U.S.NRC

United States Nuclear Regulatory Commission

*Protecting People and the Environment*

NUREG-2156

# The U.S. HRA Empirical Study

## Assessment of HRA Method Predictions against Operating Crew Performance on a U.S. Nuclear Power Plant Simulator

Office of Nuclear Regulatory Research

# AVAILABILITY OF REFERENCE MATERIALS IN NRC PUBLICATIONS

**NRC Reference Material**

As of November 1999, you may electronically access NUREG-series publications and other NRC records at NRC's Library at www.nrc.gov/reading-rm.html. Publicly released records include, to name a few, NUREG-series publications; *Federal Register* notices; applicant, licensee, and vendor documents and correspondence; NRC correspondence and internal memoranda; bulletins and information notices; inspection and investigative reports; licensee event reports; and Commission papers and their attachments.

NRC publications in the NUREG series, NRC regulations, and Title 10, "Energy," in the *Code of Federal Regulations* may also be purchased from one of these two sources.

1. **The Superintendent of Documents**
   U.S. Government Publishing Office
   Mail Stop IDCC
   Washington, DC 20402-0001
   Internet: bookstore.gpo.gov
   Telephone: (202) 512-1800
   Fax: (202) 512-2104

2. **The National Technical Information Service**
   5301 Shawnee Rd., Alexandria, VA 22312-0002
   www.ntis.gov
   1-800-553-6847 or, locally, (703) 605-6000

A single copy of each NRC draft report for comment is available free, to the extent of supply, upon written request as follows:

Address: **U.S. Nuclear Regulatory Commission**
Office of Administration
Publications Branch
Washington, DC 20555-0001
E-mail: distribution.resource@nrc.gov
Facsimile: (301) 415-2289

Some publications in the NUREG series that are posted at NRC's Web site address www.nrc.gov/reading-rm/doc-collections/nuregs are updated periodically and may differ from the last printed version. Although references to material found on a Web site bear the date the material was accessed, the material available on the date cited may subsequently be removed from the site.

**Non-NRC Reference Material**

Documents available from public and special technical libraries include all open literature items, such as books, journal articles, transactions, *Federal Register* notices, Federal and State legislation, and congressional reports. Such documents as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings may be purchased from their sponsoring organization.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at—

**The NRC Technical Library**
Two White Flint North
11545 Rockville Pike
Rockville, MD 20852-2738

These standards are available in the library for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from—

**American National Standards Institute**
11 West 42nd Street
New York, NY 10036-8002
www.ansi.org
(212) 642-4900

Legally binding regulatory requirements are stated only in laws; NRC regulations; licenses, including technical specifications; or orders, not in NUREG-series publications. The views expressed in contractorprepared publications in this series are not necessarily those of the NRC.

The NUREG series comprises (1) technical and administrative reports and books prepared by the staff (NUREG–XXXX) or agency contractors (NUREG/CR–XXXX), (2) proceedings of conferences (NUREG/CP–XXXX), (3) reports resulting from international agreements (NUREG/IA–XXXX), (4) brochures (NUREG/BR–XXXX), and (5) compilations of legal decisions and orders of the Commission and Atomic and Safety Licensing Boards and of Directors' decisions under Section 2.206 of NRC's regulations (NUREG–0750).

U.S.NRC
United States Nuclear Regulatory Commission
*Protecting People and the Environment*

NUREG-2156

# The U.S. HRA Empirical Study

## Assessment of HRA Method Predictions against Operating Crew Performance on a U.S. Nuclear Power Plant Simulator

Manuscript Completed: September 2015
Date Published: June 2016

Prepared by:
John Forester[1,7], Huafei Liao[1], Vinh N. Dang[2], Andreas Bye[3], Erasmia Lois[4], Mary Presley[5], Julie Marble[4], Rod Nowell[6], Helena Broberg[3], Michael Hildenbrandt[3], Bruce Hallbert[7], and Tommy Morgan[7]

[1]Sandia National Laboratories, Albuquerque, NM, USA
[2]Paul Scherrer Institute, Villigen PSI, Switzerland
[3]OECD Halden Reactor Project, Institute for Energy Technology, IFE, Haden, Norway
[4]U.S. Nuclear Regulatory Commission, Washington, DC, USA
[5]Electric Power Research Institute, Palo Alto, CA, USA
[6]South Texas Project, TX, USA
[7]Idaho National Laboratory, Idaho Falls, ID, USA

Erasmia Lois and Y. James Chang, NRC Project Managers

Office of Nuclear Regulatory Research

# ABSTRACT

This report documents the U.S. Human Reliability Analysis (HRA) Empirical Study (referred to as the U.S. Study in the report), which is a large systematic data collection effort supported by the U.S. Nuclear Regulatory Commission with participation of organizations from five countries representing industry, regulators, and the research community. The objective of the U.S. Study was to improve the insights developed from the International HRA Empirical Study [1–4] (referred to as the International Study) and address the limitations of that study.

Similar to the International Study, the U.S. Study evaluated the performance of different HRA methods by comparing method predictions to actual crew performance in simulated accident scenarios conducted in a U.S. nuclear power plant (NPP) simulator. There was significant agreement in the findings and conclusions between the International and U.S. studies in terms of the strengths and weaknesses of the HRA methods evaluated in both studies and in the overall findings about HRA and the identified needed improvements. In addition to identification of some new HRA- and method-related issues, the design of the U.S. Study allowed insights to be obtained on some issues that were not resolved in the International Study. In particular, because multiple HRA teams applied each method in the U.S. Study, comparing their analyses and predictions allowed separation of analyst effects from method effects and allowed conclusions to be drawn on aspects of methods that are susceptible to different application or usage by different analysts that may lead to differences in results. The findings serve as a strong basis for improving the consistency and robustness of HRA, which in turn facilitates identification of mechanisms for improving operating crew performance in NPPs.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# EXECUTIVE SUMMARY

## Background

The existence and use of a variety of human reliability analysis (HRA) methods that rely on different human performance frameworks and data, as well as inconsistent implementation from analysts, appear to be the most common sources of the variability in HRA results. To address such issues as an effort to improve the robustness of HRA, the U.S. Nuclear Regulatory Commission (NRC) participated in and supported the International HRA Empirical Study [1–4], in which HRA predictions of different analysts and methods were compared to observed crew performance at the Halden Reactor Project's HAMMLAB (HAlden huMan-Machine LABoratory) simulator. The International Study identified important strengths and weaknesses of the various HRA methods used in the study, and an important conclusion was that improving the qualitative analysis aspects of HRA methods could increase their robustness and reduce some of the sources of the variability in results that are seen in applications of different methods. However, because there was only one case in the International Study where the same HRA method was applied by multiple teams, it was difficult to clearly separate method-specific effects from variability created by the analysts' application of a given method. Thus, in addition to examining differences across methods, a major objective of the present study—which used a U.S. nuclear power plant (NPP) simulator and is thus referred to as the U.S. Study—was to test the consistency and accuracy of HRA predictions among different HRA teams using the same methods. A particular area of interest is examination of the qualitative analysis performed by different teams in applying the same method to identify shortcomings that contribute to inconsistencies in results and to determine the extent to which the shortcomings stem from analyst differences or from inherent shortcomings in the methods.

Two other potential limitations of the International Study are also addressed in the U.S. Study. First, in the International Study, the HRA analyst teams were unable to visit the Halden simulator and collect HRA-related information through interviews with plant operators and trainers and through observations of actual operating crews in the simulator, as is typically done in performing an HRA for an NPP probabilistic risk analysis (PRA). This type of information was given to the HRA teams, who applied HRA methods to predict crew performance, to the extent possible by the assessment group (who organize the experiment and compare the HRA teams' results) in the International Study to compensate the unavaibilty of interviewing plant operator and trainers directly. For additional information, the HRA teams were allowed to submit written questions that were answered by the study team and plant personnel as needed. Some of the HRA teams in the International Study felt that this significantly limited their ability to do an adequate HRA. In contrast, the HRA teams in the U.S. Study were able to visit the reference plant and collect information relevant to their HRA as it would normally be done in a PRA.

Second, there was some concern from the HRA community that because the International Study was based on the results of simulator runs using European crews at the Halden Reactor Project, the results might not be directly generalizable to what would occur with U.S. NPP crews. In addition, some of the HRA teams in the International Study thought that their expertise was more geared to understanding what U.S. crews would do and that their U.S. bias might have influenced their decision-making in applying their HRA methods. Thus, the U.S. Study provides an opportunity to determine whether the results were affected by the suspected bias.

The U.S. Study focuses the HRA on the control room personnel actions required in response to PRA initiating events. This focus was motivated by the widespread use of HRA methods within probabilistic safety/risk assessment (PSA/PRA) in the industry and the related importance of understanding and predicting human performance in accident scenarios.

**Overview of the Study Design**

The U.S. Study followed the same general approach as that followed in the International Study [1-4], with the exception that in the U.S. Study the HRA analysts were able to visit the plant, observe a crew in a training scenario in the simulator, and interview training personnel to collect information relevant to doing their HRA. The study tasks and organization are listed below:

Four High-Level Study Tasks

The study consisted of the following four high-level tasks:

- **Task 1.** The definition of the scenarios and of the human failure events (HFEs) to be analyzed and a compilation of the inputs for the HRA analyses (e.g., relevant procedures).
- **Task 2.** The analysis of the HFEs with HRA methods to produce predicted outcomes.
- **Task 3.** The production of the empirical data for comparison with HRA predictions, starting from the collection of raw data in simulator experiments conducted at a U.S. plant and followed by the analysis of the data.
- **Task 4.** Review of the HRA submittals, comparison of HRA predictions to the empirical data, along with intra-method comparisons (similarities and differences in results between analysts using the same method), and development of insights for improving HRA methods and practices.

Study Organization, Participants, and Roles

There were the following four sets of study participants:

(1) **Experimental team** (Tasks 1, 3): The experimental team was composed of training staff from a participating U.S. pressurized-water reactor (PWR) 4-loop Westinghouse NPP and research staff from the Halden Reactor Project, Idaho National Laboratory (INL), and the NRC. They collected and analyzed crew performance data on the full-scope, conventional control room training simulator of the participating NPP.

(2) **Operation crews** (Task 3): Four licensed operation crews from the U.S. NPP participated in the study. Each crew responded to three scenarios in the plant simulator, with the exception that one of the four crews was unable to complete one of the scenarios because of a simulator problem.

(3) **HRA teams** (Task 2): Each of nine HRA teams applied one of the HRA methods listed below to obtain predictions for the HFEs in the scenarios used in the study. Individuals

representing industry, regulators, and the research community participated, used one of the following four HRA methods for the U.S. Study:

- ATHEANA (A Technique for Human Event Analysis, NUREG-1624, Revision 1 [4], NUREG-1880 [5])
- SPAR-H (Standardized Plant Analysis Risk-Human Reliability Analysis, NUREG/CR-6883 [6])
- ASEP (Accident Sequence Evaluation Program Human Reliability Analysis Procedure, NUREG/CR-4772 [7])
- CBDT (Cause-Based Decision Tree [9]) & HCR/ORE (Human Cognitive Reliability /Operator Reliability Experiments [9])

(4) **Assessment group** (Overall organization and Task 4): This group (including staff from Sandia National Laboratories (SNL), Halden Reactor Project (HRP), Paul Scherrer Institute (PSI), Electric Power Research Institute (EPRI), and the NRC) had the overall responsibility of organizing and implementing the study. In the early stages of the study, this group prepared the information package (analysis inputs) for the HRA teams and answered their subsequent requests for additional information and questions concerning ambiguities in the instructions and assumptions. They also organized a plant visit for the HRA teams to collect information needed to do their HRA. After the HRA teams delivered their analyses, the assessment group reviewed and summarized the HRA teams' analysis results before making the actual comparisons with the crew data and the comparisons between the different teams using the same method. The assessment group were HRA experts who were not part of the HRA teams.

To avoid biasing the comparisons in Task 4, a study protocol was used whereby the assessment group did not receive any information about the actual crew performance in the NPP simulator until after they had reviewed and summarized the HRA teams' predicted outcomes. Similarly, the experimental team conducted their data analysis to produce the empirical data without knowledge of the HRA teams' predictions. The assessment group and the experimental team cooperated extensively on the comparison in Task 4, including interpretation and representation of the empirical data for comparison with the HRA predictions.

**Overall Methodology**

Simulation Design and Analysis Methodology

Four crews of licensed PWR operators from a U.S. plant participated in the study. Five HFEs were defined in three scenarios by the experimental team on a functional level of plant safety and based on real plant PRAs. In some cases the HFEs were defined by stricter failure criteria than a standard PRA would use, because it was important that the HFEs not only were related to those commonly used in PRA/HRA, but also were clearly observable in the simulated scenarios. That is, we wanted to see whether the operators succeeded or failed. The timing criteria used in the definitions of the HFEs were based on the timing from simulator runs and were tied to core damage or key parameter thresholds.

The experimental team used the following five phases to obtain the empirical data for comparison against HRA predictions:

(1) collection of raw data, including logs and audio and video recordings of crews' activities, and interviews with crews

(2) crew-level analysis, in which crews' performance was investigated at a detailed operational level

(3) determination of the number of HFE failures by combining the information contained in the crews' and scenarios' reviews with quantitative performance data (e.g., performance times and plant parameters)

(4) aggregate level analysis: writing of the operational descriptions (summaries of how the crews responded to the scenarios) and derivation and rating of the performance shaping factors (PSFs) by the experimental team

(5) assessment of the relative difficulty of the HFEs and their ranking

<u>HRA Methods Assessment</u>

The assessment group analyzed HRA methods and applications in this study with the following criteria. The assessment serves as the basis to delineate the strengths and weaknesses of the HRA methods and HRA in general.

- predictive power:
  - quantitative predictive power (to the extent that this can be assessed in light of the limitations of the empirical data) in terms of the absolute and relative values (i.e., HFE ranking) of human error probabilities (HEPs);
  - qualitative predictive power in terms of performance drivers and operational expressions.
- traceability of the qualitative analysis and quantification process;
- the adequacy of the guidance given by each method for the qualitative analysis and for quantification of an HFE; and the
- usefulness of the HRA results for human error reduction.

In summary, assessments of the methods' qualitative and quantitative predictive power are based on comparisons between each method's predictions and the empirical data obtained in the plant simulator. Assessments of other criteria (traceability, guidance, and insights for error reduction) also considered the documented HRA analyses submitted by the HRA teams in conjunction with the HRA predictions against the empirical data.

<u>Intra-Method Comparison</u>

Intra-method comparisons were conducted by the assessment group to evaluate the consistency of same or similar methods in qualitative and quantitative predictions when different analysts were applying them. For each method, comparisons were first made for each HFE.

Qualitative analyses and quantitative analyses were compared between the HRA teams using the same method and against the empirical data. Then, the HFE-by-HFE intra-method comparison results were synthesized and overall conclusions concerning intra-method variability were summarized.

**Overall Quantitative Results**

The mean HEPs predicted by the HRA teams are presented in the figure below alongside the Bayesian uncertainty bounds derived from the crew data (this figure is also presented in Chapter 6). On the horizontal axes, the HFEs are ordered by their difficulty ranking. Note that HFE 1B was not ranked as no empirical data were collected for this HFE but the HRA teams did make the predictions. The figure illustrates the following observations:

- There was significant disagreement between HRA teams in their predicted HEPs for each of the HFEs, but in general there appeared to be less variability between methods/teams than in the International Study and overall it appeared that the HRA predictions improved relative to the International Study.

- In most cases, the HEPs predicted by the method applications show a decreasing trend that is consistent with the difficulty ranking.

- Some HRA methods (e.g., ASEP) seemed to yield somewhat more consistent quantitative results than other methods (e.g., SPAR-H) across the HRA teams that used these methods. The reasons for this are discussed in Chapter 6 of this report.

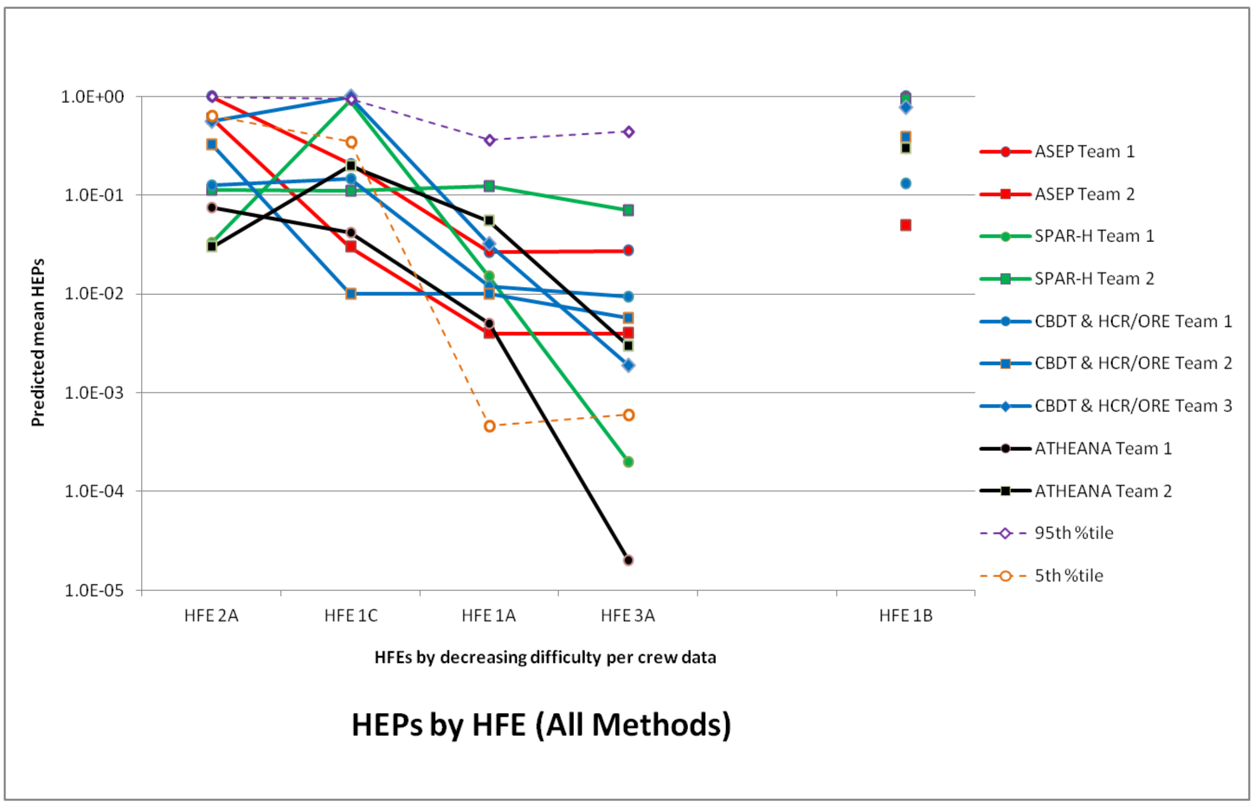



**Figure ES-1 Predicted mean HEPs by all HRA teams**

**Intra-Method Comparison Results**

In addition to comparing the HRA teams' predictions, the study also aimed to determine the causes/sources for observed discrepancies in predictions made by teams using the same method. Among other considerations, this intra-method analysis asked (1) whether the apparent discrepancies in predictions from teams using the same method can be explained by the teams' assumptions, (2) whether the discrepancies were caused by the limitations of a particular method or a function of the HRA teams (i.e., how the analysts interpret and apply a given method), and (3) whether different teams sought and obtained different information. It was also important to see whether the observed discrepancies were consistent across HFEs and investigate why these occurred.

