due to unreliability if we were to exclude

particular valves as the Bernbaum goes up.

And you can see that this is kind of the cost if you will, and it starts to take off around 10 to the minus 6. So one could use a 10 to the minus 6 cutoff on Bernbaum, and reduce the number of valves that have to be monitored, particularly for the plants with the most number of valves.

And yet not lose the contribution to the index if you will to any great extent. One could have done this from the start if you will, and perhaps used an importance measure to begin with, but if you carry it to an extreme, you might have some systems with no components in it if you will.

So I think that this is a happy medium that for those few plants that had lots of valves to monitor, you will be able to reduce the number of valves to be monitored by a measurable amount, and not miss important contribution to delta core damage frequency or the index if you will.

MEMBER ROSEN: Do you mean monitor for the purposes of the index program?

MR. DUBE: Exactly.

MEMBER ROSEN: And that these plants will continue to maintain those valves?

MR. DUBE: Right.

MEMBER ROSEN: And to continue to monitor them for maintenance rules and --

MR. DUBE: But valves are particularly difficult to monitor, because you know that when a valve fails, you know it fails. There will be a condition report of some sort.

And pumps, you know, there are graphs, and even computer generated counts on pump starts, and run hours. But valves, most plants don't have little counters that count valve strokes.

And as part of this effort, you need to count demands, as well as failures. So having a lot of valves is kind of a data collection, but this is a reasonable approach that I believe is appropriate.

MEMBER ROSEN: Are you saying that plants are going to install hardware on their valves?

MR. DUBE: No, but they have to estimate the count, and they estimate the count based on how often do they do this surveillance, and how many times on average would they stroke this valve based on normal operations. So a number of these are based on estimates, less the demands.

MEMBER ROSEN: So not hardware?

MR. DUBE: Right. So I think that the

issue of data collection burden in my mind is resolvable relatively easily. The next couple of issues are tough ones, at least maybe conceptual, but let me start with a quote from NUREG CR-6819 and I am sure that there will be lots of opinions because there are some people around this table who have done a lot in common cause.

But in this report, it says approximate causes of CCF events are no different from the approximate causes of single components failures. It is reasonable to postulate that if fewer component failures occur that fewer CCF events would occur.

My opinion of that from my experiences is that the kinds of behavior, either maintenance, procedural, human error, what have you, that may change the independent failure rate and would also lend itself to perhaps change the common cause failure rate.

Now, there is a coupling, and that if there is a change in the independent failure rate, there is in all likelihood a change in the common cause failure.

MEMBER APOSTOLAKIS: Are you saying that your standard model for multiple Greek letter or

whatever, the common cause failure contribution is the independent failure rate times, say, data for two components. Are you saying that a common cause failure term would be affected because of LAN that has been reduced or it will affect data as well?

MR. DUBE: Data may be changing, but it certainly will change the LAN.

MEMBER APOSTOLAKIS: So in which case the term would be reduced.

MR. DUBE: Right.

MEMBER APOSTOLAKIS: So it is through that that there would be a primary reduction?

MR. DUBE: Right.

MEMBER APOSTOLAKIS: But data might considerably change.

MR. DUBE: It does change.

MR. BARANOWSKY: We had data to show that it does change.

MR. DUBE: Exactly, and backing this up with data, we actually looked at a number of components, and the common cause error rate has decreased tremendously over the last decade or 15 years, and the single failure rate has gone down.

And in fact almost parallel, which kind of indicates as you said the coupling factor, which

is the data or -- over time may be changing, but it has been changing less.

MEMBER APOSTOLAKIS: It is not being inconsistent with the prevailing view that because of this major effort that was sponsored by the NRC and EPRI, or the NRC anyway, that people became more aware of the issue of common cause failures, and so they have paid more attention to the coupling factor, and they have reduced it. The coupling really itself has been reduced.

MR. DUBE: The coupling has gone down, but not as much as the overall failure rate.

MEMBER ROSEN: What's driving the fact is that these two things go together and what is driving that is improved management, safety culture if you will.

MEMBER APOSTOLAKIS: Yes.

MEMBER ROSEN: And they go together.

MR. DUBE: Right.

MEMBER APOSTOLAKIS: Yes, but the point is that Don is making is that the primary driver is the independent --

MR. DUBE: Well, they go together.

MEMBER ROSEN: NO, George that is the mathematical model. The primary driver is the guys

who work to keep it from going down. Better training, better procedures, better --

MEMBER APOSTOLAKIS: But I would agree that it is really the better training and better procedures that influences the coupling. The coupling itself.

MEMBER ROSEN: I don't think anybody knows that the coupling is there. I mean, the valves don't know the coupling is there, and whether somebody comes out and maintains it.

MR. DUBE: The coupling is something that I sometimes say to myself that statisticians calculate from the data, because there seems to be a correlation, but I am not sure people in the field are thinking, oh, .3 factor or so, and I would not get a independent failure because I might increase the common cause.

But I think that the change for the same reasons, the same improvements in maintenance practices, and procedures, and so forth. I only bring this up, because it is an important issue.

It is an issue of controversy and an issue of differences of opinion, and my second bullet says should not changes in CDF relating to changes in plant specific unreliability from single

component failures also include the effect from changes in the common cause factor rate given this coupling factor.

And if the answer to that is yes, it is a loaded question, then we need to add in the Fussell-Vesely importance from common cause for a particular common component type into the overall expression.

MEMBER APOSTOLAKIS: I have to problems. If I didn't do what you are suggesting, what would I do? Would I consider only the independent failure, the product of the failures?

MR. DUBE: The importance measure that is used in the algorithm as it is currently formulated would be just the Fussell-Vesely from independent failure of that pump.

Whereas, included in the common cause contribution would say if you change the independent failure rate and the common cause failure rate changes, and I need to capture that contribution in the Fussell-Vesely that I use in the algorithm, and the best way to show it may be to jump ahead, and clearly it has an impact on the algorithm and the index.

The screen shows this better since it is

in color, but this is the failure rate on the X-axis, failure to start. These are kind of high numbers, but just look at the concept here.

The bottom is if I just varied the independent failure rate, and how does the delta CDF calculated by the algorithm change? That is the thing that is either blue or green, or the black line on the bottom.

If I include the contribution of common cause to the Fussell-Vesely, and that as the single failure rate changes through a coupling, the common cause failure rate changes. The red shows how that affects the overall quantification.

What it means is that in practice it means that it takes somewhat fewer failures to cross the yellow white threshold in this particular case.

MEMBER APOSTOLAKIS: So right now the computer programs don't do this?

MR. DUBE: The current MSPI method is silent, is mute, on how to treat common cause, the contribution of common cause.

MEMBER APOSTOLAKIS: And what is the Fussell-Vesely importance of a component? In the calculations, it will not include the common cause failure term?

MR. DUBE: The common cause will have its own Fussell-Vesely.

MEMBER APOSTOLAKIS: As a separate --

MR. DUBE: Separately.

MR. BARANOWSKY: It may. Some will do it the other way.

MR. EIDE: Steve Eide, INEEL. The SPAR model, if you get a Fussell-Vesely for the independent failure, and you get a common cause event, that would be another Fussell-Vesely and they are not tied together in the Fussell-Vesely calculation.

MR. DUBE: Right, but you can get a group Fussell-Vesely in that, right?

MR. EIDE: Yes, you can get around that by selecting both (inaudible) common cause event, and doing a group Fussell-Vesely for that, and getting a single or combined Fussell-Vesely for that component group.

MEMBER SHACK: Well, as a performance measure, you could probably do without it based on your arguments, and if you insist on applying it to risk, you need to include it, right?

MR. BARANOWSKY: And this will reflect like if you have four pumps, three pumps, two pumps,

this will be significant.

MR. DUBE: Yes, it is an adjustment on how we or what we use for the Fussell-Vesely. Absolutely.

MEMBER APOSTOLAKIS: But this is important also for other interaction, special trip.

MEMBER SHACK: Well, yes, that has been discussed.

MR. DUBE: We derived the same issue separately.

MEMBER ROSEN: The Fussell-Vesely treatment as I recall was handled by sensitivity analyses, and it was shown that the impact of common cause was looked at through a set of sensitivity analyses.

MR. DUBE: Yes. This is one sensitivity here. We have just in the last days literally looked at some of our pilot plants the impact, and in some cases it may be a few percent, and in other cases it may be tens of percent or even more, depending on the configuration.

A 2 out of 2 situation, or in other words, two diesel generator plants, and adding in this Fussell-Vesely from common cause may not be a big adjustment. But a highly redundant plant, where

the Fussell-Vesely is very low because of the multiple density, adding in the common cause may increase that by factors of 2, 4, 5, even 10.

