STETKAR & ASSOCIATES

203 WHISPERING HILLS STREET
HOT SPRINGS, AR USA 71901
TELEPHONE: (501) 627-0500 • E-MAIL: jwstetkar@aol.com

September 20, 2019

Dr. Michael Cheok
Director, Division of Risk Analysis
Office of Nuclear Regulatory Research
Mail Stop T10-A12
U.S. Nuclear Regulatory Commission
11545 Rockville Pike
Rockville, MD 20852

Dear Dr. Cheok:

Comments on Draft NUREG-2198

Attached are my comments on the draft version of NUREG-2198, "The General Methodology of an Integrated Human Event Analysis System (IDHEAS-G)", that was discussed during the September 18, 2019, meeting of the ACRS Subcommittee on Reliability and Probabilistic Risk Assessment. They contain more details and elaboration on specific topics than the summary oral comments that I made at the meeting.

I do not request, or expect, formal responses to any of these comments and questions. They are provided only for your consideration as you finalize this important project.

Please contact me if you have any questions, or if you need any additional information.

Very truly yours,

John W. Stetkar

Attachment

Comments and Questions on
NUREG-2198
The General Methodology of an Integrated Human Event Analysis System (IDHEAS-G)
Draft, August 23, 2019
John W. Stetkar

1. Section 1, Introduction, General Comment

This section is very good!

Section 1.3, Strategic Approach to Human Reliability Analysis Method Development, and Figure 1-1 provide a good overview of the context for the IDHEAS general methodology.

Section 1.4, Overview of IDHEAS-G, appropriately emphasizes the scenario-based context of a human action and its qualitative evaluation as fundamental elements of the analysis. I particularly like the notion of the "integrative analysis" in Stage 4 of the overall process.

2. Section 2.2, Overview of the Cognitive Basis Structure

This section notes that:

> "***Teamwork*** is the macrocognitive function that focuses on how ***various teams interact*** and collaborate on a task. In the present effort, IDHEAS-G uses this macrocognitive function primarily to include coordination, collaboration, and communication ***between teams***. This macrocognitive function focuses on the emergent aspects of ***interteam interaction*** to avoid duplicating cognitive functions already described by previous macrocognitive functions." [emphasis added]

This seems to indicate that the Teamwork macrocognitive function addresses only interactions between teams of personnel (e.g., between the Main Control Room crew and local operators). It does not seem to address interactions among individuals within a team (e.g., among supervisors and operators in the Main Control Room crew) to achieve a consensus decision or plan of action.

Does the Teamwork macrocognitive function address interactions among individuals within a team?

If not, how does the methodology address the cognitive elements of that type of teamwork?

3. Section 2.2, Overview of the Cognitive Basis Structure

Figure 2-3 shows the overall cognitive basis structure of macrocognitive functions, processors, cognitive mechanisms, and performance-influencing factors (PIFs). This structure is functionally the same as that developed in NUREG-2114, Cognitive Basis for Human Reliability Analysis, (e.g., Figure 2-7 in that report). However, NUREG-2114 uses the term "proximate cause" for the linkages between cognitive mechanisms and macrocognitive functions, rather than "processor".

The notion of "proximate causes" is a fundamental element of the framework in NUREG-2114, and the term is ubiquitous in that report. Therefore, the change to "processors" in NUREG-2198 seems to imply that there is a subtle conceptual difference in the overall framework in this report. For example, that difference might have evolved from the subsequent research that is mentioned in Section 2.1.

I briefly compared the "processors" and "cognitive mechanisms" for each macrocognitive function in Section 2.3 of NUREG-2198 with the "proximate causes" and "cognitive mechanisms" for each macrocognitive function in Section 3 through Section 7 of NUREG-2114. It is apparent that the notion of "processors" is conceptually somewhat different from the notion of "proximate causes". While many of the "cognitive mechanisms" in each report are functionally similar, it is also apparent that the lists in NUREG-2198 have evolved from those is NUREG-2114.

Is the notion of a "processor" in NUREG-2198 functionally the same as a "proximate cause" in NUREG-2114?

If not, what is the difference?

If so, why does NUREG-2198 use the term "processor", rather than "proximate cause"?

Should NUREG-2198 explicitly acknowledge this conceptual difference (or different terminology) to help readers of both reports more clearly understand their relationship?

4. Section 2.3, Macrocognitive Functions, General Comment

In general, this section is very good. It clearly presents the concepts of "processors" and "cognitive mechanisms" in a way that readers can understand how those concepts are related to each macrocognitive function.

This report refers the reader to NUREG-2114 (e.g., for the underlying cognitive research and examples of the links among performance-influencing factors, cognitive mechanisms, and processors). Because the concept of a "processor" is somewhat different from the concept of a "proximate cause" in NUREG-2114, it seems useful to explicitly acknowledge and explain that difference (perhaps in Appendix A).

5. Section 2.3.1.3, Cognitive Mechanisms for Detection, D.h., Shared Cognition within a Team; Section 2.3.2.3, Cognitive Mechanisms for Understanding, U.f., Shared Cognition within a Team; Section 2.3.3.1, Cognitive Activities for Decisionmaking

The summary of cognitive mechanism D.h. seems to involve elements of the Teamwork macrocognitive function, as applied to individual members within the team (e.g., supervisors and operators in the Main Control Room). Therefore, in practice, it is not apparent how, or why, performance-influencing factors (PIFs) for this cognitive mechanism are distinguished and evaluated separately from the Teamwork macrocognitive function.

This comment also applies to the summary of cognitive mechanism U.f. for the Understanding macrocognitive function in Section 2.3.2.3.

My confusion about these distinctions is exacerbated further by the introductory discussion in Section 2.3.3.1, which notes that:

"*Decisionmaking* can be done by an individual or a coherent team (***interteam***

decisionmaking is a part of the teamwork function)." [bold emphasis added]

Section 2.3.3.2 also notes that:

> "Decisionmaking on a team may adopt various infrastructures. All of the team members may be interactively involved in each of the processors. Alternatively, the infrastructure may allocate the processors within the team. For example, in normal or emergency operations in an NPP control room, the decisionmaking infrastructure is already specified; the shift supervisor performs DM2 (manage the goals and decision criteria) through DM6 (communicate and authorize the decision) according to procedures, while other crew members assist the shift supervisor. However, procedures may not be applicable or detailed enough for a scenario, and then the entire crew may iterate the decisionmaking process and make a decision. Nevertheless, the processors model the cognitive process of *decisionmaking* for decisionmakers within a team, regardless of whether the processors are carried out by one person or multiple individuals on a team."

Thus, it seems that teamwork elements of Detection and Understanding are evaluated as part of those macrocognitive functions. Teamwork elements of inter-team Decisionmaking (i.e., coordinated decisions among two or more teams) are evaluated under the Teamwork macrocognitive function. Section 2.3.3.3 does not address a cognitive mechanism for the teamwork elements of Decisionmaking (i.e., like D.h. and U.f.). Therefore, it is not apparent how, or whether, the methodology evaluates PIFs and cognitive mechanisms for teamwork elements that affect within-team Decisionmaking (i.e., a consensus decision among members of a single team).

It is also not apparent how, or whether, Section 2.3.4.3 addresses a cognitive mechanism for the teamwork elements of Action Execution (e.g., coordination of actions that must be performed in a particular sequence by different members of a team).

Why is cognitive mechanism D.h. included in the Detection macrocognitive function, rather than the Teamwork macrocognitive function?

In practice, how are PIFs for this cognitive mechanism distinguished and evaluated separately from the Teamwork macrocognitive function?

For example, if a PIF results in poor communications among members of the Main Control Room team, is that deleterious effect "doubly accounted" in the evaluations of Detection and Teamwork for a human failure event (HFE) that involves multiple teams?

How are PIFs and cognitive mechanisms that affect teamwork elements of within-team Decisionmaking and within-team Action Execution evaluated?

6. Section 2.3.5.1, Teamwork Activities

This section notes that:

> "In contrast to interteam collaborative activities, within-team interactions are ***part of the other four macrocognitive functions***. Within-team interactions consist of those activities performed by a coherent team such as an NPP control room crew….Ineffective interaction is modeled within the failures of those macrocognitive functions." [emphasis added]

It seems somewhat artificial to relegate an evaluation of the Teamwork macrocognitive function to only activities that involve coordinated responses by multiple teams, as shown in Figure 2-8. It seems more reasonable to evaluate Teamwork as a distinct element of Detection, Understanding, Decisionmaking, and Action Execution for any human failure event (HFE), regardless of whether the HFE involves only an individual (no effect from teamwork), a single team, or coordination of multiple teams.

The preceding comment addresses my confusion about the evaluation of performance-influencing factors (PIFs) and cognitive mechanisms that affect teamwork elements of personnel performance within a single team. In particular, it seems apparent that cognitive mechanism D.h. accounts for teamwork elements of the Detection macrocognitive function, and cognitive mechanism U.f. accounts for teamwork elements of the Understanding macrocognitive function. However, it is not apparent that the framework includes explicit cognitive mechanisms for teamwork elements of the Decisionmaking and Action Execution macrocognitive functions, despite the general discussions and acknowledgement that teamwork influences those functions.

I think that additional elaboration is needed to clarify how the evaluation process consistently addresses the effects from PIFs and cognitive mechanisms that affect teamwork elements of Detection, Understanding, Decisionmaking, and Action Execution within a team of individuals. In practice, those elements of teamwork may be quite important for most of the HFEs that are evaluated in a PRA (e.g., responses by the Main Control Room operators, coordinated activities among members of a team of in-plant operators, coordinated activities among members of an emergency response team, etc. – the "within-team interactions" shown in Figure 2-8). It is important for users of this methodology to understand how it accounts for those within-team dynamics, in addition to elements of teamwork that affect coordination of activities that are performed by multiple teams.

JWS Note: As a side comment, most of the Teamwork processors that are summarized in Section 2.3.5.2 and the cognitive mechanisms that are summarized in Section 2.3.5.3 apply to performance within a single team, as well as coordination of multiple teams.

Why does the methodology explicitly evaluate the Teamwork macrocognitive function only for HFEs that involve coordination of multiple groups of personnel?

How does the methodology address PIFs and cognitive mechanisms that affect teamwork elements of within-team Decisionmaking and within-team Action Execution?

In particular, which specific cognitive mechanisms in Section 2.3.3.3 and Section 2.3.4.3 are intended to address the effects from teamwork?

7. Table 3-1, PIF Workplace Accessibility and Habitability

The Discussion and Attributes for this performance-influencing factor (PIF) emphasize rather extreme environmental conditions. A potentially important consideration for cognitive performance of personnel in the Main Control Room, Auxiliary Control Room, Technical Support Center, and Emergency Operations Facility is the status of ventilation and room cooling. Loss of all ventilation and room cooling will cause temperature and humidity to increase, resulting in personnel discomfort and perhaps distracting attention from the desired tasks (e.g., with conflicting priorities to restore ventilation or provide alternative cooling). Furthermore, people typically do not function as effectively as normal when they are hot and sweaty. These habitability conditions are not as severe as those listed in this PIF summary, but they may affect personnel cognitive performance, especially if they persist for an extended period of time.

JWS Note: I am not sure if the PIF for "Cold, Heat, Humidity" in Table 3-2 is intended to account for these conditions. However, the Attributes for that PIF emphasize "extreme" heat and cold. Most analysts would not consider loss of ventilation and room cooling to result in "extreme" heat. Therefore, I am not sure whether elevated temperature and humidity from loss of room cooling are pertinent to the PIF for "Workplace Accessibility and Habitability" in Table 3-1 or the PIF for "Cold, Heat, Humidity" in Table 3-2. The PIF in Table 3-2 may be more appropriate to account for these effects, but the Discussion and Attributes for that PIF should then better explain the context and indicate that analysts should not limit their consideration to only "extreme" conditions.

JWS Note: This comment also applies to the discussion of Environmental Context in Section E.1.2 of Appendix E. That discussion indicates that harsh conditions involve "extreme heat or cold", and the questions emphasize a search for "very" cold, hot, or humid conditions.

Should the Discussion and Attributes in this PIF summary explicitly include conditions that result from loss of ventilation and room cooling, to alert analysts to account for these effects?

8. Table 3-2, Environment- and Situation-Related PIFs Other than Workplace Accessibility and Habitability, General Question

Why do the summaries of these four performance-influencing factors (PIFs) not contain a Definition, Discussion, and better elaboration of their Attributes, like the summaries for all other PIFs?

9. Table 3-2, PIF Resistance to Personal or Vehicle Physical Movement

Do the Attributes for this performance-influencing factor (PIF) also include the need to use ladders or scaffolding to access equipment that must be operated or local instrumentation that must be checked?

If so, should the Attributes explicitly include those access considerations?

10. Table 3-3, PIF System and I&C Transparency to Personnel

Some instrumentation, control, electrical, and fluid (water, compressed air, ventilation) systems may be aligned in alternative or unusual configurations when the initiating event occurs. For example, these configurations may apply during testing, maintenance, specific shutdown plant operating states, etc. If a system is not aligned in its normal configuration, personnel may not correctly confirm that the system is operating properly, easily recognize the effects from equipment damage, or quickly determine how the system should be realigned to cope with the evolving scenario.

Do the Attributes for this performance-influencing factor (PIF) also include conditions when the system is aligned in an unusual configuration?

If so, should the Discussion and Attributes explicitly include those considerations?

11. Table 3-4, PIF Human-System Interface

The list of Attributes for this performance-influencing factor (PIF) is quite extensive and generally comprehensive. While reading the list, I thought about one potentially important Attribute that may not be captured.

In some designs, the indications provide the demanded position for a component or control function, rather than the actual equipment status. The best example of this situation was the pressurizer PORV indications at TMI, which showed that the valves were supposed to be closed, while one was actually not closed. I think that designers have become much more attentive to this issue, but I am not sure that it has been resolved completely. That may be especially true for non-safety systems or functions that affect overall human performance, but are not covered by safety-related design requirements. I am also not sure how advanced integrated digital control systems display the desired and actual status of binary (open-closed, on-off) and dynamic control functions. In any event, since the IDHEAS methodology is intended for a broad variety of applications and designs that are not necessarily consistent with modern technology, it may be prudent for the Attributes to list this consideration, so that analysts are prompted to examine it.

Should the Attributes explicitly include conditions when the indications display the demanded position for a component or control function, rather than the actual equipment status?

12. Table 3-6, Staffing

The Discussion for this performance-influencing factor (PIF) notes that:

> "Staffing consideration should not be limited only to the HFE being analyzed, but it should be considered within the ***scope of the entire event***. A limited number of staff may have to perform multiple tasks (or HFEs) simultaneously. Staffing can be inadequate when many time-critical important human actions are concurrent." [emphasis added]

This is a very good observation. However, I think that most analysts will interpret this paragraph to apply primarily to a broad consideration of the "time-critical important human actions" and human failure events (HFEs) that are modeled explicitly in the PRA. That is certainly important. However, I think that the Discussion should be expanded somewhat to specifically alert analysts to also consider other activities that are not modeled explicitly in the PRA.

For example, personnel may be allocated to mitigate failures or damage to non-safety systems that are important for overall plant investment protection or for perceived improvement of overall plant conditions, but are not modeled explicitly in the PRA. In some scenarios that involve severe plant damage (e.g., fires, floods, seismic events, etc.), operators may also need to attend to treatment and relocation of personnel who are physically injured. These types of diversions and distractions have occurred in practice, and analysts should account for them.

Should the Discussion specifically alert analysts to also consider personnel demands for other activities that are not modeled explicitly in the PRA?

13. Table 3-6, Staffing

One of the Attributes for this performance-influencing factor (PIF) is:

> "key decisionmaker's knowledge and ability are inadequate to make the decision (e.g., lack of required qualifications or experience)"

I do not understand why this Attribute is part of the PIF for Staffing. It seems more relevant to the PIF for Training or the PIF for Teamwork and Organizational Factors, since it addresses the decision-maker's ability to understand the evolving scenario, manage the overall response strategy, and allocate resources as needed.

Why is this Attribute included in the PIF for Staffing?

14. Table 3-7, PIF Procedures, Guidance, and Instructions

The scope of this performance-influencing factor (PIF) covers "Procedures, Guidance, and Instructions". However, the Discussion and Attributes focus exclusively on "procedures". Most analysts understand the term "procedures" to apply to formal step-by-step written instructions, and the Attributes reinforce that notion.

Applications of the IDHEAS methodology may include nuclear power plant accident scenarios or activities in other types of industries and facilities where only general guidance or oral instructions are available. For example, in the context of a nuclear power plant PRA, these scenarios may involve damage from fires, floods, or seismic events; scenarios that require mobilization and use of FLEX equipment; scenarios that involve use of the Severe Accident Management Guidelines (SAMGs); etc. These guidelines and instructions are often not formal "procedures" in the traditional interpretation of that term. However, they may be used to effectively mitigate an evolving scenario.

I am somewhat concerned that analysts may interpret the intent of the Attributes too narrowly and assign an inappropriately negative assessment of this PIF, simply because formal step-by-step written procedures are not available for the desired human action.

JWS Note: The Attributes cover analyst considerations of the scope and quality of procedures very well. I am primarily concerned about the narrow interpretation of "procedures".

Why do the Discussion and Attributes of this PIF focus exclusively on "procedures"?

Should the Discussion and Attributes explicitly alert analysts that this PIF also applies to the use of general guidance and oral instructions that are not organized as formal step-by-step written procedures?

If so, should the Attributes be expanded to alert analysts to specific elements to be considered when they evaluate those types of guidance and instructions?

Should the taxonomy include a separate PIF to evaluate "Guidance and Instructions" for scenarios or applications when formal step-by-step written procedures are not available?

15. Table 3-9, PIF Teamwork and Organizational Factors

The Discussion for this performance-influencing factor (PIF) notes that:

> "Teamwork factors intersect with other PIFs in the personnel category. For example, specific training on understanding the teamwork process can be considered as a part of the

> PIF 'training.' To avoid overlap, IDHEAS-G defines teamwork factors as including attributes that influence ***interteam teamwork***." [emphasis added]

Preceding comments address limitation of the Teamwork macrocognitive function to address only interactions among teams of personnel (i.e., as shown in Figure 2-8). Most of the Attributes of this PIF seem equally well-suited for the evaluation of interactions among individuals within a team to achieve a consensus decision or plan of action (e.g., among supervisors and operators in the Main Control Room crew). For example, evaluations of the sources of information, communications, locations of actions, coordination of personnel, supervisory leadership, team-based decision-making, team cohesiveness, etc. can apply equally within a crew of operators in the Main Control Room, as well as coordination of multiple crews of personnel at various locations around the site.

Of course, as with many other PIFs, not all of the Attributes summarized in Table 3-9 may need detailed analyst scrutiny for within-team interactions during a specific scenario.

With the possible exception of the fourth Attribute listed in Table 3-8, I do not think that these Attributes overlap significantly with the evaluation of the PIF for Training, as summarized in that table.

JWS Note: The Attributes cover the key elements of teamwork very well. I am primarily concerned about the narrow interpretation of "teamwork" and why these considerations do not apply for an evaluation of the performance of a single team of personnel.

Why does the Discussion indicate that these Attributes and their evaluation apply specifically and exclusively to inter-team teamwork?

Why do these Attributes not apply to an evaluation of within-team teamwork?

How do the other PIFs in this taxonomy cover these elements of within-team teamwork, and how does the guidance alert analysts to consider these Attributes when they evaluate the performance of a single team within the context of an evolving scenario?

16. Table 3-10, PIF Work Processes

The Discussion for this performance-influencing factor (PIF) notes that:

> "An organization should also maintain an effective corrective action program to address safety issues such as failure to prioritize, failure to implement, failure to respond to industry notices, or failure to perform risk analyses ***as required by regulation***." [emphasis added]

What is the significance of the highlighted phrase in this excerpt?

Why is it needed to summarize good practices, regardless of formal regulatory requirements?

17. Table 3-10, PIF Work Processes

The Attributes for this performance-influencing factor (PIF) use the following phrase:

> "lack of or ineffective instrumentation"

I found that phrase to be very confusing, considering the normal interpretation and use of the word "instrumentation". I think that the intended notion is "lack of or ineffective practices".

Is that correct?

If so, why does the report use the term "instrumentation"?

18. Table 3-11, PIF Information Availability and Reliability

The Discussion and Attributes for this performance-influencing factor (PIF) emphasize information that is obtained from sensors and displays, which are usually associated with the human-system interface (HSI). Although some of the Attributes may apply broadly to information obtained through other means (e.g., oral reports), the summary does not alert analysts how to consider the quality or reliability of that information.