Although the consistency of the quantitative results from some methods shown in Figure 1 is relatively encouraging, one question that arises is whether the qualitative predictions (performance issues, failure modes, dominant factors and failure "mechanisms") from these analyses are consistent with the quantitative consistency. In other words, do the two ASEP, three CBDT & HCR\ORE, two SPAR-H and two ATHEANA teams (respectively) obtain some consistency in their intra-method quantitative results? The results show that although there was reasonable matching in some cases, the detailed analysis that considers the qualitative predictions found significant differences in the identified basis for the HEPs in several instances. In addition, there were significant differences in the assessed contribution of the diagnosis and execution portions of the actions underlying the HEP in some cases. Thus, the common results within (and between methods for some HEPs) were not necessarily as consistent as they appeared to be. Such findings call into question the general reliability of HRA methods and support the need for the development of improved qualitative analysis, because an important goal of HRA is to find safety insights and make recommendations for potential plant improvements.

The intra-method comparisons done in this study identified the following factors that contributed to variability in predictions of applying the same HRA method (and across methods in many cases) in terms of qualitative and quantitative analyses and in terms of several general HRA practices. The contributors included both method-driven factors and analyst-driven factors and their interactions. The method-driven observations include:

- Different approaches to task decomposition (including both task analysis breakdown and task execution elements)

- Approach for estimating the time required to do the actions and treatment of the complexities associated with making the estimations

- Difficulty in understanding and treating complexity (mainly a method effect or lack of guidance)

- Different qualitative analysis approaches and carrying out different levels of the qualitative analysis. Includes going beyond the method guidance (e.g., one ATHEANA team chose to consider a range of failure paths for each HFE, which is not explicitly described in ATHEANA guidance)

- Different approaches for collecting information in interviews

The analyst-driven observations include:

- Credit for recovery – use of different criteria, sometimes driven by desire for conservatism
- Some analysts compensating for a given method's poor treatment of diagnosis
- Different impressions of what the results of interviews with trainers implied
- Appropriate use of a reasonableness check of obtained HEPs
- Use of different models to quantify execution as allowed by the method

The analyst-method interaction driven observations include:

- Attempting to make up for inadequate range of PSFs covered by a given method
- Differences in deciding which PSF to use to cover a particular effect or influence and what the weight or rating of the factors should be
- Inadequate means to tie detailed qualitative analysis to the quantitative approach led to differences between applications of the same method

**Conclusions and Recommendations**

Findings on HRA General Issues

The U.S. HRA Empirical Study and the International HRA Empirical Study [1-4] represent large systematic data collection efforts, with appropriate controls, to allow clear conclusions on the strengths and weaknesses of several HRA methods currently being used in the industry and on HRA practices in general. There was significant agreement in the findings and conclusions between the International and U.S. studies (1) in terms of the strengths and weaknesses of the methods evaluated in both studies (summarized in Section 9.1 of this report) and (2) in the overall findings about HRA and the identified needed improvements. In addition to identification of some new HRA and method related issues, the design of the U.S. Study allowed us to obtain insights on some issues that were not addressed in the International Study.

In the U.S. Study, each method was used by more than one HRA team. Comparing their predictions allowed us to separate analyst effects from method effects and draw conclusions on aspects of methods that are susceptible to different application or usage by different analysts that may lead to differences in results. It is not unreasonable to expect different analysts to make different decisions if the guidance is incomplete or subject to different interpretations. It is also unrealistic to expect that there will never be any subjectivity involved in HRA results. However, the findings of the U.S. Study point us to areas in HRA methods where additional guidance or modeling approaches could improve the robustness of HRA performance.

It was observed that there was substantially less variability in the predicted HEPs across the methods for most HFEs in the U.S. Study. In addition the difference in the intra-method HEP predictions was less than might have been expected based on general HRA performance in the International Study and other studies. Potential reasons for this include:

- There may have been some learning effects between the International Study and the U.S. Study as some of individuals on the HRA teams participated in both studies and most participants in the U.S. Study were familiar with the results of the International Study.

- It is also possible that the HRA teams were, in fact, better at predicting the performance of U.S. crews on a U.S. simulator than the performance of foreign crews, which was a concern for the International Study. However, given that similar variability in crew performance was observed in both studies, there was no strong evidence indicating there was a crew effect.

- Measures were taken in the U.S. Study to address some of the limitations in the International Study. For example, HRA analysts were given the opportunity to visit the plant and observe crews in the simulator, and interview plant trainers. This may in fact have been a contributor to the somewhat better HRA predictions in the U.S. Study, but also caused some problems contributing to variability. Certainly the HRA analysts thought it was a great benefit to their analysis to be able to visit the plant.

The key findings and conclusions obtained from both International and U.S. HRA empirical studies include the following:

- The U.S. Study echoes the International Study that qualitative analysis done to support HRA quantification and the consideration of diagnosis is an important contributor to the adequacy of HRA predictions, especially the benefits of detailed qualitative analysis of cognitive challenges in complex scenarios.

- Although a good qualitative analysis is a relative strength of some methods, qualitative analysis is a shared weakness across all methods (i.e., they all can be improved) and has a significant effect on the robustness of HRA applications.

- HRA methods differ in the extent to which they support analysts in dealing with complexity and could be improved if they are expanded to cover a broader scope of human performance issues. Given that complex scenarios normally involve relatively more cognitive challenges compared with easy scenarios, a priority for improving HRA methods is to provide means and frameworks for analysts to find and characterize contextual factors and mechanisms that can cause failures at the cognitive level. HRA methods should also offer a structured and systematic way to incorporate this information into the quantification process.

- Although extended qualitative analysis can help analysts understand the context and dynamics of complex scenarios and uncover scenario-specific performance drivers; it may not necessarily lead to appropriate HEPs in all cases because of the difficulties in translating qualitative analysis into HEP effect. Coherent coupling between the qualitative analysis and the quantitative model is needed.

- Adherence to good practices with improved guidance can improve HRA predictions.

Achievements and Overall Conclusions

The U.S. Study used a small number of operating crews and simulated accident scenarios. At least two HRA teams applied the same method to predict crew performance with standard HRA practices to the extent possible. The study produced a set of findings on HRA methodology and process and on how different analysts can use the methods in different ways to produce different results and affect the predictive power of the methods. In particular, the intra-method comparisons conducted in the study gave valuable insights into the interactions between method effects and analyst effects in causing variability in HRA results, and allowed us to find methodological enhancements needed to improve the ability of analysts to produce reliable HRA results.

In addition to the important information on HRA methodology, related achievements from the study, in conjunction with the International Study, are summarized below.

- The methodological tools developed in the studies were tailored to HRA needs and are proving to be very useful achievements. These tools include (1) the development of the experimental design, focusing on evaluating HRA methods, (2) the method for collecting crew data, (3) the method for analyzing crew performance data in terms of tasks corresponding to PRA-type HFEs and in terms of the corresponding HFE boundary conditions, and (4) the approach to data-to-method comparisons. The studies have also demonstrated that useful information on HRAs and HRA methods can be obtained without using impractically large numbers of operating crews and scenarios, which is another important achievement.

- The studies have shown the feasibility and usefulness of simulator data for HRA studies. Although simulator data was used as the empirical data for comparisons against HRA predictions, the promising results from this study support continued use of simulator data in the future, as well as encouraging analysts to use it for other purposes. The study also shows the potential of using and aggregating empirical simulator results from multiple studies to strengthen the empirical basis for both method assessment and extending the scope of methods to address some of the identified shortcomings. In summary, although there are other sources of HRA data, this study reinforced the relevance of simulator data for HRA in general.

- The scenarios developed in the studies are similar to those modeled in PRA and represent difficulty levels from basic to highly complex. They can be used as standard scenarios for other HRA benchmarking studies. Complex scenarios can be used to determine whether an HRA method may lead to underestimation in adverse scenarios. Basic scenarios can be used to establish a baseline performance. The difficulty levels can be used to test whether a method can produce HEPs with appropriate differentiation.

- The study also provided valuable empirical evidence to improve plant procedures and training programs

# ACKNOWLEDGMENTS

This study is supported by an international collaboration between the U.S. Nuclear Regulatory Commission (NRC), the Joint Programme of the Organization for Economic Co-operation and Development (OECD) Halden Reactor Project in Norway, the U.S. Electric Power Research Institute (EPRI), and the Paul Scherrer Institute (PSI) in Switzerland.

Special thanks go to Mr. Carl (Rick) Grantom of the South Texas Project who initiated and facilitated the collaboration between the utility and the NRC to support the study. In addition, South Texas Project allowed the use of their training simulator to collect the data for the study and provided numerous participants from the training department to help design and conduct the study. In particular, Mr. Rod Nowell assisted with designing the experimental scenarios and collecting and analyzing the data. His contributions to the study were invaluable.

The work of the nine human reliability analysis (HRA) teams was obviously of critical vale to the study. The authors of this report want to give a warm thanks to all of the individuals who participated on an HRA team. We have tried to evaluate their analyses as objectively as possible. The following individuals (listed alphabetically below) participated on one of the nine HRA teams, providing the HRA analyses and predictions for the scenarios of the study. In addition, many of the individuals listed below participated in a workshop to discuss the initial results of the study and thereby contributed to the findings and conclusions obtained from the study.

- * Paul Amico, Science Applications International Corporation
- Dennis C. Bley, Buttonwood Consulting, USA
- Michael Brown, U.S. NRC, Office of New Reactors, USA
- Y. James Chang, U.S. NRC, Office of Nuclear Regulatory Research, USA
- * Erin Collins, Science Applications International Corporation
- Lawrence Criscione, U.S. NRC, Office of Nuclear Regulatory Research, USA
- David Gertman, Idaho National Laboratory, USA
- Tina Ghosh, U.S. NRC, Office of Nuclear Reactor Regulation, USA
- Manuel González-Cuesta, Universidad Nacional Autónoma de México, Mexico
- Sandra Herrick, U.S. NRC, Office of Nuclear Regulatory Research, USA
- Jaroslav Holy, Nuclear Research Institute, Czech Republic
- Stanislav Hustak, Nuclear Research Institute, Czech Republic
- Jim Kellum, U.S. NRC, Office of New Reactors, USA
- Kaydee Kohlhepp, Scientech, USA
- Jan Kubicek, Nuclear Research Institute, Czech Republic
- Jeff LaChance, Sandia National Laboratories, USA
- * Tom McComas, Science Applications International Corporation
- Pamela Nelson, Universidad Nacional Autónoma de México, Mexico
- April Whaley, Idaho National Laboratory, USA
- Jing Xing, U.S. NRC, Office of Nuclear Regulatory Research, USA
- Antonios Zoulis, U.S. NRC, Office of Nuclear Regulatory Research, USA

(* Currently with Hughes Associates)

The authors gratefully acknowledge the following individuals (listed alphabetically below) who served as independent assessors to compare the HRA predictions from a particular method application to the crew performance data collected on the training simulator. They also attended the study workshop and participated in discussions of the initial study results, contributing to the findings and conclusions obtained from the study.

- Alysia Bone, U.S. NRC, Office of Nuclear Regulatory Research, USA
- Kevin Coyne, U.S. NRC, Office of Nuclear Regulatory Research, USA
- Katrina Groth, Sandia National Laboratories, USA
- Stuart Lewis, Electric Power Research Institute, USA



Sandia National Laboratories is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Idaho National Laboratory is a multiprogram laboratory operated by Battelle Energy Alliance LLC for the U.S. Department of Energy under contract DE-AC07-05ID14517.

The PSI contributions to this work were supported by the Swiss Federal Nuclear Safety Inspectorate, under DIS-Vertrag Nr. 82610.

# 1. INTRODUCTION

## 1.1 Study Background

Consistent with the U.S. Nuclear Regulatory Commission's (NRC's or the Commission's) policy statements on the use of probabilistic risk assessment (PRA) and for achieving appropriate PRA quality for NRC risk-informed regulatory decision-making, the NRC supports many activities to improve the robustness of human reliability analysis (HRA). A particular aspect of these activities is addressing variability in HRA results. The existence and use of different HRA methods that rely on different human performance frameworks and data, as well as inconsistent implementation of the methods by analysts, appear to be the most common sources of the variability in HRA results. Addressing variability in HRA has also been part of international cooperative activities sponsored by the Organization for Economic Co-Operation and Development (OECD), which has been pursuing simulator studies at the Halden Reactor Project (HRP).

To address HRA issues, the NRC participated in and supported the International HRA Empirical Study [1-4], in which HRA predictions of different analysts and methods were compared to observed crew performance data at HRP's HAMMLAB (HAlden huMan-Machine LABoratory) simulator facilities. The International HRA Empirical Study identified important strengths and weaknesses of the various HRA methods used in the study. An important conclusion was that HRA methods could be improved by enhancing the qualitative analysis portion of the methods in order to reduce some of the sources of the variability in HRA results. However, because there was only one case in the International Study where the same HRA method was applied by different teams, it was difficult to clearly separate method-specific effects from variability created by the analysts' idiosyncratic application of a given method. Thus, in addition to examining differences across methods, a major objective of the present study, which was performed on a U.S. nuclear power plant (NPP) simulator and is thus referred to as the U.S. HRA Empirical study in the report, was to test the consistency and accuracy of HRA predictions among different analyst teams using the same methods. A particular area of interest is examination of the qualitative analyses made by different teams using the same HRA method to identify shortcomings that contribute to inconsistencies in results and to determine the extent to which the shortcomings are from analyst differences or from inherent shortcomings in the methods.

Two other potential limitations of the International Study are also addressed in the U.S. Study. First, in the International Study, the HRA teams were unable to visit the Halden simulator and collect HRA-related information through interviews with plant operators and trainers and through observations of actual operating crews in the simulator, as is typically done in performing an HRA for an NPP PRA. This type of information was given to the HRA teams to the extent possible by the study team in the International Study. The HRA teams also were allowed to submit written questions that were answered by the study team and plant personnel as needed. However, some of the HRA teams in the international Sudy felt that they were significantly limited in their ability to do an adequate HRA. In contrast, the HRA teams in the U.S. Study were able to visit the reference plant and personally collect information relevant to doing their HRA as would normally be done in a PRA.

Second, there was some concern that because the International Study was based on the results of simulator runs using European crews at the HRP, the results might not be directly generalizable to what would occur with U.S. NPP crews. In addition, some of the HRA teams in the International Study thought that their expertise was more geared to understanding what

U.S. crews would do and that their U.S. bias might have influenced their decisionmaking in applying their HRA methods. Thus, the U.S. Study provided an opportunity to determine whether the results were affected by the suspected bias.

## 1.2 Overview of the Study Design

The U.S. HRA Empirical Study focuses the HRA on the control room personnel actions required in response to PRA initiating events. This focus was motivated by the widespread use of HRA methods within PRA (or probabilistic safety assessment (PSA)) in the industry and the related importance of understanding and predicting human performance in accident scenarios. The U.S. Study followed the same general approach as that followed in the International Study [1–4], with the exception that in the U.S. Study the HRA analysts were able to visit the plant, observe a single crew in a training scenario in the simulator to get a general idea of crew interaction with the plant, and interview training personnel to collect specific information about the scenarios to be analyzed in the study. An overview of the general approach for both the present study and the International Study is presented in Figure 1-1 and consists of four high-level tasks listed below:

(1) **Task 1.** The definition of the scenarios and of the human failure events (HFEs) to be analyzed and a compilation of the inputs for the HRA analyses (e.g., relevant procedures).

(2) **Task 2.** The analysis of the HFEs with HRA methods, which produced the predicted outcomes.

(3) **Task 3.** The production of the empirical data for comparison with HRA predictions, starting from the collection of raw data in simulator experiments conducted at a U.S. plant and followed by the analysis of the data. The task difficulty ranking was generated in this Task.

(4) **Task 4.** Review of the HRA submittals, comparison of HRA predictions to the empirical data, along with intra-method comparisons (similarities and differences in results between analysts using the same method), and development of insights for improving HRA methods and practices.