The thing is that 10 times is a small number, and still is not an unreasonable number. But unless --

MEMBER ROSEN: But it s a real reflection of the consideration, and if you have two, that's good, and if you have four is better, then why not 10? Well, obviously that is crossed in complexity, and you don't get the benefit is the common cause. It cuts it off.

So this has the effect of reflecting that in the analysis and in the indicator.

MR. DUBE: That is my opinion, and I think the opinion of the technical team on this.

MEMBER APOSTOLAKIS: But it would be a simple matter to find the importance for components if you have an expression from the common cause failure term that --

MR. DUBE: Yes, in practice.

MR. BARANOWSKY: Not everybody has it like that.

MR. DUBE: But in practice we think there is a way of doing it, and for the licensees to

do it is a simple mathematical approach. That is not the issue. The issue is the principal of whether we should include it or not. That is where there is no agreement.

MEMBER APOSTOLAKIS: That is why I made the comment earlier that sometimes what the code does is a boundary condition. This is the way that it should be done.

MR. BARANOWSKY: Well, this is the position that we are proposing, and we are telling you why.

MR. DUBE: And probably by the next --

MEMBER APOSTOLAKIS: Excuse me, but I am just curious, but if you have two redundant valves, and each one has a failure probability of Q, the independent failure term would be Q squared would it not?

MR. DUBE: Unavailability?

MEMBER APOSTOLAKIS: Yes.

MR. BARANOWSKY: Unreliability.

MEMBER ROSEN: You are talking about independent failure probability.

MR. DUBE: Unavailability or unreliability?

MEMBER APOSTOLAKIS: Unavailability.

MR. DUBE: Unavailability, you would probably --

MEMBER APOSTOLAKIS: Unreliability, 2 squared.

MR. DUBE: Yes.

MEMBER APOSTOLAKIS: So what is the Fussell-Vesely event? What is the Fussell-Vesely of the component since you have a square term?

MR. DUBE: There is no easy way. When something is ended, there is no easy way to calculate Fussell-Vesely right off the top of your head. If they were orange you could. The computer would tell you what it is.

MEMBER APOSTOLAKIS: Yes, it would tell you what the importance of Q squared is, but when it calculates the importance of the component, a single component, how would you do that?

MR. DUBE: A single component?

MEMBER APOSTOLAKIS: Yes, from a single valve.

MR. DUBE: It adds up all the sequence, and all the cut sets with that component, and shows the ID that you use for it, the basic event name. It divides that by core damage frequency.

MEMBER APOSTOLAKIS: Even though some

terms are squares and some are --

MR. BARANOWSKY: But they don't show up --

MR. DUBE: They don't show up squared though. X-1 times X-2. It won't show up as X-1 squared, right. It will show up as X-1 times X-2, even though X-1 and X-2 may be the same number.

Anyway, I think this is resolvable, and I think it is important, but it has been a difficult issue. The next one has to do with the support system initiated to Fussell-Vesely.

Again, 80 percent of this discussion hinges on this Fussell-Vesely and I am sure that Bill Vesely, when this term was named after him, didn't realize that it would be used in so many different ways.

But the algorithm depends on this particular measure, and that's why I put so much emphasis throughout this project that we have got to calculate this number pretty accurately because the approach depends on it, at least to a first order.

So the issue here is that the failures of components leading to support system initiator -- and, for example, loss of service water -- contribute to core damage frequency.

And when we looked at the pilot plants, about two-thirds of those 20 plants used fault trees to quantify the initiating event frequency. So if you had a loss of service water at the top of the default tree, it would have, you know, if Pump A fails, and Pump B fails, and Pump C fails, and so on and so forth.

And to the extent that they had a fault tree when the computer calculated Fussell-Vesely, it captured that contribution to the initiator, as well as a support system. But one-third, the remainder of the plants, used a point estimate.

Instead of using a fault tree, they just used a number of 10 to the minus 3 per year initiating event frequency. So that 10 to the minus 3 did not have the constituents that made it up, such as this pump failing and this pump failing.

So in the pilot program, it identified an inconsistent approach and it hinted that there might be come contribution to Fussell-Vesely left out. And so we have come up with a logical approach to address it.

For those models using point estimates, that the contribution of the initiator to core damage frequency is significant, either A, add the

support system initiator fault tree.

So we are going to have to take this point estimate and create a little fault tree, or we have come up with an adjustment factor and it is a little bit beyond -- you know, we could spend an hour on it, but an adjustment factor that will be a little bit conservative, but both myself and representatives from industry agree that it is a reasonable approach to make sure that the support system captures all of the Fussell-Vesely, and that is a long story being short on that.

The final issue, SDP and MSPI, and we kind of talked about it at the beginning and I am going to hand the baton over to Pat, because he has been following a lot of these issues.

And this is one of the final technical issues, which has to do with have we thought about this MSPI versus SDP, and when is one going to be used instead of the other. You know, what were their original purposes for, and what are the aspects of implementing it.

MR. BARANOWSKY: Actually, I will just finish off the rest of the discussion. Let's go back a little bit in history so I can tell you a few things. When SECY 99-007 was put out, it had in

there the concept that performance indicators would be used as the principal measure of performance when they were available, and when they were not available, a risk informed inspection program and significance of inspection findings would be used.

Somewhere along the way both got implemented on the same things, and it was if I don't get you here, I have got you there kind of system, which is currently in practice.

And that has been going on now for several years. That is whatever is in the SECY is not the way the program is being implemented for whatever reasons. Well, because of these concerns that we have about some of the false positives in particular, wherein one failure of the diesel generator, when you look at a short time frame, like one year, one year you might have 12 tests.

And due to the unavailability associated with that in a one year time frame, it is pretty high and you could end up with a short term risk that is on the order of 10 to the minus 6.

But if the diesel generator was surveiled for 3, 4, or 5 years, you have a track record that when failure has a different implication. So for cases where the MSPI and the

SDP differ, what we are trying to do -- and we are not done with this, but we have done some looking at this, is to see if they differ because of that kind of a premise, or some other reason.

And if they differ, it doesn't make any sense. Now, as it turns out there is actually only a few cases in the historical record where the original formulation that didn't have all of these issues addressed that Don talked about today with the original formulation, showed that a different outcome of SDP versus MSPI.

The second thing that we want to look at also is whether or not the SDP and the ASP analyses were giving similar results, because in many cases the SDP is done with the simplest technique possible, and when there is a performance issue, and folks agree that there is a performance issue, and want to move on and fix it, and not worry too much about spending a lot of time doing risk calculations.

On the other hand, we know that the MSPI and the ASP analyses are trying to spend more time on the details of the risk analysis, as opposed to a fairly short handbook kind of thing.

So what we are proposing is that we go

back and look at these things, and make sure that the validity and the appropriateness of using the MSPI in those two cases, whether our differences are such.

Our best cut right now is that they are a small, small percentage of any of the -- what I call non-green findings, whether they be by PIs or by inspection activities.

So there is just a small interface where you might get a slightly different result. Now a concern also would be that, well, gee, what does that mean. Well, that means that you think about things a little differently, just like when we didn't have the reactor oversight process, and we had SALP.

We made some findings which if you went back and overlayed the reactor oversight process approach on it, you wouldn't necessarily come up with the same findings.

Sometimes we error on this side or that side, and it is not a very super precise thing, but we think we get the really significantly poor performers in each case.

And so that is our plan, is to basically document that and present the arguments as to why at

least on the face of the outcomes that it makes sense to use a performance indicator of this type, versus the significance determination process.

There was an ACRS letter written not too long ago that basically said the significance determination process was not a good way to measure performance, but it was a good way to assess the risk significance of performance findings.

And by the way, I completely agree with that. I think it is also the only thing that we have for rare events where you can't get a string of things that you can put into a performance measure that accumulates a performance if you will to look at trends.

In that case, when there are rare events that are outside what would be expected, and you would not call them false positives -- and an example would be that you have had a LOCA.

You don't expect a LOCA in a frequency in the plant, and so that when that occurs, it is kind of outside the norms. And I think it is fair to use risk at that point, or some common cause failures which occur very rarely by the way.

You know, there are not very many common cause failures where multiple components actually

fail at plants. So you can't really trend within a plant common cause failures very well.

You can look at the whole industry over a picture period of 20 years, which we have done, but that is with a hundred plants going all the time. There is no other country that can do that besides the U.S. They just don't have the data.

So it is for those cases where we think the SDP is a good measure of some sort that it should be used, and that is for inspection findings, these long duration outages that are not captured by routine tests and so forth.

And that the PIs, where there is an accumulation of performance information, such that one would compute reliability and unavailability accumulated over time to look at trends, that is a place where the MSPI is best used.

So this is sort of a philosophy that we are overlaying on top of a practical look at what the outcomes are, and there is a little bit of heartburn to be honest with you with some of the region folks who want to use an SDP evaluation, period, for everything, and we just need to work through this issue.

And we will present the results of our

detailed look at the individual instances when we do our final report.