An evaluation of the timeliness and quality of oral information is important in many PRA scenarios when local observations and oral communications may be the only feasible way to confirm and monitor plant status. Oral transmittal of information may also be the only option for applications of IDHEAS to other industries.

Are the Attributes for this PIF intended to apply for oral information?

If not, why not?

If so, should the Discussion and Attributes explicitly alert analysts about special considerations when local observations and oral transmittal of information is the only available option?

19. Table 3-12, PIF Scenario Familiarity

The Attributes for this performance-influencing factor (PIF) provide a very good summary of analyst considerations of an evolving scenario and how it is related to personnel experience and mental models.

In practice, multiple sequential human failure events (HFEs) are often evaluated throughout the progression of an event scenario. A potentially important consideration when analysts evaluate a particular HFE is the preceding performance of personnel as the scenario evolves. For example, if personnel have taken the appropriate actions for preceding tasks, there is empirical evidence that they understand what is happening and have developed an appropriate response strategy (at least, up until that point). Conversely, if the scenario involves preceding errors or failures to perform the required tasks in a timely manner, there is evidence that the personnel are struggling with an adequate understanding of the plant status or the desired course of action.

Of course, this consideration is most relevant if the same crew of personnel performs the sequential actions. It is less relevant if different personnel are involved with each HFE.

I am not sure if the Attributes for this PIF should explicitly alert analysts to consider the status of human responses during preceding HFEs as an element of their assessment of personnel familiarity with the scenario (i.e., in the context of a subsequent HFE). On one hand, that perspective may tend to subtly disassociate analyst consideration of individual HFEs from the entire context of the scenario, which is conceptually somewhat contrary to the integrated scenario-based focus of the overall methodology. On the other hand, if personnel have performed as expected up until time T in the scenario, it seems that this information should influence an analyst's assessment of their possible performance during an HFE that is needed at time T (and likewise, if they have performed poorly).

JWS Note: Similar considerations affect the evaluation of dependencies among multiple HFEs in a scenario. Therefore, some care is needed to avoid "double counting" for either positive or negative effects when this PIF is evaluated for a subsequent HFE, and when the dependency analysis is performed for that HFE. That is why I am not sure whether the Attributes for this PIF should explicitly list this consideration. It may be better to account for the effects from preceding human successes and failures when the scenario-specific HFEs are defined and when the dependency analyses are performed.

Should the Attributes explicitly alert analysts to consider preceding personnel performance during the scenario as additional evidence of their familiarity with the scenario and the appropriate response strategy?

20. Table 3-13, PIF Multitasking, Interruptions, and Distractions

One of the Attributes for this performance-influencing factor (PIF) is:

> "distraction by other ongoing activities that ***are relevant*** to the critical task being performed" [emphasis added]

This is certainly one source of possible distractions. Experience from actual events has shown that personnel may also be distracted by failures or damage to non-safety systems that are important for overall plant investment protection or for perceived improvement of overall plant conditions, but are not modeled explicitly in the PRA. In some scenarios that involve severe plant damage (e.g., fires, floods, seismic events, etc.), operators may also need to attend to treatment and relocation of personnel who are physically injured. These concerns introduce conflicting strategic and time priorities for decision-makers and constraints on the assignment of limited personnel resources. These types of diversions and distractions have occurred in practice, and analysts should account for them. That is why it is essential that the integrated scenario narrative must describe the entire context of the plant damage, and not focus only on systems and equipment that are modeled explicitly in the PRA, and the distinct human actions that are needed to cope with only those failures.

Why do the Attributes not explicitly also alert analysts to account for distraction by other ongoing activities that ***are not directly relevant*** to the critical task being performed (e.g., damage to systems and equipment that are not modeled explicitly in the PRA, personnel injuries, etc.)?

21. Table 3-16, PIF Time Pressure and Stress

Some training protocols emphasize the importance of making assertive, immediate decisions, and they reward personnel for rapid correct responses. This technique was very prevalent in the past. I think that it is now less common. However, I have encountered it in the training for some types of nuclear power plant accident scenarios. It may also be prevalent in current training for other industries that may use the IDHEAS methodology.

This type of training can instill an inappropriate sense of urgency, reluctance to question initial impressions, and resistance to deliberative team consultation. Although this influence on cognitive performance is related to the training programs, it is not necessarily best evaluated as part of the performance-influencing factor (PIF) for Training. For example, the training program may emphasize the appropriate responses for the scenario that is being evaluated. However, the overall format of the training program may instill an inappropriate sense of urgency for personnel responses during many scenarios. It seems that this effect is better evaluated under the PIF for perceived time pressure.

Should the Attributes for this PIF explicitly alert analysts to account for training protocols that instill an artificial sense of time pressure and urgency for decision-making?

22. Section 3.3, Details of Performance-Influencing Factor Structure, Concern for Personnel Safety

I think that the set of performance-influencing factors (PIFs) should include a separate PIF for Concern for Personnel Safety. None of the PIFs that are listed in Section 3.3 seem to directly address this influence on human behavior.

In practice, this PIF would most likely apply during scenarios that involve plant damage from internal hazards (fires, floods, etc.), external events (seismic events, floods, high winds, aircraft crashes, etc.), impending or actual core damage, large releases of radiation or toxic chemicals, etc. It accounts for additional stress on supervisors, operators in the Main Control Room, and local operators, due to concerns about their own personal safety and possible harm or known injuries to their co-workers. It also accounts for plant personnel concerns about possible harm or the unknown status of their families and loved ones during extreme external events that affect the surrounding area.

The effects from this PIF may be manifested by personal fear, cognitive distractions, an enhanced sense of urgency, additional time delays for cognitive response and action implementation, supervisory reluctance to send personnel into specific plant locations, operator reluctance to perform local actions, etc.

This PIF is different from the summarized PIFs for Workplace Accessibility and Habitability; Environment- and Situation-Related PIFs Other than Workplace Accessibility and Habitability; and Multitasking, Interruptions, and Distractions. It is also different from the typical assumption that personnel will not enter or pass through a plant location that has experienced severe damage. That assumption addresses issues of physical access and known dangers. It does not account for individual and team-level cognitive performance effects from distraction, possible fear, conflicting personal priorities, etc. that are captured by this PIF.

Table 3-16 summarizes the PIF for Time Pressure and Stress. The Discussion of that PIF notes that:

> "Other stresses and anxieties, such as concern for families in emergency conditions, fear of potential consequences of the event, and worrying about personal safety, can also increase the level of psychological stress and affect performance."

That PIF is listed in the general category of Task-Related Performance-Influencing Factors. That task-centered focus does not clearly alert analysts to consider the overall cognitive effects from the considerations that are discussed in this comment. I am concerned that the definition and overall discussion of that PIF and its listed attributes do not adequately highlight these effects, and prospective users of the methodology may inadvertently overlook them during relevant scenarios. I think that the overall taxonomy should include a separate PIF to explicitly account for these considerations and alert analysts to think about them.

JWS Note: I am not quite sure how to label this PIF. I used the term "personnel safety", rather than "personal safety", to indicate that it can affect supervisory decisions and to indicate that it extends to concerns about co-workers, families, etc. The literature may contain a better term for this notion.

Why does the methodology not include a separate PIF to account for this type of influence?

23. Section 4.2.2, Development of Operational Narrative; Section 4.2.2.2, Additional Scenarios; Appendix E, Section E.2, Development of the Operational Narrative; Appendix E, Section E.2.2, Identification of Additional Scenarios

The last paragraph of Section 4.2.2 notes that:

> "A ***PRA model has the baseline scenario*** including the narrative and event progression diagram. The objective of an HRA operational narrative is to identify and document information specific to human performance along with the PRA model." [emphasis added]

This statement is certainly true. However, I think that it does not adequately alert analysts to the conceptual treatment and importance of the "additional scenarios" that are shown briefly in Figure 4-4.

I think that most analysts will interpret the term "baseline scenario" to mean the nominal expected plant response to an initiating event, without additional complications from equipment failures and personnel errors (e.g., the top sequence in an event tree model). In that conceptual construct, all other scenarios that are depicted explicitly in the PRA logic model are deviations from the baseline scenario. For example, a scenario that involves failure of main feedwater is a deviation from the baseline scenario when main feedwater is available. A typical PRA model would then contain only one baseline scenario and numerous "additional scenarios".

On the other hand, the term "baseline scenario" might mean a distinct expected plant evolution, based on the identified combinations of equipment responses and personnel actions. In that conceptual construct, the failure of main feedwater characterizes a different baseline scenario from the scenario when main feedwater is available. A typical PRA model would then contain numerous baseline scenarios with characteristics that are determined by the combined effects from plant thermal-hydraulic response, timing, possible mitigation options, and important influences on human performance.

Section 4.2.2.2 notes only that:

> "To perform a risk assessment, the PRA identifies possible event scenarios leading to undesired consequences. Failure of the system or required important human actions may generate new scenarios deviating from the baseline scenario. Additional scenarios are identified by asking 'what if' questions on the failures and consequences of the systems responses and important human actions. The focus on identifying additional scenarios is to develop a high-level risk perspective of the system responses and important human actions that, if failed, would change the scenario progression."

This brief summary of the notion of "additional scenarios" does not adequately describe how those deviations from the baseline scenario are modeled in the PRA or how analysts should evaluate human performance in those scenarios. In practice, one important purpose of identifying additional scenarios is to determine whether characteristics of system responses or human performance in those scenarios merit distinct and explicit evaluation in the PRA models.

In some cases, the PRA team may decide that the differences should be represented by distinct scenarios in the PRA event trees and fault trees. In those cases, human performance should be evaluated in the context of those scenarios according to the same methods and guidelines that are applied for analyses of the baseline scenario. For example, distinct human failure events (HFEs) are defined, evaluated, and quantified, accounting for the scenario context.

In other cases, the PRA team may decide that it is not necessary to explicitly account for differences from the baseline scenario by defining distinct new scenarios in the PRA event trees and fault trees. In those cases, the identified possible deviations from baseline scenario conditions introduce a source of uncertainty in the evaluations of the performance-influencing factors (PIFs) and human error probabilities (HEPs) for the defined baseline scenario HFEs. For example, an analyst might need to consider additional PIFs or a broader range of PIF attributes to account for the range of possible deviations within the baseline scenario.

Of course, in practice, both notions of "additional scenarios" apply in a practical analysis. At this level of the methodology overview, I think that it is important to introduce these concepts, so that analysts begin to understand them and how they affect a practical analysis. Otherwise, analysts may have the impression that "detailed" implementation of the IDHEAS methodology applies only for a small number of baseline scenario HFEs, without an appreciation of why the identification of "additional scenarios" is important to the overall PRA or how to account for those deviations from the baseline conditions after they are identified.

JWS Note: Section 4.2.4 notes that important human actions are identified in the baseline scenario, and new important human actions may be identified in "deviation scenarios".

Section E.2 of Appendix E notes that:

> "Altogether, several representative scenarios (including the baseline scenario) may be identified, and together, they represent different potential evolutions of the event. A ***complete operational narrative is required for the baseline scenario***. Because all the scenarios with the same initiating event share similarities in important human actions and system responses, a ***complete operational narrative is not generally needed for each scenario***. The operational narratives for other scenarios should emphasize differences from the baseline scenario. If a scenario contains ***dramatic differences*** in important human actions or system responses as compared to the baseline scenario, then developing a complete operational narrative for the new scenario is recommended." [emphasis added]

The discussion in the introduction to Section E.2 provides a somewhat better distinction between the concepts of a "baseline scenario" and other scenarios that involve deviations from the baseline conditions. However, despite the warning in the last sentence of this excerpt, I am concerned that analysts will focus only on the guidance that "a complete operational narrative is required for the baseline scenario" and "a complete operational narrative is not generally needed for each scenario". Of course, I do not advocate development of separate detailed operational narratives for every event sequence that involves a human action. In practice, I think that an operational narrative is needed whenever the PRA team concludes that it is necessary to define a new HFE, even if that HFE is only a different variant of the "baseline scenario" HFE (e.g., it may require a different HEP, due to the "dramatic differences" that are mentioned in the last sentence).

Of course, in practice, this process is somewhat iterative. For example, analysts may not recognize the need to define a new HFE as a variant of the "baseline scenario" HFE until they carefully examine the scenario-specific influences that affect the variant human performance. The operational narrative of the variant scenario helps the analysts to identify and understand those differences and the need to define the new HFE.

Section E.2.2 of Appendix E does not provide much help in this regard. The examples very briefly illustrate a few types of considerations that may identify "additional scenarios". However, the guidance seems to imply that an operational narrative is needed only for the "baseline scenario" conditions. The last paragraph in that section acknowledges the practical limitations for creating numerous detailed scenario narratives. However, it does not help analysts to understand when and why deviations from the "baseline scenario" may be significant enough to warrant a detailed narrative (e.g., when a new variant HFE is needed).

I think that both Section 4.2 of the main report and Section E.2 of Appendix E should clarify the distinctions among "baseline", "deviation", and "additional" scenarios, and clarify the types of considerations that warrant development of a scenario-specific operational narrative.

What is the conceptual notion of a "baseline scenario" in this construct?

For example, in the context of this comment, would failure of main feedwater result in a deviation from nominal conditions within the baseline scenario, or would failure of main feedwater characterize a different baseline scenario?

Why does this section not explicitly note that the process of identifying "additional scenarios" may result in changes to the PRA models and the need for a detailed evaluation of human performance in the context of those scenarios?

Should this section introduce the two notions of deviations from the baseline scenario that are mentioned in this comment (i.e., identification of distinct new scenarios in the PRA models, and identification of sources of variability and uncertainty in the evaluation of baseline human performance)?

24. Section 4.2.3, Identification of Event Context, General Comment

This section is very good!

25. Section 4.3, Modeling of Important Human Actions, General Comments

This section contains a very good presentation of the basic elements of this stage of the analysis process, practical guidance, and useful insights. It also contains a good discussion of the relationships between these elements of the analysis and the overall cognitive framework that is described in Section 2 and Section 3.

Preceding comments address applicability of the Teamwork macrocognitive function for only interactions between teams of personnel, and not interactions among individuals within a team. Except for the last item in Table 4-3, the discussions and guidance in this section seem to limit the assessment of Teamwork to only activities that require coordination of multiple teams.

Section 4.3.3.2 provides a good summary of the functional derivation and hierarchy of the high-level, middle-level, and detailed cognitive failure modes (CFMs). Of course, others might propose detailed CFM lists that are somewhat different from those in Tables 4-6 through 4-9. However, the text clearly indicates that these lists are the authors' suggestions, and individual analysts may use them directly or adjust them for specific applications.

26. Table 4-6, Detection CFMs

Four of the detailed cognitive failure modes (CFMs) for "Fail to perceive, recognize, or classify information" are:

D3-2 Key alarm not perceived
D3-3 Key alarm incorrectly perceived
D3-5 Cues not perceived
D3-6 Cues misperceived (e.g., information incorrectly perceived; failure to perceive weak signals; reading errors; incorrectly interpret, organize, or classify information)

I do not understand why it is necessary to distinguish between a "key alarm" and other "cues". In particular, I do not understand why the CFM structure distinguishes D3-2 from D3-5, and distinguishes D3-3 from D3-6. For example, it seems that combined CFMs could apply for "key alarm or other cues". In practice, some scenario-specific cues may involve distinct audible or visual alarms, while others may require closer inspection of displays, indicators, and trends. However, the quality and availability of the aggregate set of information that should prompt personnel response should determine whether the relevant information is perceived at all, or whether it might be misperceived.

Unless there is a fundamental reason for this distinction between "alarms" and "cues", I am concerned that analysts may not appreciate why these are separate CFMs, and they may criticize the detailed list as being too complex and subtle.

JWS Note: This comment may also affect the first line in Table 4-11.

JWS Note: This comment also applies to Table H-2 and Example 3 in Section H.1.3 of Appendix H.

Why it is necessary for these detailed CFMs to distinguish between a "key alarm" and other "cues"?

27. Table 4-9, Action Execution CFMs

One of the detailed cognitive failure modes (CFMs) for "Fail to assess action plan and criteria" is:

E1-1 Action is not initiated

One of the detailed CFMs for "Fail to coordinate action implementation" is:

E3-2 Fail to initiate action

These CFMs seem to inappropriately evaluate the same influence. In other words, they do not seem to satisfy the characteristic of "non-overlapping" CFMs, as described in Section 4.3.3.1.

JWS Note: This comment may also affect the last line in Table 4-11.

JWS Note: This comment also applies to Table H-5 in Appendix H.

Why does the list contain CFM E3-2, in addition to CFM E1-1?

28. Table 4-9, Action Execution CFMs

One of the detailed cognitive failure modes (CFMs) for "Fail to coordinate action implementation" is:

> E3-3 Fail to perform status checking required for ***initiating*** actions [emphasis added]

It seems that the middle-level CFM for "Fail to assess action plan and criteria" accounts for personnel failures to initiate the required action. The middle-level CFM for "Fail to coordinate action implementation" seems to account for failures to coordinate activities that must be performed in a sequential or integrated manner.

JWS Note: This comment also applies to Table H-5 in Appendix H.

Should this detailed CFM apply for "Fail to perform status checking required to ***coordinate*** actions"?

29. Section 4.3.4.3, Modeling Performance-Influencing Factor States

The last paragraph in this section notes that:

> "Modeling PIFs with binary or several arbitrarily defined states requires grouping the attributes into arbitrarily defined states based on their effects on HEPs. In modeling PIFs, the analyst needs to clearly define the meaning of each state. Because the effects of PIF attributes on HEPs generally ***vary continuously***, the 'poor' state can represent any place between no impact and maximal impact. As a result, the HEP for the poor states can ***vary greatly***. Thus, modeling of PIF states should comply with the following guidance: When modeling PIFs with binary or several discrete states, the ***definitions of each state must be specified, used consistently in HEP estimation***, and documented as the contingency for using the estimated probabilities." [emphasis added]

I understand the basic intent of this section. However, I think that most readers will struggle with it and ask "What do they expect me to do?" As noted in a preceding comment, the material in Appendix I, Assessment of Performance-Influencing Factors, does not help to answer that question.

Section 4.3.3.2 provides a suggested set of detailed cognitive failure modes (CFMs) that are derived from the overall model for human cognition and the authors' experience. That section is very good, because it provides some practical guidance and examples that should help to promote consistency among users of the IDHEAS methodology.

Section 3 describes the 20 performance-influencing factors (PIFs) in the IDHEAS methodology and their attributes. I think that it would be very useful for the report to provide a nominal scale for the states of each PIF, with examples that illustrate applicable attribute conditions for those states. For example, the authors might select a nominal scale of 0 to 10, with 0 corresponding to the base state of the PIF and 10 corresponding to the worst conceivable effects on the PIF. The authors could then describe examples of conditions that correspond to intermediate states, so that analysts have some conceptually consistent anchors. Of course, the examples cannot be exhaustive, and they cannot cover every conceivable combination of attributes or conditions. However, they can provide useful anchor points for analysts to better understand the intention of terms like "limited", "reduced", "poor", "difficult", "unfamiliar", etc. ("Oh, that's what they mean by a 3 and a 7!")

Like the detailed CFMs, the scale and examples for the PIF states would be based on human cognition research, literature, and the authors' experience. As with the detailed CFMs, the guidance should indicate that the PIF scales are the authors' suggestions, and individual analysts may use them directly or adjust them for specific applications.

The PIF scales would not resolve the translation of a set of scenario-specific PIF states into a numerical human error probability (HEP). However, they would promote improved consistency in the assessment of each PIF, which is a very important element of the analyses. Improved consistency in the PIF assessments should also support more consistent derivation of the HEPs, regardless of how that quantification process is accomplished.

JWS Note: I read Section 4.3.4.3 and wrote this comment before I read Section 4.4.3.2 and Section 6. Considering the quantification model that is proposed in Section 4.4.3.2 and the framework that is discussed in Section 6.2, I think that it is extremely important for the report to provide example scales and guidance for the PIF states.

JWS Note: It seems that Appendix I or Section J.1 of Appendix J could be reconstituted to present these examples, without any major disruptions of the overall report format and flow.

Why does the report not provide a nominal scale for the states of each PIF, with examples that illustrate applicable conditions for those states?

30. Section 4.4.3.2, IDHEAS-G Human Error Probability Quantification Model, Estimating Base HEPs for Every CFM

This section notes that:

> "The quantification model requires ***base HEPs for the poor states of the three base PIFs*** (*information availability and reliability, scenario familiarity, and task complexity*). The generalized data have human error rates or estimated HEPs for most attributes of these PIFs." [bold emphasis added]

I am somewhat confused by this statement.