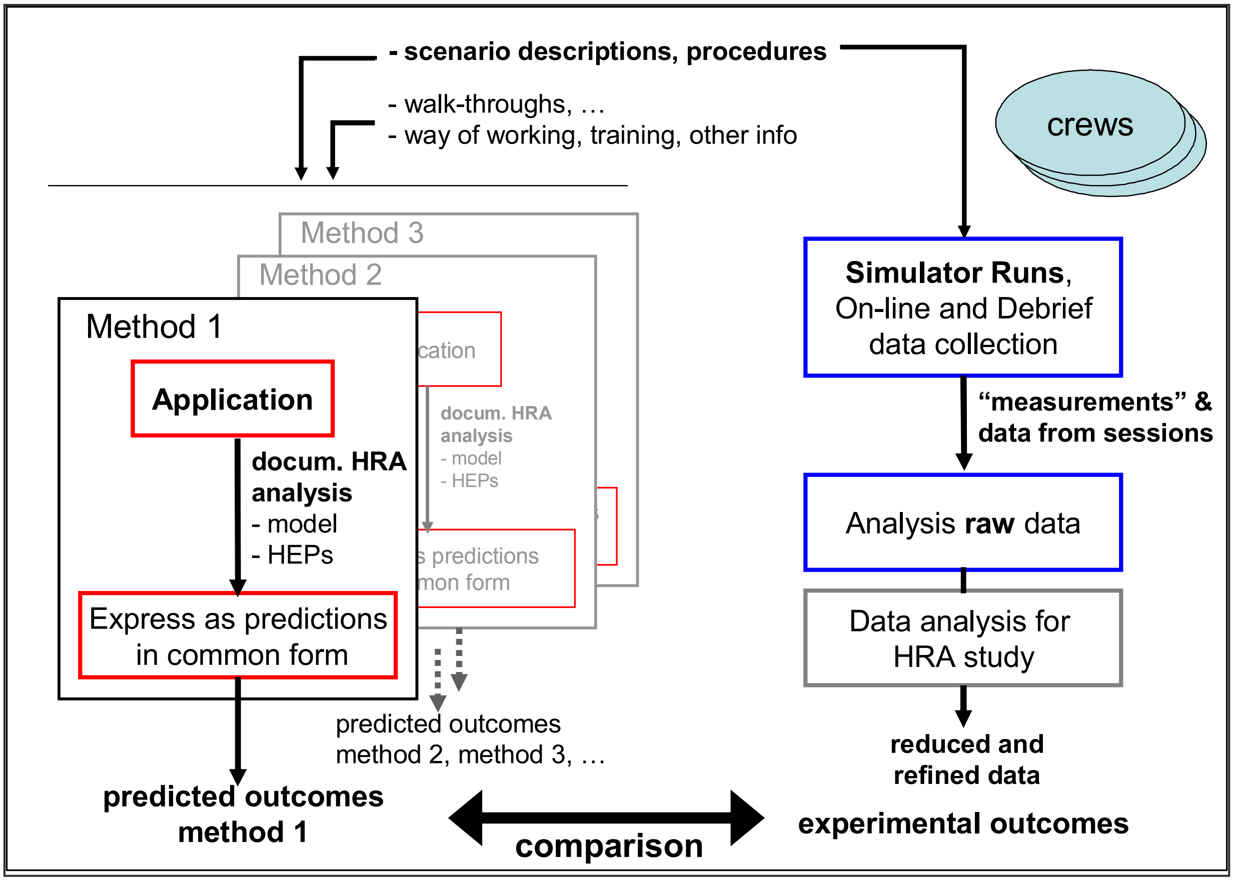



**Figure 1-1 Overview of the U.S. HRA Empirical Study**

Task 1 was the compilation of the inputs for the HRA analyses. As shown at the top of Figure 1-1, these inputs included not only the descriptions of the scenarios and of the HFEs to be analyzed but also information on the relevant procedures, operator training, conduct of operations, and other aspects of the performance context. The performance of the predictive HRA analyses (Task 2) is shown on the left. The production of the empirical data, Task 3 (right-hand side of Figure 1-1), consisted of three subtasks: (1) the simulator experiment itself, in which the operator crews responded to the scenarios while observations and other data were collected; (2) a first data analysis stage aimed at producing an understanding of the individual crews' performance; and (3) an HRA-oriented data analysis, which aggregated the set of crew performance data to characterize the overall performance level of the crews related to each HFE and the performance drivers. Task 4 compared the predicted outcomes with the empirical outcomes, and required them to be expressed in a format that allowed the predictions of the methods to be compared to the empirical data. Task 4 also included the intra-method comparisons made to separate analyst effects from method effects to the extent possible, and identify factors that contribute to variability in HRA results.

# 2. STUDY DESIGN

## 2.1 Study Organization, Participants, and Roles

There were four sets of study participants:

(1) **Experimental team** (Tasks 1, 3): The experimental team was composed of training staff from a participating U.S. pressurized-water reactor (PWR) 4-loop Westinghouse nuclear power plant (NPP) and research staff from the Halden Reactor Project (HRP), Idaho National Laboratory (INL), and the U.S. Nuclear Regulatory Commission (NRC). They collected and analyzed crew performance data on the full-scope, conventional control room training simulator of the participating NPP.

(2) **Operation crews** (Task 3): Four licensed operation crews from the U.S. NPP participated in the study. Each crew responded to three scenarios in the plant simulator, with the exception that one of the four crews was unable to complete one of the scenarios because of a simulator problem.

(3) **HRA teams** (Task 2): Each of nine human reliability analysis (HRA) teams applied an HRA method to obtain predictions for the human failure events (HFEs) in the scenarios defined for the study. Individuals representing industry, regulators, and the research community participated, used one of four HRA methods selected for the study. Each team had mixed organizations. The different methods are discussed in the next section.

(4) **Assessment group** (Overall organization and Task 4): This group (including staff from Sandia National Laboratories (SNL), HRP, Paul Scherrer Institute (PSI), Electric Power Research Institute (EPRI), and the NRC) had the overall responsibility of organizing and carrying out the study. In the early stages of the study, this group prepared the information package (analysis inputs) for the HRA teams and answered their subsequent requests for additional information and questions concerning ambiguities in the instructions and assumptions. They also organized a plant visit for the HRA teams to collect information needed to do their HRA. After the HRA teams delivered their analyses, the assessment group reviewed and summarized the HRA teams' analysis results before doing the actual comparisons with the crew data and the comparisons between the different teams using the same method. The assessment group were HRA experts who were not part of the HRA teams.

For a number of its tasks, the assessment group worked closely with the HRP, INL, NRC, and reference plant staff to prepare the information package, answer operational questions about the simulations, and to make the comparisons. To avoid biasing the comparisons in Task 4, a study protocol was used whereby the assessment group did not receive any information about the actual crews' performance in the NPP simulator until after they had reviewed and summarized the HRA teams' predicted outcomes. Similarly, the experimental team conducted their data analysis to produce the empirical data without knowledge of the HRA teams' predictions. The assessment group and the experimental team cooperated extensively on the comparison itself, including interpretation and representation of the simulator data for comparison with the HRA teams' predictions.

## 2.2 HRA Teams, Methods and Experimental Treatment

Nine teams participated in the study and each team applied an HRA method to obtain predictions for the HFEs in the simulator scenarios described in Chapter 5. The teams and their methods are listed below. In summary, two teams used ATHEANA (A Technique for Human Event Analysis, NUREG-1624, Rev. 1 [4], NUREG-1880 [5]); two teams used SPAR-H (Standardized Plant Analysis Risk-Human Reliability Analysis, NUREG/CR-6883) [6]; two teams used ASEP (Accident Sequence Evaluation Program Human Reliability Analysis Procedure, NUREG/CR-4772) [7]; two teams (CBDT & HCR/ORE Teams 2 and 3) used the EPRI HRA Methodology with the HRA Calculator version 4.1.1 [8] (with CBDT (Cause-Based Decision Tree [9]), HCR/ORE Human Cognitive Reliability /Operator Reliability Experiments [9] and Technique for Human Error Rate Prediction (THERP) [10]), and one team (CBDT & HCR/ORE Team 1) used a hybrid CBDT & THERP & ASEP method.

**Table 2-1 HRA Teams and Their Methods**

| HRA Methods | HRA Teams |
|---|---|
| ATHEANA | • ATHEANA Team 1<br>• ATHEANA Team 2 |
| SPAR-H | • SPAR-H Team 1<br>• SPAR-H Team 2 |
| CBDT & HCR/ORE | • CBDT & HCR/ORE Team 1<br>• CBDT & HCR/ORE Team 2<br>• CBDT & HCR/ORE Team 3 |
| ASEP | • ASEP Team 1<br>• ASEP Team 2 |

### 2.2.1 Information Package

Each HRA team received an information package (see Appendix A), which included the following items, to be used as the basis for the application of an HRA method:

(1) overview of the information package and instructions for the HRA teams
(2) study outline
(3) scenario descriptions and HFE definitions (HFEs to be quantified)
(4) forms for HRA team responses

More generally, the package supplied information about the organization of the study, information about the specific scenarios simulated in the reference plant, and the expected output from the HRA teams' analyses (described below).

In addition, while analyzing the scenarios, the HRA teams had the opportunity to ask for clarifications or additional information from the assessment group (supported by the experimental team), who had the overall responsibility for the organization and implementation of the study. The questions from the individual teams and answers from the assessment group were not generally shared with all HRA teams to avoid the influence of extra information on other teams' analysis results, with the exception of a few instances in which it was deemed appropriate to share the information it was decided to be shared with all HRA teams (e.g., an error in the timing information originally given for one of the HFEs).

### 2.2.2 Plant Visit

A prerequisite for HRA analyses in a probabilistic risk analysis (PRA) is the familiarity of the analysts with the background, training and experience of the performers (the crews) and the performance conditions (e.g., human-system interface and job aids such as procedures). In the study, all HRA teams had an opportunity to familiarize themselves with the plant during a 2-day plant visit. During the visit, the HRA teams toured the plant simulator and were allowed to ask general questions about operations in the control room and the panel layout and usage. The teams also observed a crew of operators in an actual simulator training scenario (not one of the test scenarios), but were not allowed to ask questions to interrupt operators' normal training. In addition, each HRA team had two, 2-hour sessions of interviews with different plant trainers to ask specific questions on the scenarios described in Chapter 5. . They were allowed to collect any information that they thought would be relevant to their HRA analyses, such as procedure usage, relevant cues and timing information, and training related to the scenarios.

Two plant visits took place, with approximately half of the HRA teams present for each visit. The HRA teams toured the simulator and observed a training scenario as a group. The training scenarios and the crews were not the same at each visit. However, each team interviewed the trainers separately and interviewed a different trainer at each plant visit. The intent was to allow the HRA teams to get possibly different perspectives on the scenarios and confirm information previously received. The trainers were not prepared on the scenarios before the interviews for two reasons. First, HRA analysts are normally expected to explain scenarios to plant personnel during interviews. Second, the study team was not able to prepare the trainers because of time constraints. The HRA teams were allowed to submit follow-up questions to the assessment group, which sought information from plant staff as needed.

### 2.2.3 HRA Documentation

There are differences in the underlying models, the number of performance shaping factors (PSFs), the definition of their scope, and the terminology used in different HRA methods. In addition, the documentation of HRA analyses in a PRA is typically oriented to tracing how the information on the performance conditions obtained in the qualitative analysis has been incorporated into the estimation of the HFE failure probability, rather than to predicting specific outcomes in terms of behaviors and actions. To address the terminological differences as well as to provide predicted outcomes that could be compared with the simulator outcomes, the HRA teams were asked to deliver their predictions in three parts:

(1) Form A (see Appendix A): An "open form" questionnaire where the teams reported for each HFE (1) the human error probability (HEP), (2) the driving factors, and (3) the "operational expressions." The questionnaire asked essay-type questions, which permitted the teams to respond using the terminology of the applied method. Specifically, the teams were asked to specify the important failure and success performance drivers (driving factors) related to each HFE, as well as to discuss them in terms of expected crew behaviors and responses. These behaviors and responses are the means by which the driving factors express themselves in NPP operation-specific and scenario-specific terms, and are therefore referred to as "operational expressions." They were particularly helpful in unambiguously interpreting the teams' statements related to the driving factors.

(2) Form B (the Diaries) (see Appendix A): The intent of Form B was to obtain additional information on the overall process that the HRA teams used for their qualitative and quantitative analyses.

(3) Supplemental material specific to each method, which included a "normal" documentation of their HRAs, as in a PRA.

In many cases, the methods themselves do not require the analysts to produce the results exactly the way they were requested in this study. Although asking the analysts to document their results in the way described above may have required the teams to go beyond the specifics of the method to some extent, it can be argued that it was not much more than what is normally expected when doing an HRA. Certainly a major goal of an HRA is to identify when the crews' response will be good (that is, anticipation of high success rates) and when and why potential problems might arise in crew performance: that is, the application of any of the HRA methods should be based on a reasonable understanding of what can occur in the various accident scenarios. Thus, it was expected that, in general, HRA teams should be able to provide the needed information and that this information could be compared to the simulator results and the results of the other methods.

## 2.3 Defining Human Failure Events

The definitions of the scenarios and HFEs in the study are presented in Chapter 5. It should be noted that a number of issues arise in defining HFEs for studies involving benchmarking HRA predictions against empirical data collected from simulators. The following discussion about these issues is borrowed but adapted somewhat from a similar discussion in the international HRA study [4]. It is adapted to reflect the efforts made in this study to reduce ambiguity in HFE definitions, particularly with respect to the timing criterion for a successful crew response.

To ensure that all HRA teams would produce predictions for identically defined HFEs, this study predefined all HFEs for the HRA analysts. These definitions were based on the definitions of similar HFEs from real plant PRAs and were defined on a functional level (i.e., "fails to perform X before Y" or "fails to perform X within *t* minutes"). In some cases the HFEs were defined with stricter success criteria than a standard PRA would use. For example, failure for some actions might result in a near-immediate plant state change, such as onset of core damage. This type of action could be defined in this study as it would be in a PRA. However, some other actions may not have such a near-immediate and/or irreversible plant state change, but might instead have a lesser effect, such as delaying the overall timing of the operators' response or influencing the evolution of the scenario (e.g., system pressure ends up at a level that is not preferred), without necessarily affecting the final outcome of interest (e.g., core damage). Therefore, for actions without immediate and/or irreversible plant state changes, the corresponding HFEs were defined by superimposing timing criteria.

In some cases, PRA event trees model accident sequences by defining several HFEs that are carried out in sequence. Failure of any specific HFE within the sequence, or even all HFEs observable in a typical simulator session of 1 to 2 hours, might not cause core damage. In a steam generator tube rupture (SGTR) event, for instance, the HFEs include identification and isolation of the affected steam generator, forced cooldown of the reactor coolant system, depressurization of the system, and termination of safety injection when appropriate. These actions collectively stop the leakage past the tubes, significantly decrease the radiation being released to the secondary side of the plant, and allow for control of the plant to eventually achieve safe shutdown. Failure in any or even all of these actions does not necessarily cause core damage.

On the other hand, because these actions should be conducted expeditiously in order to stop the primary-to-secondary leakage, decrease the radiation release, and avoid core damage without resorting to less desirable means (e.g., bleed and feed), it is reasonable to associate failure of any single action with a time criterion. These definitions impose a time by which the actions should be taken, as well as the precise steps that have to be taken to complete the desired action (e.g., close main steam line isolation valves as part of the overall isolation action). This allows analysts to observe whether the actions are done within the desired time frames, and whether all critical steps are indeed carried out.

Another reason to adopt stricter failure criteria is that, when assessing the effect of performance shaping factors (PSFs), human performance should ideally be treated at a fine level of granularity. As there can be several human actions that are carried out in a particular sequence to achieve one specific goal, the contexts or PSFs for these actions vary. The procedure might for example give guidance of varying clarity, the complexity of the subtasks can also vary, and, if the series of tasks is extended in time, stress might increase with it.

In the International Study, the timing criteria used in the definitions of the HFEs were obtained from expectations of operators' performance and/or inferred from plant thermal-hydraulic expectations for the scenarios. For a few of the HFE definitions, time limitations were based on trainers' expectations, rather than more common PRA criteria (e.g., "25 minutes from rupture" rather than "overfill and opening of the steam generator (SG) relief valve" as failure criteria for the HFE "Broken SG identification and isolation"). It was recognized by the study team at the time that failure criteria based on training expectations and past experiences with the participating crews could lead to somewhat artificial definitions of failure, as training expectations and past experience may be loosely defined.

To avoid the limitation discussed above, none of the success criteria in the present study were based on trainers' expectations. All timing criteria for HFE success were tied to core damage or to key parameters such as the pressure (and the associated time) at which a power-operated relief valve (PORV) would be expected to open in an SGTR scenario if the correct depressurization steps were not done. Because the times associated with when a particular parameter would be expected to be reached can vary based on the scenario initial conditions and on timing of operator actions during the progression of the scenario, the times used in this study are only estimates. However, the estimates were based on the timing from simulator runs at the plant. Therefore, they can be considered a reasonable approximation of real times. It is also acknowledged that since the times selected are only estimates based on a selected set of assumptions (e.g., scenario development with no operator actions performed), it is possible that the crews could manage to maintain the plant in a stable state over time frames longer than the estimated times and complete the actions after the time criterion, which would be considered as a success from a plant perspective. However, it was counted as a failure in the study since we asked the HRA teams to predict the HEPs for failing to complete the action within the specified time frame.

Another argument to include timing criteria involves the expectations of the HRA methods to be applied: for instance, many methods use a time reliability function to estimate the diagnosis failure probability portion of the HEP. Use of this function requires a definitive time by which the desired action must be completed (typically referred to as the "allowable time"). Then, based on the time estimated for completing the action, the analyst can block out a time by which the diagnosis must occur. Using the time reliability function to determine the diagnosis time yields the diagnosis failure probability. To use such methods, the HFEs had to be defined in terms of time, although we recognize that the relation to the plant thermal-hydraulic expectations for the scenarios could have been made more explicit in the example above.

# 3. EMPIRICAL DATA COLLECTION AND ANALYSIS

## 3.1 Crew Composition

The data were collected on a U.S. plant full-scope training simulator (Westinghouse 4-loop pressurized-water reactor (PWR)) with a conventional control room. Four crews of five licensed crew members participated. For three of the crews the composition was one Shift Manager (SM), one Unit Supervisor (US), one Shift Technical Advisor (STA), and two Reactor Operators (ROs). In one crew the SM was not present and instead they had three ROs. A typical crew in the plant comprises six members: one SM, one US, one STA, and three ROs (including one auxiliary reactor operator (ARO)). However, it is not unusual for only five members to be in the control room at a given time and is consistent with the plant's Conduct of Operations.