MEMBER APOSTOLAKIS: What do you mean by --

MR. BARANOWSKY: By what?

MEMBER APOSTOLAKIS: You said they want to use an SDP evaluation for everything. What exactly does that mean?

MR. BARANOWSKY: For every item that could go into a performance indicator where a performance issue was identified. So if a valve failed and there was a performance issue identified with the failure of the valve, then there are some people who want to run an SDP on that every time, even if we are tracking valve performance using reliability and availability indicators.

MEMBER SIEBER: On the other hand that is mostly instigated by licensees, who say I am agreeing to a white threshold, but my calculation shows that it is not risk significant. So they ask for the SDP.

MR. BARANOWSKY: I think it goes both ways, but in most instances it is not valves as much as it is maybe a diesel generator, because the SDP looks at a one year period of time remember, and the

performance indicators cover a 3 year period of time.

Now, maybe that is an issue that ought to be looked at to see if they should have the same period of time.

In that case, you would be surprised at how much closer they could eventually come to the same outcomes.

MEMBER SIEBER: Actually, the SDP looks at the event, and then says if you come up with a failure it is probably unavailable half of the time since the last test. So it is less than a year and it sort of elevates the importance of that single event, compared to what it would have been averaged in over 3 years worth of data. That's what I think.

MR. BARANOWSKY: I think it gives a risk significance and importance of that finding or that incident. I mean, I wouldn't say negative things about the process because it is modeled after the accident sequence precursor program, which does the same thing.

But what we don't do with the accident sequence precursor program is take a single accident sequence precursor and go, oops, we had a major failure in poor regulation last year because we had

a precursor.

And then when we don't have one the next year, we go but we did very good the following hear. No, we have to take a string of these things, because we know that looking at one of them can't give you a performance trend.

And it is the same problem that goes with trying to use the significance determination process for things where you can have a string of issues, and look at them because the interval is such that you could have more than one hit if you will in that time frame.

MEMBER SIEBER: Let's say you were -- it seems to me that if you end up with an inspection finding that you should go to the SDP, as opposed to, say, that this modifies the performance index or this component.

On the other hand, if it is revealed through the performance index, you ought to use the thresholds that are appropriate to the performance index.

MR. BARANOWSKY: I think we are saying that.

MEMBER SIEBER: Well, it is not clear, because I think that you could run them one way or

another, depending on -- you know, if you are writing them, and you say this is the way it is going to be, then it will be that way.

On the other hand, if you have a choice, people will make the choice that causes the least amount of grief.

MR. DUBE: We agreed ahead of time on this.

MR. BARANOWSKY: And we do agree on this issue, and so what I am doing is giving you our tentative conclusion just based on some logical thinking and looking at the differences in these things. Now also to help me talk about this is Mark Satorius.

MR. SATORIUS: I am Mark Satorius with the staff. I was just going to point out that most of the examples that Pat is talking about are event driven, where we have an event response, and we do an inspection, and the result of that is that you can't know what you know.

So you identify certain performance issues during these event responses inspection, and those are relatively limited. But those are the ones I think that -- and wouldn't you agree, Pat, that is where you are going to get this overlap more

than anything else?

MR. BARANOWSKY: Yes. But there has been some concern because if you talk about a diesel generator failure and you look at the unavailability of that failure over a one year period of time, that is going to give you a different perspective than if you look at it over 3 years.

And what we were talking about earlier were this one-half lambda tau term in terms of unreliability is not going to be equal to the probability of a failure on demand. You have to have a time period sufficiently wrong and T has to go to infinity for observations in order for those two to be equal.

I guess anyone who has done any reliability 101 or whatever has derived that equation.

MEMBER APOSTOLAKIS: But I like your third bullet, which says PIs measures changes in performance. Now, what you mean from what you said is that the performance of this valve. You are not comparing with peers, right?

MR. BARANOWSKY: You are comparing with what?

MEMBER APOSTOLAKIS: With PI

performance. I mean, the same valve somewhere else?

MR. BARANOWSKY: No.

MEMBER APOSTOLAKIS: But still though if the PIs measure changes in performance, then it seems to me naturally that the thresholds should be performance based and not risk.

The risk calculations that you are doing can be a valuable input to the process of developing the performance based thresholds, but I agree with you that PIs measure changes in performance.

So the PIs and the SDPs are doing two different things.

MR. BARANOWSKY: Well, I would ask that you have an open mind in that we think that we have come up with a blend on here, and just take a step back.

MEMBER APOSTOLAKIS: They both contribute to the decision, that's true.

MR. BARANOWSKY: And I think that you will see that it is one way as Don said, you have got these front and back stops, and you have to figure out where you put these things. And you adjust them or some things based on how risk comes into the picture.

MEMBER APOSTOLAKIS: Yes.

MR. BARANOWSKY: But we just don't go with risk which allows a way, way wider span, okay?

MEMBER SIEBER: Well, you have the front stop and you have the back stop, and the adjustment factor in the middle.

MR. DUBE: Exactly.

MEMBER SIEBER: I think you scratch your head and is this really real, you know.

MR. BARANOWSKY: One of the things that I didn't mention at the beginning of the talk was that the white paper by the way, and which I did say was sort of a kick-off concept, we are going to document every one of these things in a written report.

And that will be coming in the fall before we ask to have the next meeting. So you will have a fair amount of time to see this stuff laid out a little bit more than just a few viewgraphs.

MR. DUBE: In following up on what George said on the issue, the alternative to having an algorithm if you will, and which basically calculates the number of failures and the unavailability to the threshold would be a multi-dimensional, big super matrix that says BWR-2 plant, and BWR-3 plant, combustion engineering plant with

pores, and combustion engineering plant with no pores. How many failures and it takes into account all the differences in design, and the differences in performance.

I mean, if one were to address plant specific aspects to the threshold, and even then I am not sure that expert judgment would come up with the right answer, because you have to take into account the variability of the design, the vintage of the plant, and so forth.

What this does is reduce this multi-dimensional matrix of thresholds to an algorithm that in essence calculates what that threshold is, but within certain limits. I mean, that is the way that I kind of view it, and that's it.

MEMBER SIEBER: Well, the first hurdle is to understand what it is that yo have done, and the second hurdle is to decide whether it meets the need or not.

MR. BARANOWSKY: The whole activity that we have gone through is fairly complex, and I am not saying that it isn't, because we have invested things that people had not thought of 2 years ago when we thought we knew quite a bit about risk-based performance indicators.

But what we have come up with is a number of simplifying approaches to address all the complexities which are not as simple as adding one and one, but they are not as complex as doing an ECCS calculation either.

So you don't have to redo your PRA, and you don't have to redo your HRA, and you don't have to do any of that stuff, as long as the concepts -- front stops and back stops -- using the importance measures in a simple equation, those things are all brought into it, and it is a pretty straightforward and grind it out.

So let me just summarize here, because then I want to just about some future activities. So the MSPI, as you have seen, it is highly risk informed, and it has plant specific design and plant specific data. We think that these maximum-minimum limits are a pretty big deal on making it kind of rational.

MEMBER APOSTOLAKIS: But are they any different from the performance based thresholds that we have requested? They are the same thing aren't they?

MR. BARANOWSKY: I am saying that I think that this is consistent with the detailed

discussions that I read that are behind the specific -- like one sentence position that you took to be honest with you.

MEMBER APOSTOLAKIS: Yes, and so that is what I am saying. That we are consistent.

MR. BARANOWSKY: I think so.

MEMBER ROSEN: For example, when we had a hard time with 23 SCRAMS, this deals with that.

MR. BARANOWSKY: This deals with it.

MEMBER APOSTOLAKIS: And then you would say there is a maximum.

MR. BARANOWSKY: Not for 23 SCRAMS it doesn't deal with it, but it could deal with it. And we have lots of technical --

MEMBER ROSEN: Well, if we ever went to this, that action matrix thing would not show this, right, with 23 SCRAMS?

MEMBER APOSTOLAKIS: It would not show the 23, no, because those guys would intervene and put a back stop.

MEMBER SIEBER: That would be a back stop.

MEMBER ROSEN: Back stop, okay.

MEMBER SIEBER: And then you would put in adjustment factors.

MEMBER APOSTOLAKIS: Are you suggesting, Mr. Sieber that we should do everything on a risk basis?

MEMBER SIEBER: No, I'm not.

MEMBER APOSTOLAKIS: Better me than you.

MR. BARANOWSKY: We expect to complete our analysis and then simulation analyses, and complete our analysis of these issues, and then do some simulations by the end of the summer to look at how all of these things fit together, because we have not really looked at them all together. So we need to do that.

And then if some new issues arise, we will address them, but we are fairly confident on a technical basis that if you will accept some of the philosophical thinking that went in here, we can probably address any residual things that might pop up in that regard.