I understand that the proposed quantification model is based on the assumption that combinations of the states of the three "base" performance-influencing factors (PIFs) for "Information Availability and Reliability", "Task Complexity", and "Scenario Familiarity" are the primary determinants for a nominal human error probability (HEP). The nominal HEP is then modified further by the effects from a linear combination of numerical weights from the other 17 PIFs. For example, in Figure 4-8, the blue curve shows how the nominal HEP varies as a function of the combined effects from the states of the three base PIFs. The green and orange curves are conceptual depictions of how the nominal HEP would be modified further by the effects from two other PIFs.

This quantification model requires a "calibration curve" that shows how the nominal HEP varies as a function of the combined effects from the three base PIFs. That is the blue curve in Figure 4-8. It is not apparent how that calibration can be derived from only "base HEPs for the poor states of the three base PIFs". That information does not seem sufficient to determine how the nominal HEP varies as a function of all combinations of the base PIF states. For example, in a 3-discrete-state construct for each base PIF, there are 27 possible combinations of "good, nominal, poor" or "low, medium, high". In principle, each of those 27 combinations should be correlated to an empirically-derived HEP to develop the blue "calibration curve".

JWS Note: I read the paper from PSAM 14 that is cited in Reference 52. It does not provide an empirical relationship between the "base" PIF states and HEPs that are derived from operating experience, simulator training, or human testing.

JWS Note: Section 6.2 discusses this general concept in somewhat more detail. For example, Figure 6-3 shows the basic notion of a "calibration curve", as discussed in this comment, in the context of the effects from a single PIF. The discussion in Section 4.4.3.2 emphasizes the significance of the combined effects from the three "base PIFs". Therefore, a "calibration curve" that is analogous to Figure 6-3 would be needed to relate the composite effects from those PIF attributes and their interactions, which may not necessarily behave according to the linear assumption.

What specific "generalized data for human error rates" have been used to derive the HEP calibration curve for variations in the combined states of the three base PIFs?

Where is that derivation and its supporting data documented?

31. Section 4.4.3.2, IDHEAS-G Human Error Probability Quantification Model, Calculating the HEP of a CFM for a Given Set of PIF States

Equation 4.7 is reproduced below.

$$P_{CFM} = P_{CFM_{Base}} \cdot \left(1 + \sum_{i=1}^{n} w_i\right) \cdot C \cdot \frac{1}{Re}$$
$$= \frac{P_{CFM_{Base}} \cdot (1 + w_1 + w_2 + \cdots + w_n) \cdot C}{Re}$$

This section notes that:

> "…Re is a factor that accounts for the potential recovery from failure of a task, and it is set to 1 unless there are empirical data suggesting otherwise."

This formula indicates that the numerical recovery factor Re can be applied to the overall probability for the cognitive failure mode (CFM), regardless of the states (and corresponding weights) of the contributing performance-influencing factors (PIFs). This seems to be an inappropriate and overly-simplistic numerical treatment of "recovery". For example, during a particular scenario, specific PIFs may have different influences on the feasibility or reliability of potential recovery from a cognitive error. Therefore, it does not seem reasonable to simply apply a single numerical recovery factor to the composite effects from all PIFs.

I think that any methodology that is proposed in this report should focus on the fundamental process for evaluating human performance. Formulas that combine multiple PIF weights to estimate a human error probability (HEP) should not imply that a simple numerical factor can be applied to account for "recovery". The concept of modeling and accounting for potential recovery from a cognitive error should be left for further research.

JWS Note: I intentionally did not elaborate further on the general notion of "recovery" and the use of simple numerical factors to adjust an HEP. In my opinion, if the PIFs and their composite effects are evaluated properly, there is no conceptual justification or need for an additional "recovery" factor.

What is the technical basis for use of a single numerical recovery factor, as shown in Equation 4.7?

How is that concept supported by an understanding of how each PIF functionally influences the probability for the CFM?

32. Section 4.4.3.2, IDHEAS-G Human Error Probability Quantification Model

The notion that the effects from the performance-influencing factors (PIFs) for "Information Availability and Reliability", "Task Complexity", and "Scenario Familiarity" are the primary (or "base") determinants for a nominal human error probability (HEP), which is then modified further by the effects from a linear combination of numerical weights from the other 17 PIFs, is an interesting concept that has not been proposed in earlier versions of the IDHEAS methodology.

Of course, this is simply another conceptual way to numerically combine PIFs and relate them to a set of empirically-derived HEPs. Therefore, as with all other quantification methods that have been reviewed by the NRC staff, careful consideration and testing are needed to develop confidence that it provides results that are qualitatively and quantitatively reasonable, and it facilitates consistent estimates of the risk significance of various human actions and their contributors.

JWS Note: I think that the basic concepts that are summarized in the discussions of "Modeling PIF States", "Modeling the Impact of PIF States on HEPs", and "Calculating the HEP of a CFM for a Given Set of PIF States" (without recovery factor Re) are reasonable and consistent with the original intent of this report. They can be applied conceptually to any HEP quantification scheme. Furthermore, in a more general interpretation of Equation 4.7, the value for $P_{CFMBase}$ could be the HEP that applies with all PIF states at their "best" values (i.e., it need not necessarily be the HEP that is derived from only the three "base" PIFs, as proposed by the authors' quantification model).

Has this model been reviewed and critiqued by other experts in human cognitive behavior and human reliability analysis outside of the NRC staff?

Has this model been tested in any practical pilot applications for which empirical HEP estimates are available?

If this quantification scheme has not been examined and tested by independent experts, why is it summarized in this "NRC-approved" methodology?

In particular, considering the original intent of this report, why was the IDHEAS-G methodology extended to recommend this HEP quantification model?

33. Table 4-13, Summary of IDHEAS-G

One of the Outputs for Stage 1 is:

- "Event context - System, personnel, and task context"

Table 4-12 shows how the 20 performance-influencing factors (PIFs) are allocated conceptually among influences that are related to considerations of the Environment and Situation, System, Personnel, and Task. That taxonomy is very reasonable.

Why does this Stage 1 Output in Table 4-13 not mention the environment- and situation-related context of the scenario?

34. Section 5, Time Uncertainty Analysis, Omission of Key Information and Guidance

There is a critical omission of key concepts and information that are necessary for readers to understand the analyses that are summarized in this section. In particular, neither this section of the main report, nor any of the appendices, describe the basic construct and elements of the timeline that forms the basis for this analysis. For example, that fundamental information is described in Section 3.2.1 and Figure 3-1 of NUREG-2199.

Furthermore, Section 4.1.1 in this report indicates that one important element of the general human reliability analysis (HRA) process is to "perform time analysis and assess feasibility of the important human actions". That element of the analyses is also shown in Figure 4-1. However, neither this section of the main report, nor any of the appendices, describe the basic elements of a feasibility assessment or provide guidance for performance of that assessment. For example, that fundamental information is described in Section 3.1 of NUREG-2199. Guidance for the performance of a feasibility assessment is also provided in Section 3.5 and Section 4.3 of NUREG-1921, in the context of a fire analysis.

Section 7.3 acknowledges that this report does not include guidance for feasibility assessment. As noted above, I think that is a critical omission which should be corrected. The basic concepts and practices of performing a feasibility assessment are not associated with a specific methodology or application. They are a fundamental element of any analysis, and they belong in the general methodology. Since previous drafts of the methodology contained a good discussion of the considerations and guidance for feasibility assessment, it seems rather easy to resurrect that information and include it in this report. I do not think that it is appropriate to refer the reader to NUREG-1921 for that guidance, as suggested in Section 7.3.

The timeline, concepts, and guidance for performance of a feasibility assessment and performance of a time uncertainty analysis were described in previous drafts of the IDHEAS methodology. The current version of the report should provide this critical information and guidance.

JWS Note: In principle, the feasibility assessment is part of the qualitative evaluation of human performance, while the time uncertainty analysis quantifies one contribution to the overall HEP. However, it may be best to describe these elements of the HRA in the same section of the report, since they use the same timeline diagram and concepts (e.g., like Section 3 of NUREG-2199). Of course, a feasibility assessment also involves several other considerations, in addition to the timing information.

Why does this section of the report not describe the concepts, show the timeline diagram, and define the parameters that are the basis for the time uncertainty analysis?

Why does the IDHEAS-G methodology not contain guidance for the performance of a feasibility assessment?

Should this section of the report provide the combined guidance for performance of a feasibility assessment and a time uncertainty analysis?

35. Section 5.1, Time Uncertainty Model

This section notes that:

> "P_t is the error probability resulting from time uncertainties in the time available and time needed to perform an action, ***assuming that actions are performed at a normal pace without complications***." [emphasis added]

I am concerned that some analysts and plant experts may interpret the highlighted phrase as guidance to intentionally exclude some considerations that may affect the uncertainty in the time that is needed to perform the action (i.e., parameter T_n). That is not consistent with the intent of fully characterizing the uncertainty in that time.

The intent of this assessment is to account for all sources of variability and uncertainty that may apply during the baseline human response scenario, and during any other variants of that scenario which use the same human failure event (HFE). If those assessments are restricted to only an evaluation of the "normal pace without complications", it is very likely that the analysts and plant experts will not adequately consider the full range of conditions that may apply and the uncertainties in T_n that are associated with those conditions. Therefore, I do not think that it is appropriate to qualify the basic definition of P_t with these limitations.

Of course, the uncertainty in T_n must be evaluated in the context of the event scenario and the corresponding HFE. In the guidance for estimating T_n, it is certainly appropriate to caution analysts that the uncertainty in that time should not account for conditions which apply during other scenarios (e.g., for the same functional response, but a differently-defined HFE for that response). However, once the scenarios that use the same HFE are defined, the uncertainty for T_n should account for all conditions within the scope of that HFE which could affect the time that is needed to perform the action, without inappropriate conceptual restrictions on the considerations that may affect that uncertainty.

This section also notes that:

> "Time needed represents the time taken for the actions in the HFE to be completed, including information detection, diagnosis, decisionmaking, executing the action, and the time needed for teamwork if multiple teams or distributed individuals are involved. This

> assumes that all of these activities are performed at a ***normal pace without complication*** (i.e., the ***outliers are excluded***)." [emphasis added]

The same comment applies to the highlighted phrases in this excerpt. In fact, analysts and plant experts should be encouraged to consider possible "outlier" conditions that may affect the overall uncertainty in T_n.

JWS Note: In practice, if the variability of conditions within a set of scenarios that use the same HFE result in very large uncertainties in T_n, with a corresponding risk-significant contribution from P_t, analysts may then decide to subdivide the scenarios and define variants of the original HFE to account explicitly for those influences. That is a normal part of the iterative process of defining HFEs in the PRA models.

Why is it necessary for the basic definition of P_t to contain these qualifications?

How does the analysis guidance ensure that these qualifications do not bias the analysts and plant experts to inappropriately underestimate the uncertainties in parameter T_n?

Why does the guidance indicate that analysts and plant personnel should not account for possible "outlier" conditions that may affect the overall uncertainty in T_n?

36. Section 5.2, Guidance on Estimating the Distribution of Time Available

This section notes that:

> "The analytic approach starts by reviewing the preliminary risk results to identify the ***dominant risk contributors***. The calculations can help analysts identify ***areas where uncertainty analysis is needed*** and where more sophisticated analyses should be performed to better define the success criteria. This phased approach makes ***uncertainty analysis feasible***." [emphasis added]

I disagree very strongly with the implications of this guidance. It seems to tell analysts that uncertainty analysis is needed only for the "dominant risk contributors", which are identified from analyses that do not account for uncertainty. Uncertainty analysis is always "feasible", even if it involves only a subjective quantitative assessment of the uncertainties, based on expert judgment.

The NRC State-of-the-Art Reactor Consequence Analysis (SOARCA) project has clearly shown the importance of evaluating and quantifying uncertainties as an integral part of the risk assessment process (i.e., as an integral element of the parameter estimations). In particular, that project has shown that the uncertainties may determine which specific scenarios are important contributors to overall risk. A retrospective assessment of uncertainty for only the "dominant contributors" from a "point estimate" calculation may not identify the correct risk contributors, and it is not consistent with the general guidance for performing a risk assessment.

It is certainly reasonable to indicate that analysts may use a variety of methods to evaluate the uncertainty in parameter T_a, without performing numerous resource-intensive thermal-hydraulic simulations. However, it is not appropriate for "NRC-approved" guidance to imply that uncertainties should be evaluated as an after-thought or as an add-on that is not an integral part of the analysis process.

Why does the "NRC-approved" guidance in this section imply that uncertainty analysis for parameter T_a is needed only for the "dominant risk contributors", which are identified from analyses that do not account for uncertainty?

37. Section 5.3, Guidance on Estimating the Distribution of Time Needed

The first bullet item in this section notes that:

> "Average crew response time should be obtained, as well as an estimate of the time by which the slowest operating crews would be expected to complete the actions."

I think that the guidance should emphasize the importance of developing a complete uncertainty distribution, and not focus only on the central tendency and the "upper tail" of that distribution. Therefore, I think that the guidance should indicate that an average crew response time should be obtained, as well as estimates of the times by which the fastest and slowest operating crews would be expected to complete the actions.

JWS Note: This comment also applies to the discussion of biases for "Underrepresentation / Incomplete Representation of the Range of Times" at the end of this section. Since that discussion focuses mostly on under-estimation of T_n, it is appropriate to emphasize estimates for the slowest time. However, it also addresses the range of times, so it seems appropriate to mention the fastest time as a measure of the lower bound of that range (and perhaps as an additional anchor point to identify possible unrealistic optimism in the expert estimates).

Why does this guidance not indicate that the initial uncertainty distribution should also estimate the fastest crew response time?

38. Table 5-1, Typical Factors Contributing to T_n

The entries for the Action Execution macrocognitive function include:

> "Implement the action steps or continuous action and required timing of steps."

> "Confirm completion of the actions and wait for system feedback."

Experience has shown that some analysts and plant personnel may focus only on the time that is needed to initiate a required action. For example, they may account for the time that is needed to perform the manipulations to start a cooldown and depressurization. The success criteria for the respective human failure event (HFE) typically require that the entire action must be completed before plant conditions reach a threshold which alters the event progression. For example, the entire cooldown and depressurization must be completed before level reaches a specific setpoint.

Consideration of uncertainties in the time that is needed to fully complete the required action can be important in some scenarios. For example, the operators may need to reduce temperature by 250 °F at a rate that does not exceed 100 °F per hour. Thus, a minimum of 2-1/2 hours would be needed to complete a cooldown at the maximum allowed rate. However, plant-specific and scenario-specific constraints may not facilitate a continuous cooldown at the maximum allowed rate. For example, degraded heat removal system flow rates, higher cooling water temperatures, or limited heat transfer rates may slow the maximum achievable cooldown rate as temperature is reduced, and thereby extend the total amount of time that is needed to achieve the desired conditions. These constraints can be sources of uncertainty in the estimates for T_n, depending on the range of conditions that may apply during the scenarios that use the same defined HFE.

In principle, the cited considerations for a "continuous action" and "completion of the actions" should prompt analysts and plant personnel to think about these sources of uncertainty. Unfortunately, experience from actual analyses has shown that they occasionally do not.

Should the guidance or an example in this section explicitly prompt analysts to account for sources of uncertainty that affect the entire amount of time that is needed to achieve the HFE success criteria (i.e., especially for continuous actions that may have scenario-specific constraints)?

39. Table 5-1, Typical Factors Contributing to T_n

A time uncertainty analysis is also needed for integrated human actions that require teamwork. Although the timelines and inter-relationships among various teams may be rather complex, methods are available to display and account for combinations of series, parallel, and functionally dependent activities. Therefore, it is not appropriate to imply that an evaluation of the uncertainty in parameter T_n is not needed, or that the evaluation need not consider the effects from the Teamwork macrocognitive function. In fact, the time needed for effective Teamwork may be the most important source of uncertainty in those analyses.

Why does Table 5-1 not include any factors that contribute to the time required for the Teamwork macrocognitive function?

40. Table 5-2, Uncertainty Factors that Modify the Distribution of T_n

One of the considerations for "Plant Condition" is:

> "Other ongoing activities that compete for resources."

I am concerned that most analysts and plant personnel will narrowly interpret "resources" to address people, equipment, etc. None of the considerations in this table seem to explicitly address scenario-specific sources of distractions, interruptions, possibly conflicting priorities, stress, etc. that may divert supervisors' and operators' attention away from the desired course of action for the defined human failure event (HFE). In practice, many of those sources of uncertainty for T_n may arise from conditions that are not modeled explicitly in the PRA, and are documented only in the scenario narratives.

Should the considerations in this table explicitly prompt analysts to account for plant-wide conditions that may distract supervisors' and operators' attention or introduce competing demands and delays, despite the availability of adequate personnel and equipment "resources" to perform the PRA-modeled action?

41. Section 5.3, Guidance on Estimating the Distribution of Time Needed

This section notes that:

> "Estimation of T_n is based on the ***baseline scenario*** and its context identified in the scenario analysis of an HRA." [emphasis added]

I am concerned that analysts may interpret this guidance to inappropriately restrict their full consideration of conditions that may affect the uncertainty in T_n.

A preceding comment on Section 4.2.2 addresses guidance for the identification and treatment of "additional scenarios" and deviations from the "baseline scenario". In some cases, analysts may conclude that the potential effects on human performance in those deviations warrant definition of a distinct variant of the baseline human failure event (HFE). The analyses of that new HFE then account explicitly for those scenario-specific influences. In other cases, analysts may conclude that it is not necessary, or practical, to define a new HFE. In those cases, analyses of the "baseline" HFE should account for variations of the conditions that apply during any of the scenarios which use that HFE (i.e., the nominal "baseline scenario" and all other scenarios that use the same HFE).

In a practical PRA, a single HFE is often used to quantify the human error probability (HEP) for the same functional action within a set of many individual scenarios (or "event sequences"). The evaluation of that HFE must then account for variations in the performance-influencing factor (PIF) attributes and timing among the entire set of scenarios that use the nominal HFE. In these common situations, the assessment of uncertainty in T_n should not account only for the context of the "baseline scenario". It should also account for deviations from that context that may occur during any of the other scenarios that use the "baseline" HFE.

JWS Note: As noted in a preceding comment, in practice, if the variability of conditions within a set of scenarios that use the same HFE result in very large uncertainties in T_n, with a corresponding risk-significant contribution from P_t, analysts may then decide to subdivide the scenarios and define variants of the original HFE to account explicitly for those influences. That is a normal part of the iterative process of defining HFEs in the PRA models.

JWS Note: The discussion in the rest of this paragraph seems intended to address the notion in this comment. It alerts analysts to consider "context variability" as a source of uncertainty in T_n. However, the discussion of EOP branching points is a bit distracting from the practical considerations that are noted in my comment. Without more elaboration and reference to the normal practice of using the same HFE for multiple scenarios, I am concerned that analysts will focus too narrowly on the first sentence and consider only the nominal "baseline scenario". Thus, I think that the discussion should provide more practical insights or an example that illustrates the notion of how analysts should consider "context variability" within a set of scenarios that use the same HFE.

Should the guidance in this section more clearly address how and why analysts need to consider sources of uncertainty in T_n that arise from variations among all scenarios that use the same HFE, in addition to the nominal "baseline scenario"?

42. Section 5.3, Guidance on Estimating the Distribution of Time Needed

A preceding comment on Table 5-1 addresses the omission of factors that contribute to the time required for the Teamwork macrocognitive function. It seems that explicit consideration of those effects should be relevant to the discussion of biases for "Underestimation for Complex Scenarios".

Should the discussion of biases for "Underestimation for Complex Scenarios" explicitly address consideration of time that is needed for effective teamwork?

43. Section 5.3, Guidance on Estimating the Distribution of Time Needed

The discussion of "Underestimation of the Effects of Interruption" notes that:

> "A related issue is ***time available*** for personnel to perform actions." [emphasis added]

I do not fully understand the intent of this sentence in the context of this discussion. I think that it may also confuse other readers and distract from the basic ideas in this paragraph.

In particular, the methodology and guidance in Section 5 use the term "time available" and the variable T_a to account for the amount of time that is available for personnel to complete the desired action before plant conditions reach a threshold which alters the event progression. It is very important to maintain a consistent interpretation and use of that term throughout the guidance. In that context, it is not apparent how, or why, the "time available" affects these particular considerations of possible bias in the estimates for the time that is needed to perform the action (i.e., T_n).

What is the intent of this sentence in the context of this discussion?

In particular, does the "time available" in this sentence pertain to parameter T_a in the time uncertainty analysis?