## 3.2 Scenario Presentation Order

Three of the four crews ran all three scenarios of the simulation (see Chapter 5. for description of the scenarios), but one crew was unable to complete Scenario 3 (steam generator tube rupture (SGTR)) because of a simulator problem. To control for potential confounded learning effects caused by the order of the simulated scenarios, the crews ran the scenarios in different orders. The crews did not have advance knowledge of the scenarios that they would face.

## 3.3 Data Collection

Data were collected during each simulator session through the following: (see Appendix B for the procedures for data collection):

1. notes from observers from the experimental team of crew actions
    - two observers in the control room
    - four observers in the back of the control room and in the simulator booth

2. notes from observers from the plant of the crew actions (timeline and comments; predefined actions)
    - two trainers in the control room
    - one trainer in the simulator booth

3. simulator logs
    - alarms
    - process parameters (selected)
    - simulator actions

4. audio and video recording
    - Data were collected after each simulator session

5. critical decision interviews conducted by three observers from the experimental team with:
    - crew decision makers (SM and US)
    - two plant trainers

During this debriefing interview, each phase of the scenario was discussed with the crew. The crew decisionmakers were asked to describe the scenario evolution from their point of view.

The Halden Reactor Project (HRP) observers then asked them a series of pre-defined, scenario-specific questions aimed at gaining further insight into the reasoning of the crew decisionmakers, their assessment of the situation, and their awareness of alternative courses of actions. After the talk-through for each HFE, the trainers assessed the performance of the crew using an adapted version of the NRC Simulator Crew Evaluation Form, NUREG-1021, Form ES-604-2 [11] (this form is routinely used by the trainers during normal training sessions) (see Appendix C). At the end of the debriefing interview, the US was asked to rank the difficulty of the human failure events (HFEs).

6. Interviews conducted by three observers from the experimental team (different from those doing the critical decision interview) with the following crew members (conducted in parallel with the critical decision interview and focused on experienced difficulties in the scenarios and on communication):
    - STA and ROs
    - one operator trainer

7. performance shaping factors (PSF) questionnaire
    - All crew members

## 3.4 Data Analysis

The data analysis approach and process, which followed those used in the International HRA empirical study, are described in the following subsections. See [2–3] for details.

### 3.4.1 Scenario Development and Crew Responses

The analysis was done by the HRP staff (with some input from Idaho National Laboratory (INL) staff) who observed the data collection and by a plant trainer. The scenario development and crew responses were summarized for all crews for each scenario by reviewing:

(1) notes taken during the scenarios by the experimental team and one or two trainers
(2) notes from the critical decision interviews
(3) simulator data logs and trend curves created from these logs
(4) when necessary, video recordings were consulted to clarify specific issues

### 3.4.2 Drivers of Performance

For each main task in a given scenario, including the HFEs and some additional tasks, the experimental team attempted to determine what had made the crew succeed, not succeed or not perform as optimally as they could have (for example: What made the crew stop the reactor coolant pumps (RCPs)? What made the crew stop the RCPs late in the scenario?). Information was compiled on these questions from notes taken during the debriefing interviews. Thus, a list of statements from the SM and US, and some statements from the trainers, were available for each of the questions. From these statements the drivers of performance that seemed to be the most important were determined.

### 3.4.3 Performance Shaping Factors (PSFs)

The PSFs were evaluated by two analysts of the experimental team and a trainer from the plant for each HFE, and were based on the analysis of crew performance, interviews, scenario analysis, and informed by the NRC Simulator Crew Evaluation Form (NUREG-1021, Form ES-604-2 [11]) (see Appendix C).

Each PSF was evaluated taking into consideration two questions: Was the PSF present, and did it have an influence on the task in question (the HFE)? Based on the rating system developed for the International HRA empirical study [1–4], the PSFs were categorized as:

(1) N/P = Nominal or Positive
(2) 0 = Not a driver
(3) ND = Negative driver
(4) MND = Main negative driver

PSFs that directly influenced performance of the HFE were categorized as either ND or MND depending on how strongly they affected the performance. PSFs that were positive or good enough to not negatively influence the HFE were categorized as N/P. Some PSFs may have varied across the crews, but if a direct effect on performance could not be observed, these were categorized as 0.

The PSFs evaluated are described in Table 3-1. The selection and definitions of the PSFs were based on the HRA Good Practices document (NUREG-1792) [12], and also included factors that the experiment team considered necessary to explain the behavior of the crews in the simulator scenarios.

These PSFs are also discussed in the reports from the first phases of the International HRA empirical study [1-2]. Note that in the U.S. Study there were some differences in the PSFs evaluated compared to the International Study. In the present study, the PSF Time Pressure was not evaluated because of the difficulty in assessing subjective time pressure and the even greater difficulty in establishing a causal link between this factor and performance. The two PSFs Training and Experience were combined into one category in the U.S. Study. In the International Study, a distinction was made between these where one was used for general training on the type of situation being studied (e.g., SGTR), and the other to cover training on the specific situation in the scenario (e.g., an SGTR with masked radiation indications). In the current data analysis, any differences between general and specific training were noted in the discussions of the PSFs identified as affecting crew performance or in the summaries of PSFs identified in the HRA analyses.

**Table 3-1 Definitions of Performance Shaping Factors (PSFs)**

| PSF | Definition |
|---|---|
| **Adequacy of time** | The adequacy of time relates to the difference between the time available and the time required. The time available is estimated based on an expected evolution of the scenario, which defines when performing the action modeled by the HFE can no longer be effective in reaching the success criteria. The time required is an estimate of the time needed by the crews to perform the cognitive and execution components of the task.<br>The adequacy of time affects the assessment of the human error probability (HEP) simply because there may be a shortage of time to get the actions done as well as by allowing opportunities for checking the performance of the action, detecting errors, and correcting these errors. |
| **Stress** | Effect of high workload, perceived time pressure, urgency, perceived threat on performance, perceived severity of consequences, perception/effect of losing overview and control over the situation. |
| **Scenario complexity** | Difficulty of situation assessment and diagnosis. Related to ambiguous situations (e.g., masking), diagnosis complexity, and the need to decipher and combine numerous indications, alarms, and other sources of information in order to assess the situation.<br>The number of simultaneous goals influences both scenario complexity and execution complexity. If it involves prioritization, it probably should be listed under 'Scenario Complexity,' which deals with decisionmaking, planning, etc. If it involves the management and coordination of tasks, it should probably be listed under 'Execution Complexity.'<br>In many PSF frameworks, this PSF relates to the indications of conditions (availability of cues, ease of perceiving these cues, the difficulty of interpreting these indications). |
| **Indications of conditions** | Availability and clarity of key indications and/or alarms. This is affected by the availability of instrumentation and, given that the instrumentation is available, the salience of cues, signal-to-noise, ambiguity of cues. In some cases, also the availability of system feedback for execution.<br>This factor is often treated in 'scenario complexity' although the latter has a larger scope. |
| **Execution complexity** | Difficulty of performance (implementation) of the task (not including situation assessment, diagnosis, etc.). The number of steps to be done, whether the task is associated with a single variable or multiple variables, nonlinear response of the system, so that you need "to have a feel" in order to adequately control, and whether special sequencing or coordination of multiple performers is required are features of a task that increase the execution complexity.<br>The number of simultaneous goals influences both scenario complexity and execution complexity. If it involves prioritization, it probably should be listed under 'Scenario Complexity,' which deals with decisionmaking, planning, etc. If it involves the management and coordination of tasks, it should probably be listed under 'Execution Complexity.' |

| PSF | Definition |
|---|---|
| **Training/ Experience** | The degree of familiarity with the scenario and the actions to be performed that can be expected based on the training of the crews. Includes both theory/knowledge (classroom) and practice (e.g., in training simulator).<br>The factor should consider not only the amount or general quality of training but also the applicability of the training in the specific scenario (i.e., how helpful the training received will be in the scenario). In rare cases, the training may even be counterproductive.<br>Note: HRA analyses deal primarily with training as it concerns the behavior of the nuclear power plant (NPP) and the appropriate situation-specific response. In data analysis, training also includes training on how to solve problems in general, etc. |
| **Procedural guidance** | Support provided by the procedure for performing the situation assessment (decisionmaking) and execution of the specific task being analyzed. In the context of the scenario of interest, steps that are ambiguous, unclear (including layout), or not detailed and situations where the way to proceed through the procedure is unclear contribute to a poor rating for this factor. |
| **Human-Machine Interface** | Ergonomics, including the presentation and labeling of process parameters, the availability of feedback following an action on a component or system, and the interface for acting on components or systems. |
| **Work processes** | Refers to the way of working and mechanics of work, (e.g., the care taken in reading procedures and generally in doing the task work). Task work is referred to as the work directly with the process as opposed to teamwork work which is about the collaborative aspects of work. Task work can be analyzed at a more individual level than teamwork,<br>In a predictive analysis, this factor indicates how well the expected work processes match the given scenario and how sensitive the task may be to work practices.<br>In analyzing an actual performance, this factor is rated poor if individual work is not thorough, and if the general handling of the procedures is less than adequate. Note that in fast-moving scenarios, "good" work processes may have a negative effect on task success.<br>In the given study reactor operator and auxiliary reactor operator sometimes do process work together as a close unit. In these cases this is analyzed as work process and not teamwork. |

### 3.4.4 Difficulty Ranking of HFEs

After having completed all three scenarios in the study, the USs from the crews were asked to rank the difficulty of the HFEs from the most difficult to the least difficult (see Section 3.3). The rankings were collected from three of the four crews that participated in the study. The crew from which the ranking was not collected did not participate in all three scenarios. The rankings made by the USs were compared with the failure rates of the HFEs and although the failure rates are based on a limited number of crews, the rankings were consistent with those results.

# 4. METHODOLOGY FOR ASSESSING AND COMPARING HRA PREDICTIONS AGAINST SIMULATOR DATA AND THE INTRA-METHOD COMPARISONS

This chapter describes the methodology used to compare and assess the HRA predictions from each team against the empirical data from the three scenarios used in the study. It also describes the methodology used for the intra-method comparisons, where the results from different human reliability analysis (HRA) teams using the same method are examined. Chapter 3. described the data analysis method used to obtain the empirical data from the crews and the resulting data are presented in Chapter 5. . The human error probabilities (HEPs) predicted by the HRA teams are presented in Chapter 6. . The results of the comparisons of HRA predictions with the empirical data for all of the human failure events (HFEs) and an overall assessment of each HRA method application based on the methodology described in this chapter are presented in Appendix G. Intra-method comparisons based on the results of the comparisons and assessments of the HRA predictions against the empirical data are presented for each HFE in Appendix H and the results are summarized in Chapter 7. (intra-method comparisons ) and 8. (factors contributing to intra-method predictive differences).

## 4.1 Overall Process for Comparison and Assessment of the HRA Predictions Against the Empirical Data

The assessment and comparison process consisted of the following steps:

(1) *Summarizing the qualitative predictions made in the HRA analyses.* The qualitative predictions consist of the factors identified as negative drivers for an HFE and the associated failure mechanisms or modes, in relation to the tasks expected or required to be performed (i.e., a description of an HFE in operational terms). Because different HRA analysts could express similar predictions differently because of differences in the levels of details of their analyses and in their terminology, the objective of this step is to ensure that those predictions referring to the same issues or factors are described consistently. This is accomplished by having the assessment group (those summarizing the predictions within the submitted HRA analyses) use a common set of performance shaping factors (PSFs), with the definitions introduced in Chapter 3. To avoid introducing biases in interpreting the predictions, the assessment group summarized the predictions without knowledge of the crew performance in the scenarios in the simulator. The details of this process are described in Section 4.1.1 below.

(2) *Review of the assessors' summary of qualitative predictions by the HRA team that completed the HRA.* An important element of the summarization is to express the predictions in a common terminology. Because this may involve a translation from the method's own terminology to a common terminology, this step was done to ensure that the characterization used in the comparison accurately represents the intent of the HRA teams who completed the analyses. In other words, the HRA teams had the opportunity to review the summaries to make sure they agreed with any translations and the final representations made by the assessors.

(3) *Comparisons of HRA qualitative predictions with the empirical data.* The comparisons were made for each individual HFE using the summaries generated in Step 1 above.

(4) *Comparisons of HRA quantitative predictions with the empirical data.* HEPs for individual HFEs were provided in the HRA. The HRA teams' HFE rankings were generated based on the estimated HEP values. The HEPs and the HFE rankings were both examined in the comparisons with the empirical data, which included the qualitative difficulty rankings obtained from the unit supervisors, as described in Chapter 3. With respect to Steps 1 and 2 above, note that no analogous summarization was done for the quantitative predictions.

(5) *Assessment/comparison summary.* The assessment of each method is made jointly by the assessment group and HRA teams. It includes the results of Step 3 summarized across all HFEs, the results of Step 4, and assessments of the traceability, adequacy of method guidance, and usefulness of the method for error reduction. The results of Steps 3 and 4 served as a basis for an assessment of the qualitative predictive power and quantitative predictive power, respectively, of each application. The criteria for these assessments are discussed below.

Note that the summary in Step 5 was not intended to give an overall assessment of each method based on weighting all of the assessment criteria taken together. Rather, ratings were made by the assessment group only for the separate individual criteria (see Section 4.2 for additional discussion). The intent is to give the reader an initial impression of how well the application of each method did in this study and how well the method did in terms of meeting certain desired HRA criteria.

Another purpose for assessing the predictive power of the method applications and examining aspects such as traceability, method guidance, and insights for error reduction was to facilitate the identification of the strengths and weaknesses of the various HRA methods and HRA in general. That is, if a method did particularly well or particularly poorly in terms of predicting crew performance, what was it about the method or its application that seemed to lead to this result? Similarly, one might ask if the effect appeared to be related to the method itself, its guidance, or the different HRA teams' use of the method. The intra-method comparisons described below addressed some of these issues.

Details on the processes used for the comparisons and assessments described above are described in the sections below and the results are presented in Appendix G.

### 4.1.1 Process for Summarizing the HRA Submittals to Obtain the Qualitative Predicted Outcomes

As discussed in Step 1 of Section 4.1 above, the results of the HRA analyses were summarized in a way to allow direct comparison with the empirical data. This section describes the process used to accomplish the summaries.

As set forth in Section 2.2.3, the HRA submissions typically included the following three types of information:

(1) *Form A* (see Appendix A) represents summary information with a particular emphasis on finding the main drivers in terms of PSFs, causal factors, other influence characterizations explicitly identified through the HRA method being used, and a description in operational terms of the difficulty (or ease) of performance for the tasks associated with each HFE.

(2) *Form B* (the diaries) (see Appendix A) gives information on the overall process used by the HRA teams to do their qualitative and quantitative analyses.

(3) *Supplemental material specific to each method*, which included information such as task analytic reviews of operating procedures, analysis worksheets specific to the HRA method, and documentation of assumptions that were made by the HRA team.

Form A summaries served as the main basis for comparing HRA predictions against the empirical data, but the assessment group drew upon other parts of the submission as needed. At its core, each HRA method attempts to capture factors relevant to the human performance of the HFE with particular focus on the factors that may drive the crews to fail. This was the primary basis for the present comparison of HRA methods to the data — the extent to which the HRA method accurately and completely predicted those factors that shaped performance of the crews in the reference plant simulator. The discussion in Form A was to reflect the basis for the HEP obtained for the HFE and be expressed in terms of the "factors" or characterizations explicitly identified as important from the application of the HRA method. The terminology of the HRA method was to be used by the HRA teams to the extent possible, but as discussed above and in more detail below, in some cases translation into a common terminology was needed.

Not all HRA methods use the same set of factors to represent the crew data and quite often they use somewhat different terminology. Therefore, in order to be able to compare the driving factors identified by the HRA methods with the crew data, the driving factors identified in the HRA teams' analyses were translated by the assessment group into the same set of PSFs in

These PSFs are also discussed in the reports from the first phases of the International HRA empirical study [1-2]. Note that in the U.S. Study there were some differences in the PSFs evaluated compared to the International Study. In the present study, the PSF Time Pressure was not evaluated because of the difficulty in assessing subjective time pressure and the even greater difficulty in establishing a causal link between this factor and performance. The two PSFs Training and Experience were combined into one category in the U.S. Study. In the International Study, a distinction was made between these where one was used for general training on the type of situation being studied (e.g., SGTR), and the other to cover training on the specific situation in the scenario (e.g., an SGTR with masked radiation indications). In the current data analysis, any differences between general and specific training were noted in the discussions of the PSFs identified as affecting crew performance or in the summaries of PSFs identified in the HRA analyses.

Table 3-1 of Section 3.4.3. Similar to what was done with the simulator data, the PSFs were rated based on the extent to which they were predicted to affect crew performance. These ratings were based on the discussions provided by the HRA teams in Form A with respect to the important driving factors and the operational stories of what would be affecting the crews in completing the relevant actions. In addition, the assessment group also reviewed other documentation given by the HRA teams to assess the extent to which the various factors appeared to contribute to the total HEP for each HFE. The following scale was used by the assessors to rate the impact of the various factors:

- MND = Main negative driver
- ND = Negative drivers
- 0 = Not a driver (effect could not be determined)
- N/P = Generally positive effect and contribution to the total HEP being small (Note that some methods use the term "Nominal" to denote a default set of positive circumstances and our use of the N/P rating is consistent with that terminology)
- N/A = Not addressed by the method

A table of PSFs was created by the assessment group for each HFE and each method to summarize PSF ratings. The table also included the specific details of how each PSF was captured in the HRA analysis, and how it was being represented in the terminology of the table. In addition, the assessment group also summarized the HRA teams' description of how the crews would perform operationally in the simulator scenarios, difficulties that might arise, and any other relevant factors from the HRA teams' qualitative analyses.

Because the assessors' summaries of the HRA teams' analyses were necessarily somewhat subjective, the summaries were reviewed by the respective HRA teams (Step 2 in Section 4.1) to seek feedback (e.g., comments on the assessment team's translation of the HRA method's driving factors into the common terminology and the assigned ratings). The assessment team reviewed changes suggested by the HRA teams and discussed them with the HRA teams as needed to reach a consensus on the judgments. This summarization task was completed before either the HRA teams or the assessment team saw the simulator data.

The results of these summaries allowed the qualitative results for each HRA method to be directly compared with the simulator data, which were represented in a similar format.