So that is sort of the technical bottom line here. Now, that does not address all things regarding implementation, although this says tentative implementation schedule, let's look at a few things that are not really covered here that are also implementation related.

First of all, we are going to do the

technical work as a I stated, and at the end of August we will have the technical issues done, and we have done all the SPAR enhancements for the 20 plants that are in the pilot program, and it will run all the things.

The pilot actually ends in September. It gives us a certain amount of time afterward the data collection to complete analysis and evaluation of the data. Then comes the big effort to see, well, what does this all mean in terms of the success criteria which Mark Satorius mentioned in a few items earlier?

Will it complete our table top analysis of the MSPI and the SDP issues and other implementation issues such as we will ask ourselves are we able to change the guidance for boundaries and data collection to eliminate some of the inefficiencies that occurred during the pilot.

Are we able to change the inspection guidance to eliminate some of the inefficiencies that occurred during the pilot. And I don't know if there are other issues, but we are going to have to work on guidance and what the costs in terms of burden of this thing is.

And I am sure that technically that this

is far superior to the current PI, but one has to make a decision as to whether or not implementation wise we are ready for it. We have issues regarding the PRAs and so forth, which we are not going to use this program as a wedge to go in and make the ASME standard work.

MEMBER SIEBER: Why not?

MR. BARANOWSKY: Because we have already got a cadre of people doing that, and I don't want to get them unemployed. But there are some insights that we have here that I think affect the bottom line of PRA and the qualitative outcomes of PRA that might help focus some of these things.

MEMBER SIEBER: I think that you are addressing things in your comparisons that the standard doesn't really deal with.

MR. BARANOWSKY: Well, they are in there inherently, but they should be explicitly in my opinion.

MEMBER SIEBER: I think if you follow the standards that you may end up with a high quality PRA, and then whatever discrepancies would in fact be (inaudible), but that is not consistently the case.

I could imagine now that you could go

through these standards, the PRA reviews, or peer reviews, and come out maybe okay, and still have some strictly inherent problems in your PRA. So this is just another way to look at that and I think it is good information.

MR. BARANOWSKY: Well, we are counting on the industry and the NRC working to implement the standard. That is an assumption that we have. It doesn't necessarily have to be all perfectly done in the beginning, because we have ways of identifying which plants we have the biggest questions about as I told you earlier.

MEMBER SIEBER: Well, you are going to reconcile the SPARs models to the plant PRAs anyway.

MR. BARANOWSKY: Yes.

MEMBER SIEBER: And that appears in a number of programs that once you do that, and you can rely on yours, or you can rely on theirs, provided that you know what the limitations are for each calculation model for the intended purpose.

MR. BARANOWSKY: I think ours are good for doing audit checks and for doing simulations to look at issues. They are really excellent for that.

MEMBER SIEBER: Let me ask a question. You have a schedule of things that you are going to

do, and so one of the things that the white paper talks about is a revision to NEI 99-02, which probably isn't out yet, but needs to come out in order for this to be a complete packet, and when will that happen? Maybe Tom, if he is still here, could tell us.

MR. BARANOWSKY: And Tom, when you get up to talk about that, as long as you are getting up, I had expressed an earlier interest in the industry's reaction to this. And maybe you could cover that as well.

MR. HOUGHTON: Sure. I am Tom Houghton from NEI. The NEI guidance document will be out in a draft a couple of weeks after we have decided these issues. According to that schedule that earlier fall effort before the go-no go is going to include industry also going through and putting all of these changes into the models and see what the results are that come out.

Because as Pat said, we don't know what the cumulative effect of all of these different activities that we are doing will have. But we think that in a couple of weeks after that these things will be wrapped up.

MEMBER SIEBER: You mean that fast?

Well, that's good.

MR. HOUGHTON: And then need to be out so that people can really focus on --

MR. SATORIUS: But the draft is already out because we used it to run the pilot. So I was not sure if that was clear.

MEMBER SIEBER: That's right. That is REV-2 isn't it?

MR. SATORIUS: Well, no, this was -- what we did was that we pulled the format directly from REV-2 and then modeled it specifically for the pilot. So we already have a document that we are working on.

MR. HOUGHTON: That's right. The section of 99-2 REV-2, which relates to mitigation systems, is what is going to be replaced with this MSPI, and that that draft that Mark was talking about is that placement.

We didn't change it during the pilot because we didn't want to confuse everybody who was trying to report data for the 6 months of the pilot. So it stayed fix until we make these decisions, and then we will implement them into the document.

As far as the program is concerned, industry supports this program. We think it has the

advantages of resolving some of the complexity between maintenance rule definitions and WANO definitions.

We think that it is going to resolve some of the complications for system managers in trying to determine cascading, which we won't be doing anymore. That will make it more consistent with the way that people do maintenance rules.

We won't have this question of fault exposure and you get into lots of theoretically fault exposure, which is not an issue, but when you get into questions of would that failure mechanism reveal itself in a monthly test or an annual test, it gets quite confusing sometimes.

And that makes it very difficult. It is not the theory that is the problem. It is the issue of would this failure mechanism be exhibited in a test that is only an hour long, versus a full 24 hour run. Things like that.

MEMBER APOSTOLAKIS: In this business of one-half lambda tau, I thought that was used as a means also of seeing what the impact on the unreliability would be if I changed the inspection interval.

MR. BARANOWSKY: That's with the

assumption that you have a long time period to make the calculation of lambda. If you calculate lambda every year using one year's worth of data, one year you get a low lambda, because you had no failures.

The next year you get a huge lambda because you had a failure. Then the next year you had no failures and you get a low lambda.

MEMBER APOSTOLAKIS: Well, 3 years, and I use 4 years, 5 years. I have a lot.

MR. BARANOWSKY: Then it becomes equivalent to basically the probability of failure on demand. They start to equate to each other.

MEMBER SIEBER: The longer --

MR. DUBE: That's right.

MR. BARANOWSKY: Well, one-half lambda tau becomes equal to the probability of failure on demand.

MEMBER SIEBER: The longer the period, the less significant is a single failure.

MR. DUBE: That's right.

MEMBER SIEBER: So there has to be a limit.

MR. BARANOWSKY: That's an approximation for small lambda, constant lambda, integrated, zero to infinity.

MEMBER APOSTOLAKIS: Let's say I am doing this on a monthly interval, and then someone who wants to successfully argue or wants to argue successfully that they can go to two months, are you saying then that your data then are invalid?

MR. BARANOWSKY: No.

MEMBER APOSTOLAKIS: Because they were connected under conditions of only one month?

MR. BARANOWSKY: No, what we do is just keep counting the demands.

MEMBER APOSTOLAKIS: But your basic calculation now is that you have an entirely new situation. I mean, you collect the data and you formulated a distribution that was based on the fundamental assumption of monthly tests, and I am telling you that I am going to do them every quarter. Can you really use that distribution again and start updating it with the new data?

MR. BARANOWSKY: I don't think that the days distribution are that sensitive. Remember, the ones that we are using are based on industry information and updated with plant specific. So what that means is that we have got weekly, monthly, quarterly stuff all mixed in there.

MEMBER APOSTOLAKIS: I am not saying

that what you are doing is wrong. I am just trying to figure out all the implications with the one-half lambda tau.

MR. HOUGHTON: Well, we would have a mixture until such time that the sliding 3 year average moves over and it would be a little inconsistent.

MEMBER APOSTOLAKIS: Well, you are producing a probability of failure per demand, and that is independent of time, correct?

MR. HOUGHTON: Yes.

MEMBER APOSTOLAKIS: It was developed under the assumption -- well, not assumption. It was reality that the tests are monthly. And if I change the interval and make it quarterly, do I start from scratch, or do I start from somewhere else?

At least with the one-half lambda tau, I had a way of going out and changing them to 3 months, and coming back and saying, yes -- and which is also stupid to say that their unavailability is multiplied by three.

MR. BARANOWSKY: No, but you are making the assumption that --

MEMBER APOSTOLAKIS: That it is

different, which it may not be.

MR. BARANOWSKY: The assumption there is that there is not a demand dependent element to the failure rate.

MR. DUBE: And it is really that it has both, has both components in reality.

MR. BARANOWSKY: In reality, we know from actually taking data in several studies that have shown there are both elements that are in there.

MEMBER APOSTOLAKIS: But you have not taken data because it probably does not exist, but from one licensee who does it every three months, and another one who does it every month, and compare them and say there is no difference, because everybody does it monthly.

MR. BARANOWSKY: But what we had to do was take a licensee who does a monthly test, and another one who does it quarterly, and another one who does it weekly, --

MEMBER APOSTOLAKIS: The same test?

MR. BARANOWSKY: For the same equipment.

MEMBER APOSTOLAKIS: But the same test.

MEMBER APOSTOLAKIS: The same test.