Is this sentence needed to explain the basic ideas in this paragraph?

44. Section 6.1, Human Error Data, General Comment

This is a good discussion of cautions, considerations, and limitations that apply when analysts adapt available numerical estimates for use in a specific methodology or application. It also provides a good discussion of the context and use of expert judgment.

45. Section 6.2, Data Generalization and Integration

I understand the basic concepts in this framework and the notions of how the three tables would be used to correlate the states of performance-influencing factor (PIF) attributes for estimating the probabilities for specific cognitive failure modes (CFMs) and the consequential composite human error probabilities (HEPs). This is a very ambitious effort. I am somewhat concerned that its discussion in this report may mislead readers regarding the degree to which the tables, correlations, and supporting data have been developed.

For example, the introduction to Section 6.2 notes that:

> "Thus, the NRC staff used IDHEAS-G to generalize various sources of human error data and then integrated the data to inform HEP estimation."

> "The NRC staff used IDHEAS-G to generalize human error data in the following steps, as shown in Figure 6-2:"

Based on these statements, and other similar statements about what has been accomplished, I am a bit concerned that readers may conclude that much of this effort has been completed, and the tabular relationships are ready (or almost ready) for use in practical applications. If that is not the case, I think that the discussions in Section 6.2 should clarify the fact that these constructs provide a conceptual framework, but the actual work to populate the tables and test the quantification methodology in practical applications has not yet been performed.

In other words, I think that the authors should be very careful to avoid language that may "over-sell" the proposed methodology or imply that it has reached a level of maturity that has not yet been achieved. This does not detract from the conceptual elegance of the methodology or its potential usefulness. However, it provides an appropriate basis for readers' expectations.

To what extent have the conceptual elements of the methodology and the tabulations that are discussed in Section 6.2 been developed and tested for practical applications (e.g., to estimate HEPs that are influenced by the composite effects from several PIFs)?

46. Section 6.3, Demonstration of Human Error Data Generalization and Integration, General Comment

This section and the tables are useful to illustrate the concepts that are outlined in Section 6.2. I intentionally did not comment on specific numerical values in Table 6-2, Table 6-3, or Table 6-4, because I am much more concerned about readers' potential misinterpretation of the authors' intent for generic use of those values.

47. Section 6.3.1, Generalizing Expert Judgment of Human Error Probabilities for IDHEAS At-Power Application to IDHEAS-G HEP Table

I am very concerned that readers of this report will misinterpret the intent of the numerical values in Table 6-2 and their implied "NRC-approved" use for general applications of the IDHEAS methodology. That concern is exacerbated by the title of this section, which indicates that the human error probabilities (HEPs) in Table 6-2 have been "generalized" for IDHEAS-G.

According to the overall scope and intent of this report, I think that most readers will conclude that any tabulated HEP estimates apply generically for a broad variety of potential applications of the IDHEAS-G methodology. In particular, the numerical values have been examined thoroughly by the authors, and they should not be inappropriately biased due to their derivation from limited data sources or very specialized analyses. Therefore, unless the discussions and guidance in this report are extremely clear and contain very explicit prohibitions, readers will interpret and use any tabulated numerical values as "NRC-approved generic data" that apply for any potential application of the IDHEAS methodology.

The entries in Table 6-2 illustrate the basic concept of the HEP Table. As noted in the text, they are derived only from the expert estimates in NUREG-2199. Therefore, those particular examples do not satisfy the desired generality of the proposed methodology and tabulations. In particular, they suffer from the focus of the experts' assessments in the specific context of that application, as cautioned by the discussion of expert judgment in Section 6.1.

The last two bullet items in Section 6.2.1 note that the HEP Table should document:

- "brief narrative of the task or types of failure in the data source, including the work domain (e.g., nuclear, aviation) and type of data source (e.g., experiment, training simulation, event database)"

- "uncertainties in the data source and in the mapping to IDHEAS-G CFMs and PIFs"

The discussion of "Integrate the Data to Inform the Base HEPs" in Section 6.2.1 also provides very relevant insights about generalizing and integrating data from a broad variety of sources to derive "consensus HEPs" (my term) for specific combinations of cognitive failure modes (CFMs) and associate performance-influencing factor (PIF) attributes. The discussion of "Quantification of PIF Effects" in Section 6.1 cautions about the importance of a clear understanding of binary PIF states, such as "good" and "poor".

Table 6-2 does not contain information that describes the task contexts for the expert assessments. More importantly, it does not document uncertainties in those assessments or conclusions about their general applicability for the identified cognitive failure modes (CFMs) in other applications. Without that information, it is very likely that readers will conclude that the authors intend that these HEPs can be used generically.

JWS Note: This comment also applies for Table 6-3 and Table 6-4.

Do the authors of this report intend that the HEP values that are listed in Table 6-2 should be used for other applications of the IDHEAS methodology?

If so, what is the technical basis for that conclusion, and how is it consistent with the general principles and guidance in Section 6.1 and Section 6.2.1?

Why does Table 6-2 not document the "uncertainties in the data source and in the mapping to IDHEAS-G CFMs and PIFs"?

If the authors intend Table 6-2 to only illustrate how one source of HEP estimates can be represented in the HEP Table, and they do not intend that those particular HEP estimates should be used in any general applications of the methodology, why do the text in Section 6.3.1 and the annotations in Table 6-2 not specifically (and perhaps emphatically) indicate that the entries are "For Illustrative Purposes Only"?

48. Section 6.3.2, Integration of the Data in the HEP Table to Inform the Base Human Error Probabilities

Despite my initial confusion by the title of this section, it is apparent that the entries in Table 6-3 are not related directly to the numerical values and expert assessments from NUREG-2199 that are listed in HEP Table 6-2. It is evident that the entries in Table 6-3 are derived from several unrelated studies that have examined the possible effects from specific attributes for the performance-influencing factor (PIF) for Task Complexity.

That is certainly fine for the purposes of illustrating the conceptual content and sources of information that may be used to map these PIF impacts. It also subtly reinforces the notion that the tables in Section 6.3 are intended only for illustrative purposes, and not for an actual application of the IDHEAS methodology.

JWS Note: The text indicates that "Table 6-3 shows some generalized human error data for the base PIF task complexity". The preceding comment addresses my concerns about readers' interpretation of the intended use of tabulated numerical values in this report.

Why does Table 6-3 not document the specific study that supports each of these assessments?

49. Section 6.3.2, Integration of the Data in the HEP Table to Inform the Base Human Error Probabilities

Section 6.3.2 discusses development of the PIF Impact Table. Table 6-4 shows an example of part of the PIF Impact Table for "Multitasking, Interruption, and Distraction". Table 6-3 is an example of an intermediate level of information processing that provides a link between the HEP Table and the PIF Impact Table. The intermediate nature of Table 6-3 is not readily apparent from the discussion, and readers might misinterpret it as an example of a PIF Impact Table. (I did initially, until I read the introduction to Table 6-4.)

It is apparent that the assessments of the PIF attributes for "Task Complexity" in Table 6-3 are not related directly to the assessments of the PIF impacts and weights for "Multitasking, Interruption, and Distraction" in Table 6-4. However, it is not apparent how the information in those tables is related conceptually. In particular, it is not apparent how the types of information that are summarized in Table 6-3 are used to derive the PIF weights in Table 6-4. A specific example would be useful to illustrate that process (i.e., an example application of Equation 4.6, based on the types of information in Table 6-3). Unfortunately, that is not very easy, due to the different PIFs that are summarized in these tables.

How are Table 6-3 and Table 6-4 related conceptually?

Should Section 6.3.2 contain an example that shows how the information in Table 6-3 can be used to derive PIF weights for one or two attributes of Task Complexity?

50. Section 6.3, Demonstration of Human Error Data Generalization and Integration

Section 6.2 defines three distinct and related tables that organize mapping of the states of performance-influencing factor (PIF) attributes to the probabilities for specific cognitive failure modes (CFMs) and the consequential composite human error probability (HEP). Those tables are:

- HEP Table, described in Section 6.2.1
- PIF Impact Table, described in Section 6.2.2
- PIF Interaction Table, described in Section 6.2.3

The discussion in Section 6.3.1 and Table 6-2 pertain to an example of the HEP Table.

The discussion in Section 6.3.2, Table 6-3, and Table 6-4 pertain to an example of the PIF Impact Table.

Section 6.3 does not contain a discussion or an example of a PIF Interaction Table.

Are Section 6.2.3 and Table 6-1 intended to be the only discussion and example of a PIF Interaction Table?

Why does Section 6.3 not contain a discussion or an example of a PIF Interaction Table?

51. Section 6.3.3, Mapping between SACADA Database and IDHEAS-G, General Comment

I understand how the SACADA database is related to the taxonomy in the IDHEAS cognitive performance model. I did not study details of the information in Table 6-5 and Table 6-6.

52. Section 7, General Discussion and Comments, General Comment

This section is a good summary and wrap-up of the report. At this summary level, I think that there is only one other topic that may warrant additional discussion in Section 7.2.

The IDHEAS-G Human Error Probability Quantification Model is described in Section 4.4.3.2. The proposed quantification model is based on the assumption that combinations of the states of the three "base" performance-influencing factors (PIFs) for "Information Availability and Reliability", "Task Complexity", and "Scenario Familiarity" are the primary determinants for a nominal human error probability (HEP). The nominal HEP is then modified further by the effects from a linear combination of numerical weights from the other 17 PIFs. Section 6 contains a rather extensive discussion of the basic concepts of a process for developing the data and functional relationships that correlate the states of PIF attributes into a translation scale for estimating the probabilities for specific cognitive failure modes (CFMs) and the consequential composite HEPs within the framework of that model.

At a very basic level, this quantification model is founded in the concepts of the IDHEAS cognition model for human performance. However, to my knowledge, the functional construct of that model (i.e., the assumptions that three specific PIFs fundamentally determine a "base" HEP, and that the composite effects from all other PIFs can be evaluated as a modifier of that "base" HEP) has not been tested in any practical applications or compared to empirical evidence to verify and validate that it provides an appropriate way to account for these influences. Therefore, I think that Section 7.2 should discuss the current status of the quantification model and the need to validate its basic assumptions and relational structure through pilot applications that can be benchmarked with empirical human performance data.

Why does Section 7.2 not discuss these considerations for future research related to the proposed quantification model?

53. Appendix A, Cognitive Mechanisms Underlying Human Performance and Reliability, General Comment

I do not have any additional comments or questions about these tables.

54. Appendix B, Links of Performance-Influencing Factor Attributes to Cognitive Mechanisms

I did not study the details of these tables. I skimmed through a few examples, and they seem generally reasonable.

Of course, these specific links were developed by the authors of this report, and others might have somewhat different detailed constructs. However, in the context of the IDHEAS methodology, these tabulations summarize the identified functional relationships.

Based on the material in Section 2 and Section 3 of the main report, I understand the basic cognitive model and why individual performance-influencing factor (PIF) attributes may affect multiple cognitive mechanisms, processors, and macrocognitive functions. However, with that background, it is not readily apparent why these detailed links are tabulated in this manner, or how the links are used in practice to support qualitative or quantitative evaluations of a human failure event (HFE). In principle, I can see how they might be useful as a tool for understanding how an analyst's critical assessment of a specific attribute for a particular PIF affects the respective HFE macrocognitive functions (e.g., as a forensic tool for understanding contributors to risk-significant HFEs). For example, in Table B-1, environmental conditions that adversely affect accessibility or habitability are manifested by reduced perception (D.b.) for Detection and by limited movement (E.a.), spatial precision (E.l.), and timing precision (E.m.) for Action Execution.

Figure 3-2, Section 3.3, and Section 3.4 of the main report refer the reader to this appendix. Unfortunately, based on the material in Section 2 and Section 3 of the main report and the introduction to this appendix, I do not fully understand the primary intent for these detailed tabulated relationships. In other words, I do not understand how an analyst will use them in a practical human reliability analysis (HRA).

I think that it would be useful to provide a reader who has finished Section 3 of the report, and turned to this appendix, some additional context for these tabulations and how they support either the evaluation of an HFE or an understanding of the most important contributors to failure of the HFE.

Should the introduction to this appendix provide additional background information or insights about how these links are used to support performance of a practical HRA or examination of the HRA results?

In other words, how are these links used in a practical HRA?

55. Appendix C, Insights into Performance-Influencing Factors from the Cognitive Literature, General Comment

These are interesting abbreviated summaries that illustrate the effects from specific performance-influencing factor (PIF) attributes. It is good that the examples are selected from activities other than nuclear power plant operations, to reinforce the general applicability of the overall cognitive model. Section C.3 provides a good summary of the intent of these examples. I do not have any comments or questions about them.

56. Appendix D, Cognitive Basis for the Combined Effect of Performance-Influencing Factors on Human Error Probabilities

This is an interesting sample of studies that generally support a logical construct which adds, rather than multiplies, the numerical effects from multiple performance-influencing factor (PIF) weights.

The next-to-last bullet item notes that:

> "The individual and combined weights (w_a, w_b, w_{ab}) of PIFs are typically in the range of 0.5–5 and rarely exceed a value of 10."

I think that this observation might be somewhat misleading, and it might inadvertently bias analyst expectations about the combined effects from multiple scenario-specific PIFs. For example, it seems to indicate that the combined effects from multiple PIFs typically result in less than a factor of 10 increase from the baseline human error probability (HEP).

The examples in this appendix are derived from variations in only two PIFs. In practice, several PIFs may be evaluated for a particular human failure event (HFE). It is also noteworthy that one of the six examples illustrates an HEP increase of about a factor of 30, accounting for the combined effects from the two measured PIFs.

In practice, I think that the notion of an "error-forcing context" in some specific PRA scenarios can result in rather large increases in the HEP, compared to the conditions during a nominal response scenario. For example, despite the assessment of a rather low baseline HEP, the composite effects from multiple challenging PIFs may result in a scenario-specific HEP that approaches unity. I think that it is important for the background and the authors' insights in this report to reinforce the need for scenario-specific analyses, acknowledge these compound scenario-specific influences, and not inadvertently create expectations that their overall numerical effect may be rather modest.

For example, a possible undesired consequence from the cited observation might be that analysts will not perform a systematic critical scenario-specific assessment of all relevant PIFs and attributes, because they "know" that the composite effects will not have a very significant influence on the quantified HEP. They might then assert that increasing the baseline HEP by a nominal factor of 10 is adequately "conservative" without a detailed analysis, based on the observation in this summary.

JWS Note: The notion that the PIF weights have a relatively small effect on the HEP is also contrary to the material in Section 4.4.3.2 and Figure 4-8 which suggest that the composite effects from the three "base" PIFs for "Information Availability and Reliability", "Task Complexity", and "Scenario Familiarity" can alter the HEP by a factor of almost 10,000.

Why does the summary of this appendix contain this observation about the numerical effects from combined PIF weights?

If this observation is retained, should the summary also explicitly caution analysts that the combined effects from multiple PIFs that are evaluated during challenging scenario-specific conditions may have a much larger numerical effect on the estimated HEP (and a scenario-specific analysis is always needed)?

57. Appendix E, Section E.1.2, Search for Scenario Context, System Context

The general discussion in this section is very good. However, the sample questions seem to focus analyst attention on the status of specific systems and functions that are needed to mitigate the "safety issue". Those considerations are certainly very important to the scenario context. On the other hand, as noted in the general discussion, the analysts should also examine and document the status of systems that are not directly associated with event mitigation, but which may divert personnel attention from the desired course of action, create conflicting priorities, introduce unexpected time delays, etc.

Should the sample questions explicitly include these considerations, to clearly prompt analysts to examine and document the entire context of what is happening in the plant concurrently with the effects on the identified event mitigation systems?

Should the sample questions explicitly prompt analysts to search for unexpected system responses that are a direct consequence of the event (e.g., fire-induced spurious actuations) or a consequence of automatic control functions (e.g., increasing or reducing flow, level, temperature, pressure, etc. in response to failed or spurious input signals)?

58. Appendix E, Section E.1.2, Search for Scenario Context, Personnel Context

A preceding comment discusses possible consideration of a distinct performance-influencing factor (PIF) for "Concern for Personnel Safety". Even if the authors decide to not adopt that separate PIF, it seems that the analysts should examine and document scenario conditions that affect the personal safety of individuals who must perform the desired actions, cause possible harm or known injuries to their co-workers, or cause possible harm or unknown status of their families and loved ones (e.g., during extreme external events that affect the surrounding area).

Should the general discussion and sample questions for Personnel Context explicitly address this issue?

59. Appendix E, Section E.1.2, Search for Scenario Context, Example 1, Environmental Context

This item notes that:

> "Environmental factors are not considered for main control rooms unless the event is a control room fire."

I disagree very strongly with this statement.

As noted in a preceding comment, event scenarios may involve loss of Main Control Room (MCR) ventilation, room cooling, and normal lighting. These scenarios can occur from a variety of initiating events and system failures.

Fires in locations other than the MCR may result in smoke entering the MCR through ventilation systems.

Seismic events may cause partial damage in the MCR (e.g., suspended ceilings, overhead light fixtures, toppling storage shelves, etc.).

Some plant-specific and site-specific scenarios may result in toxic gases entering the MCR from locations inside the plant or from accidents that occur nearby the plant.

JWS Note: The example is intended to apply specifically for "NPP internal, at-power events". Loss of MCR ventilation, cooling, and lighting can certainly occur during those events. Furthermore, the excerpt inappropriately indicates that analysts need to consider MCR environmental conditions only for fires that occur in the MCR.

What is the technical basis for this assertion?

Why is it included as "NRC-approved" guidance in this report?

60. Appendix E, Section E.1.2, Search for Scenario Context, Example 1, System Context

This item notes that:

> "Personnel operate the reactor systems from a main control room using ***well-designed*** HSIs. System time available for personnel to perform actions ***usually has adequate time margins***." [emphasis added]

The highlighted phrases inappropriately bias an analyst's objective consideration of scenario-specific conditions. For example, the human-system interface (HSI) may not be "well-designed" to facilitate personnel performance in the context of a specific scenario, and the scenario-specific time margins may not be "adequate". These considerations are precisely why the guidance should prompt analysts to perform a systematic, comprehensive, objective examination and documentation of the scenario context, without pre-existing biases that may cause them to inadvertently overlook scenario-specific effects.

What is the technical basis for these assertions?

Why do they apply universally to every plant and to any scenario that may occur?

Why are they included as "NRC-approved" guidance in this report?

61. Appendix E, Section E.1.2, Search for Scenario Context, Example 1, Personnel Context

This item notes that:

> "Fatigue and stress are not assumed to be severe enough to affect crew performance for most internal, at-power events."

Conclusions about personnel fatigue and stress should result from an objective assessment of the scenario-specific context and the corresponding states of the relevant performance-influencing factors (PIFs).

Furthermore, this assertion notes that adverse effects from fatigue and stress are "not assumed" for "most" events. The purpose of a systematic, comprehensive, scenario-specific assessment is to search for those scenarios when these effects may have an important influence on human performance.

Why is this assertion included as "NRC-approved" guidance in this report?

62. Appendix E, Section E.1.2, Search for Scenario Context, Example 1, Task Context

This item notes that:

> "Human actions are those allocated to control room operators in internal, at-power events. Human tasks are prespecified in control room normal or emergency procedures. The tasks mainly involve the macrocognitive functions of *detection, understanding*, and *action execution*. *Decisionmaking* is limited to choosing among alternative strategies and implementing strategies based on procedures."

I disagree strongly with the assertion that the scope of human actions in these applications is limited to only actions that are performed in the Main Control Room (MCR). I am very familiar with many PRA models for internal initiating events during full-power operation which include numerous local operator actions. Assessment of the Teamwork macrocognitive function is needed for coordination of those local actions.

The discussion also seems to indicate that analysts need not pay much attention to the Decisionmaking macrocognitive function, because that element always involves only selection of a procedure-directed strategy which will always match the evolving scenario conditions. This is far from true, especially for complex scenarios and for analyses that extend to a Level 2 PRA.

What is the technical basis for these assertions?

Why do they apply universally to every scenario that may occur?

Why are they included as "NRC-approved" guidance in this report?

63. Appendix E, Section E.1.2, Search for Scenario Context, Example 1 and Example 2, General Comment

I do not have any specific comments on Example 2. I understand that the general intent of these examples is to contrast the kinds of considerations that may be needed during an evaluation of each type of scenario. However, I am concerned that readers will interpret the very narrow focus of Example 1 (as highlighted by my comments) as "NRC-approved" guidance that may cause them to inappropriately overlook potentially important scenario-specific influences. I think that the fundamental message from the methodology should be that analysts need to examine, understand, and document the scenario-specific conditions that have a potentially important effect on human performance, without regard to the PRA model-related treatment of those scenarios or pre-conceived notions about what considerations might be universally unimportant.