### 4.1.2 Structure of Comparisons for Each HFE

The results of the qualitative comparisons of the method predictions to data (Step 3 in Section 4.1) are presented in Appendix G and provide the following for each HFE for each HRA team:

(1) a summary of the team's qualitative analysis (operational description)

(2) the quantitative predictions, including their HEPs, uncertainty, and associated insights

(3) a summary table of the driving factors based on the HRA method

(4) a comparison of the predicted drivers to the empirical data

(5) a comparison of the qualitative analysis to the empirical data

(6) a brief discussion of the extent to which the HRA quantification accounted for the factors predicted to affect performance and what effect, if any, the identified PSFs had on the HEPs

A comparison summary addressing the predictive power of each method application in the present study and its strengths and weaknesses is also presented in Appendix G. The assessment criteria are discussed below.

### 4.1.3 Structure of Assessment/Comparison Summary of Each Method as Applied by Each HRA Team

The assessment/comparison summary (Step 5 in Section 4.1) relied on the information from Steps 3 and 4. The process for Step 3 is described above. Step 4 was done in the context of developing the assessment/comparison summary and relied on the results of the overall quantitative results presented in Chapter 5.

The application of each HRA method was assessed against the following criteria based on the HRA teams' predictions for all HFEs.

- predictive power:
    - overall predictive power
    - qualitative predictive power in terms of driving factors (drivers) of performance
    - qualitative predictive power in terms of operational expressions, and
    - quantitative predictive power (to the extent that this can be assessed in light of the limitations of the empirical data)

- traceability of the qualitative analysis and quantification process

- adequacy of the guidance given by each method for the qualitative analysis and quantification

- usefulness of the HRA results for human error reduction

The assessment of each HRA team's application of the HRA method addresses each of the criteria above in statements that yield a qualitative rating from poor to good (5-point Likert scale) of each individual criterion and include the main aspects of how the method performed against the criterion. This assessment takes into account all of the HFEs in the scenarios. In addition, the assessment includes a commentary on the key strengths and some of the weaknesses of the method. The five points of the Likert scale are 'poor,' 'moderately poor,' 'fair,' 'moderately good,' and 'good.' The specific aspects considered in assessing each criterion are discussed further in Section 4.3. The categories used in the assessment are shown in Table 4-1.

It should be emphasized that a single overall assessment "summing" the assessment of the separate criteria is not done. Such an assessment would require an explicit weighting and ranking of the assessment criteria, defining critical and minimum levels of fulfillment of the criteria, and consideration of the resources required for the application of the method and of the needs of the HRA/risk assessment application. The summary assessment includes a section that addresses some of the overall strengths and weaknesses of each method from the perspective of the assessment group, to the extent that these are not clear in the discussion of the separate criteria.

**Table 4-1 Structure of Assessment Summary of Each HRA Team's Application of the HRA Method**

| **Criterion** | **Process step** | **Criteria** |
|---|---|---|
| **Predictive Power** | Overall predictive power | The overall predictive power is assessed, based on the comparisons between the predictions for each HFE and the empirical data.<br>*See 4.3.1 for discussion.* |
| | Qualitative predictive power - comparison of drivers | Assessment of<br>• how well the method predicted the specific performance issues and drivers identified in the empirical data<br>• whether the method predicted factors and issues that were not supported by the empirical data<br>*See 4.3.2.1 for discussion of specific aspects of comparison and assessment.* |
| | Qualitative predictive power - comparison of operational expressions | Assessment of how well the method predicted the failure mechanisms (in operational terms) observed in the empirical data<br>*See 4.3.2.2 for discussion of specific aspects of comparison and assessment.* |
| | Quantitative predictive power - comparison of the quantitative method predictions with the empirical data. | Listed from highest to lowest priority:<br>• potential optimism of the most difficult HFEs<br>• consistency of the ranking of the HFEs (by predicted HEP) with the reference difficulty ranking<br>• predicted HEPs relative to the confidence/uncertainty bounds of the empirical data<br>• quantitative differentiation of the HFEs by HEP<br>*See 4.3.3 for discussion of specific aspects of comparison and assessment.* |
| **Traceability** | Assessment traceability | • traceability of the basis for quantification inputs<br>• traceability of quantification<br>*See 4.3.4 for further discussion.* |
| **Guidance** | Assessment of guidance | • guidance for the qualitative analysis<br>• guidance for modeling of the HFE and decomposition (if applicable)<br>• guidance for the quantification<br>*See 4.3.5 for further discussion.* |
| **Error reduction** | Insights for error reduction | *See 4.3.6 for discussion.* |
| **General conclusion and other remarks** | | General conclusion and other remarks on method strengths and weaknesses. |

## 4.2 <u>Criteria for Assessment/Comparison Summary of Each Method</u>

### 4.2.1 Overall Predictive Power

The overall predictive power was assessed based on the evaluation of qualitative predictive power and quantitative predictive power, which are discussed in Sections 4.3.2 and 4.3.3.

### 4.2.2 Assessment of Qualitative Predictive Power

The qualitative predictive power was assessed in terms of the following three aspects:

(1) How well did the team's application of the method predict the specific performance issues and drivers identified in the empirical data?

(2) Did the method predict factors and issues that were not supported by the empirical data and in terms of operational expressions?

(3) How well did the method predict failure mechanisms in operational terms that were identified in the empirical data?

These aspects are discussed in the next two sub-sections.

*4.2.2.1 Assessment Criteria in Terms of Drivers*

The HRA qualitative predictive power in terms of drivers was assessed against the empirical data based on the following two criteria.

(1) *Prediction of the drivers and associated performance issues identified in the empirical data.* Given that the crews are well trained, it was expected that the HEP would be low. This implies that human fail to meet the success criteria (i.e., human error) may not be observed or the data-based HEP from the experiement is not practical because the limited number of crew in this study. In this situation the quality of an HRA method is evaluated by whether the method was able to predict performance drivers and performance issues identified in the empirical data. In other words, in addition to correctly identifying a driving factor, a method should also be able to explain why the predicted driver contributes negatively to crew performance. Some methods may identify a correct driver without identifying specific performance issues or with a wrong rationale, which raises concerns on the validity of the methods. Hence, such methods would be rated lower with respect to this criterion than those that did correctly identify the performance issues. Given the differences in factor definitions among the methods, the emphasis on the drivers in operational terms and in terms of specific issues bypasses possible ambiguities with the assignment of issues to specific PSFs (the "translation" problem).

(2) *Predicted factors and issues that were not supported by the empirical data*. In contrast to the preceding criterion, this one focuses on the factors and issues predicted by the HRA teams. Did the HRA method predict drivers and performance issues that were not observed in the simulator or shown not to be a performance issue for the crews? The assessors took into account the fact that crew performance tends to be fairly good (i.e., low HEPs) and that there may be issues that are correctly predicted but simply not

observed given the sample size. In other words, the assessors did not necessarily penalize the analysts for factors that were not observed in the empirical data.

#### *4.2.2.2 Assessment Criterion in Terms of Operational Expressions*

*Prediction of failure mechanisms in operational terms*. Although HRA analysts need to understand how crews will approach a given task in order to predict the HEP, some methods rely strongly on these operational aspects and many methods predict specific modes or mechanisms of failure. This criterion deals with the accuracy of these predictions. Did the HRA analysis characterize correctly how the crews would fail or where they would have problems? It can be seen that the "driving factors and issues" criteria above focus on the problematic performance conditions while this criterion focuses on how degraded or failed performance manifested itself.

### 4.2.3 Assessment of Quantitative Predictive Power (including ranking)

The comparison of the method's quantitative predictions with the empirical data addressed both the absolute values of the HEPs predicted by each HRA team and the ranking of the HFEs based on magnitude of the predicted HEPs (across the HFEs). However, the small sample of observations results in large uncertainties in the reference HEPs derived from the failure rates in the empirical data so that the accuracy of the predicted HEPs is difficult to assess. In addition, in many PRA applications, the relative values of the HEPs (i.e., the ranking of the HFEs) are sufficient to draw conclusions and derive safety insights. Thus, the subcriteria in the bullet list below were used to assess quantitative predictive power and are listed with the highest priority first and in order of decreasing priority.

- HFEs where several failures were observed in the empirical data can be regarded as very difficult tasks that should have correspondingly high HEPs. If an HRA method produced low HEPs for such HFEs, the submission was examined in more detail in order to find the reasons for the discrepancies and any indications of systematic method optimism.

- Consistency of the ranking of the HFEs based on the predicted HEPs with the difficulty rankings given by the Unit Supervisors during collection of the empirical data (i.e., were the predicted HEPs and the difficulty ranking in the same order) (see discussion in Section 3.4.4). Despite the large uncertainties in the reference HEPs (in terms of what is the true error rate), it was possible to obtain a strong consensus on which HFEs appeared to be more difficult, with the expectation that the probability of failure was higher.

- Predicted HEPs relative to the confidence/uncertainty bounds of the empirical data. Were the HEPs within the bounds, which in this study were estimated by a Bayesian update that used the observed performances as evidence (see Chapter 6). The uncertainty bounds predicted by the HRA teams for each HEP were not used in the current comparison.

- Quantitative differentiation of the HFEs by HEP. Were the predicted HEPs for the most difficult HFEs significantly larger than those predicted for the least difficult HFEs? The quantitative predictive power of the method is judged to be reduced if the predicted HEPs all fall within a narrow band.

### 4.2.4 Assessment of Traceability

The assessment of traceability examined

- The basis for the quantification inputs obtained by the HRA teams in the application of an HRA method. For instance, the assessment examines how the ratings of PSFs were derived from the qualitative analysis or the identification of the failure mechanisms associated with operational narratives. In both cases, the assessment looks at how the HRA method and the documentation of the application of the method (of the HRA results) establish the link between the qualitative analysis and the quantification inputs (the PSF ratings). How did the issues and factors identified as relevant and important to HFE failure translate into PSF ratings or identified failure mechanisms?

- The quantification. This part of the assessment of traceability looks at the link between the quantification inputs and the HEP values. Is expert judgment involved in deriving the HEPs from the quantification inputs? If so, how large is the role of expert judgment? Alternatively, is the quantification based on a mathematical, fully repeatable algorithm?

### 4.2.5 Assessment of Adequacy of Method Guidance

The assessment of method guidance examines:

- The guidance for the qualitative analysis. Some of the relevant questions are: To what extent does the method give guidance for doing the qualitative analysis, and how does this guidance contribute to a comprehensive assessment of the performance shaping factors or contextual factors in terms of how they may affect the probability of HFE failure? Does the method guidance clearly describe the required or expected scope of the qualitative analysis? To what extent does the guidance for the qualitative analysis appear to support inter-analyst consistency?

- The guidance for HFE modeling and decomposition (if applicable).

- The guidance for quantification. For those methods where PSF ratings are used to translate the qualitative analysis into quantification, what guidance is available to support the rating of the factors? For those methods where quantification includes expert judgment, what guidance or aids are available to support the expert judgment process and its consistency?

### 4.2.6 Insights for Error Reduction

This assessment addresses the degree to which the qualitative analysis and evaluation of performance influences addressed by the HRA method produce information that would allow insights for how to reduce error. In other words, does the analysis of driving factors and understanding of potential failure mechanisms support the identification of potential fixes in areas where errors might occur (e.g., procedural or training improvements). The overall ability of the method to produce this information was judged.

## 4.3 Intra-Method Comparison Methodology

Intra-method comparisons were conducted collaboratively by the HRA analysts and assessors to evaluate the consistency of same or similar methods in qualitative and quantitative predictions when different analysts were applying them. In addition to comparing the HRA teams' predictions, the study also aimed to determine the causes/sources for observed discrepancies in predictions made by teams using the same method. Among other considerations, this intra-method analysis asked (1) whether the apparent discrepancies in predictions from teams using the same method can be explained by the teams' assumptions, (2) whether the discrepancies resulted from the limitations of a particular method or a function of the HRA teams (i.e., how the analysts interpret and apply a given method), and (3) whether different teams sought and obtained different information. It was also important to see whether the observed discrepancies were consistent across HFEs and investigate why these occurred.

For each method, intra-method comparisons were first done for each HFE using the information in Appendix G on the comparisons of the HRA predictions against the empirical data. Qualitative analyses and quantitative analyses were compared between the HRA analyst teams using the same method and against the empirical data. Then, the HFE-by-HFE intra-method comparison results were synthesized and overall conclusions concerning intra-method variability were summarized in Chapter 7 for each method, with the factors leading to the variability summarized in Chapter 8. When the causes/sources for the discrepancies between the same methods could not be identified solely based on the information in Appendices G and H, the HRA teams' report (e.g., Form A), HRA analysis documentation, diaries, and interviews with the plant trainers were examined. The intra-method comparison process is briefly described below.

### 4.3.1 Comparing Qualitative Analyses

The intra-method comparisons focused on the following two aspects in comparing the qualitative analyses that were made by each HRA team that used the same method for each HFE.

(1) Discrepancies in operational stories. Did all HRA teams correctly characterize the performance issues or failure mechanisms in operational terms that were identified in the empirical data? Were there any discrepancies across the teams? Were the teams' predictions on procedural paths, scenario evolutions, etc. consistent with each other? What are the causes for the discrepancies or inconsistencies?

(2) Discrepancies in driving factors. Did all teams predict the driving factors identified in the empirical data with the correct rationale? Were there any discrepancies across the teams? What are the causes for the discrepancies?

As mentioned above, consistent discrepancies across HFEs warrant a deep investigation as to why they were occurring.

### 4.3.2 Comparing Quantitative Analyses

For quantitative analysis, intra-method comparisons looked at (1) the discrepancies between the teams' total HEPs (diagnosis HEP plus execution HEP) and (2) the discrepancies in diagnosis and execution HEPs separately when applicable.

1. Total HEPs.
    - Absolute value. How did the total HEPs of the different HRA teams compare to the crew failure rates and the uncertainty bounds of the empirical data? Were the total HEPs of the different HRA teams comparable? Were the total HEPs of one HRA team consistently more conservative (higher) or optimistic (lower) than those of the other team(s) across HFEs? What caused the discrepancies? How are the discrepancies connected to the qualitative analyses?

    - Relative value. Did the total HEPs of all HRA teams using the same method reflect the HFE difficulty rating? Was there reasonable differentiation in the teams' total HEPs? What are the causes for the discrepancies? How are the discrepancies connected to the qualitative analyses?

2. Diagnosis and execution HEPs (when applicable).
    - Absolute value. Were the diagnosis or execution HEPs of different HRA teams using the same method comparable? Were the diagnosis or execution HEPs of one HRA team consistently more conservative or optimistic than those of other team(s) across HFEs?

    - Relative value. Were the contributions of diagnosis and execution HEPs to the total HEPs consistent with the conclusion in the team's qualitative analysis about whether the total HEPs were diagnosis-HEP dominated or execution-HEP dominated? Did the differences in diagnosis or execution HEPs show the same pattern as that of the differences in the total HEPs?

#### 4.3.3 Comparison summary

In the intra-method comparison summary (Chapter 7), the findings obtained from HFE-by-HFE comparisons were synthesized, and overall conclusions were made concerning the sources of intra-method variability in terms of method effects, analyst effects, and the interactions between them. The identified sources are discussed in detail in Chapter 8.

# 5. DESCRIPTION OF SCENARIOS, HFES, AND EMPIRICAL SIMULATOR RESULTS[1]

In this chapter, the three scenarios used in the study are described along with associated human failure events (HFEs), critical actions, and relevant procedures. Following the description of each scenario, a summary of the simulator results for each HFE (see Appendices D, E and F for detailed crew performance) is also presented with an analysis of performance drivers and performance shaping factors (PSFs).

## 5.1 Scenario 1—Total Loss of Feedwater (LOFW) Followed by Steam Generator Tube Rupture (SGTR)

### 5.1.1 Scenario Description and Human Failure Events

*5.1.1.1 Plant Technical Information*

- There are three main feedwater pumps: 11, 12 and 13.
- There are four auxiliary feedwater (AFW) pumps: 11, 12, 13 and 14. AFW pump 14 is turbine-driven and the other three are motor-driven.

Relevant Procedures:

- 0POP04-FW-0002 “Steam Generator Feed Pump Trip”
- 0POP05-EO-EO00 (E-0) “Reactor Trip Or Safety Injection”
- 0POP05-EO-ES01 (ES-01)“Reactor Trip Response”
- 0POP05-EO-F003 “Heat Sink Critical Safety Function Status Tree”
- 0POP05-EO-FRH1 (FR-H1) “Response to Loss of Secondary Heat Sink”
- 0POP05-EO-ES11 (ES-11) “SI Termination”
- 0POP05-EO-EO10 (E-10) “Loss of Reactor or Secondary Coolant”
- 0POP05-EO-F004 “Integrity Critical Safety Function Status Tree”
- 0POP05-EO-FRP1 (FR-P1) “Response to Imminent Pressurized Thermal Shock Condition”
- 0POP05-EO-EO30 (E-30) “Steam Generator Tube Rupture”

[1]The information in Chapter 5 and Appendices D to F is obtained from the Halden Reactor Project. See Halden report titled “Crew response to safety relevant tasks in scenarios of varying complexity – A study of performance and performance drivers in a US training simulator”, prepared by Helena Broberg, Michael Hildebrandt and Rodney Nowell in February, 2011. In addition, Bruce P. Hallbert and Tommy Morgan of Idaho National Laboratory (INL) made significant contributions to the collection and analysis of the crew data. Additional discussion of the crew results for a parallel study can be found in NUREG/CR-7163 (A Formalized approach for the collection of HRA data from nuclear power plant simulators).

- 0POP05-EO-EC31 (EC-31) “SGTR with loss of reactor coolant – subcooled recovery desired”

*5.1.1.2 Total Loss of Feedwater (LOFW)*

At the beginning of the scenario the plant was operating at 100% power. The shift technical advisor (STA) was not in the control room but could be called, and would then arrive within 5 minutes after the call.

Two minutes into the scenario, all main feedwater pumps are tripped (loss of main feedwater pump 11, and subsequent trip of feedwater pumps 12 and 13 within the next 10 seconds), and the startup feedpump cannot be started. If the crew fails to manually trip the reactor, it will automatically trip on low SG level (i.e., 20% in narrow range (20% NR)) within 50-60 seconds.