MR. BARANOWSKY: But as close as we can

come up with, and looked at that.

MEMBER APOSTOLAKIS: But then if you see no difference, then I agree.

MR. BARANOWSKY: I am not saying that there has been no difference. What we are saying is that the true novel is one that has got a demand dependent element to it, as well as a time dependent element to it.

And when you plot a curve, you get a linear curve on these things usually, and you can come up with a proportionality factor that relates demands versus the run, versus the time dependent failure mechanisms. It is pretty complicated.

By the way, it is the second order effect in risk for most of these intervals that we are talking about, which I am not worried about in light of other inaccuracies. And we are talking a second or third decimal place of the risk equation.

MR. HOUGHTON: In terms of complexity, we think that we are making sausage right now in trying to develop this indicator. And there is complexity in it, but we think that it is going to be simpler when the program is in place and will be simpler for the utilities, because they are going to be just reporting demands, failures, and hours at

power when that equipment is unavailable.

And so once the computer algorithm is set up, that data can flow in without a lot of complication and a lot less what if's by the system engineer.

MEMBER SIEBER: And they don't even have to understand it.

MR. HOUGHTON: Another part of the complexity has been as Don and Pat have said, is what are the system boundaries and what are the components that are active and so on and so forth.

And we see that being able to be resolved through the other 80 units in a good change management plan, where there are a series of workshops, and where people can get together with what we learned from the process and develop those such that when the whistle blows to start the program that we don't have a lot of discussions about why is this valve in and why is that valve not in, and why didn't you model this, and why this or that.

That can be fixed so that this turning in complexity doesn't have to happen when it is implemented. So those are reasons why we think this is a better way to go and I think the only real

disadvantage from our point of view is the initial gathering of data.

Other than that, we think it is a better indicator, and it has less opportunities for arguing about when did it really fail, and what was going on. So we are in favor of it and in favor of the approaches that Don and Pat are talking about, about a front stop and a back stop, which will solve those problems.

And we were able to agree on what a reasonable number of SCRAMs was in the first one.

MEMBER SIEBER: 27.

MR. HOUGHTON: Three.

MEMBER APOSTOLAKIS: The green and white.

MR. HOUGHTON: The green and white.

MEMBER APOSTOLAKIS: I still think that expert judgment should play a role there and we have pioneered all these methods and don't just negotiate.

MEMBER ROSEN: That's what it was, expert judgment.

MEMBER SIEBER: It should be between smart people.

MEMBER ROSEN: Informed people.

MR. HOUGHTON: And we did have a rule where we tried to use the 95th percentile of performance.

MR. BARANOWSKY: Good. Thank you.

MEMBER SIEBER: Thank you.

MEMBER APOSTOLAKIS: So when is the letter going to be?

MR. BARANOWSKY: We are going to get that and get revised NEI guidance, and a report on all of this technical work, and the assessment of the success criteria. That will all be done in the fall. I think that is the package that comes here.

And then after that we have an ACRS meeting to go over and explain what we decided to do and see if you endorse that. Then we want a letter.

MEMBER APOSTOLAKIS: Is there going to be another subcommittee meeting or just straight to the full committee?

MR. BARANOWSKY: What do you think?

MEMBER SIEBER: If the documents are clear enough, I don't think we would need to have another subcommittee meeting unless you have changed the principles that you are going to use.

MR. DUBE: I don't think the principles have changed. Some of the details will.

ACTING CHAIRMAN BONACA: Yes, that's right. One of the committees, the fact is one of the discrepancies is being resolved on this particular performance --

MEMBER SIEBER: Well, I am sure that you are going to make a presentation to the full committee of an hour or two.

MR. BARANOWSKY: Yes, but it won't be going into detail like we did here.

MEMBER SIEBER: Right.

MR. BARANOWSKY: I mean, you will have to accept the report as giving you that information.

MEMBER ROSEN: But our own staff can do it once they get the package, and they can look at that and see what was said in prior letters, and help us understand whether this has been responsive to our points of view.

MR. BARANOWSKY: And if someone will feed back to us issues that you would like for us to cover as a result of that, we can make sure that those are in our presentation.

MEMBER APOSTOLAKIS: So how long will we have the package before the full committee?

MR. BARANOWSKY: Oh, quite a while.

MEMBER APOSTOLAKIS: Okay. Good.

That's good.

MS. WESTON: It has to be at least 30 days.

MR. BARANOWSKY: It will be more than 30, I'm sure.

MEMBER APOSTOLAKIS: Well, we don't meet in January anyway.

MR. BARANOWSKY: I think we were really thinking in February.

ACTING CHAIRMAN BONACA: That's it?

MR. BARANOWSKY: That's it.

ACTING CHAIRMAN BONACA: Any other comments or questions? One question I had was regarding this firewall. You did by the (inaudible) and you did get a lot of lessons learned, and many of them I am sure are just the plant specific, and adjustments that you had to make and too much plant specific PRAs, or vice versa in some cases.

In some cases, you must have learned some lessons that can be reflected on the other SPAR models. Are you going to have a lessons learned about it?

MR. BARANOWSKY: We have lessons learned, and maybe Pat O'Reilly, who actually runs the SPAR model development program can tell us how

he would use this information to go through the rest of the SPAR models.

MR. O'REILLY: I am Pat O'Reilly from the Office of Research. As Pat pointed out we are doing the 11 plants, the 20 units that are in the pilot program, and based on that and what we have already learned from our on-site QA reviews at every plant site, we have a number of issues which we know apply across the board as SPARs that are standardized, and that this the important thing.

And so we know from this pilot program that Harrison exercised that there are a number of issues that will be implemented across all the PWR models, for instance, and BWR models, and in some cases across all 72 models. So we have learned enough from that so that we don't have to go through and do a detailed comparison exercise for the other 61 models that aren't included here.

MEMBER SIEBER: Does every plant have a PRA that is suitable for this comparison?

MR. O'REILLY: You find a wide spectrum of PRAs out there, some which are very well done, very robust, very complete, and others which are about mediocre. They have some information, and there are some that just had the minimum that were

Generic Letter 88-20. They stuck to the letter of the law there.

MEMBER SIEBER: And in which percentage would that minimum set be?

MR. O'REILLY: That is a tough question, because some of our visits are complicated. The PRA is not necessarily of poor quality, but the staff that is there now is not the staff that worked on the development of the PRA, and there has been no technology transfers between the people that did the PRA and those that are there now. So that is an additional handicap.

MEMBER SIEBER: So they are basically clueless.

MR. O'REILLY: In some cases that is not a bad description.

MEMBER SIEBER: Okay.

MR. BARANOWSKY: We are going to be putting that together as something for us to use and possibly pass on to the quality activities.

MEMBER APOSTOLAKIS: And you said you would give us this other document, which was the key -- you said you were fully aware of where the sensitive parts of the PRAs were, model uncertainties.

MR. BARANOWSKY: We will probably have in our final report some listing of these things, because that is an issue of what does this all mean in terms of this program. So we need to cover that.

MS. WESTON: We are not going to get it before then?

MR. BARANOWSKY: I don't have a specific program activity to produce a report on this before this report, and whether it can or can't be done, I just can't say in this meeting. We are resource limited.

MEMBER SIEBER: That's a nice way to put it.

MR. BARANOWSKY: That's a fact. I request budget and I am told what I can get and we are working at 116 percent.

ACTING CHAIRMAN BONACA: With that, I adjourn this meeting.

(Whereupon, at 4:28 p.m., the meeting was concluded.)

CERTIFICATE

This is to certify that the attached proceedings before the United States Nuclear Regulatory Commission in the matter of:

Name of Proceeding: Advisory Committee on Reactor Safeguards Reliability and PRA and Plant Operations Subcommittees

Docket Number: n/a

Location: Rockville, MD

were held as herein appears, and that this is the original transcript thereof for the file of the United States Nuclear Regulatory Commission taken by me and, thereafter reduced to typewriting by me or under the direction of the court reporting company, and that the transcript is a true and accurate record of the foregoing proceedings.

Yvonne Chang
Official Reporter
Neal R. Gross & Co., Inc.