I do not think that these examples are necessary to reinforce the basic concepts in Section E.1.2. Those concepts are presented rather well in the discussions of each element of the scenario context. As noted in my comments, in some cases, those discussions could be expanded a bit to capture somewhat broader considerations of specific issues. That would also subtly reinforce the need for a scenario-specific assessment, regardless of the scope of the PRA models.

Rather than spending additional effort to further refine or provide more elaboration in Example 1, the authors might consider simply discarding both examples.

64. Appendix E, Section E.2.1.1, Scenario Narrative

The discussion in this section is good. However, it is focused almost exclusively on information that is related directly to the "safety issues", as they are depicted in the PRA models of the scenario.

For example, the discussion of "Scenario Flow and End State" notes that:

> "The purpose is to provide a good understanding of the scenario flow with sufficient details to perform a detailed HRA. The scenario flow is represented by a number of scenes.

> Principally, a scene describes a safety issue. A scene could have subscenes or a specific topic related to the safety issue. Collectively, the scenes cover all safety issues."

This information is certainly very important. However, it does not necessarily capture all of the potentially important influences on human performance, and the guidance does not explicitly prompt analysts to search for those influences and document them in the narrative. Experience from actual events has shown that these other influences can be important. Restricting the focus of the narrative to describe only the scenario "safety issues" that are depicted in the PRA models may cause the analyst to overlook these influences and perhaps develop an overly optimistic perspective about the desired personnel response.

In principle, the scenario narrative should describe everything that is happening in the plant, because that is the actual context within which personnel must respond. Of course, that ideal can rarely be achieved in a practical analysis. However, the narrative should describe all conditions that may have a potentially important effect on human performance, even if those conditions are not included explicitly in the PRA models. That description helps the analysts to identify and evaluate the states of relevant performance-influencing factors (PIFs) that account for distractions, interruptions, multi-tasking, conflicting priorities, time pressure, stress, etc. It also helps others to understand what conditions were considered by the analysts and to question the reasons for possible omissions.

JWS Note: In the film analogy, that other information would explain why the hero suddenly climbed a tree, when the primary plot line expected him or her to take MacGyver-like actions to repair a broken vehicle. The hero climbed the tree because a grizzly bear had wandered into the background of the scene. It's perhaps a poor analogy, but it shows why it is important to clearly acknowledge that the bear is part of the scenario, even if it subsequently wanders off-camera and is never seen again.

Why does this section not emphasize the need for analysts to examine the totality of what is happening in the plant and document any conditions that may have a potentially important influence on personnel response, even if those conditions are not related directly to the equipment, human actions, and "safety issues" that are modeled explicitly in the PRA?

65. Appendix E, Section E.3.2, Resolution of Uncertainties or Contradictions

The last sentence in the last quoted excerpt in this section notes that the interviewed analysts stated that they would "… discuss the uncertainty with other HRA and PRA analysts and with the operating personnel to agree on how to deal with this in the HRA (usually by making conservative assumptions), and then document this qualitatively in their analyses".

This final observation seems to leave readers with the impression that uncertainties in human reliability analysis (HRA) are typically addressed either qualitatively or by the use of "conservative assumptions". I think that this report should emphasize quantitative evaluation of uncertainties throughout the HRA process, to the greatest extent possible. That is consistent with NRC guidance on the evaluation of uncertainties (e.g., NUREG-1855). It also provides a more realistic assessment of the scenario-specific human error probabilities (HEPs) and their contributors, compared to a "conservative" analysis.

Should this section conclude with an explicit recommendation to quantify uncertainties throughout the HRA process, to the greatest extent possible?

66. Appendix E, Section E.4, Interaction with Other Disciplines for Analysis, General Comment

This section is very good!

67. Appendix F, Section F.1, Identification of Pre-Initiator Important Human Actions

This section notes that:

> "Component reliability is calculated based on component performance data which include latent failures contributed by human error. Therefore, the pre-initiator IHAs are ***typically not modeled separately***." [emphasis added]

It also notes that:

> "***If the analysts want to separately model*** the pre-initiator IHAs, the analysts should ensure no overlap between the component reliability contributed by human error and by the other causes. For nuclear reactor safety, the pre-initiator IHAs, ***if separately modeled***, are typically the actions of calibration and alignment, affecting multiple trains or multiple systems." [emphasis added]

I disagree very strongly with these statements.

It is certainly true that some types of maintenance and calibration errors may not be discovered by post-maintenance testing. Those errors are eventually revealed by equipment failures. Their prevalence depends on the plant-specific maintenance practices and the effectiveness of the plant-specific post-maintenance inspection and testing protocols. Most PRAs do not separately quantify these causes for equipment failure, because it is too resource-intensive to extract them from the composite equipment performance records.

All PRAs explicitly identify, model, and quantify many pre-initiator human actions. Examples are errors that do not restore equipment to their normal alignments after maintenance and testing activities, miscalibration of instrumentation and signal processing logic, etc. A typical PRA model may contain dozens of these pre-initiator human failure events (HFEs). In some cases, depending on the plant-specific design, maintenance protocols, and testing frequencies, these types of errors can be important contributors to risk. It is not appropriate for "NRC-approved" guidance to imply that these pre-initiator actions are "typically not modeled" or that analysts have an option to overlook them.

JWS Note: The rest of this section discusses identification and modeling of precisely the types of pre-initiator actions that are noted in this comment. However, the highlighted text from the introduction sets an inappropriate context and analyst expectations for the entire topic.

JWS Note: This comment also applies to the second bullet item in Section F.6.

Why does this section indicate that "pre-initiator IHAs are typically not modeled separately"?

68. Appendix F, Section F.1, Identification of Pre-Initiator Important Human Actions

This section quotes an excerpt from EPRI 3002008094, "Data and Modeling of Pre-Initiator Human Failure Events in Probabilistic Risk Assessment". I do not have a copy of that report. It is not apparent why this excerpt is relevant to a search for plant- and system-specific maintenance and testing activities which may involve human errors that leave equipment misaligned.

The excerpt refers to the risk assessments that are performed as part of the Maintenance Rule programs. Those risk assessments typically examine the effect on plant risk while the equipment is removed from service for maintenance. They do not typically evaluate the effect on risk if personnel fail to correctly realign the system for normal operation after the maintenance is completed. For example, a plant-specific post-maintenance test may start a pump and confirm that it develops adequate pressure, but the test may not functionally confirm that all valves in the pump discharge flow paths are open (e.g., by measuring flow in all paths that are isolated during the maintenance activity). Thus, the post-maintenance test is not fully effective for discovery of personnel errors that leave a valve closed, and the PRA should explicitly model those possible errors.

Why is this excerpt from the EPRI report relevant to the identification of maintenance and testing activities which may involve human errors that leave equipment misaligned?

69. Appendix F, Section F.1, Identification of Pre-Initiator Important Human Actions

This section lists four screening criteria for pre-initiator human actions from the SHARP1 guidance. I do not agree fully with the implications of Criteria (1), (3), and (4).

For Criterion (1), scenarios may involve failures of the actuation signals or power supplies that are needed to correctly configure the misaligned component. In those scenarios, the human error that left the component misaligned contributes directly to its functional failure. In these situations, it is better to model the human error in combination with the logic and support systems that are needed to correctly configure the component, if it is misaligned, including hardware failures of the component itself. These considerations can be important for risk-significant events that involve failures of instrumentation, actuation and control signals, AC power, and DC power.

For Criterion (3) and Criterion (4), it is not apparent how analysts can judge that the reconfiguration frequency is sufficiently "low" or that the verification frequency is sufficiently "high" to justify a conclusion that the risk contribution can be screened out. In many cases, it is easier to simply model these effects explicitly, rather than try to justify why they are numerically "insignificant".

Should this section alert analysts to these considerations?

70. Appendix F, Section F.1, Identification of Pre-Initiator Important Human Actions; Section F.6, Summary

I am concerned that the last two bullet items Section F.1 may be misinterpreted. The first item addresses errors that affect multiple components simultaneously (e.g., miscalibration of multiple redundant sensors). The second item more generally addresses consideration of errors that may affect a single component. I am concerned that readers may inappropriately interpret these items in an "and" context. In other words, they may interpret the guidance to imply that they should search for errors that have the characteristics of the second bullet, and only search for those errors which can disable multiple components simultaneously. I think that the guidance should be clarified to avoid that potential misinterpretation.

The second bullet item in Section F.6 strongly reinforces my concern. It notes that:

> "If modeled in reactor risk assessment, the pre-initiator IHAs typically are the human actions of calibration or system alignment and affect ***multiple trains or multiple systems***; the initiator IHAs typically are the human actions of working on the support system that could affect ***multiple trains or multiple systems***." [emphasis added]

This is not consistent with general human reliability analysis (HRA) guidance or normal practice. In particular, pre-initiator errors should be modeled if they can leave one or more components disabled. Depending on the equipment and the plant-specific practices, the probability of those errors may be substantially higher than the probability of the respective equipment hardware failures (even for individual components).

Errors that contribute to initiating events may also affect only one component. For example, a normally-running cooling water pump may fail, and the operators may fail to start the standby pump manually. The resulting initiating event frequency is determined by the combination of a hardware failure and a single operator error, rather than an operator error that simultaneously disables both pumps.

Why does this "NRC-approved" guidance imply that analysts should search only for errors that can disable multiple components simultaneously?

71. Appendix F, Section F.1, Identification of Errors of Commission

The first bullet at the end of this section notes that:

> "The action directly disables or does not start the system, subsystem or component needed to provide the system function required in the scenario."

Without further elaboration, the "does not start" part of this summary seems to primarily characterize an error of omission (EOO), rather than an error of commission (EOC). The summary also seems to focus narrowly on actions that stop equipment that is needed for mitigation of the nominal event scenario.

An important consideration in the cited excerpts is that an EOC alters the event scenario progression in a way that is not anticipated by the PRA models for the nominal event sequence, and is not evaluated by the identified EOO human failure events (HFEs). For example, the excerpt from NUREG-1921 notes that an EOC might transform a transient scenario into a consequential LOCA scenario. This context is different from inadvertently disabling or failing to start equipment that is needed to mitigate conditions in the nominal scenario.

JWS Note: The third bullet item contains some of this notion in the context of "a cascading effect on the scenario progression", but it remains focused on actions to turn on or turn off pumps.

Should the first bullet item explicitly address the notion that an EOC alters the scenario progression from that anticipated by the PRA models for the nominal (or "baseline") event sequence?

72. Appendix F, Section F.5, Composition of Important Human Actions, General Comment

This is a good example. Figure F-2 clearly shows how decisions to simplify and combine intermediate actions can affect the analyses of the time during which DC power is available.

73. Appendix G, Task Analysis, General Comment

This appendix is generally very good. The concepts are easy to understand, and the examples appropriately illustrate the basic thought process (with the possible exception of some parts of the example in Section G.4.2, as discussed below).

74. Appendix G, Section G.4, Task Analysis Techniques

The introduction to this section notes that:

> "Kirwan and Ainsworth [105] examined many task analysis techniques and recommended ***10 techniques*** alone or in combination for HRA purposes. Taylor and Le Darz [100] piloted ***five task analysis techniques*** in an HRA case study and demonstrated that a ***combination of different task analysis techniques is typically more effective*** than the use of a single technique….This section introduces ***several task analysis techniques*** that have proven useful for HRA." [emphasis added]

This section seems to discuss only two task analysis techniques. It contains a rather brief (but useful) discussion of the "Hierarchical (Functional) Task Analysis and Tabletop Analysis" technique. Almost all of the section is devoted to a detailed discussion of the "Crew Response Diagram" (CRD) technique.

Why is the "NRC-approved" guidance limited to only these two examples?

Why does this section not provide at least brief summaries of other task analysis techniques that are recommended and tested in the reference literature?

75. Appendix G, Section G.4.2, Crew Response Diagram

The discussion of "Determine the Relevant Cues and their Timing" notes that:

> "It is particularly important to identify the cues that lead personnel to enter the correct procedure(s) for the important human action."

This is certainly true. However, for complete understanding of the scenario context and conditions that may affect the performance-influencing factor (PIF) assessments, it is also important to identify cues that may divert attention, cause confusion, or distract personnel from the desired procedural response, and their timing relative to the relevant cues.

The guidance in Appendix F discusses the identification and evaluation of errors of commission (EOC). Therefore, it is also important to identify specific cues that may trigger those actions (e.g., as an alternative to the desired response), and their timing.

Should this section also mention these considerations in the context of the crew response diagram (CRD) construct?

76. Appendix G, Section G.4.2, Crew Response Diagram, Figure G-7, Crew Response Diagram for HFE-3 in the Example of NPP Fire Event

Figure G-7 shows the example Crew Response Diagram (CRD) for the fire event. The summary indicates that no critical tasks are associated with Node 0, Node 1, or Node 2. I am not sure whether those brief conclusions are appropriate for an example that is intended to illustrate how analysts should systematically consider possible alternative responses and the effects from those alternatives.

I am certainly not familiar with the procedures and training at the example plant. For the purpose of this comment, I will (perhaps optimistically) assume that the crew always enters Emergency Operating Procedure (EOP) Path-1 whenever a reactor trip occurs. Therefore, it might be reasonable to indicate that Node 0 is "for information only". However, I think that the example should better explain the analysts' reasons for that conclusion.

At Node 1, there seems to be a decision point to enter procedure EPP-4, based on the conclusion that no safety injection (SI) is needed. It is not apparent what alternatives might arise if the crew concludes that an SI is imminent, or if they do not correctly select procedure EPP-4. Therefore, it is not apparent why no critical cognitive tasks are associated with Node 1.

The event summary indicates that the initiating event occurred at 18:52. The timeline in Figure G-8 indicates that an automatic SI occurred at 19:00, due to an excessive cooldown. Figure G-8 also indicates that the crew was actively involved with other actions during that 8-minute period, such as starting charging pumps B and C. (They were also involved with actions related to the electrical fault, the fire, and their effects on other systems.)

The summary of Node 2 indicates that the crew re-entered EOP Path-1 at 19:00 when they detected the SI signal. To achieve success at Node 2, it seems that the crew must detect the SI signal and correctly decide to re-enter Path-1. It is not apparent what alternatives might arise if the crew did not re-enter Path-1 at that time (i.e., if they continued with the normal progression in EPP-4). For example, the discussion of Node 3 seems to indicate that the crew is prompted to check RCP thermal barrier cooling at a specific diamond step in EOP Path-1. Therefore, it is not apparent why no critical cognitive tasks are associated with Node 2.

JWS Note: The action to start the charging pumps at 18:53 may be relevant to the CRD, because it indicates that the crew was initially aware of the importance of charging flow (and perhaps seal injection flow). The success criteria for HFE-3 specifically require the operators to re-open component cooling water (CCW) valve FCV-626. The CRD is developed to address those specific success criteria. However, if the functional success criteria were to restore RCP seal cooling, that function could be accomplished by restoring seal injection or CCW flow. In that broader functional context, it seems that the initial actions to start the charging pumps are relevant to the overall analysis, and the CRD logic should be revised accordingly. For the purpose of this focused example and my comments, I will not pursue this issue further. However, it illustrates how the CRD structure and its depiction of the crew responses are affected by the specific HFE definition and success criteria.

Why are no critical cognitive tasks associated with Node 1?

How would the crew response differ if they do not enter EPP-4 at Node 1?

Why are no critical cognitive tasks associated with Node 2?

How would the crew response differ if they do not re-enter Path-1 at Node 2?

In particular, how and when would the crew be prompted to check the status of RCP seal cooling if they remained in EPP-4?

Why does the example not further explain why no critical cognitive tasks are associated with Node 0, Node 1, and Node 2 (i.e., to clearly illustrate the desired analyst thought process and needed documentation)?

77. Appendix G, Section G.4.2, Crew Response Diagram, Figure G-7, Crew Response Diagram for HFE-3 in the Example of NPP Fire Event

Figure G-7 shows the example Crew Response Diagram (CRD) for the fire event. One of the critical tasks listed for Node 4 is:

> "Check elapsed time since all RCP Seal Cooling was Lost"

Node 5 seems to address the decision to transfer to Step 10 of AOP-018, based on whether RCP seal cooling has been lost for longer than 15 minutes. Therefore, it is not apparent why the activity to check the elapsed time is a critical task for Node 4, rather than Node 5.

As it is represented in the CRD, the failure path from Node 4 would occur if the crew fails to verify seal injection flow, or if they fail to check the elapsed time for loss of all seal cooling. The summary of Node 7 indicates that the RCP bearing high temperature alarms may alert the crew to loss of seal cooling, even if they do not identify the loss of component cooling flow in Node 3 and the loss of seal injection flow in Node 4.

If the crew does not verify the duration in Node 4, it is not apparent how the potential recovery path from Node 7 would return them to the entry conditions for Node 5. In other words, if the operators do not verify the duration in Node 4, it seems that they would also fail to verify the duration in Node 5, and never enter Step 10 of AOP-018. Thus, it seems that the activity to check the elapsed time for loss of all seal cooling is functionally a critical task for Node 5, rather than Node 4.

Why is the activity to check the elapsed time for loss of all seal cooling identified as a critical task for Node 4, rather than Node 5?

78. Appendix G, Section G.4.2, Crew Response Diagram, CRD Timeline

The timeline in Figure G-8 is a very good depiction of the scenario evolution and the expected timing of the activities that are evaluated for HFE-3. The bullet items after Figure G-8 summarize timing information that was obtained through discussions with the plant staff.

Section 5 of the main report contains guidance for evaluating the effects from uncertainties in the time that is available for an important human action and the time that is needed to complete all elements of the action. I think that it is very important for all examples in the report to reinforce the notion that uncertainties should be identified, documented, and quantified as a necessary and integral part of the information collection process.

JWS Note: The example notes that the time available to restore seal cooling is 19 minutes.

Based on analyses of RCP seal performance, there is also uncertainty in that estimate. However, for the purpose of this comment, I will focus only on uncertainties in the estimated times for performance of each activity, as derived from the operator interviews.

Why does this example not document and quantify the uncertainties in the time estimates that were provided by the plant staff?

79. Appendix G, Section G.4.2, Crew Response Diagram, CRD Timeline

This comment is related to the preceding comment. However, it addresses a consideration that is related to the functional effects from uncertainties in the time estimates, but not necessarily related directly to the analyses that are described in Section 5 of the main report. In particular, it is related to possible alternatives to the Crew Response Diagram (CRD) logic structure, or the evaluation of specific nodes in the CRD.

The timeline in Figure G-8 shows that the crew will reach the diamond step in Path-1 approximately 30 minutes after the initiating event (i.e., at 19:22). At that time, no component cooling water (CCW) flow is available to the RCP seals and no seal injection flow is available. In particular, CCW flow was lost at 18:52, and adequate seal injection flow was lost at 19:19. Thus, if the operators follow the procedural guidance at that time, they have an opportunity to confirm that CCW flow is not available (CRD Node 3), and they have an opportunity to confirm that no seal injection is available (CRD Node 4).

Suppose that the crew reaches the Path-1 diamond step at approximately 20 minutes after the initiating event (i.e., at 19:12). At that time (or any other time before 19:19), no CCW flow is available to the RCP seals, but seal injection flow is available. If the operators follow the procedural guidance at that time, they have an opportunity to confirm that CCW flow is not available (CRD Node 3), but they will conclude that adequate seal cooling is available. Thus, they may not take actions to reopen valve FCV-626 until some other compelling cue alerts them to re-check the status of seal cooling (e.g., the RCP bearing high temperature alarms at 19:24 and 19:30). In this evolution of the scenario, it seems that the failure path from Node 4 would always apply when the operators reach that step, and the operators would always need to rely on the recovery cues in Node 7 to prompt their entry into AOP-018 and to achieve eventual success of HFE-3. Thus, it seems that the estimated time to reach the Path-1 diamond step could have a functional effect on the CRD logic structure, and it would almost certainly have an effect on the evaluation of the critical tasks for Node 4 (and possibly Node 7).

JWS Note: This comment illustrates a situation where it is not necessarily "conservative" to always assign a long time to reach a specific point in the procedures. In some scenarios, the operators may not take the desired actions if the combined prompting conditions do not exist at the time when they pass through a specific decision step, but the conditions evolve at a later time.

Should the discussion of this example alert analysts to these considerations about the functional effects from uncertainty in the time estimates and how they may influence the evaluation of specific nodes in the CRD?