At autostart, AFW pump 14 will overspeed and cause damage that cannot be repaired. AFW pump 11 will have a seized shaft and trip and will not be available. AFW pump 13 will start but the shaft will shear and no flow will be indicated.

AFW pump 12 will start automatically and indicate full flow, but this flow will not reach (feed) the steam generator (SG B) because a recirculation valve is mispositioned (it is open). There is no indication of the valve’s position in the control room, thus the open recirculation valve will mask the fact that no AFW at all is going into the SGs, and the SG levels will go down. In reality, criteria to start procedure FR-H1 are met. But, because of the indicated flow from AFW pump 12, the plant computer will not show a red path on the heatsink status tree. The crew will have to realize that the indication of AFW flow from pump 12 is false and then decide to go to FR-H1.

According to procedure FR-H1, feed and bleed (F&B) shall be established when the wide range (WR) level on any two SGs are less than 50%. (In this case F&B refers to *primary F&B,* with feed from safety injection and bleed through the pressurizer power-operated relief valves (PZR PORVs)*, and not F&B through the SGs.*)

To establish AFW flow to the SGs, the crews can (in principle):

- dispatch a plant operator (PO) to check and close the open recirculation valve (feed SG B). In the simulation, one of the instructors acted as a plant operator (and other plant personnel that are not in the control room) and could be reached on the phone; or
- cross-connect AFW flow from pump 12 to SG A, C or D. The cross-connection can be done from the main control room.

However, if the crew sends a PO to the recirculation valve before start of F&B, the PO will delay closing the valve until F&B is established. Similarly, if the crew tries cross-connecting before F&B, the breaker for the power to the valve would open (as part of the study) and the valve would remain closed.

After F&B, the valve breaker would be reclosed by a PO. If the crew were to cross-connect after F&B, the valve would open.

*5.1.1.3 HFE 1A and HFE 1B*

HFE 1A: Failure to establish F&B within 45 minutes of the reactor trip, if the crew initiates a manual reactor trip before an automatic reactor trip.

HFE 1B: Failure to establish F&B within 13 minutes of the reactor trip, if the crew does not manually trip the reactor before an automatic reactor trip occurs.

The actions to start F&B include:

- Actuate safety injection (SI)
- Open both of the PZR PORVs

*5.1.1.4 Steam Generator Tube Rupture (SGTR)*

After F&B has been established, the crew will be able to establish AFW flow to one or several SGs by either closing the recirculation valve and/or cross-connecting the flow from the running AFW pump to the other SGs.

As soon as the crew has established AFW flow, the trainers will initiate a tube rupture in the first SG that is fed. The crew will want to fill an SG to be able to exit FR-H1, and the tube rupture may be masked by AFW flow to the SG, as long as it is being fed. The leak size of the ruptured tube is about 500 gallon per minute (gpm) at 100% power, but the flow will depend on the differential pressure between the reactor coolant system (RCS) and the ruptured SG. There is initially no secondary radiation because there is only a minimum steam flow. The blowdown (BD) and sampling is secured because of the SI.

By the time the crews fill the SG(s) enough to exit FR-H1, they may have problems with the RCS integrity status tree and be forced to enter procedure FR-P1, which will delay the possibility of transitioning to the SGTR procedure E-30.

*5.1.1.5 HFE 1C*

HFE 1C: Failure of crew to isolate the ruptured steam generator and control pressure below the SG PORV setpoint to avoid SG PORV opening. *The time window to perform the required actions is estimated to be approximately 40 minutes.*

The actions include:

- Isolate the ruptured SG (close feedwater and main steam isolation valves)
- Maintain SG pressure below the setpoint by cooling down the RCS (cooling the secondary system by dumping steam and depressurizing the RCS).

### 5.1.2 Performance of HFEs 1A and 1C

**HFE 1A:**
<u>Success Criterion</u>: Establish F&B within 45 minutes after reactor trip.
<u>Results</u>: Zero of four crews failed.

Based on training, the crew is expected to manually trip the reactor before its automatic trip in the loss of main feedwater (MFW) scenarios. All crews manually tripped the reactor within 1 minute after loss of MFW, and established F&B well within 45 minutes of the reactor trip (Table 5-1).

**Table 5-1 Crew Performance on HFE 1A**

| Crew | Time to reactor trip after LOFW [Sec.] | Time to F&B initiation after reactor trip [Min.] | SG level (WR) at F&B [%] |
|---|---|---|---|
| Q | 37 | 17:14 | 41 |
| R | 29 | 15:45 | 46 |
| S | 37 | 12:04 | 45 |
| T | 51 | 10:03 | 37 |

**HFE 1B:**

Success Criterion: If the reactor trips automatically, establish F&B within 13 minutes of reactor trip

Results: no data since no crew had an automatic reactor trip.

**HFE 1C:**

Success Criterion: Isolate the ruptured SG and control pressure below the SG PORV setpoint within 40 minutes after SG rupture.

Results: Three of four crews failed.

Three crews (Q, S & T) isolated the ruptured SG and controlled pressure below the SG PORV setpoint to prevent the SG PORV from opening (Table 5-2). Thus, the three crews succeeded from a plant perspective, but only Crew S accomplished the desired actions within less than 40 minutes. Crew S isolated the ruptured SG approximately 27 minutes after the SGTR and secured the final HH safety injection pump approximately 40 minutes after the SGTR, therefore the RCS pressure was controlled and SG PORVs did not open. Crews Q and T isolated the ruptured SG and controlled RCS pressure (secured the HH pump) to prevent the PORV from opening, but they did not complete the actions within 40 minutes of the SGTR. The other crew (R) isolated the ruptured SG according to Procedure E-30 but failed to reduce the RCS pressure because they failed to secure the final HH pump, as a result the SG PORV did open approximately 40 minutes after the SGTR.

**Table 5-2 Crew Performance on HFE 1C**

| Crew | Reach 100% SG level (WR)? (Yes/No) | SG PORVs open? (Yes/No) | Isolate ruptured SG and control pressure within 40 minutes after SGTR? (Yes/No) |
|---|---|---|---|
| Q | Yes | No | No |
| R | Yes | Yes | No |
| S | No | No | Yes |
| T | No | No | No |

It should be noted that the HRA criterion specifies that the crews must "Maintain SG pressure below the setpoint by cooling down the RCS (cooling the secondary by dumping steam and depressurizing the RCS)." The crews were required to stop the HH safety injection pumps to meet this criterion.

See Appendix D for details of crew performance.

### 5.1.3 Scenario Development and Crew Responses in Scenario 1

In Scenario 1, the STA was not present in the control room at the beginning of the scenario. The crews were made up of a shift manager (SM), a unit supervisor (US) and two reactor operators (ROs), except for crew Q who had no shift manager (SM) present but three ROs. The crews' performance is summarized in the following subsections, and the timing of main events and crew actions in Scenario 1 is listed in Table5-3.

**Table 5-3 Timing of Main Events and Actions for All Crews in Scenario 1**

| Action/Event | Crew Q | Crew R | Crew S | Crew T | Source* |
|---|---|---|---|---|---|
| Start of scenario | 0:00:00 | 0:00:00 | 0:00:00 | 0:00:00 | SAML |
| LOFW (FP 2 trip) | 0:02:01 | 0:02:07 | 0:02:14 | 0:02:09 | SAML |
| Reactor trip | 0:02:38 | 0:02:36 | 0:02:51 | 0:03:00 | SAML |
| Trip, time after LOFW | 0:00:37 | 0:00:29 | 0:00:37 | 0:00:51 | SAML |
| Start E-0 (immediate actions) | 0:02:45 | 0:02:35 | 0:03:00 | 0:03:12 | OBS |
| Start procedure ES-01 | 0:06:40 | 0:08:25 | 0:08:30 | 0:06:10 | OBS |
| STA arrives in control room | 0:08:40 | 0:08:53 | 0:11:32 | 0:09:02 | OBS |
| Start FR-H1 procedure | 0:14:50 | 0:15:18 | 0:12:11 | 0:10:34 | OBS |
| Actuate SI | 0:18:36 | 0:16:52 | 0:13:37 | 0:12:11 | SAML |
| Open second PZR PORV (start F&B) | 0:19:52 | 0:18:21 | 0:14:55 | 0:13:03 | SAML |
| WR level SGA at start F&B | 41 | 46 | 45 | 37 | PPL |
| WR level SGB at start F&B | 45 | 48 | 47 | 41 | PPL |
| WR level SGC at start F&B | 41 | 46 | 45 | 37 | PPL |
| WR level SGD at start F&B | 44 | 48 | 47 | 40 | PPL |
| Close recirculation valve | 0:21:28 | 0:19:21 | 0:21:04 | No | SAML |
| Time for tube rupture | 0:22:27 | 0:20:28 | 0:23:37 | 0:28:01 | SAML |
| Tube rupture in SG no | SG B | SG B | SG B | SG C | SAML |
| WR level SG A at SGTR | 29 | 39 | 41 | 23 | PPL |
| WR level SG B at SGTR | 36 | 42 | 50 | 26 | PPL |
| WR level SG C at SGTR | 29 | 39 | 41 | 26 | PPL |
| WR level SG D at SGTR | 31 | 39 | 44 | 25 | PPL |
| Cross-connect AFW | 1:03:05 | 0:35:02 | 0:37:11 | 0:25:44 | SAML |
| Stop AFW to ruptured SG | 0:38:02 | 0:37:20 | 0:34:30 | 0:51:22 | PPL |
| Stop AFW to ruptured SG (time after SGTR occur) | 0:15:35 | 0:16:52 | 0:10:53 | 0:23:21 | PPL |
| WR level in ruptured SG when AFW stopped | 71 | 75 | 70 | 70 | PPL |

| Action/Event | Crew Q | Crew R | Crew S | Crew T | Source* |
|---|---|---|---|---|---|
| NR level in ruptured SG when AFW stopped | 65 | 87 | 67 | 44 | PPL |
| Adjust SG PORV setpoint | No | 0:50:41 | No | 2:00:47 | SAML |
| Isolate ruptured SG | 0.33:36 | 0:53:29 | 0:48:09 | 2:02:33 | SAML |
| Isolate ruptured SG (time after SGTR) | 0:11:09 | 0:33:01 | 0:24:32 | 1:34:32 | SAML |
| Stop last HHSI pump | 1:45:41 | No | 1:00:42 | 1:02:45 | SAML |
| Stop last HHSI pump (time after SGTR) | 1:23:14 | No | 0:37:05 | 0:34:44 | SAML |
| Exit FRH1 procedure | 1:50:05 | 0:43:40 | 1:03:45 | 1:10:50 | OBS |
| Start FRP1 procedure | 1:51:25 | No | 0:46:15 and 1:04:02 | 1:11:10 | OBS |
| Exit FRP1 procedure | No | No | No | 1:58:31 | OBS |
| Start E-30 procedure | No | 0:49:17 | No | 1:59:26 | OBS |
| SG PORV open | No | 1:03:16 | No | No | PPL |
| SG PORV open (time after SGTR) | N/A | 0:42:48 | N/A | N/A | SAML |

* SAML = Simulator Action Monitor Log, PPL = Process Parameters Log, OBS = Observer's notes, FILM = video camera of screens

*5.1.3.1 Reactor Trip*

All four crews manually tripped the reactor. The fastest crew (R) did it 29 seconds after the total LOFW, and the slowest crew (T) took 51 seconds. An automatic trip is expected within about 50-60 seconds. All crews improved the situation by tripping the reactor manually instead of waiting for an automatic trip. Figure 5-1 illustrates how the timing of a reactor trip determines how low and how fast the SG levels drop. The SG levels did not drop as far and as fast as they would have, had the crews waited for the automatic trip. As a result, the crews had more time to restore feedwater and establish F&B.

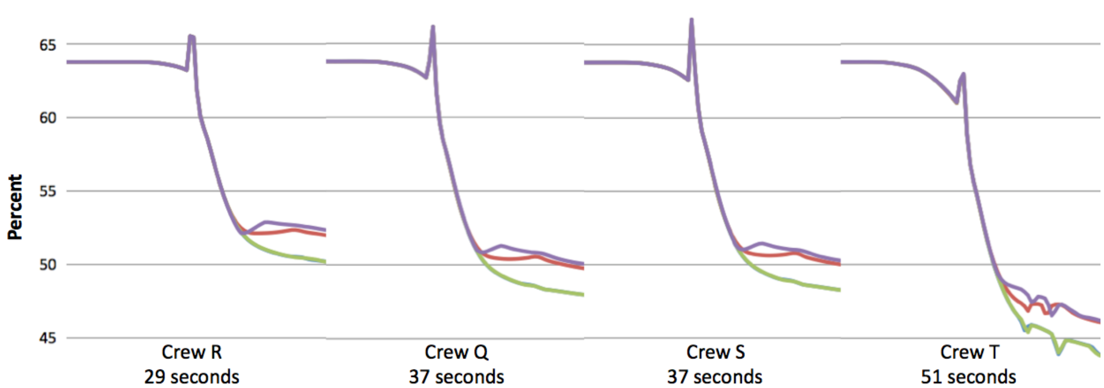



**Figure 5-1 Wide range SG level reactions at reactor trip (from the fastest trip on the left to the slowest trip on the right)**

*5.1.3.2 Identification of Total LOFW and Start of F&B*

Because the crews all tripped the reactor manually, they needed to initiate F&B within 45 minutes after the reactor trip according to the HFE 1A success criterion (see

Section 5.1.3.1). As shown in Table 5-1 Crew Performance on HFE 1ATable 5-1, all four crews met the time limit and started F&B between 10-17 minutes after the reactor trip.

After loss of MFW, operators are expected to check AFW status. The difficulty in the scenario was to realize that all AFW was lost because the recirculation valve on AFW pump 12 (feeding SG B) was mispositioned open. Therefore, even though there is indication of AFW flow there is no water flow into the SGs.

Combination of low water levels in all SGs and the loss of AFW is an entry criterion to procedure FR-H1, which calls for starting F&B. The STA was not in the control room at the beginning of the scenario but all crews called for their STAs, who arrived 8-11 minutes into the scenario. The plant had an operating experience of AFW recirculation valves being mispositioned so all crews had been trained on similar situations before the experiment. This experience might have helped the crews' responses in detecting loss of all AFW. All crews detected that the level in SG B continued to decrease even though AFW flow was indicated, and they questioned and wanted to verify if they actually had AFW to SG B. Three of the four crews (Q, R and S) suspected that the recirculation valve was open and thus called a PO to check the valve. These crews decided to start FR-H1 between 11-15 minutes into the scenario (or 8-13 minutes after the reactor trip) after the STAs verified that they had met the entry criteria to the procedure. Crew T did not discover the open recirculation valve. After having started F&B they decided to cross-connect the AFW 12 flow to feed SG C.

When the crews (i.e., Q, R, and S) started FR-H1, the WR levels in the SGs were already below 50%, which met the criterion to start F&B. The action to start F&B is in FR-H1 Step 2 and on the Conditional Information Page (CIP). Three crews (i.e., R, S, and T) initiated F&B from the CIP and one crew (Q) from Step 2. One crew (S) missed a procedure step to stop RCPs before starting F&B, but the STA discovered the error within a minute and then they tripped the RCPs.

#### *5.1.3.3 Reestablishment of AFW and Start of SGTR*

As mentioned above, three crews (Q, R and S) sent a PO to check the recirculation valve on the AFW pump for SG B. The PO closed the valve after F&B had been started, making the AFW flow go into SG B. Thus, the three crews were given a tube rupture in SG B. Crew T did not discover the open recirculation valve. After having started F&B they were able to feed SG C by cross-connecting the AFW flow. So they were given a tube rupture in SG C.

#### *5.1.3.4 Identification and Isolation of SGTR*

The crews were filling the ruptured SG in order to stop F&B in procedure FR-H1 (the goal is 14% narrow range (NR) under normal conditions and 34 percent NR at adverse containment). The tube rupture was masked as long as the AFW water was injected into the SG. With absence of main steamline radiation alarms, the rising SG level was likely to be explained by the crews as resulting from injection of AFW water.

The crews stopped the AFW pump when the NR level in the SG was 44 percent (Crew T), 65 percent (Crew Q), 67 percent (Crew S) and 87 percent (Crew R). They stopped AFW on the basis of having a high enough level to stop F&B and because they wanted to cross-connect to start feeding the other SGs. At this point, they were not yet aware of the tube rupture. But after having stopped AFW the crews quickly noticed that the level continued to rise.

All crews suspected a tube rupture because of the rising SG level and then backed up their diagnoses by checking steamline radiation, which was slightly elevated but not high enough to trigger an alarm. Three of the four crews (Q, S, and T) identified the tube rupture while they were working in procedure FR-H1. The other crew (R) had just left FR-H1 and held a crew brief that identified the tube rupture.

Crews Q, S, and T discovered that they had conditions that met a red criterion on FR-P1 (Response to Imminent Pressurized Thermal Shock Condition) at about the time when they suspected a tube rupture (while still working in FR-H1). Crew S decided to put FR-H1 on hold to perform actions in FR-P1. They then went back to FR-H1, but returned again to FR-P1 after having finished FR-H1. Crews Q and T continued in FR-H1 and transferred to FR-P1 after having exited from FR-H1. These three crews (i.e., S, Q, and T) were unable to go immediately to the tube rupture procedure E-30 as they were tied up in the FR-procedures. Crew R had exited FR-H1 and was directed to E-10 when they identified the tube rupture. Then, they went to procedure E-30 from the E-10 CIP.