MITIGATING SYSTEMS PERFORMANCE INDEX

PRESENTATION TO ACRS SUBCOMMITTEES ON RELIABILITY AND PRA, AND PLANT OPERATIONS

PATRICK W. BARANOWSKY (301-415-7493) (pwb@nrc.gov)
DONALD A. DUBE (301-415-5472) (dad3@nrc.gov)
OPERATING EXPERIENCE RISK ANALYSIS BRANCH
OFFICE OF NULCEAR REGULATORY RESEARCH
U.S. NUCLEAR REGULATORY COMMISSION

JULY 8, 2003

PURPOSE AND OBJECTIVE OF MEETING

- Update progress on Mitigating Systems Performance Index (MSPI) development
- Address ACRS comments
- ACRS letter in 2004

Table of Contents

- MSPI Overview
- ACRS comments from May 2002 briefing
- White Paper on Methodology
- Status of Pilot
- Key Technical Issues and Approach to Resolution
  - Independent verification and PRA Quality
  - "Invalid Indicator" (one failure to WHITE)
  - "Insensitive Indicator" (many failures to WHITE)
  - System boundary issues
  - Data collection
  - Common Cause Failure
  - Support system initiators
  - Relationship of SDP and PIs
- Summary
- Implementation Timeline

MITIGATING SYSTEMS PERFORMANCE INDEX (MSPI)

- MSPI evolved from a feasibility study of Risk-Based Performance Indicators (RBPI)
- A highly risk-informed simplification to RBPIs that addresses recognized PI issues:
  - treatment of demand failures and fault exposure time
  - definition of unavailability and maintenance rule consistency issues
  - plant specific risk-informed performance thresholds
  - cascade failure treatment of cooling water support systems
- MSPI monitors risk impact of changes in performance of selected mitigating systems
  - accounts for plant specific design and data
  - scope consistent with current PIs: level-1, at power risk
  - covers unavailability and unreliability consistent with PRA modeling
  - uses detailed scope and calculation specifications to achieve consistency
  - performance thresholds consistent with basis for current PIs

ACRS SUBCOMMITTEE BRIEFING OF MAY 2002: MAJOR COMMENTS AND QUESTIONS

- MSPI approach is "in the right direction and solves many of the problems ... in the current mitigating system"
- What has been learned from the Pilot?
- Should the PI deal with risk? Concern with the example of large number of SCRAMS to RED in the current ROP.
- Should the setting of the baseline be based on a plant's own historical performance, or measured against the industry performance?
- Are there sufficient data in EPIX?

MSPI WHITE PAPER

- Presents Fundamental Concepts and Related Issues

- Mathematical Formulation
  - Importance measure relationships
  - Treatment of unreliability and unavailability
  - Relatively simple calculation

- Benefits
  - Proper accounting of demand unreliability
  - Plant specific design and data

- Limitations
  - Significance of multiple concurrent failures
  - Conditions not capable of being discovered during normal surveillance tests

- Key Issues
  - Differences between MSPI and SDP
  - False negative/false positive indications
  - Validation

MSPI PILOT OBJECTIVES

- Exercise MSPI Guidance:
  - System boundary and component identification
  - Data collection
  - MSPI computation

- Validation and Verification:
  - Issue identification & special studies (Table Top)
  - SPAR comparisons
  - Relevance of SDP

- Perform Temporary Inspections

Status of Pilot Program

- Held July 2002 Workshop prior to Pilot implementation.
- Issued Program Guidelines in NEI 99-02 Revision and Regulatory Issue Summary in September 2002.
- Licensees collected performance data Sept 2002 through Feb 2003 and submitted to NRC on monthly basis.
- NRC Issued Temporary Instruction and performed inspections of Pilot Plant implementation.
- Held January 2003 Workshop for mid-course assessment.
- Identified technical issues and have re-directed research efforts to resolve all key issues.
- Initiated a major effort to reconcile differences in SPAR and Plant PRAs for 11 distinct models (all 20 units in Pilot).

Key Technical Issues

- Independent verification – significant differences between SPAR model and Plant PRA.
- One component failure resulting in system indication turning to WHITE ("Invalid Indicator").
- Large number of failures to turn system to WHITE ("Insensitive Indicator").
- Identification of system boundaries.
- Data collection burden.
- Treatment of Common Cause Failure contribution to Fussell-Vesely.
- Support system contribution to Fussell-Vesely.
- Relationship of SDP and PIs.

Independent Verification

Issues

- Replicate MSPI submittals using SPAR models.
- Expectation going into Pilot that SPAR models in good agreement with Plant PRAs.
- "High level" agreement not necessarily indicative of agreement for cut-sets at 1E-6/yr and lower.
- Importance measures of components monitored in MSPI often differed by one to two orders of magnitude, especially cooling water support systems.

Resolution

- Quality Improvement: Reconcile differences at lower cut-sets levels, and change SPAR and/or Plant PRA *where justified.*
- Major effort to enhance all 11 distinct SPAR models (20 nuclear units).

Summary of Results for Braidwood

- Internal events CDF
  - Plant PRA = 3.0E-5 /yr
  - "Old" SPAR Model = 7.5E-5 /yr
  - "New" SPAR Model = 3.1E-5 /yr

- On average, FV/UR was too low in the old SPAR Model by about a factor of 10.

- With enhancements to SPAR Model, on average agreement within factor of about 2 (high or low).

(assumes same success criterion for feed & bleed)

Braidwood	FV/UR ratio	
Component	old SPAR/Plant PRA	new SPAR/Plant PRA
RHR Pump 1B	0.08	1.38
AFW Pump 1B	4.72	1.30
EDG 1A (T&M)	0.20	0.75
SW Pump 1B	0.05	1.57
VCT Outlet Isol valve	0.11	0.99

Final Results for Palo Verde

- FV/UR's on average within plus/minus 25%

Parameter	Plant PRA	SPAR Enhanced	SPAR 3i
CDF	1.4E-05	1.2E-05	1.79E-05
Truncation	2E-11	1E-11	1E-15
Contributors			
Transients	35%	39%	26%
Loss of 4KV	29%	24%	0
SGTR	12%	6%	30%
Loss of MFW/COND	7%	9%	0
LOOP	7%	11%	34%
LOCA	4%	3%	7%
Loss of DC	2%	4%	1%
Other	3%	4%	1%
ISLOCA	<1%	<1%	<1%

Invalid Indicators

Issues

- For some components with high importance measure, one component failure can result in ~ 1E-6/yr delta CDF (i.e. WHITE).
- Can not determine an adverse performance trend from one failure (a form of "false positive").

Resolution

- Use performance-based "Frontstop": Minimum number of failures within a system before performance indicator turns WHITE.
- Adapted from Risk-Informed Tech Spec initiative.
- Will resolve Invalid Indicator issue and substantially reduce *false positives.*

Invalid Indicators - Preliminary Results

Licensees' Plant PRA Model	Number of Component Types with Invalid Indicators				
	EAC	HPI	HRS	RHR	SWS / CCW
Braidwood 1	Valid	Valid	Invalid	Valid	Valid
Braidwood 2	Valid	Valid	Invalid	Valid	Valid
Hope Creek	Valid	Invalid	Valid	Valid	Valid
Limerick 1	Valid	Valid	Valid	Valid	Valid
Limerick 2	Valid	Valid	Valid	Valid	Valid
Millstone 2	Valid	Invalid	Invalid	Valid	Invalid
Millstone 3	Valid	Valid	Invalid	Valid	Valid
Palo Verde 1	Valid	Valid	Invalid	Valid	Valid
Palo Verde 2	Valid	Valid	Invalid	Valid	Valid
Palo Verde 3	Valid	Valid	Invalid	Valid	Valid
Prairie Island 1	Valid	Valid	Valid	Valid	Valid
Prairie Island 2	Valid	Valid	Valid	Valid	Valid
Salem 1	Invalid	Valid	Valid	Invalid	Invalid
Salem 2	Invalid	Valid	Valid	Invalid	Invalid
San Onofre 2	Valid	Valid	Valid	Valid	Valid
San Onofre 3	Valid	Invalid	Valid	Valid	Valid
South Texas 1	Valid	Valid	Valid	Valid	Valid
South Texas 2	Valid	Valid	Valid	Valid	Valid
Surry 1	Valid	Valid	Valid	Valid	Valid
Surry 2	Valid	Valid	Valid	Valid	Valid

Valid (all components within system are valid) (grey shading)
Invalid (one or more components are invalid) (black shading)

Insensitive Indicators

Issues

- For some components with low importance measures, may require many failures to exceed delta CDF of 1E-6/yr.
- Deterministic criteria to include sufficient number of components within system even if low importance.
- Approximately 11% of systems have at least one insensitive component.

Resolution

- Use performance and statistically-based "backstop": Maximum number of allowed failures before performance is considered "degraded" and indication is WHITE.
- Adapted from Risk-Informed Tech Specs where maximum allowed outage time ("completion time") is 30 days regardless of analysis allowing delta CDF of 5E-7/yr.