Should the example explicitly show how the CRD logic structure or the evaluation of Node 4 (and possibly Node 7) would be altered if the crew reaches the Path-1 diamond step before 19:19?

80. Appendix H, Identification of Cognitive Failure Modes, General Comment

This appendix is very good. The concepts are easy to understand, and the examples appropriately illustrate the basic thought process. Section H.2 contains a very good discussion of cautions for pruning and adapting the set of detailed cognitive failure modes (CFMs) that are evaluated in a specific application. The examples illustrate how an analyst should justify decisions about elimination or adaptation of specific CFMs.

81. Appendix H, Section H.1.3, Behavioral Characteristics of Processor Failure as Detailed Cognitive Failure Modes, Example 1

The examples in this section illustrate how the detailed cognitive failure modes (CFMs) account for the situations that a processor is "not achievable", "achievable, but personnel do not perform it", or "achievable, but personnel perform it incorrectly". Example 1 is noteworthy, because it is the only example that does not list the complete set of detailed CFMs for the respective processor. In particular, Example 1 does not list or discuss detailed CFM D1-2:

D1-2 Wrong mental model for detection (e.g., incorrect planning on when, how, or what to detect)

Why does Example 1 not list and discuss detailed CFM D1-2?

82. Appendix I, Section I.1, The Process of Assessing Performance-Influencing Factors

Step (1) in this section notes that:

> "Many PIFs may not be ***relevant*** to the context; therefore, they are ***not selected***. If a PIF is not selected, a rationale must be given for why it is not relevant to the application." [emphasis added]

I am somewhat concerned that this terminology may not provide the appropriate context and desired guidance for analysts. The PIFs described in Section 3 of the main report are defined and organized in a way that is intended to systematically account for all elements of human cognitive performance. Therefore, it seems that all 20 PIFs are always "relevant" to every human failure event (HFE). Of course, an analyst may conclude that a subset of the 20 PIFs has potentially more significant effects on human performance during a particular scenario, and the analyst may then decide to focus attention on only those PIFs. The analyst must then justify why the other PIFs, and their composite effects, do not merit further scrutiny.

I think that this is the functional intent of the excerpt. However, I am concerned that analysts may too quickly and too simply dismiss individual PIFs as globally "irrelevant", rather than explain the rationale and cognitive basis for their selection. In particular, in my mind, all 20 PIFs are always "relevant" to every HFE. If I decide to focus attention on only 6 of those PIFs for a particular HFE, I must then justify why the other 14 PIFs, and the composite effects from those 14 PIFs, do not have a significant influence on altering the nominal expected human cognitive performance (i.e., the baseline HEP). Depending on the analyst's interpretation of the term "relevant", that perspective may result in different decisions and their justification.

JWS Note: The discussion of PIF attribute selection in Step (2) contains the appropriate notion. In principle, all attributes are "relevant". However, for practical considerations, the analyst must decide which specific attributes may not "contribute significantly" to the scenario-specific analysis and justify that rationale.

Rather than emphasizing the notion of "relevant" PIFs, is it better to indicate that an analyst may judge that some PIFs, including their composite effects, may not have a significant influence on altering the nominal expected human cognitive performance during the scenario that is being evaluated? The analyst must then provide the rationale and justification for why those PIFs do not merit further detailed evaluation.

83. Appendix I, Section I.2, Demonstration of Assessment of Performance-Influencing Factors in Nuclear Power Plant Control Rooms

The introduction to this section notes that:

> "This section documents some considerations for PIFs in NPP control room operations as an example to ***demonstrate assessment of PIFs in a specific application***. The information is intended to facilitate readers' understanding of PIFs. The example does not provide an assessment of all PIFs in the list. Instead, this example assesses ***only some PIFs that may challenge the macrocognitive functions***." [emphasis added]

The preceding comment addresses my concerns about quickly dismissing consideration of specific performance-influencing factors (PIFs) because they are not "relevant" to a particular human failure event (HFE). This section reinforces those concerns.

The general cognition model indicates that every human cognitive response involves the macrocognitive functions of Detection, Understanding, Decsionmaking, and Action Execution. In my broad construct, the macrocognitive function of Teamwork also applies for within-team dynamics. The 20 PIFs described in Section 3 of the main report are defined and organized in a way that is intended to systematically account for all elements of human cognitive performance. The tabulations in Appendix B show how individual PIF attributes can affect multiple cognitive mechanisms, processors, and macrocognitive functions.

As an example of how an analyst should systematically consider and evaluate the PIFs in a practical analysis, I think that it is very important for this appendix to demonstrate the assessment of all 20 PIFs. If it is judged that some PIFs are not potentially important for this particular application, then the example should show how an analyst should document and justify that conclusion. Otherwise, the example implicitly endorses the notion of selectively focusing on a specific set of PIFs, without adequate scrutiny of the others and justification for why they are not examined in detail.

Why does this example not systematically address all 20 PIFs and justify the analyst's decision why specific PIFs are judged to not "challenge the macrocognitive functions"?

What is the overall intent and purpose of this section of Appendix I?

How does this appendix provide guidance and examples of practical PIF considerations that are consistent with a systematic, comprehensive, scenario-based analytical thought process?

84. Appendix I, Section I.2.1, Performance-Influencing Factor Considerations for Detection

This comment is a continuation of the preceding comment, using the example of the Detection macrocognitive function to illustrate specific reasons for my concern.

The tables in Appendix B list the following performance-influencing factors (PIFs) which have one or more attributes that are linked to the Detection macrocognitive function.

- Work Place Accessibility and Habitability
- Work Place Visibility
- Noise
- Cold / Heat / Humidity (overarching effects)
- System and I&C Transparency to Personnel
- Human-System Interface
- Staffing
- Procedures, Guidance, and Instructions
- Training
- Teamwork and Organizational Factors
- Work Processes (overarching effects)
- Information Availability and Reliability
- Multi-tasking, Interruptions, and Distractions
- Task Complexity
- Mental Fatigue (overarching effects)
- Time Pressure and Stress

Thus, in the taxonomy of the IDHEAS cognition model, these 16 PIFs are "relevant" to the Detection macrocognitive function. An analyst must then examine each PIF in the context of the specific scenario that is occurring and the specific human failure event (HFE) that is being evaluated to determine whether the PIF merits special attention (i.e., whether it might improve or adversely affect the expected baseline human performance).

The example in this section seems to indicate that analysts should consider only the PIFs for "Human-System Interface" and "Procedures, Guidance, and Instructions" when they examine scenario-specific issues that may affect Detection in the Main Control Room. The example does not explain the authors' rationale why none of the other PIFs might "challenge the macrocognitive function" for Detection. Therefore, I am concerned that readers of this example, noting its emphasis by the methodology authors, will use it as evidence for why it is acceptable to quickly focus attention on only a couple of "obvious" PIFs, without the systematic assessment of scenario-specific influences that should be advocated throughout the report and its examples.

In particular, it is not apparent why this example does not discuss considerations for any of the other 14 potentially influential PIFs, in a manner similar to the general discussions of the two highlighted PIFs.

JWS Note: Of course, similar comments apply to the specific PIFs discussed in the examples for Understanding, Decsionmaking, and Action Execution. Furthermore, a discussion of the PIFs for Teamwork is notably absent. I intentionally did not comment on specific details of the selected PIFs for any macrocognitive function, because I am more concerned about how prospective analysts will interpret these examples as guidance for inappropriately narrow assessments, without adequate justification for their specific PIF selections.

Why does this example not describe the authors' rationale for examining only 2 of the 16 PIFs that may affect Detection in the Main Control Room?

If this example is intended to apply generically for any scenario that requires Detection in the Main Control Room, what is the authors' rationale for the conclusion that none of the 14 other PIFs will ever merit consideration as a potentially important influence on Detection?

85. Appendix I, Assessment of Performance-Influencing Factors, General Comment

Neither the main text nor any other appendix refers the reader to Appendix I. I read it after Section 3 and Appendices B, C, and D, because it seemingly contains additional information and examples that help an analyst to understand how to assess performance-influencing factors (PIFs). However, after reading Appendix I, I am not sure why it is included in the report. It is not evident that it provides substantially useful information that is not covered elsewhere in the report. The preceding three comments address my concerns that the examples in this appendix might be interpreted as endorsement of a process that too quickly dismisses large numbers of PIFs without adequate scenario-specific consideration and justification. Rather than spending additional effort to provide further elaboration, the authors might consider simply discarding this appendix.

86. Appendix J, Section J.3, Estimation of Human Error Probability Distribution

Section 4.4.3.2 in the main report briefly discusses the conceptual construct of the "IDHEAS-G Human Error Probability Quantification Model". Preceding comments on that section address my questions about that model and whether it should be endorsed in this report. Appendix J does not contain any information about that conceptual construct, any information about its anchors to empirical human performance data, or any guidance about how the model might be applied in practice.

Why does Appendix J not contain any discussion of the quantification model that is discussed in Section 4.4.3.2 of the main report?

87. Appendix J, Section J.3, Estimation of Human Error Probability Distribution

The last sentence in the introduction to Section J.3 notes that:

> "The next three sections briefly describe three approaches for estimating HEPs: data generalization and extrapolation, Bayesian estimation, and expert judgment."

Section J.3.1 discusses Bayesian Estimation, and Section J.3.2 discusses Expert Judgment. I could not find a discussion of an approach for "data generalization and extrapolation". Section J.4 briefly discusses some considerations for the interpretation and use of empirical data. However, that section does not address an approach for estimating human error probabilities (HEPs) from the data.

JWS Note: I think that Section 6 of the main report addresses this approach for estimating HEPs. However, since it is mentioned in the introduction to Section J.3, it seems that this appendix should provide some discussion, elaboration, or examples of that approach (or not mention it at all).

What is the basic concept of HEP estimation by "data generalization and extrapolation"?

Why does Section J.3 not discuss that approach?

88. Appendix J, Section J.3.1, Bayesian Estimation

This section notes that:

> "An informative prior is based on ***historical data*** on the distribution of the parameter of interest." [emphasis added]

In the most general form of the Bayesian process, an underlying prior distribution is first derived from the state of knowledge of informed domain experts. A two-stage Bayesian process next updates that state of knowledge with relevant generic industry experience. The process then updates that generic probability distribution with plant-specific or application-specific evidence.

- The expert state of knowledge might be characterized as "uninformed". In that case, the prior distribution provides equal probability that the actual value of the parameter lies anywhere within an unconstrained (or often constrained) range of values.

- The expert state of knowledge might be characterized as "informed". In that case, the shape of the prior distribution is derived from the experts' assessments of their confidence (probability) that the actual value of the parameter lies within more finely defined portions of the relevant range of values.

Thus, it is not true that an "informative" prior must be derived from "historical data". That notion is very important for human reliability analysts to understand, because many of the human error probability (HEP) estimates in the literature are derived primarily from expert judgment. The IDHEAS methodology also discusses the use of expert judgment to develop HEP estimates. In practice, those expert estimates might then constitute an "informed" prior for Bayesian updates as more relevant empirical data are compiled.

Why does this section indicate that an "informative prior is based on historical data"?

In particular, why does this section not acknowledge that an "informative prior" can also be based on expert judgment?

89. Appendix J, Section J.3.1, Bayesian Estimation

This section notes that:

> "A noninformative prior is one that expresses ignorance as to the true value of the parameter being estimated. In general, a noninformative prior is one in which ***data dominate*** the Bayesian update." [emphasis added]

I do not understand the meaning, intent, or technical basis for the highlighted concept in the second sentence. In each stage of the Bayesian process, the effects from the observed data depend on the range and the shape of the underlying prior distribution and the amount of data that are available.

What is the meaning, intent, and technical basis for the highlighted concept in the second sentence from this excerpt?

Why is that sentence needed to explain the basic notion of a non-informative prior?

90. Appendix J, Section J.3.1, Bayesian Estimation

This section notes that:

> "The frequently used probability distribution functions for the prior include the normal, lognormal and beta distributions. It is generally considered that normal or lognormal distribution is better in modeling physical phenomena while the beta distribution is better for modeling probabilities. Yet, there have been no data-based studies comparing the applicability of these functions in modeling HEPs."

In many cases, the most informative representation of the uncertainty in a parameter value may be provided by a discrete probability distribution that does not have a defined analytical form. Discrete probability distributions are used extensively in PRAs. For example, the NRC State-of-the-Art Reactor Consequence Analysis (SOARCA) project quantifies several important uncertainties using that format. Experience has also shown that discrete probability distributions are often the best and most efficient format to represent the uncertainty from an expert elicitation process.

Why does this section not explicitly acknowledge that a discrete probability distribution, without a pre-defined range or analytical form, may be used to characterize the uncertainty when one of the cited functions does not adequately represent the evidence or expert assessments?

91. Appendix J, Section J.3.2, Expert Judgment

This section is very good. The summarized process is certainly relevant and important when domain experts develop estimates for human error probabilities (HEPs) that will be used as anchor points or fundamental references for an application-specific quantification method (e.g., as in NUREG-2199). However, I am concerned that readers may conclude that the recommended process is much too complex and resource-intensive for practical use in a human reliability analysis (HRA).

In particular, the general concept of "expert elicitation" also applies to the derivation of information and estimates from interviews with plant personnel (e.g., probability distributions for the timing of events, time needed to perform specific tasks, etc.). Many of the basic expert elicitation principles apply for the conduct of those interviews and the use of the experts' estimates. However, the formal process that is described in this section and shown in Figure J-1 is certainly not intended for elicitation of that information.

Should this section explicitly acknowledge that the process shown in Figure J-1 is not needed for the elicitation of expert information during interviews with plant personnel?

Should this section highlight which basic principles of the elicitation process should be followed when estimates of parameter values and associated probabilities are derived from interviews with plant personnel?

92. Appendix K, Section K.1.1, Dependency Types

This section summarizes three general types of dependency: Consequential Dependency, Resource-Sharing Dependency, and Complacency Dependency.

It seems that a fourth general type of dependency exists. It does not seem to fit into these three categories, as they are explained in the text and the illustrative examples. For the purpose of this comment, I will call it (perhaps too simply) Cognitive Dependency. For example, there may be dependence among the performance of several tasks because the personnel performing those tasks have an incorrect mental model of the scenario, inadequate training for the scenario-specific conditions, preconceived expectations, a bias that steers them to an inappropriate preferred course of action, etc. This type of dependency seems especially relevant for the concepts and guidance in the IDHEAS methodology. It is also addressed in contemporary guidance for the identification and evaluation of important factors that affect dependencies, such as Section 6.2 and Figure 6-1 in NUREG-1921.

JWS Note: The first sentence in Section K.2.1 refers to "cognition demands (e.g., common mental models for understanding and decisionmaking)" as a factor that affects dependence. The discussion of "Cognitive Foundation for Dependency" in Section K.3.1 explicitly addresses this concept.

JWS Note: The Davis-Besse example might illustrate this type of dependency, rather than a consequential dependency, due to the common "mindset" and the observation that "the station did not appreciate" the possible effects from boric acid corrosion.

Why does this section not discuss this general type of dependency?

Which of the three types of dependency that are discussed in this section account for this concept?

93. Appendix K, Section K.1.2, Dependency in Pre- and Post-Initiators

The excerpt from NUREG-1792 notes that:

> "Dependencies among the ***pre-initiator*** human failure events, and hence the corresponding human error probabilities in an accident sequence, should be quantitatively accounted for in the PRA model." [emphasis added]

The text in this section also notes that:

> "The PRA model should quantitatively account for dependencies among the ***post-initiator*** HFEs and hence the corresponding HEPs in an accident sequence by virtue of the joint probability used for the HEPs." [emphasis added]

The discussion does not address the treatment of dependencies between human failure events (HFEs) that contribute to the occurrence of an initiating event and HFEs that are evaluated in the models for response to that event. These types of dependence may have a very important influence on the overall evaluation of human performance throughout the scenario.

For example, an initiating event may involve failure of a normally-operating cooling water pump, followed by personnel failures to start the standby pump before temperatures exceed equipment trip setpoints. (This might be an extended period of time, depending on the plant-specific design, equipment trip setpoints, etc.) The reasons why the personnel did not start the pump, and the time during which that error persisted, can affect the evaluation of personnel actions that are needed to cope with the loss of cooling event. That is especially true for actions that are required within a relatively short time after the initiating event (i.e., the plant trip) occurs.

In practice, the same methods are used to identify, evaluate, and quantify the HFEs that contribute to initiating events and the HFEs in the post-initiator response models. Those actions are conceptually and functionally the same. They are actions that are taken in response to a triggering cue, are needed to achieve a desired outcome, and must be accomplished within a functionally-determined time window. Therefore, the same techniques for identifying and evaluating sources of dependence among post-initiator HFEs apply for treating the dependencies between initiator HFEs and post-initiator HFEs.

I think that it is important for this section to explicitly address the need to quantitatively account for these dependencies, because they have often been overlooked by other methodologies.

Why does this section not explicitly identify the need to quantitatively account for dependencies between HFEs that contribute to the occurrence of an initiating event and HFEs that are evaluated for response to that event?

94. Appendix K, Section K.2.1, Factors to Assess Dependence Influence

This section emphasizes the THERP guidelines for treatment of dependencies. Furthermore, most of the section is devoted to a discussion of factors that affect dependencies among pre-initiator actions.

JWS Note: As a side comment, the introduction to the last set of bullet items in this section indicates that they apply for pre-initiator dependence. However, the considerations seem to apply for an evaluation of post-initiator dependence. I did not check NUREG-1842 for the source of that list. The authors should confirm whether it applies for pre-initiator or post-initiator actions.

This section briefly mentions the guidance in NUREG-1921 and notes that:

> "The Fire HRA Guidelines [75] present a traceable table that defines the correspondence of the dependency factors to dependency levels, and the table follows this general principle. The Fire HRA guidance does not provide justifications for specific level assignment. Nevertheless, using the table would greatly reduce the variability that different analysts may produce."

Section 6.2 and Figure 6-1 in NUREG-1921 provide a structured process for consideration of the factors that may influence dependencies. Of course, one might question or disagree with details of specific assumptions or recommendations in that guidance. However, the conceptual framework provides some structure and a source of consistency for these evaluations. It is also noteworthy that NUREG-1921 is "relatively" recent (published in 2012), and it is a joint NRC/EPRI publication.

JWS Note: Section 7.3 of NUREG-2199 does not contain the decision framework and guidance from NUREG-1921. It simply refers to NUREG-1921 as an existing dependency model.

Why does this report not explicitly present and discuss the decision framework that is shown in Figure 6-1 of NUREG-1921 (e.g., to supplement the discussion of the THERP guidance)?

95. Appendix K, Section K.3, Insights on Advancing the State of Practice for Dependency Analysis, General Comment

This section is very good!

96. Appendix K, Section K.3.3, Insights into Modeling Dependency in IDHEAS-G

The simple procedure example discussed in the context of Figure K-1 notes that:

> "It is reasonable to assume that HFE 1 failure creates a cognitive state such that skipping a key procedure step is more likely in HFE 2. Therefore, HEP22 is expected to be larger than HEP21. This is because HEP21 includes only ***independent failures***, but HEP22 includes both independent failures and the failure caused by dependency." [emphasis added]

The discussion continues to indicate that the different context for HEP22 can result from consequential dependency, complacency dependency, or resource-sharing dependency. A preceding comment on Section K.1.1 addresses the notion of "cognitive dependency". Those cognitive effects might actually result in a lower value for HEP21, compared to an independent assessment of action HFE2 without preceding success of HFE1. In particular, if personnel have taken the appropriate actions for HFE1, there is empirical evidence that they understand what is happening and have developed an appropriate response strategy (at least, up until that point in the scenario). Therefore, in principle, the evaluation of some performance-influencing factors (PIFs) for HEP21 should account for that "positive" cognitive coupling. Of course, the evaluation of HEP21 may need to address some sources of cognitive errors, and it should account for the probability of un-recovered implementation "slips". However, the numerical value for HEP21 may be lower than the "independent" value for HEP2, if action HFE2 were evaluated without considering the preceding personnel successes.

It seems that this notion of "positive" cognitive dependence would be useful to further reinforce the subsequent discussion of the examination and treatment of the context for HFE2.

Should this section also discuss the notion of "positive" cognitive dependence in the context of HEP21 in Figure K-1?

97. Appendix K, Section K.4, Minimum Joint Error Probability, General Comment

This is a good, and appropriate, discussion of this topic.

98. Appendix L, Uncertainty Analysis and Documentation, General Comment

The discussions and general guidance in this appendix are very good.