Isolating a ruptured SG includes isolating AFW, closing a number of valves (e.g., MSIVs, SG dump valves, and chemical sampling valves), and adjusting the SG PORV setpoint. All crews had already isolated the steamline earlier in the scenario. Crews Q and S secured AFW to the ruptured SG but did not adjust the SG PORV setpoint. Because RCS pressure never reached the setpoint, this omission did not have any negative consequence. In addition, two crews (i.e., Q and S) did not start using procedure E-30 because they did not finish procedure FR-P1 during the scenario even though they knew there was an SGTR. Crew S isolated the ruptured SG within 40 minutes after the SGTR. Crew T exited procedure FR-P1 and entered were directed ES-11 almost 2 hours into the scenario. They then started E-30 from the CIP in ES-11 and performed all of the isolation steps, including securing AFW and adjusting the SG PORV setpoint, but this was completed after the 40 minutes time frame. Crew R, who did not meet the FR-P1 entry criterion, became aware of the tube rupture during the brief after they had left FR-H1 and entered E-30 to secure AFW and adjust SG PORV setpoint within the 40 minutes time limit. In summary, all four crews isolated the ruptured SG. Two of them (Q and S) based the decision to isolate the ruptured SG from knowledge [2] and the other two (R and T) from procedure guidance.

Crews Q, S, and T had stopped all high head safety injection (HHSI) pumps before leaving procedure FR-H1. They then went to procedure FR-P1 that tells them not to raise the RCS pressure. Although Crew Q top filled the ruptured SG, the SG PORV was never opened because the RCS pressure was maintained low. These three crews succeeded in preventing SG pressure to exceed the SG PORV setpoint. Among these three crews, crews S and T secured HHSI within 40 minutes after the SGTR. Crew Q secured all HHSI pumps about 83 minutes after the SGTR.

Crew R did not stop the last HHSI pump before leaving procedure FR-H1, and consequently had problems controlling RCS pressure. They did not stop the HHSI pump because they incorrectly assumed they had an "active loop" and followed the procedure accordingly (see appendix D for a more detailed description). (Active loop means that the SG must have a water level greater than 14 percent NR and a natural circulation has been established). Crew R

---

[2] Note that this expression "from knowledge" does not mean that they do not have support in their procedures for doing this. As stated in "procedural guidance" in the table in section 5.1.4.2, they do have support from ZA-18 (EOP users guide) for taking this action while they are inside another procedure (FR-H1).

performed the isolation steps in E-30 but the high pressure still caused the SG PORV to open. The ruptured SG was at 100 percent WR and NR level.

When crew R checked the criteria to stop SI in E-30, RCS pressure was going down because of the open SG PORV. Even though the crew had detected steam flow from the ruptured SG they did not connect it with the decreasing RCS pressure. Instead, they followed the procedure verbatim and, based on an instantaneous reading of RCS pressure trend, transferred to EC-31 and did not stop SI. The RCS pressure remained high and they continued to release through the SG PORV for about another 10 minutes.

### 5.1.4 PSFs for Scenario 1 (HFEs 1A and 1C)

*5.1.4.1 PSFs for HFE 1A "Establishing F&B within 45 minutes"*

Table 5-4 lists the PSFs identified by the experimental team for HFE 1A.

**Table 5-4 Performance Shaping Factors for HFE 1A**

| PSF | Rating | Comment |
|---|---|---|
| **Stress** | 0 | |
| **Adequacy of time** | N/P | The crews tripped the reactor manually and had approximately 45 minutes to start F&B. |
| **Team dynamics** | 0 | |
| **Work processes** | 0 | One crew (S) missed stopping the RCPs in the procedure, but stopped it within a minute of noticing they had skipped it. |
| **Communication** | 0 | |
| **Scenario complexity** | ND | The loss of all AFW was masked, but all crews detected it. The crews used some time to evaluate the situation. If the recirculation valve had not been open, they would have started FR-H1 and F&B earlier. |
| **Indications of conditions** | ND | The loss of all AFW was masked by an open recirculation valve. Entry into FR-H1 was not indicated in the plant computer. |
| **Human-Machine Interface (HMI)** | N/P | |
| **Training and experience** | N/P | Crews were well trained on an open recirculation valve and have experienced it in the plant. All crews train on LOFW at least every 2 years. |
| **Procedural guidance** | ND | The Critical Safety Functions status trees do not address misaligned valves. |
| **Execution complexity** | N/P | |
| N/P = Nominal/Positive, 0 = Not a driver, ND = Negative driver, MND = Main negative driver | | |

*5.1.4.2 PSFs for HFE 1C "Isolate the ruptured steam generator and control pressure below the SG PORV setpoint within 40 minutes after SGTR"*

Table 5-5 lists the performance shaping factors (PSFs) identified for HFE 1C.

**Table 5-5 Performance Shaping Factors for HFE 1C**

| PSF | Rating | Comment |
|---|---|---|
| **Stress** | 0 | |
| **Adequacy of time** | ND | Since the SGTR is masked by AFW for a while after the rupture, the crews had fewer than 40 minutes (the HRA criterion) to diagnose the SGTR, work through the procedures, and accomplish the required actions. Given that even the fastest crew barely made the time criterion, it was decided that adequacy of time was a negative driver for this HFE. |
| **Team dynamics** | 0 | The STA in the crew (R) that did not prevent the PORV from opening could have provided more information about the loop not being active and about the reason for the low RCS pressure during the transfer from E-30 to EC-31. |
| **Work processes** | 0 | The crew (R) that did not prevent the PORV from opening incorrectly exited FRH1 without stopping the last HHSI pump that caused the pressure to be high and the SG PORV to open. The other three crews were able to follow procedures correctly, but the time assumed available for the HFE was inadequate. It was not clear that work processes were an issue for any of the crews. |
| **Communication** | 0 | |
| **Scenario complexity** | MND | The tube rupture was initially masked by the AFW flow to the ruptured SG and the low steam flow prevented radiation alarms.<br>When the tube rupture started, the crews already had an emergency situation and were working in FRH1. |
| **Indications of conditions** | ND | The tube rupture was initially masked by the AFW flow to the ruptured SG and the low steam flow prevented radiation alarms. |
| **HMI** | N/P | |
| **Training and experience** | N/P | The crews train on various tube ruptures several times every year. They have trained on both at-power trip SGTR and post-trip SGTR where they have delayed secondary radiation. The crews had not been trained on this particular scenario. |
| **Procedural guidance** | ND | Three crews could not go to E-30 because they were held up in the higher priority procedures FR-H1 and FR-P1. Two crews (Q and S) isolated the ruptured SG from knowledge, having support from ZA-18 (EOP users guide). The other two crews were procedurally driven to go to E-30. One of them (R) incorrectly exited FRH1 without stopping the last HHSI pump. |
| **Execution complexity** | N/P | |
| N/P = Nominal/Positive, 0 = Not a driver, ND = Negative driver, MND = Main negative driver | | |

### 5.1.5 Drivers of Performance

All crews manually tripped the reactor because if they had not, the reactor would have tripped automatically. They were also helped by training and by understanding that they would need as much inventory as possible.

The crews identified the loss of all AFW because they did not obtain the expected response (level in SG B lowering even though flow was indicated). The ROs detected the unexpected symptom and communicated the deviances to the US. Previous training also helped the crews; they have had operating experience where the AFW flow was indicated but not fed into the SG because of open AFW recirculation valves.  When the STAs arrived they helped to validate the loss of heat sink.

Starting F&B is driven by procedure FR-H1.  The guidance in the procedure steps and on the CIP helped the crews to start F&B.  The crews were trained to start F&B with input of SG WR levels.

One crew (S) failed to stop the RCPs before they started F&B, which was an error in procedure execution.  They knew that they needed to go to F&B and the US did not read the CIP criterion aloud.  Although they should not have experienced stress because they knew where they had to go, they suffered from self-imposed urgency.  The mistake was quickly discovered by the STA when reviewing the CIP notes.

The tube rupture was masked by concurrent AFW flow injecting into the broken SG.  Furthermore, early main steamline isolation delayed radiation indications.  When the crews stopped AFW to the ruptured SG, the level continued to rise, which made the crews suspect a tube rupture.  Board awareness and continuous monitoring of the SG level and radiation in steamlines helped the crews with the identification.  Moreover, the crews' training on recognizing what is normal swell helped them diagnose that the rising SG level was by SGTR rather than swelling.  Teamwork, in the form of crew updates and briefs, also helped the crews' diagnoses.  A trainer added that the diagnosis might have been made earlier if some crews had stopped AFW to the SG as soon as they could.

Two of the crews (Q and S) isolated the ruptured SG from knowledge since they could not leave the procedure they were in (Crews Q and S did not use E-30).  There is procedure support for responding to an SGTR in ZA18, and it was applied from knowledge/memory.  The main steamlines were already isolated and the crews decided to isolate AFW.  They did not change the SG PORV setpoint but were aware that the concern was to control RCS pressure.  Crew (T) was directed to E-30 from ES-11 and Crew R entered E-30 from E-10.

The crews thought that it was challenging to have a tube rupture when they could not leave the FR-procedures.  They (Crews Q, S, and T) could not go to E-30 soon enough even though they were aware of the SGTR, because their priority was to go through the FR-procedures.  Some crew members expressed the feeling that they were in a no-win situation and experiencing fatigue from a long and complicated scenario.  Nonetheless, even though the crews could not transfer to E-30 soon enough, the actions in FR-H1 to reduce SI helped them control the pressure and avoid SG PORV opening.

Crew R, which failed to prevent the SG PORV from opening, believed they had an active loop. (Active loop means that the SG has a water level of greater than 14 percent NR and an established natural circulation.)  Consequently, they did not follow the instructions in FR-H1 to stop all HHSI pumps.  Crew members stated that they believed they had a leak, but could not depressurize because they believed everything was solid.  When they got to the depressurization step in E-30 they could not perform it because of the high PZR level, and they had also entered EC-31 from E-30 before stopping SI (they exited E-30 because of the temporarily decreasing RCS pressure caused by the open SG PORV).

## 5.2 Scenario 2—Loss of CCW and RCP Sealwater

### 5.2.1 Scenario Description and Human Failure Event

*5.2.1.1 Plant Technical Information*

Component Cooling Water (CCW)

- CCW pump 1A, powered by E1A
- CCW pump 1B, powered by E1B
- CCW pump 1C, powered by E1C

RCP sealwater

- charging pump 1A, powered by E1C
- charging pump 1B, powered by E1A
- positive displacement pump (PDP), powered by 1G8-bus (remains energized), cooled by air (does not use CCW)

*5.2.1.2 Loss of CCW and RCP Sealwater*

In the beginning of the scenario the plant is operating at 100 percent power and all five crew members are in the control room. The CCW B train is out of service so that CCW pump B is unavailable.

Two minutes into the scenario, the distribution panel 1201 fails. As a consequence, the crew has to establish manual control of following controlling channels:

- A and B SGs
- PZR level control
- rod control
- nuclear instrumentation (NIS)
- PZR pressure control

For this scenario, it is of particular importance for the crew to establish manual control of feedwater flow to SGs A and B.

The failed distribution panel is unrelated to the loss of CCW and sealwater but increases the complexity of the scenario. It masks the status of CCW and sealwater by keeping the crew busy because of the number of alarms.

The feedwater regulation valve on SG A remains fully open and cannot be operated manually, feeding the SG. If the crew does not trip the reactor, there will be an automatic turbine trip on high SG level (87 percent), which would cause a reactor trip.

When the reactor trips, one AFW pump cannot start because of the loss of the distribution panel 1201. In addition, Bus E1C will have a bus lockout caused by a bus fault (the busbar is de-energized and the DG breaker cannot be closed), and the CCW pump 1A breaker will trip because of a failed and seized shaft. As a result, there are no CCW pumps in service (the pump B is out of service, pump A is tripped, and pump C is de-energized), and no charging pump is running (pump A is de-energized). If charging pump 1B is started, it will trip 2 minutes after the reactor trip.

According to procedure 0POP04-RC-0002 “Reactor Coolant Pump Off Normal,” any RCP that experiences a simultaneous loss of seal injection flow and loss of CCW flow to thermal barrier shall be stopped within 1 minute after determining that both RCS seal injection flow and thermal barrier cooling were lost. The risk of a seal failure increases after 1 minute.

Procedure 0POP04-RC-0002 “Reactor Coolant Pump Off Normal” and procedure ES-01 (reactor trip response) both have guidance to start the positive displacement pump (PDP) when CCW and seal injection are lost. However, the PDP can only be started if the RCP seal temperatures are below 230 degrees Fahrenheit (F), and reaching these procedure steps takes some time. In accordance with the Westinghouse RCP vendor manual if seal inlet or seal inlet bearing reach 230 degrees F, the potential for seal damage is too great to risk placing seal injection in service.

*5.2.1.3 HFE 2A*

HFE 2A: Failure of the crews to trip the RCPs and start the PDP to prevent RCP seal loss of coolant accident (LOCA).

Success requires that the crew:

Trip the RCPs after the loss of CCW and start the PDP to provide seal injection before sealwater inlet or lower sealwater bearing temperatures exceed 230 degrees F (per ES-01 Step 6 or 0POP04-RC-0002: Reactor coolant pump off-normal) to avoid potential (not necessarily immediate) RCP seal LOCA. Time to reach 230 degrees F is about 7-9 minutes from the loss of CCW.

*5.2.1.4 Relevant Procedures*

- 0POP04-VA-0001 “Loss Of 120 VAC Class Vital Distribution”
- 0POP04-RC-0002 “Reactor Coolant Pump Off Normal”
- 0POP05-EO-EO00 “Reactor Trip Or Safety Injection”
- 0POP05-EO-ES01 “Reactor Trip Response”
- 0POP09-AN-02M3 “Annunciator Lampbox 2M03 Response Instructions” (page 10, CCW PUMP 1A(2A) TRIP)
- 0POP09-AN-04M7 “Annunciator Lampbox 4M07 Response Instructions” (page 3, RCP 1A(2A) SEAL WTR INJ FLOW LO)

### 5.2.2 Performance of HFE 2A

**HFE 2A:**

Success Criterion: Trip the RCPs after the loss of CCW and start the PDP to provide seal injection before sealwater inlet or lower sealwater bearing temperatures exceeding 230 degrees F.

Results: Four of four crews failed.

As shown in Table 5-6, all crews stopped the RCPs. Three crews (R, S, and T) stopped them before the RCP seal temperatures were above 230 degrees F, and one crew a minute after. However, all crews exceeded the 1-minute criterion in POP4-RC2 for stopping the RCPs. None of the crews started the PDP.

**Table 5-6 Performance on HFE 2A**

| Crew | Stop all RCPs [time after loss of CCW and sealwater] | Start PDP | RCP seal temp >230 [time after loss of CCW and sealwater] |
|---|---|---|---|
| Q | 0:08:39 | Not started | 0:07:33 |
| R | 0:06:45 | Not started | 0:07:23 |
| S | 0:04:49 | Not started | 0:07:52 |
| T | 0:07:29 | Not started | 0:08:30 |

See Appendix E for details of crew performance.

### 5.2.3 Scenario Development and Crew Responses

The crews' performance in Scenario 2 is summarized in the following subsections, and the timing of main events and crew actions is listed in Table 5-7.

**Table 5-7 Timing of Main Events and Actions for All Crews in Scenario 2**

| Action/Event | Crew Q | Crew R | Crew S | Crew T | Source* |
|---|---|---|---|---|---|
| Start of scenario | 0:00:00 | 0:00:00 | 0:00:00 | 0:00:00 | SAML |
| Reactor trip | 0:02:54 | 0:02:45 | 0:02:51 | 0:03:03 | SAML |
| Loss of CCW and sealwater | 0:02:57 | 0:02:47 | 0:02:53 | 0:03:05 | SAML |
| Start E-0 (immediate actions) | 0:03:05 | 0:02:50 | 0:02:55 | 0:03:16 | OBS |
| Start procedure ES-01 | 0:08:05 | 0:08:30 | 0:09:40 | 0:07:15 | OBS |
| Detect no CCW or sealwater | 0:09:15 | 0:08:50 | 0:07:05 | 0:09:20 | OBS |
| Stop all RCPs | 0:11:36 | 0:09:32 | 0:07:42 | 0:10:34 | SAML |
| RCP seal temp. > 230°F | 0:10:30 | 0:10:10 | 0:10:45 | 0:11:35 | FILM |
| Start POP4 RCP-procedure | 0:10:00 | 0:13:00 | 0:11:50 | No | OBS |
| Start PDP | No | No | No | No | OBS |
| * SAML = Simulator Action Monitor Log, OBS = Observer's notes, FILM = video camera of screens | | | | | |

*5.2.3.1 Failure of Distribution Panel and Reactor Trip*

All four crews quickly recognized the failing distribution channel and took actions to take manual control of the affected equipment. They subsequently detected that the feedwater regulation valve to SG A could not be put in manual, causing the SG level to rise, and that they were going towards an automatic trip. Three of the four crews (Q, R and S) manually tripped the reactor before the automatic trip occurred.

*5.2.3.2 Loss of CCW and Sealwater*

Overall, all four crews tripped the RCPs but failed to start the PDP. Three crews (R, S, and T) tripped the RCPs before the RCP sealwater temperatures were greater than 230 degrees F, but none of them met the 1-minute criterion to stop RCPs specified in procedure 0POP04-RC-0002.

The quickest crew (S) detected the loss of CCW and sealwater 4 minutes after it occurred, and the other three crews (Q, R and T) detected it after 6 minutes. Crews R and S tripped the RCPs within 1 minute after detection of the loss of CCW and sealwater, Crew T tripped just after 1 minute, and Crew (Q) tripped over 2 minutes. In summary, the crews tripped the RCPs between 5-9 minutes after the loss of CCW and sealwater, which was too late to stop the RCPs within the time limit.

Crew Q directed an RO to start procedure 0POP04-RC-0002 and tripped the RCPs as directed from the procedure. The other three crews (R, S, and T) ordered the trip from knowledge when they detected the loss of CCW and sealwater.

According to 0POP04-RC-0002 “Reactor Coolant Pump Off Normal” and ES-01, the PDP pump must not be started if the RCP seal temperatures are 230 degrees F or higher. In the scenario, the temperature criterion was met at between 7-9 minutes after the loss of CCW and seal injection. Late detection of the loss of CCW and sealwater not only made the crews trip the RCPs late, but also gave them insufficient time (between 1-4 minutes) to start the PDP before the RCP sealwater reached the criterion temperature.