Insensitive Indicator Study: Failures to Reach G-W Threshold (San Onofre)

System	Component	Failure Mode	Unit 2 Plant PRA Model	
			# Failures to White	FV/UR
EAC	EDG	FTS	> 20	0.06
		FTLR	> 20	0.06
		FTR	> 20	0.06
HPI	MDP-SBY	FTS	7	1.93
		FTR	6	1.93
	MOV	FTO/C	1	4.46
HRS	MDP-SBY	FTS	14	1.90
		FTR	10	1.90
	TDP-AFW	FTS	7	2.45
		FTR	3	2.45
	MOV	FTO/C	11	4.32
RHR	MDP-SBY	FTS	> 20	0.02
		FTR	> 20	0.02
	MOV	FTO/C	1	7.49
SWS	MDP	FTS	5	2.61
		FTR	17	2.61
	AOV	FTO/C	5	0.51
CCW	MDP	FTS	3	4.26
		FTR	10	4.26

Insensitive Indicators – Preliminary Results

Licensees' Plant PRA Model	Systems with Insensitive Indicators (> 20 failures to White)				
	EAC	HPI	HRS	RHR	SWS / CCW
Braidwood 1					
Braidwood 2					
Hope Creek					
Limerick 1					
Limerick 2					
Millstone 2					
Millstone 3					
Palo Verde 1					
Palo Verde 2					
Palo Verde 3					
Prairie Island 1					
Prairie Island 2					
Salem 1					
Salem 2					
San Onofre 2					
San Onofre 3					
South Texas 1					
South Texas 2					
Surry 1					
Surry 2					

Valid (all components within system are valid)
Insensitive (one or more components are insensitive)

Identification of System Boundaries

Issues

- Numerous questions as to which MSPI system to assign particular components and sub-systems for monitoring.
- Definition of "trains" is based on number of parallel heat exchangers, pumps, or flow paths, whichever is fewer.
- Numerous design configurations which do not fit neatly into definition, especially parallel/series branch lines.

Resolution

- Submit questions for clarification to "Frequently Asked Questions" forum on the NRC web-site.
- Revise NEI 99-02 to include improved guidance.
- Conduct training and lessons-learned workshops prior to full implementation.

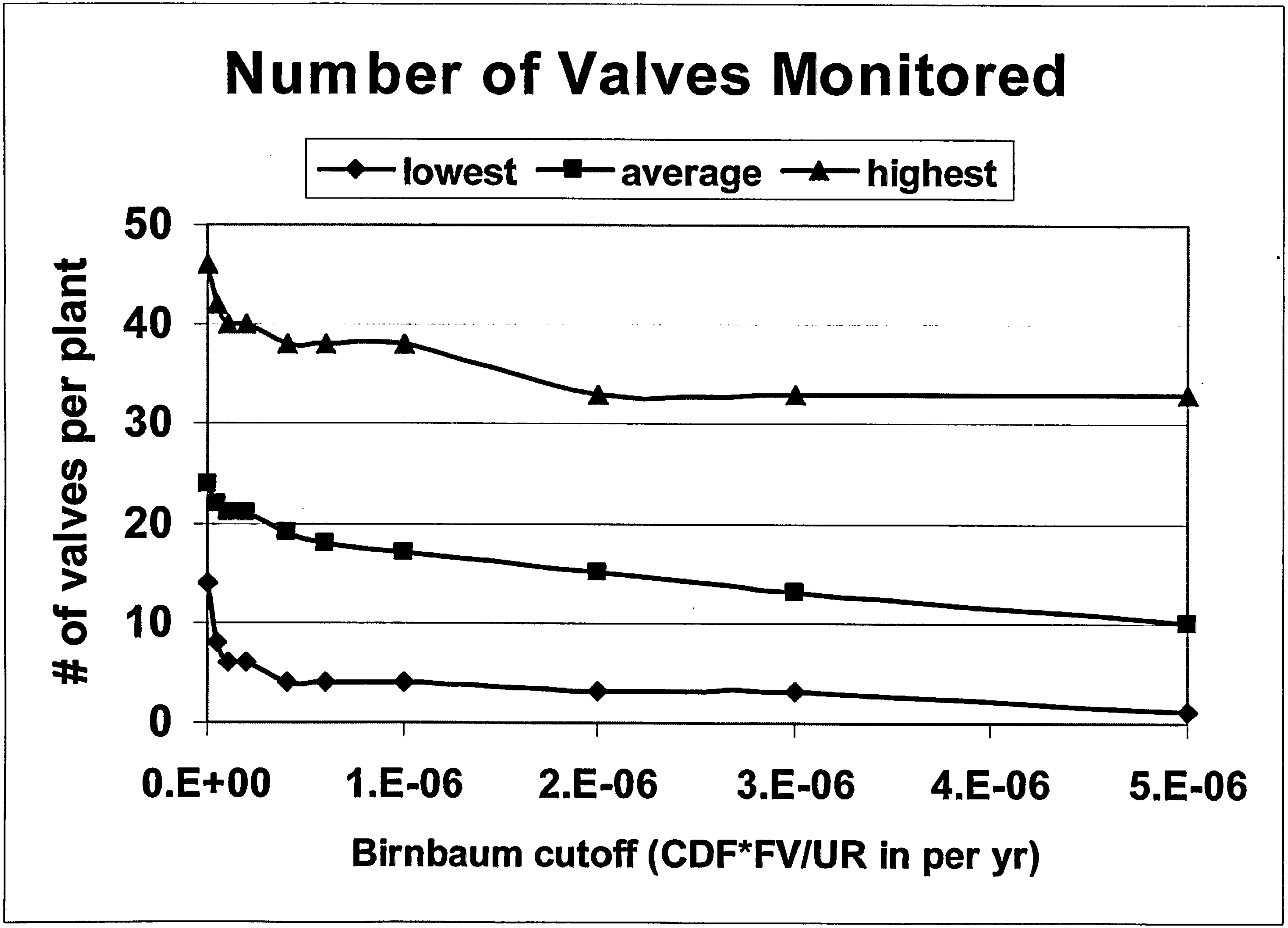
Data Collection Burden

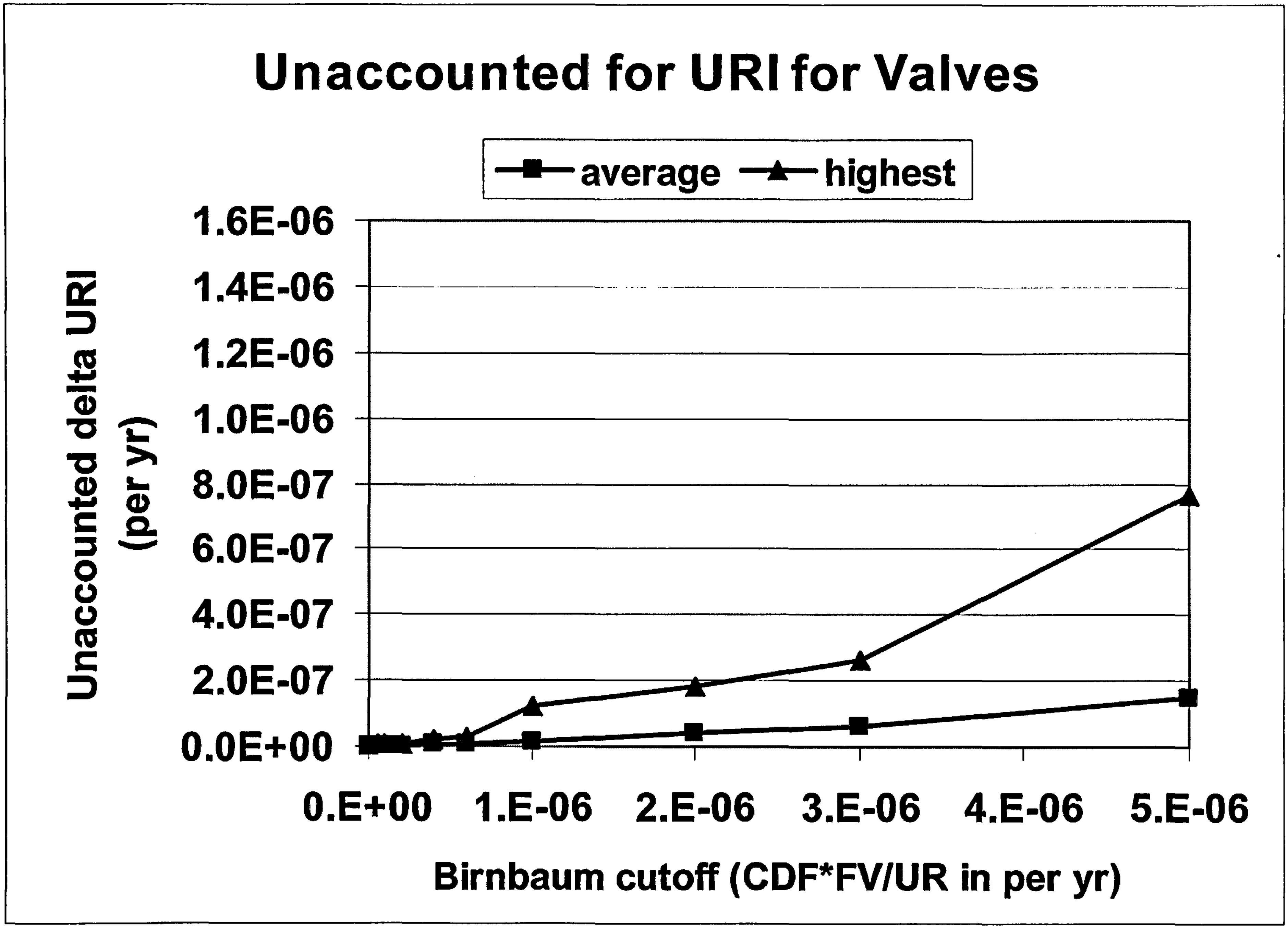
Issues

- Average number of components to be monitored is about 50 per plant.
- Because of deterministic criteria for inclusion, some plants have large number of valves to monitor.
- Pilot Program identified some issues with estimation of demands, tabulation of data, and data entry.