99. Appendix L, Section L.2.1, Model Uncertainty

This section notes that:

> "Key model uncertainty is related to an issue for which no consensus approach exists and where the choice of approach is known to influence the HRA outcomes (e.g., introduction of new HFEs, alternative choice of critical tasks, and introduction of new CFMs and PIFs)."

The analytical form of the quantification model and its related assumptions are also a source of key model uncertainty. For example, three different quantification algorithms (e.g., THERP, HCR-ORE, and the proposed IDHEAS model) may produce significantly different human error probability (HEP) estimates for the same human failure event (HFE) with the same evaluated performance-influencing factors (PIFs), simply due to differences in how each model evaluates the relationships between those PIFs and the HEP. The fourth bullet item in this section acknowledges this source of key model uncertainty. However, the text does not discuss how it can be addressed in practice. For example, analysts may use the different algorithms to quantify a set of defined HFEs that have the same scenario context and PIFs to determine the range of estimated HEPs and to identify whether the use of a particular algorithm might introduce a systematic bias in the analysis results. Or they may examine the results from available benchmark studies that have performed similar comparisons.

Should this section discuss methods or provide an example to show how this source of key model uncertainty might be quantified?

100. Appendix L, Section L.2.2, Parameter Uncertainty

A specific source of this type of uncertainty in the context of the IDHEAS methodology (or any other methodology, for that matter) might arise from the analysts' assessment of the states of specific performance-influencing factor (PIF) attributes. For example, suppose that a PIF is characterized by a simple 3-discrete-state construct of "good, nominal, poor" or "low, medium, high". In a particular scenario, the analysts may be uncertain whether to assign a "nominal" or "poor" state to that PIF. That assignment could have a significant numerical impact on the respective human error probability (HEP).

In this situation, the range of numerical effects from this source of parameter uncertainty can be quantified by evaluating the HEP for both PIF states. In a more comprehensive assessment, the analysts can assign probabilities that each state applies (e.g., P1 that the state is "nominal" and P2 (= 1-P1) that the state is "poor"). The overall HEP, and its uncertainty distribution, is then the probabilistically-weighted combination of the results from the two assessments.

Should this section discuss methods or provide an example to show how this source of parameter uncertainty might be quantified?

101. Appendix M, Section M.1.1.1, Operational Narrative, General Comment

This is a very good narrative of the H.B. Robinson fire event. It appropriately documents the relevant plant-wide effects from an operational perspective, and it communicates the variety of diverse issues that the operators were addressing, in addition to the desired response to restore RCP seal cooling. It also documents the off-shift staffing on a Sunday night, the initial absence of the shift manager (SM) and shift technical advisor (STA) when the event occurred, the fact that the shift supervisor (SS) was not a normal member of the crew and did not stand watches on a regular basis, the fact that the balance of plant (BOP) operator's attention was occupied with response to the fire, and the fact that the SM and STA were at least partially preoccupied with assessment of the applicable emergency activation level (EAL) criteria.

The intermediate notes are an excellent way to show readers how each part of the narrative satisfies specific needs for the scenario context and the associated human reliability analysis (HRA).

102. Appendix M, Section M.1.1.1, Operational Narrative

The last sentence of the paragraph immediately before the discussion of "RCP Seal Injection – Charging Flow" notes that:

> "On time restoration of the seal injection, seal cooling, ***or RCP trip*** would protect the RCP seals from damage." [emphasis added]

I am not familiar with the design of the H.B. Robinson RCP seals when the fire occurred. During the scenario, RCP B was de-energized by the loss of power at Bus 4. I think that RCP A and RCP C remained running, but I could not quickly confirm that from the event summaries that I have. In any event, unless an RCP has modern passive shutdown seals, tripping the RCP does not necessarily prevent eventual seal damage. It may increase the time until seal damage occurs, but stopping the pump is not sufficient, by itself, to prevent damage if no seal injection flow and thermal barrier cooling are available.

The highlighted phrase is potentially very important for the human reliability analysis (HRA), because it describes the functional success criteria for prevention of seal damage (i.e., restore seal injection ***or*** restore thermal barrier cooling ***or*** trip all RCPs).

JWS Note: Figure M-2 shows the instruction to "Stop All RCPs". That action is intended to extend the amount of time that is available to restore seal injection flow or thermal barrier cooling before seal damage begins. Thus, a detailed PRA model of this scenario might include a human failure event (HFE) for operator actions to trip the running RCPs. If that action is successful, the time available (T_a) to restore seal cooling would be X minutes. If that action is not successful, T_a would be a shorter Y minutes. The sequence of events in Attachment 2 to Reference [92] for this report does not explicitly indicate whether, or when, the operators tripped RCP A and RCP C. They may have tripped the pumps shortly after they entered AOP-018 at 19:34, but that is only my speculation, and that is at least 2 or 3 minutes after they re-opened FCV-626. (I wrote this comment when I reached this point in the narrative, so I do not know how the HFEs are eventually defined. In any event, the intent of this comment is to prompt attention to the highlighted phrase from the excerpt, which implies that tripping the RCPs is sufficient, by itself, to prevent seal damage.)

JWS Note: This comment also applies to the "what-if" question in Section M.1.1.4, which seems to imply that it is necessary for the operators to trip the RCPs. It also applies to Section M.1.2.1, which indicates that tripping the RCPs is a "critical task" to prevent seal failure.

Why does this narrative imply that operator actions to trip the running RCPs would prevent seal damage (i.e., without restoration of seal injection or thermal barrier cooling)?

103. Appendix M, Section M.1.1.1, Operational Narrative, Boundary Conditions

The fourth listed boundary condition is:

> "A ***latent failure*** caused the VCT automatic makeup at 24.4 inches (62 cm) to not be functional." [emphasis added]

This does not seem to be consistent with the scenario narrative.

During normal operation, letdown flow and RCP seal return flow are directed to the volume control tank (VCT). Makeup to the VCT is needed occasionally to compensate for a very small amount of RCP seal leakoff flow that is directed to the reactor coolant drain tank, for small inventory imbalances in the chemical and volume control system (CVCS) processes, and for normal reactor coolant system leakage (if any). There is no indication that normal VCT makeup was not functioning properly during the days before the electrical fault occurred. The scenario narrative and the timeline in Table M-1 indicate that letdown flow was isolated automatically, as designed, when a pressurizer low level signal occurred shortly after 18:52. At 18:53, the operators started charging pump B and charging pump C. At 18:57, the VCT low level signal occurred, which should have initiated automatic makeup.

The narrative notes that automatic makeup to the VCT did not occur "due to the earlier electric fault that disarmed the automatic VCT makeup". This seems to indicate that the unavailability of automatic VCT makeup was a direct consequence of the initial electrical fault and the subsequent electrical system realignments, rather than a "latent failure".

JWS Note: As noted in the fifth boundary condition, the cause for failure to transfer the charging pumps' suction to the RWST was certainly related to a "latent" condition in the control circuits for the RWST suction valve.

JWS Note: For general information, the sequence of events in Attachment 2 to Reference [92] for this report indicates that the operators started the charging pumps at 18:53. It indicates that letdown isolated automatically at 18:54, in response to 14% level in the pressurizer. That sequence of events does not list the time when VCT level reached the automatic makeup setpoint. Both Table M-1 and the sequence of events in Reference [92] indicate that the safety injection (SI) signal occurred at 19:00, and VCT level reached the RWST transfer setpoint at 19:00. The SI signal may have increased flow from charging pump C, which was in automatic control. Valve CVC-310A opened fully at 19:18 or 19:19, which apparently increased charging flow further and reduced seal injection flow, because more flow was aligned to the charging line. Thus, the failure of valve CVC-310A did not exacerbate the VCT level response before the RWST transfer setpoint, but it apparently affected the available amount of seal injection flow after 19:19. Only the scenario conditions before 18:57 are relevant to the demand for automatic VCT makeup at the normal makeup setpoint (i.e., 24.4 inches).

Why does this boundary condition indicate that automatic VCT makeup was disabled due to a "latent failure"?

104. Appendix M, Section M.1.1.2, Scenario Timeline, General Comment

The timeline in Table M-1 is an excellent summary of the scenario evolution.

105. Appendix M, Section M.1.1.3, Relevant Operating Experience

This section contains a good summary of the March 1968 fire event at San Onofre Unit 1. It illustrates the complexity of the fire damage and the time that was needed to stabilize plant conditions (including eventual restoration of the cooldown).

The Browns Ferry Unit 1 fire in March 1975 also involved numerous fire-related effects, complex operator actions, and considerable time to stabilize plant conditions.

Should this section also summarize the Browns Ferry fire as another example of relevant operating experience?

106. Appendix M, Section M.1.2.1, Task Context

This section notes that:

> "The critical tasks were restoring the RCS seal cooling ***and*** seal injection in time to prevent RCP seal damage, tripping RCPs, and terminating SI." [emphasis added]

Restoration of thermal barrier cooling or restoration of seal injection flow will prevent seal failure.

JWS Note: A preceding comment on Section M.1.1.1 addresses the need for the "critical task" to trip the RCPs.

Why does this summary indicate that the task to restore thermal barrier cooling and the task to restore seal injection flow are "critical tasks" that are both required to prevent seal damage? In other words, why are these tasks described through "and" logic?

What is the technical basis for the assertion that failure to restore thermal barrier cooling ***or*** failure to restore seal injection will result in seal damage?

107. Appendix M, Section M.1.2.3, System Context, General Comment

This section contains a good discussion of the effects from the de-energized power supplies, and the intermittent and disabled displays and annunciators.

108. Appendix M, Section M.1.2.4, Personnel Context, General Comment

This section contains a good discussion of the personnel context.

109. Appendix M, Section M.1.3, Identification of Important Human Actions

I am very concerned that the discussion in this section, the definitions of the important human actions (IHAs), and the assigned success criteria for those actions are not appropriate for an example of a realistic assessment of the risk from this scenario. In particular, I disagree with the assertion that tripping the RCPs is a necessary action, and the assertion that this action alone will prevent seal damage during this scenario.

I am not familiar with the detailed design of the H.B. Robinson charging and seal injection flow paths. Based on my knowledge of other similar plant designs and the scenario narrative, it seems that valve CVC-310A is intended to allocate the charging pumps discharge flow between the RCS charging line and the RCP seal injection header. If CVC-310A is fully open, only the minimum design injection flow is available to the RCP seals. In most plants, that amount of flow is adequate to prevent seal failure. (The pressure differentials and the flow splits with the charging flow control valve fully open typically meet minimum seal injection flow design criteria. Of course, the H.B. Robinson design might be different.) However, with only minimum design injection flow, long-term seal degradation might occur because that flow may not fully flush all potential impurities away from the seal faces. That is typically why the operators are instructed to trip the RCPs as a precautionary action if seal injection flow is at its minimum value.

In this scenario, when CVC-310A opened fully at 19:19, minimum injection flow subsequently remained available to the RCP seals. If the scenario continued to evolve without any intervention, all seal injection flow would have been lost when the volume control tank (VCT) was drained completely and the charging pumps cavitated (e.g., at about 19:37, or somewhat later). Of course, all seal injection flow would also have been lost if the operators had recognized the loss of all suction and tripped the charging pumps, but they did not. Complete loss of all seal injection flow is certainly a challenge to RCP seal integrity if no thermal barrier cooling is available. However, tripping the RCPs under those conditions will not prevent seal damage. The operators must restore thermal barrier cooling, or they must restore seal injection flow, regardless of whether or not the RCPs are running.

It seems that the authors may be misinterpreting the guidance in AOP-018 to trip all running RCPs as a necessary, and sufficient, task to prevent seal damage, rather than a precautionary action to preclude long-term damage, possibly delay short-term damage, and possibly provide more time to restore thermal barrier cooling or seal injection. Thus, as noted in a preceding comment, tripping the RCPs might extend the available time window (i.e., T_a) to restore seal cooling, but that action alone cannot achieve successful termination of this scenario without other personnel responses.

In summary, as the example is described in this section, the applied success criteria require that the operators must trip the RCPs to prevent seal damage. That seems to be excessively conservative. The analysis further assigns success of that action alone as adequate to prevent seal damage. That is optimistic, and it is not justified by available analyses of expected seal performance. I do not think that these success criteria are justified technically or are appropriate for an example of a realistic assessment of the potential risk from this scenario.

JWS Note: If no component cooling water (CCW) flow is available to an RCP, heat is not removed from the oil coolers for the pump radial bearing, pump thrust bearing, and pump motor bearings. If the pump remains running without oil cooling, the bearings will eventually be damaged, resulting in loss of a very expensive piece of equipment. Loss of CCW flow to the bearing oil coolers is the reason for the bearing high temperature alarms that occurred during the H.B. Robinson fire event. I am fairly confident that the guidance to trip the RCPs in AOP-018 is intended primarily to prevent damage to the bearings (i.e., for plant investment protection), and not to mitigate potential thermal damage to the pump seals. That being said, prolonged operation of an RCP without bearing oil cooling could conceivably result in very high pump vibrations and failure modes that are severe enough to cause mechanical damage to the seals, due to abrasion of the seal faces. Some PRA models account for this potential cause for seal damage, but it is not associated with loss of all cooling to the seals, themselves. Thermal damage to the elastomers and O-rings in the seal package will occur from combined loss of CCW flow and seal injection flow, even if the pump is stationary. That is why the operators must restore CCW or seal injection flow, regardless of whether or not the pump is tripped.

What is the basis for the assertion that the opening of valve CVC-310A resulted in a sufficient reduction of seal injection flow to cause inadequate cooling of the seals and impending seal damage?

How much seal injection flow is available to each RCP with normal reactor coolant system pressure, discharge flow from only one charging pump, and valve CVC-310A fully open?

In particular, based on the H. B. Robinson design and supporting analyses of its RCP seals, how long can the RCPs continue to run with no thermally-induced seal failures under conditions with no thermal barrier cooling, normal reactor coolant system pressure, discharge flow from only one charging pump, and valve CVC-310A fully open?

What is the technical basis for the instruction to trip the RCPs in AOP-018? In other words, what is the functional intent of that action?

What is the technical basis for the example analysis requirement that the operators must trip the RCPs to prevent seal damage in this scenario (i.e., for the conditions that apply at 19:20 in the event progression)?

What is the technical basis for the conclusion that the action to trip the RCPs, by itself, is sufficient to prevent seal damage (i.e., without also restoring thermal barrier cooling, or ensuring that a source of charging pump suction water is available)?

If the authors have made simplifying or "conservative" assumptions about the applied success criteria, why does the example not clearly describe those assumptions and their technical justification (e.g., to fully explain the analysis and to demonstrate how analysts should document their assumptions)?

110. Appendix M, Section M.1.3, Identification of Important Human Actions

This section notes that:

> "Typically, the time available was approximately ***13 minutes*** (from when all RCP seal cooling and injection were lost) based on studies performed by Westinghouse [94], [127], [128]. Therefore, FCV-626 needed to be re-opened by 19:32 (= 19:19 + 13 minutes) to prevent RCP seal failure." [emphasis added]

I think that the 13-minute time window from the Westinghouse analyses applies for conditions when the RCPs are stationary, no thermal barrier cooling is available, and no seal injection flow is available (e.g., a station blackout). Those conditions do not apply in this scenario shortly after valve CVC-310A opens at 19:19. In particular, in this scenario, I think that RCP A and RCP C are running at that time, RCP B is stationary, no thermal barrier cooling is available, and seal injection flow is at the value that is determined by the flow splits with valve CVC-310A open fully (i.e., seal injection flow is not zero).

Why does this analysis use the cited 13-minute time window for these scenario-specific conditions?

If the authors have made simplifying or "conservative" assumptions about the available time window, why does the example not clearly describe those assumptions and their technical justification (e.g., to fully explain the analysis and to demonstrate how analysts should document their assumptions)?

111. Appendix M, Section M.1.4.1, Identification of Critical Tasks

This section lists several alarms that are associated with the loss of thermal barrier cooling and increasing RCP temperatures.

Why are the volume control tank (VCT) level alarms not also relevant cues to alert the operators to impending loss of seal injection flow?

112. Appendix M, Section M.1.4.1, Identification of Critical Tasks

The discussion of AOP-018 in this section notes that:

> "Steps 3 and 4: to trip all RCPs. This action would ***prevent*** RCP seal failure." [emphasis added]

Preceding comments address the example analysis success criteria, which require that the operators must trip the RCPs.

Does AOP-018 explicitly indicate that this action will "prevent" seal failure?

If not, is this action intended to "delay" seal failure?

113. Appendix M, Section M.1.4.2, Characteristics of the Critical Tasks

Table M-2, Table M-3, and Table M-4 provide good examples of the intended scope and types of information that should be documented to support the analyses of each task.

The entry for "Others" in Table M-2 notes that:

> "A couple hundred alarms are triggered within a few minutes after the initiating event."

This appropriately alerts the analysts to the large number of alarms that are present. However, the operators may also be preoccupied or distracted by trying to determine the status of systems, equipment, and functions that are unknown, due to the effects from the power failures and unreliable annunciators (i.e., not only by the presence of many alarms, but also by the absence of a complete picture of the plant status). The scenario narrative and other information in the example summary indicate that this was the case during the actual event (e.g., due to lack of rod bottom lights, reliable temperature readings, etc.).

Why does the entry for "Concurrent Tasks" or "Others" in Table M-2 not include other effects from the power supply failures, intermittent alarms, and failed displays and annunciators that could require focused scrutiny or divert attention away from the specific cues that are listed for this task?

114. Appendix M, Section M.1.5, Cognitive Failure Modes and Performance-Influencing Factors

This example is intended to demonstrate how an analysis should be performed in practice. The general IDHEAS methodology emphasizes that analysts should document the bases for their assumptions and the reasons for their decisions. That documentation is very important to justify the technical foundation for each analysis. It also provides a discipline that should improve consistency among analyses (or more clearly show the reasons for observed inconsistencies), and a point of reference for others to better understand what was done, and why.

The discussions in this section simply list the "dominant CFMs" and the "applicable PIFs" without any justification for those conclusions or documentation of the analysts' rationale why other CFMs and PIFs are not included in the analysis. The discussions of the listed CFMs and PIFs are very good. However, it is not evident why only those selected items are relevant for each analysis. Furthermore, these examples implicitly provide "NRC-approved" guidance that it is not necessary for analysts to justify and document these conclusions.

In addition to the rather short list of CFMs and PIFs for each task, the discussion of Critical Task 3 simply states that "there are no obvious adverse PIFs to this critical task".

JWS Note: This general comment also applies to Section M.2.5.

Why do the examples for each critical task not document the analysts' rationale and the bases for their selection of the "dominant CFMs" and the "applicable PIFs"?

In other words, why are other CFMs not important, and why are other PIFs not applicable?

115. Appendix M, Section M.1.6, Time Uncertainty Analysis

In general, this section is very good.

The discussion of "Procedure-Guided and Late Self-Awareness Cue Detection" notes that:

> "It is estimated that the crew would reach to the procedure step to check the cue 'RCP thermal barrier HI or LO flow alarms illuminated' at 30 minutes after the initiating event."

I am not familiar with the PATH-1 procedure at H.B. Robinson. However, this seems to be a rather long time. In particular, in the context of the actual event, it indicates that the operators would not typically reach this point in the procedure until about 22 minutes after their re-entry into PATH-1, following the safety injection (SI) that occurred about 8 minutes after the initiating event.

Is this nominal 30-minute delay based on the analysts' assessment of where this specific step is located in PATH-1, the intervening actions before that step, and scenario-specific factors that delay the operators' progression through the procedure?

In other words, what is the basis for this estimate, and how is this estimate related to the specific conditions that apply during this scenario?

Would the same estimate apply when the PATH-1 procedure is used during other event scenarios?

116. Appendix M, Section M.1.6, Time Uncertainty Analysis

This section notes that:

> "Assuming that 36 and 46 minutes represent the 5th and 95th percentiles of a ***normal distribution***, respectively, with a mean of 40 minutes, time required can be represented as a normal distribution with a mean of 40 minutes and standard deviation of 3 minutes 37 seconds." [emphasis added]

A normal distribution cannot have a 5th percentile value of 36 minutes, a median (mean) of 40 minutes, and a 95th percentile value of 46 minutes. If 36 minutes represents the analysts' estimate for the "shortest time", 40 minutes represents the "best estimate", and 46 minutes represents the "longest time" for T_n, then the analysis should not fit a normal distribution to those estimates.

What are the analysts' estimates for the "shortest time", "best estimate", and "longest time" for T_n?

What are the bases for each of those estimates?

What is the shape of the probability distribution over that range of estimates?

117. Appendix M, Section M.2.1.1, Operational Narrative, General Comment

This is a good example of an operational narrative for a scenario that would be modeled in a PRA (i.e., compared to the description of an actual event, as in Section M.1.1.1).