Resolution

- Consolidated Data Entry program through INPO will consolidate and ease reporting.
- Use cutoff of 1E-6/yr on Birnbaum for valves to reduce number to be monitored.
- Recommend improved software interface for data entry for full implementation.


Number of Valves Monitored
lowest
average
highest
# of valves per plant
50
40
30
20
10
0
0.E+00
1.E-06
2.E-06
3.E-06
4.E-06
5.E-06
Birnbaum cutoff (CDF*FV/UR in per yr)


Unaccounted for URI for Valves
average
highest
1.6E-06
1.4E-06
1.2E-06
1.0E-06
8.0E-07
6.0E-07
4.0E-07
2.0E-07
0.0E+00
Unaccounted delta URI (per yr)
0.E+00
1.E-06
2.E-06
3.E-06
4.E-06
5.E-06
Birnbaum cutoff (CDF*FV/UR in per yr)

Contribution of Common-Cause to Fussell-Vesely Importance Measure

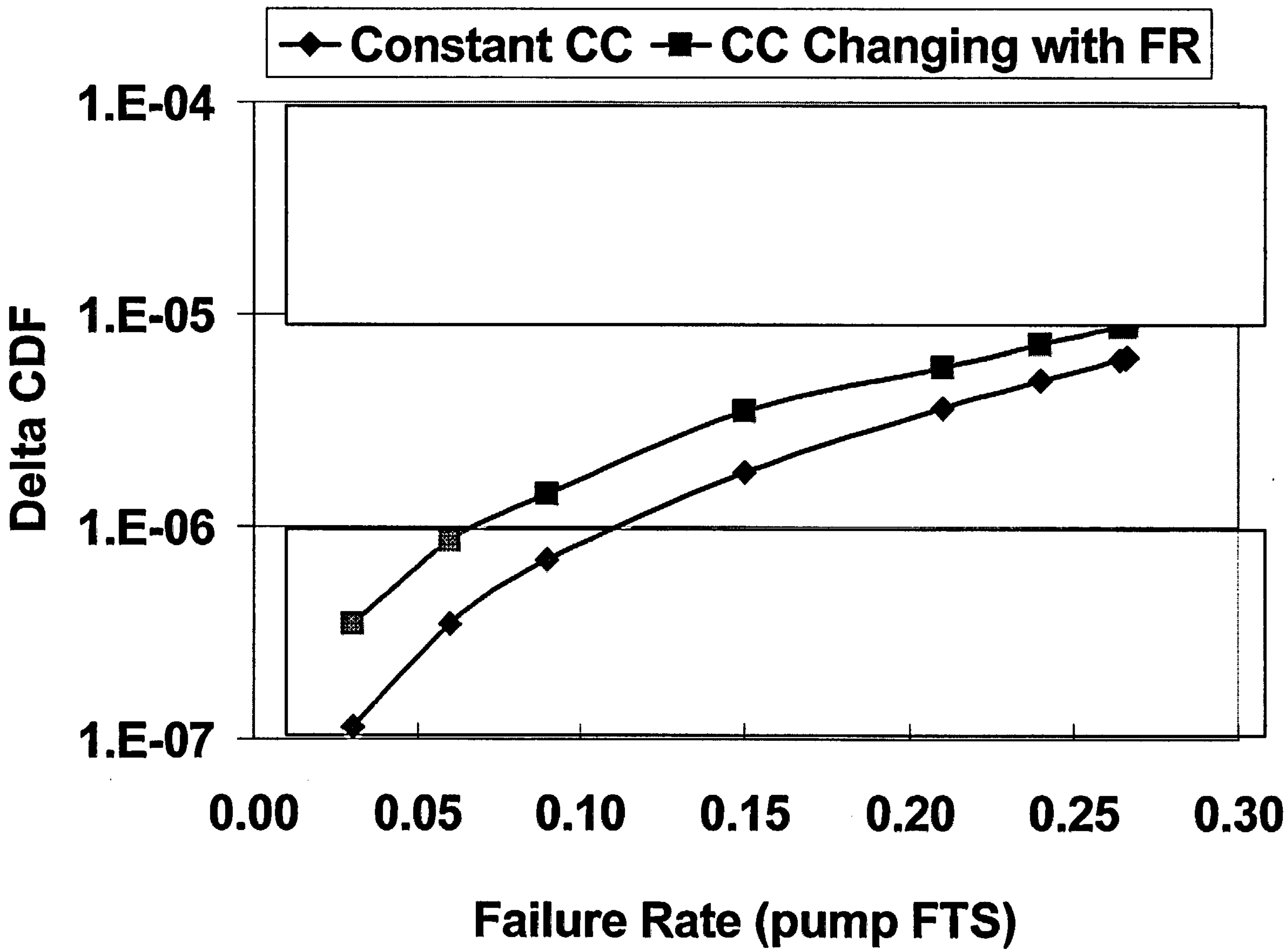
Issues

- As noted in NUREG/CR-6819 CCF Event Insights:
  "Proximate causes of CCF events are no different from the proximate causes of single component failures...it is reasonable to postulate that if fewer component failures occur, fewer CCF events would occur."
- Thus, in the MSPI formulation: Should not changes in the CDF owing to changes in plant-specific unreliability from single component failures also include the effect from changes in CCF rate, given the coupling factor?

Resolution

- Add-in the Fussell-Vesely importance from the common cause failure for a component type in the FV/UR value for a component.

Common Cause Sensitivity
(Hope Creek SACS)



Contribution of Support System Initiators to Fussell-Vesely Importance

Issue

- Failures of components leading to a support system initiator (e.g. loss of service water) contribute to CDF.
- About two-thirds of plant PRAs use fault trees to quantify initiating event frequency; the remainder use a point-estimate frequency, based on plant and/or industry experience.
- For a basic event, use of fault trees give higher FV than when a point-estimate frequency is simply provided.

Resolution

- For those models using point-estimates, if the contribution of the initiator to CDF is significant, either: (a) add support system initiator fault tree(s), or (b) use an adjustment factor to FV.

SDP AND MSPI COMPARISONS

Issue

- Comparison of original MSPI formulation with SDP historical results for in-scope inspection findings (i.e. normal surveillance demand failures) had differences in several instances.

Resolution

- Reassess MSPI analysis of historical data with "final" technical positions implemented.
- Determine validity and appropriateness of any residual MSPI and SDP differences.
- Document validity basis of "final" formulation and use in ROP assessments for in-scope application.

RELATIONSHIP OF SDP AND PIs

- SDP measures risk significance of events or conditions that exist individually or concurrently and that can be due to performance related deficiencies; it is an episodic risk significance measure.

- SDP is a good measure to use when setting priorities for timeliness of corrective actions, need for effective route cause determination, and identification of events with risk significant value for inspection follow-up to verify corrective action adequacy and broader plant-specific or generic implications.

- PIs measure changes in performance; PIs can accommodate performance based and risk informed thresholds; it is a cumulative performance measure.

- MSPI measures change in risk associated with change in mitigating systems performance (within scope limitations).

- Tentative Conclusion: Use MSPI to measure performance within scope limitations.

SUMMARY

- MSPI is a risk-informed performance indicator using plant-specific design configuration and equipment performance data.
- Minimum and maximum limits to the number of component failures before the component is deemed degraded (non-GREEN) will be effected (frontstop and backstop concepts).
- Numerous technical issues have arisen during the Pilot. Proposed solutions for the key issues expected to be fully developed by end of summer.
- Additional implementation issues will be addressed as the need arises.

Tentative Implementation Schedule

August '03	Resolution of all major technical issues; Completion of SPAR model enhancements.
September '03	End of Pilot Program.
Fall '03	Assessment of Pilot Program Success Criteria; "Table-top" comparison to SDP; Address Implementation issues; Go/No-Go Decision.
Late '03 / Early '04	Follow-up Briefings before ACRS.
Winter/Spring '04	Revised Guidance and Inspection Manuals; Lessons-Learned Training and Workshops.
~ Mid-to-late '04	Full Implementation.