118. Appendix M, Section M.2.1.1, Operational Narrative, Staffing

Is anyone injured by the earthquake damage (e.g., in-plant operators or other plant staff)?

If so, how might those conditions affect the availability of injured local operators who are needed to directly support event mitigation actions, or how might they divert attention of shift supervisors who must coordinate actions to stabilize and evacuate other injured personnel?

If the analysis is performed under the assumption that nobody is injured, or that the injuries do not divert attention or possible priorities away from the desired actions, what is the basis for that assumption?

119. Appendix M, Section M.2.1.1, Operational Narrative

A **Note** in this discussion states:

> "Because this is a ***hypothetical event***, the scenario narrative describes the baseline scenario with the emphasis on how personnel and systems are expected to respond to the situation. The potential instrumentation failures and latent failures that could significantly affect operator performance are ***not discussed in the hypothetical events*** mainly because ***such failures are specific to actual events, not hypothetical events***."

I do not understand what this means in practice, or why this note is needed. The narrative indicates that DC-powered instrumentation and displays are available in the Main Control Room (MCR). Except for equipment powered by DC-supplied inverters, it also seems evident that AC-powered instrumentation and displays are not available. Thus, the "baseline scenario" analysis should account for that degraded and disabled instrumentation.

According to the general IDHEAS methodology, analysts should systematically search for failures of equipment and personnel actions that cause deviations from the "baseline scenario", characterize the conditions during those "additional scenarios", and determine how to evaluate the risk associated with those conditions. Thus, in a PRA, the analysts should certainly examine other scenarios that may involve equipment and instrumentation failures that are not caused directly by the seismic damage or the station blackout (SBO). Those scenarios are not more "hypothetical" than any other scenarios that are analyzed in the PRA, including the "baseline scenario". Therefore, I do not understand this distinction between "actual events" and "hypothetical events", or why the analyses of "hypothetical events" should not consider and evaluate the effects from other failures.

In practice, the models for an "actual event" often account only for the conditions that occurred during that particular event. Thus, I understand why a retrospective analysis of an "actual event" might not consider any other failures. In contrast, the prospective models for a full-scope risk assessment should account for all conceivable combinations of failures. (Of course, that ideal completeness is rarely achieved in practice.) The example in Section M.2 demonstrates a prospective analysis of a PRA scenario. Thus, according to the IDHEAS methodology, it seems that this analysis should certainly consider and evaluate additional "hypothetical events", as part of a complete assessment of the possible event scenarios and their associated risk.

JWS Note: Section M.2.1.4 discusses the identification of "additional scenarios" that are deviations from the "baseline scenario".

Is this note intended to mean that the entire analysis process for this example should not account for any other instrumentation failures, beyond those that are directly associated with the seismic damage and SBO?

Why does this discussion distinguish between the consideration of additional failures in the models for "actual events", but not in the models for "hypothetical events"?

Why does the note imply that a PRA analyst should not consider "hypothetical event" scenarios that involve other possible failures?

Why is this note needed for this discussion of the operational narrative for this example?

120. Appendix M, Table M-5, Timeline of the Baseline Scenario

The summary of events at 00:05 notes that:

> "**H(SS):** Requests field operators to report to the MCR for task assignments and to obtain master keys."

The summary of events at 00:15 notes that:

> "**H(MCR):** Distribute master keys in MCR to field operators (FO)."

This timeline seems to imply that the Shift Supervisor (SS) is immediately aware of the fact that the station blackout (SBO) has disabled the security systems in a way that requires the operators to use a master key to enter local plant areas. I do not know how prevalent this type of information is in practice. For example, unless this condition is noted explicitly in the plant procedures and emphasized in training, personnel may not realize that they are locked out of certain locations until they try to enter them as part of their task assignments. In that case, they must then notify the SS of the security failures, and the SS must summon the personnel back to the Main Control Room (MCR), distribute the keys, and send them back to work. That situation adds much more confusion and time delays than are implied by immediate recognition of the security failures and issuance of the keys during the initial task assignments.

JWS Note: The discussion in the first paragraph after the "hypothetical event" **Note** seems to indicate that the SS recognition is "guided by the LOOP procedure". It is important that the procedure clearly contains the explicit guidance to distribute the keys, in addition to the functional task assignments for the local operators.

What specific procedure, training, experience, or other guidance alerts the operators and the SS that the security system is disabled during this scenario?

In other words, at 5 minutes after the initiating event occurs, how does the SS know that the master keys must be issued to local operators?

What are the effects on the scenario timeline and the time needed to complete local actions if the local operators do not discover the need for keys until they first try to enter specific locations to perform their assigned tasks?

121. Appendix M, Table M-5, Timeline of the Baseline Scenario

The summary of events at 00:15 notes that:

> "**H(FO*):** Locally start the EDGs per procedure X, attachment A (T_{exe} = 45 min.)"

> "**H(FO):** Perform initial DC load shed per SE-11, Att. T (T_{exe} = 45 min.)"

The footnote to Table M-5 notes that:

> "*FO: Field operator could be equipment operator, fire brigade, chemist, digital Instrumentation and control technician, reactor protection technician, health physics, and security personnel, etc."

According to the timeline, it is apparent that the task to troubleshoot and attempt to start the diesel generators locally and the task to perform the initial DC load shed are performed in parallel. Only equipment operators can perform these tasks (with possible troubleshooting assistance from instrumentation and control technicians).

According to the nominal shift manning for this example, are these tasks competing for the same field operators?

If not, what is the justification for the conclusion that sufficient qualified personnel are available to perform these tasks in parallel, within the nominal 45-minute execution time?

122. Appendix M, Table M-5, Timeline of the Baseline Scenario

The summary of events at 05:45 notes that:

> "**H(FO):** Commence control room venting per FSG-030."

Unless this plant has very unusual provisions for DC-powered ventilation systems, all Main Control Room (MCR) ventilation and cooling is lost when the station blackout (SBO) occurs.

Why are actions to establish alternative MCR ventilation and cooling delayed until almost 6 hours after the initiating event?

What is the MCR heatup profile during this period?

How might the elevated temperatures affect personnel cognitive performance or enhance priorities to perform this task earlier (e.g., in parallel with, or in lieu of, other tasks)?

123. Appendix M, Section M.2.1.4, Identify Additional Scenarios, General Comment

This is a good summary of considerations for the identification of possible additional scenarios.

124. Appendix M, Section M.2.1.4, Identify Additional Scenarios, Containment

Depending on the affected systems or functions, if specific containment penetrations are not isolated automatically and cannot be isolated manually (or are not isolated within a scenario-specific time window), that condition could affect the event scenario progression before core damage occurs. (Of course, it could also affect the scenario progression in the Level 2 PRA models, but this example is limited to a Level 1 analysis.)

Why does this example not include a question about what if specific containment penetrations that affect event mitigation systems, functions, or event scenario progression are not isolated automatically?

125. Appendix M, Section M.2.2.1, Task Context

This section notes that:

> "Primary containment control procedure. The procedure entry condition is the containment pressure greater than 2 psig. Based on the plant simulation model calculation, this entry condition is reached at about ***ten minutes*** after the earthquake." [emphasis added]

Why does the scenario timeline in Table M-5 not list this condition?

126. Appendix M, Section M.2.2.2, Environment and Situation Context

Unless this plant has very unusual provisions for DC-powered ventilation systems, all Main Control Room (MCR) ventilation and cooling is lost when the station blackout (SBO) occurs. It seems very likely that all ventilation and cooling is also lost for the Technical Support Center (TSC) and the Operations Support Center (OSC). The increasing temperature and humidity may also affect personnel cognitive performance in those locations.

Why does this section not discuss the effects from loss of ventilation and cooling for the MCR, TSC, and OSC?

Why does this section not discuss the likelihood that rather strong aftershocks may continue for several hours after the initial earthquake (i.e., throughout the duration of this scenario)?

127. Appendix M, Section M.2.2.3, System Context, General Comment

This section contains a good discussion of the system context.

128. Appendix M, Section M.2.2.3, System Context, DC Power

This scenario involves communications and coordination of activities in the Main Control Room (MCR), Technical Support Center (TSC), and Operations Support Center (OSC). Therefore, I think that it is very important to document the DC power dependencies and their effects on the availability of lighting, communications, and plant status displays (if any) in each of those locations.

JWS Note: The electrical loads in the TSC and OSC are typically not considered to be "safety-related". I have seen many plant designs where the TSC and OSC are powered from the non-safety batteries. I also seem to recall at least one plant where those loads were intentionally shed for battery conservation as part of the instructions for fire event responses. (I am not completely positive about my recollection of the exact plant or that plant-specific procedure. I cannot recall if the TSC and OSC loads were fed from safety-related buses at that plant, and it was a fire response procedure that applied before the Fukushima event. Furthermore, I do not know if current plant-specific procedures apply different DC load shedding priorities for fire events and for losses of AC power due to other causes, like this ELAP scenario.) In any case, this example should demonstrate that the analysts must understand and clearly document these plant-specific dependencies. Of course, the analysis should also explicitly account for the fact that the TSC or OSC loses DC power at a certain time during the scenario, if that is the actual situation.

How does the station blackout (SBO) affect the availability of lighting, communications, and displays in the TSC and OSC?

In particular, are the lighting, communications, and displays in those locations powered from buses that are fed by the safety-related batteries?

If not, what is the available duration of the DC power supplies for those locations?

If those loads are normally supplied by the safety-related batteries, are they disconnected during the initial load shed or during the deep load shed?

In other words, how does the analyst have assurance that lighting, communications, and displays in the TSC and OSC remain available throughout the duration of this scenario?

129. Appendix M, Section M.2.2.4, Personnel Context

This section notes that:

> "Personnel injury may occur, and the non-technical staff will be evacuated from the site. These are assumed not to affect event mitigation."

As noted in a preceding comment, the injuries might directly affect in-plant operators who are needed to support mitigation of this scenario. Injuries to other plant personnel may also require direct involvement by operations supervisors to coordinate efforts to stabilize their conditions and to coordinate their medical evacuation. Concerns about injuries to co-workers throughout the plant may also distract the attention of operators and supervisors, and cause delays in their performance of the desired responses.

Since this event is caused by a very severe earthquake, it is essentially certain that the area surrounding the plant has sustained extensive damage. Thus, it seems likely that personnel cognitive performance may also be affected by their concerns about possible harm or the unknown status of their families and loved ones.

This example is intended to demonstrate how an analysis should be performed in practice. The general IDHEAS methodology emphasizes that analysts should document the bases for their assumptions and the reasons for their decisions. That documentation is very important to justify the technical foundation for each analysis. It also provides a discipline that should improve consistency among analyses (or more clearly show the reasons for observed inconsistencies), and a point of reference for others to better understand what was done, and why.

What is the basis for this assumption?

Why does this example not discuss possible influences from personnel concerns about the status of their families and loved ones?

130. Appendix M, Section M.2.3, Identification and Definition of Important Human Actions

This section notes that:

> "Note that the debris in the transportation route needs to be removed before the action can be performed. Removing debris is considered as a separate important human action because it is performed by a different group of people and affects the deployment of all FLEX equipment."

When I read this section, I was initially very concerned that the analysis of the important human actions (IHAs) to deploy and connect the FLEX generator did not account for the time that is needed to remove debris from the transportation route. However, when I finally read Section M.2.6, I discovered that it does.

To alert analysts to the importance of considering the integrated scenario timing and functional relationships among the defined IHAs, should this section explicitly note that the analysis of the selected IHAs accounts for the time that is needed to clear debris before the FLEX generator can be moved to its operating location?

131. Appendix M, Table M-6, Task Characteristics of Detect ELAP Procedure Instruction to Deploy a FLEX Generator

According to the scenario timeline in Table M-5, entry into the ELAP procedure should occur about 1 hour after the initiating event. At that time, the plant may be in a rather chaotic condition, and it is likely that rather strong aftershocks will be occurring.

Why does the entry for "Concurrent Tasks" or "Others" in this table not identify the occurrence of aftershocks, concerns about injuries to in-plant co-workers, activities to evacuate injured and non-injured plant staff, and concerns about the safety of family and loved ones as sources of distractions or conflicting priorities that may divert the crew's attention from the desired task?

132. Appendix M, Section M.2.5, Cognitive Failure Modes and Performance-Influencing Factors, Starting and Continuously Operating the FLEX Generator

This section notes that:

> "The ***primary CFM*** is the following:
>
> - E4-2 – Fail to perform ***simple actions***: This refers to failing to start the generator or failing to open / close the breakers as needed during the operation." [emphasis added]

In practice, it is likely that the operators will need to sequence specific loads onto the generator, manage and control the generator load, and perhaps alternate loads according to the evolving scenario conditions. Therefore, this task seems to involve more than the cognitive failure mode (CFM) for failure to perform "simple actions".

Table 4-9 lists the following CFMs for Action Execution:

E4-3 Fail to execute complex action (e.g., execute a complex action with incorrect timing or sequence, execute actions that do not meet the entry conditions)
E4-3A Fail to execute control actions
E4-3B Fail to execute long-lasting actions

JWS Note: A preceding comment addresses the general need for analysts to document and justify their rationale for their selection of the "dominant CFMs" and the "applicable PIFs". However, this example seems especially noteworthy.

Why does this task not also involve CFM E4-3?

133. Appendix M, Section M.2.6, Time Uncertainty Analysis

I performed a quick confirmatory check of the three final calculations in this section. They seem to be correct. However, considering the unusual observation that is discussed below, I want to make sure that I understand what was done and that the distribution arithmetic is correct.

The discussion of "Time Available for Performing the Action" is very good. It clearly describes the scenario conditions that would determine the nominal functional time windows of 2 hours, 5 hours, and 7 hours. It is noted that the ELAP procedure is entered at time t = 1 hour after the initiating event occurs, and debris removal will start at that time.

The discussion of "Time Needed" indicates that the estimated mean amount of time that is needed to clear the debris is 2 hours, and the estimated mean total amount of time that is needed to transport the generator to the location of the in-plant bus connections and to connect the generator is 45 minutes. Thus, it seems that a mean total time of about 2-3/4 hours are needed to clear the debris, move the generator to the in-plant connections, and make those connections.

The scenario timeline in Table M-5 seems to be consistent with these estimates. It indicates that the ELAP procedure is entered at 01:00, and the actions to deploy the FLEX generator begin at 03:00.

It seems evident that the debris removal uncertainty distribution does not account for the initial 1-hour delay before entry into the ELAP procedure, which prompts the activities to start debris removal. In particular, the lower bound of that uncertainty distribution extends well below 1 hour, which implies that no time would be needed to clear the debris. That is not consistent with the conditions during this scenario, which explicitly require that debris must be removed. Thus, based on the scenario timeline and the uncertainty distributions, it seems that the mean time at which the generator would be connected to the in-plant buses is:

- 1 hour time delay to enter the ELAP procedure +
- 2 hours mean time to clear the debris +

- 45 minutes mean time to transport and connect the generator

Thus, it is estimated that the generator should be fully connected at a mean time of approximately 03:45 during the scenario evolution.

The last paragraph before the discussion of "Time Needed" notes that:

> "If debris removal is needed, the time available for deploying the FLEX generator would be (T_{SW} – one hour for entering ELAP procedure – the time needed for removing debris)."
>
> "In fact, if T_{SW} is ***two hours*** and debris is needed, the action becomes ***infeasible*** because there is not enough time to perform it." [emphasis added]

This conclusion seems to be fully consistent with the scenario timeline and the estimated mean times discussed above.

For the condition when T_{SW} is two hours, the overall quantification results indicate that the probability for failure to re-repower the 480 VAC buses before the batteries are depleted is 0.96. Thus, in principle, the action is actually conceptually "feasible", but there is only a 4% probability that it can be completed within that 2-hour time window.

I think that the uncertainty calculations are correct. If they are, this part of the example is very interesting. In particular, the 2-hour results show that a feasibility assessment which uses only the nominal (mean) values for T_a and T_n might occasionally reach a pessimistic conclusion, without considering the full uncertainties in those times. That is quite unusual. In practice, simple "point-estimate" feasibility assessments can more often lead to optimistic conclusions, because they typically do not consider the probability that T_n may exceed T_a (i.e., as demonstrated by the example results when T_{SW} is 5 hours and 7 hours).

JWS Note: The 2-hour example results certainly may not make much practical difference to the overall PRA, since 0.96 is rather close to 1.0. However, this example might subtly help to garner some conceptual support for the performance of a time uncertainty analysis among analysts who simply assert that it is too complex or not worthwhile.

Did I correctly interpret the three contributions to the estimated total time for T_n in this analysis?

Do the calculations and results in this section correctly account for the 1-hour time delay before debris removal is initiated?

In particular, is the applied uncertainty distribution for completion of the debris removal a normal distribution with a mean value of 3 hours and a standard deviation of 1 hour (i.e., to account for the delay before entering the ELAP procedure)? Is the applied uncertainty distribution for transportation and connection of the generator a normal distribution with a mean value of 45 minutes and a standard deviation of 15 minutes? Is the total uncertainty distribution for T_n the sum of those distributions?

If the calculations are correct, should the discussion of these results explicitly highlight this unusual observation (i.e., that a quantitative treatment of uncertainties can also show that a "point-estimate" feasibility conclusion is occasionally pessimistic, depending on the scenario-specific context)?

134. Appendix M, Section M.2.6, Time Uncertainty Analysis

While I was composing the preceding comment and re-checking the calculations, I noticed that the assigned uncertainty distribution for the time that is needed to clear the debris is not physically consistent with the basic boundary conditions for this analysis.

In particular, there is a 1-hour time delay before entry into the ELAP procedure. The estimated time that is subsequently needed to clear the debris is then characterized by a normal probability distribution with a mean value of 2 hours and a standard deviation of 1 hour. If the scalar 1-hour time delay is simply added to the mean value of that distribution, the resulting normal probability distribution has a mean value of 3 hours and a standard deviation of 1 hour. As noted in the preceding comment, I believe that this distribution was used for the final calculations.

However, this uncertainty distribution is not physically consistent with the basic assumptions and constraints for this scenario. I particular, that distribution has a probability of about 2.3% that the debris is cleared before time 01:00 (i.e., debris removal is completed before it starts!). Although that probability is "small", it is not physically consistent with the basic assumption that debris removal will not begin before 01:00. Therefore, the assigned probability distribution is not appropriate for a demonstration of how analysts should carefully account for the uncertainties and understand the implications from those uncertainties.

Based on these observations, I think that it is very important that the authors should carefully re-evaluate the assigned probability distribution for the time that is needed to clear the debris, to avoid this type of physically inconsistent conclusion. For example, rather than using a normal distribution to account for the uncertainty in this time, it may be better to use a skewed distribution (e.g., like a beta, gamma, or possibly lognormal) to avoid any probability that the completion time occurs before 01:00. If the authors desire to retain a normal probability distribution with a standard deviation of 1 hour, then that entire distribution should be delayed in time by at least 1 hour after the start of the scenario. For example, the estimated mean time for completion of the debris removal might be somewhere between 04:30 and 05:00, with the earliest possible completion time at about 01:30. Of course, the actual distribution depends on the authors' assessment of the plant-specific and scenario-specific conditions. However, in no case should the earliest estimated time to complete debris removal occur before 01:00, plus some nominal time that would apply for the "best possible" conditions.

During their revision of this analysis, the authors should also examine the lower bound of the combined uncertainty distributions to confirm that there is no probability that the time for completion of the sequential actions to clear the debris, transport the generator, and connect the generator occurs earlier than 01:00. The minimum time for completion of the combined actions (i.e., the earliest time after 01:00) should also be consistent with the authors' understanding of the best possible conditions that can facilitate those actions.

JWS Note: Of course, if a different uncertainty distribution is used for the time to clear the debris, it seems quite likely that the unusual results that are discussed in the preceding comment will disappear. In other words, it seems likely that the 2-hour example will show a probability of 1.0 that the operators fail to connect the generator within the 2-hour time window. However, depending on the actual uncertainty distributions, there could conceivably be a small non-zero probability for overall success before time 02:00. Therefore, I retained the preceding comment, just in case it still applies for the revised analysis. However, it is certainly not worthwhile to emphasize that insight if the calculated success probability is less than 1%.

Why does this example apply a probability of about 2.3% that the debris may be cleared before time 01:00 during the scenario evolution?

In other words, why does the uncertainty analysis not account for the scenario-specific constraint that debris removal does not begin before 01:00, and that some positive amount of time is needed to clear the debris after that point in the scenario?

How will this analysis be revised to eliminate any improperly optimistic assessments of the lower-bound completion times?