The crews could not move through procedures fast enough to start the PDP in time, even though they discovered the loss of CCW and sealwater before the RCP seal temperatures reached 230 degrees F. Three crews (Q, R and S) started 0POP04-RC-0002 “Reactor Coolant Pump Off Normal.” Members of two of the crews (R and S) suggested starting the PDP but this action was not completed within the available time. Crew T did not start procedure 0POP04-RC-0002 because they did not prioritize the seals in the situation.

### 5.2.4 Performance Shaping Factors for Scenario 2, HFE 2A

Table 5-8 lists the performance shaping factors (PSFs) identified by the experimental team for HFE 2A.

Table 5-8 Performance Shaping Factors for Scenario for HFE 2A

| PSF | Rating | Comment |
|---|---|---|
| **Stress** | 0 | Some observations of stress |
| **Adequacy of time** | ND | If the crews had had more training, there would have been enough time |
| **Team dynamics** | 0 | Some examples of poor communication, but overall as expected |
| **Work processes** | ND | Monitoring control boards and acknowledging alarms is not adequate. Procedure POP4 was read by the ROs, and they sometimes had difficulties in executing the procedure. |
| **Communication** | 0 | |
| **Scenario complexity** | MND | Many things happening at the same time made it difficult to detect the priority items. |
| **Indications of conditions** | ND | A lot of Train A indications were not available because of Panel 1201 failure. |
| **HMI** | N/P | |
| **Training and experience** | MND | The crews were not familiar with this kind of scenario and the needed prioritization. The loss of CCW and sealwater to RCPs is usually trained in an EC-00 (loss of offsite power) scenario once every 2 years. |
| **Procedural guidance** | ND | There is procedural guidance for starting PDPs, but the crews could not reach the critical actions in time. |
| **Execution complexity** | 0 | Depending on what procedure guidance they used and who read the procedure. (An RO may struggle with steps that a US has no problem with.) |
| N/P = Nominal/Positive, 0 = Not a driver, ND = Negative driver, MND = Main negative driver | | |

### 5.2.5 Drivers of Performance

All four crews quickly detected and identified the failing distribution panel and took appropriate actions. Training on this particular situation as well as plant experience helped the crews.

All crews decided to manually trip the reactor because they realized that there was no success path to avoid the trip. (One crew had an automatic trip just after the decision to trip.) The decision came from training that when an automatic reactor trip is inevitable then the reactor should be manually tripped as soon as possible.

All four crews detected the loss of CCW and sealwater several minutes after it occurred, which was too late to trip the RCPs based on the 1-minute requirement of POP4-RC2. Factors that helped them with the identification were training (to understand the significance of indications and alarms), board awareness, and the work practice of pausing after immediate actions to check if there was anything that needed to be addressed quickly. However, the complexity of the scenario delayed crew detection of the loss of all CCW and seal injection. The crews were distracted by the alarm cascade and required actions after the Panel 1201 failure and the reactor trip. Better board scans and checking the alarms could have made the crews discover the loss of CCW and seal injection earlier.

Three of the four crews (R, S, and T) stopped the RCPs from knowledge and one crew (Q) from procedure. The crews that stopped the RCPs from knowledge were helped by training. However, all the crews were late in stopping the RCPs because of the scenario complexity and short allowable time.

None of the four crews started the PDP. The crews either did not find procedure guidance to start the PDP or it took them too long to go through the procedures. The step that requires starting the PDP is in procedure ES-01, but they did not reach the step in time to complete the action within the allowable time. E-0 and ES-01 were the highest priority procedures in the scenario. The other procedure that could have helped them, POP4-RC2, was not the highest priority procedure and the ROs that handled the procedure lacked experience and training using it. In addition, the complexity of the scenario made the task difficult. They identified the situation late and had a problem recognizing the urgency of starting the PDP. The crews did not have training on this specific scenario as they were trained on loss of CCW or sealwater only in loss of offsite power scenarios, and therefore did not expect a concurrent loss of CCW and sealwater.

## 5.3 <u>Scenario 3 – SGTR</u>

### 5.3.1 Scenario Description

#### *5.3.1.1 Plant Technical Information*

- All participating crew members are in control room.
  - shift manager
  - unit supervisor
  - shift technical advisor
  - two reactor operators

- The plant is operating at 100 percent.
- Core burnup is 19,000 megawatt days per metric ton uranium (MWD/MTU) (EOL).

#### *5.3.1.2 SGTR*

Scenario 3 is a standard SGTR scenario without added complications. About 1 minute after the start of the scenario, a tube rupture occurs in the Steam Generator (SG) C. The leak size is about 500 gpm at 100 percent power.

#### *5.3.1.3 HFE 3A*

HFE 3A: Failure of crew to isolate the ruptured steam generator and control pressure below the SG PORV setpoint before SG PORV opening. The time window to make the required actions is estimated to be 2 to 3 hours.

The actions include:

- isolate the ruptured SG (feedwater and main steam isolation valves closed)
- maintain RCS pressure below the setpoint by cooling down the RCS (cooling the secondary by dumping steam and depressurizing the RCS)

*5.3.1.4 Relevant Procedures*

- 0POP05-EO-EO00 “Reactor Trip Or Safety Injection”
- 0POP05-EO-EO30 “Steam Generator Tube Rupture”

### 5.3.2 Performance of HFE 3A

Success criterion: Isolate the ruptured SG and control pressure below the SG PORV setpoint before SG PORV opening.

Results: Three out of three crews succeeded. Crew S did not participate in this scenario because of a simulator problem.

As shown in Table 5-9, all crews successfully maintained RCS pressure below setpoint.

**Table 5-9 Crew Performance on HFE 1C**

| Crew | Time to isolate SG | Maintained RCS pressure below setpoint |
|---|---|---|
| Q | 18:48 | Yes |
| R | 20:45 | Yes |
| T | 14:10 | Yes |

See Appendix F for details of crew performance.

### 5.3.3 Scenario Development and Crew Responses

All crews quickly detected the radiation alarms. They determined that the leak was only in SG C based on increased radiation indications for that SG. Two crews (Q and R) increased charging flow to determine the leak size and two crews (Q and T) entered POP4-RC4 (Steam Generator Tube Leak). All crews tripped the reactor between 1 and 3 minutes after the radiation alarms, and proceeded to E-30 within 8–12 minutes after the trip. After the initiating event, they isolated the SG between 14–21 minutes, and cooled down and depressurized within 35–42 minutes.

**Table 5-10 Timing of Main Events and Actions for All Crews in Scenario 3**

| Action / event | Q | R | T | Source |
|---|---|---|---|---|
| Start of scenario | 0:00:00 | 0:00:00 | 0:00:00 | OBS |
| Radiation alarms | 0:00:50 | 0:01:25 | 0:01:08 | OBS |
| Reactor trip | 0:03:45 | 0:03:38 | 0:02:21 | OBS |
| AFW 13 PTL | 0:05:00 | 0:06:02 | 0:03:38 | OBS |
| Enter E-30 | 0:13:06 | 0:15:30 | 0:10:10 | OBS |
| Adjust PORV | 0:16:29 | 0:17:52 | 0:12:02 | OBS |
| Close MSIV & bypass | 0:17:32 | 0:18:42 | 0:12:54 | OBS |
| Close AFW OCIV (SG isolated) | 0:19:38 | 0:22:10 | 0:15:18 | OBS |
| Stop depressurization | 00:40:10 | 0:43:10 | 00:36:00 | OBS |
| * SAML = Simulator Action Monitor Log, OBS = Observer’s notes, FILM = video camera of screens | | | | |

### 5.3.4 Performance Shaping Factors for Scenario 3, HFE 3A

Table 5-11 lists the performance shaping factors (PSFs) identified by the experimental team for HFE 3A.

**Table 5-11 Performance Shaping Factors for HFE 3A**

| PSF | Rating | Comment |
|---|---|---|
| **Stress** | 0 | |
| **Adequacy of time** | 0 | |
| **Team dynamics** | N/P | Some examples of good teamwork. |
| **Work processes** | 0 | |
| **Communication** | 0 | |
| **Scenario complexity** | 0 | |
| **Indications of conditions** | N/P | Radiation alarms and indications helped the crews identify the SGTR quickly. |
| **HMI** | N/P | |
| **Training and experience** | N/P | Frequently trained scenario, no added complications. |
| **Procedural guidance** | N/P | Procedures supported this base-case scenario well. |
| **Execution complexity** | 0 | |
| N/P = Nominal/Positive, 0 = Not a driver, ND = Negative driver, MND = Main negative driver | | |

### 5.3.5 Drivers of Performance

The crews had no major difficulties in this scenario. Factors that supported their performance were indication of conditions, procedures, training and teamwork.

## 5.4 <u>Difficulty Ranking of All HFEs</u>

The HFEs were ranked in terms of their difficulty levels by three of the four unit supervisors who participated in the study. As show in Table 5-12, they all ranked the HFEs in the exact same order.

1 = Most difficult
4 = Least difficult

**Table 5-12 Unit Supervisors' Difficulty Rankings of the HFEs**

| HFE | Task | Crew Q | Crew R | Crew T |
|---|---|---|---|---|
| HFE 2A | Stop RCPs and start PDP in Scenario 2 | 1 | 1 | 1 |
| HFE 1C | Identify and isolate ruptured steam generator in Scenario 1 | 2 | 2 | 2 |
| HFE 1A | Start bleed and feed in Scenario 1 | 3 | 3 | 3 |
| HFE 3A | Identify and isolate ruptured steam generator in Scenario 3 | 4 | 4 | 4 |

The Unit Supervisors' rankings correlate with the failure rates of the HFEs. The evaluated difficulty of the HFEs in the right column of Table 5-13 was agreed upon by the two data analysts and the trainer from the plant who were informed by the US ranking and the HFE failure rates.

**Table 5-13 Evaluated Difficulty of the HFEs**

| HFE | Task | US rank | Failure rate | Difficulty rating |
|---|---|---|---|---|
| HFE 2A | Stop RCPs and start PDP in scenario 2 | 1 | 4/4 | Very difficult |
| HFE 1C | Identify and isolate ruptured steam generator in scenario 1 | 2 | 3/4 (given 40 minute time criterion) | Difficult |
| HFE 1A | Start bleed and feed in scenario 1 | 3 | 0/4 | Fairly difficult to difficult* |
| HFE 3A | Identify and isolate ruptured steam generator in scenario 3 | 4 | 0/3 | Easy |
| * HFE1A was made less difficult by the crews being well trained on similar scenarios in which the recirculation valve was open. | | | | |

# 6. OVERVIEW OF EMPIRICAL AND PREDICTIVE QUANTITATIVE RESULTS

A comparison of the predictive quantitative results from the human reliability analysis (HRA) applications to the empirical results is presented in this chapter. More detailed discussions on intra-method quantitative differences and their relationship with the empirical data are presented in Chapter 7. Before proceeding however, it should be noted that because of the limited number of scenarios, human failure events (HFEs), and teams involved in this study, on the basis of these results alone, no methods can be identified as clearly superior to other methods either in an overall way or with respect to predicting performance on particular types of HFEs. However, it is generally agreed that the results of this study raise some significant issues in methods as well as practices of the methods as a whole. For example, it was observed that two or more teams came up with, in some cases significantly, different human error probability (HEP) estimates or qualitative predictions. And although, one may be able to explain the reasons for the produced estimates and associated differences, the explanations do not eliminate the need for the probabilistic risk analysis (PRA)/HRA community to address the issues identified here and be aware of the potential strengths and weaknesses of the various methods for doing an HRA. In addition, the findings from this study with respect to the methods and HRA in general are consistent with those from the International HRA Empirical Study [1-4], while providing additional information about why differences in results can be found both within and across the methods.

These findings are especially important because incorporation of a weak HRA in a PRA can have significant effect on the use of PRA in regulatory space. For example, inappropriate HEP ranking could produce incorrect dominant sequence results, thus distorting the risk profile of a plant, and focus the derivation of insights and identification of the most significant needed improvements on the wrong issues.

Consistent method-to-method or analyst-to-analyst variability indicates a lack of robustness in HRA methods, which jeopardizes the efficiency and effectiveness of the use of HRA in the regulatory process. However, as evident from the discussion in the sections below, addressing HRA issues is not necessarily more difficult than addressing other PRA issues. For example, issues such as modeling common-cause failures and addressing uncertainties in PRA have been a focus of the NRC and the industry and this focus has led to significant advances in the state-of-the-practice. Such advances can also be achieved in HRA with thoughtful use of the findings from studies such as this one.

## 6.1 Bayesian Results for the Empirical HEPs

A Bayesian update with a non-informative prior (Jeffrey's prior) [14-15] was done to obtain the empirical distributions of the empirical HEPs, to be used as reference values in the part of the comparison that addressed the HRA methods' quantitative predictions. A Bayesian approach is preferred because of the small sample size for each HFE.

The distributions of the empirical HEPs are shown in Table 6-1 and in Figure 6-1. Note that no empirical HEP distribution was estimated for HFE 1B because, as previously explained, no data were gathered during the experiment for this HFE since the conditions for this HFE were not met and, therefore, no crews worked under those conditions). The rightmost column of Table 6-1 provides the difficulty ranking of each HFE assessed from the qualitative evaluation of the observed performances (see discussion in Section 4.2). In Figure 6-1, the HFEs are ordered from left to right by decreasing difficulty, according to the qualitative ranking. The predicted HEPs are compared against the bounds, in other words, whether the mean HEPs are inside or outside the empirical bounds. The mean and median values of the empirical HEP distribution are provided in the table for additional information; however, the HEPs are only compared with respect to upper and lower bounds.

**Table 6-1 Empirical HEP Distributions Obtained in Bayesian Update***

| HFE | No. of observations | No. of failures | 5th percentile | Median | Mean | 95th percentile | Qualitative difficulty ranking (1 = most difficult) |
|---|---|---|---|---|---|---|---|
| 1A | 4 | 0 | $4.62 \times 10^{-4}$ | $5.20 \times 10^{-2}$ | $1.00 \times 10^{-1}$ | $3.62 \times 10^{-2}$ | 3 |
| 1B | 0 | N/A | N/A | N/A | N/A | N/A | N/A |
| 1C | 4 | 3 | $3.49 \times 10^{--1}$ | $7.28 \times 10^{-1}$ | $7.00 \times 10^{-1}$ | $9.54 \times 10^{-1}$ | 2 |
| 2A | 4 | 4 | $6.38 \times 10^{-1}$ | $9.48 \times 10^{-1}$E-1 | $9.00 \times 10^{-1}$ | 1.00 | 1 |
| 3A | 3 | 0 | $6.03 \times 10^{-4}$ | $6.74 \times 10^{-2}$ | $1.25 \times 10^{-1}$ | $4.44 \times 10^{-1}$ | 4 |

* The means and medians for HFEs 1A and 3A may seem conservative because of the use of the Jeffrey's prior. However, the purpose of the Bayesian update is to obtain the uncertainty bounds of the empirical data rather than to assess the means and medians.

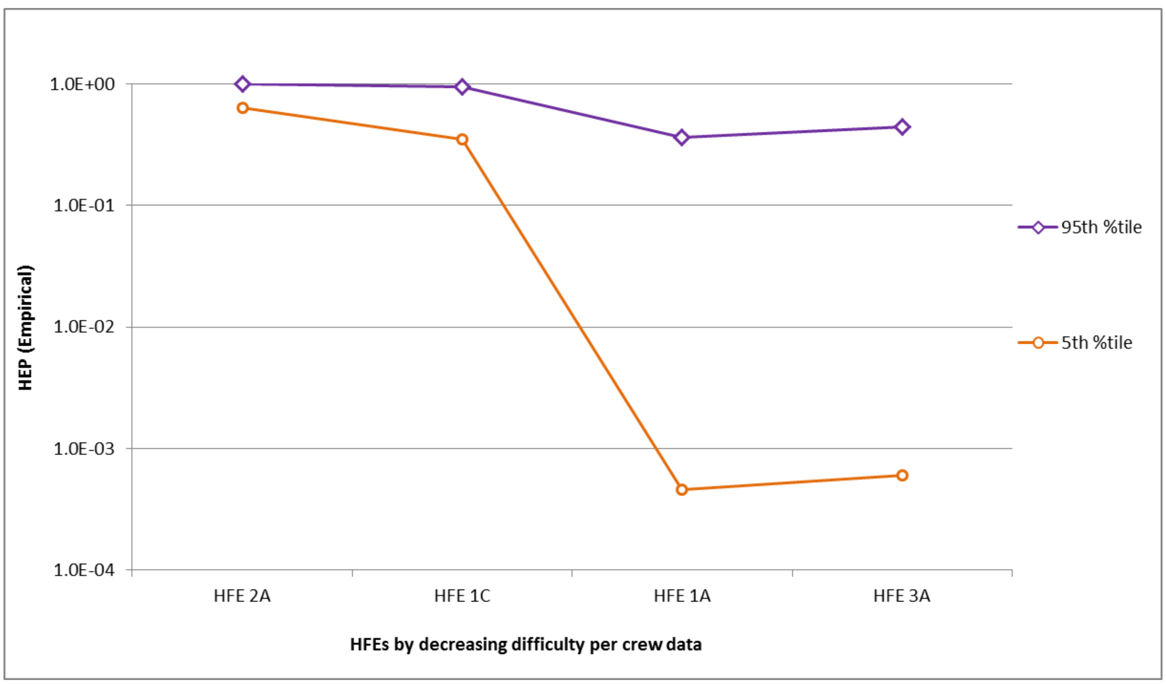



**Figure 6-1 Bayesian uncertainty bounds of empirical HEPs**

Although no failures were observed for either HFE 1A or HFE 3A, HFE 3A was ranked less difficult than HFE 1A. This is because there was no evidence of performance issues for HFE 3A but some minor performance issues were observed for HFE 1A. The issues observed for HFE

1A were judged to be significant enough to warrant a higher level of difficulty from the unit supervisors' perspectives. However, the HEP distributions for these HFEs are practically identical and do not reflect any differences. Given the small number of observations, even if small differences in difficulty could affect the probability of failure, they may not have been detectable.

## 6.2 Quantitative Results from the HRA Team Predictions (Including Bayesian Uncertainty Bounds)

The predicted mean HEPs of all HFEs estimated by all HRA methods used in the study are presented in Figure 6-2. On the horizontal axis, the HFEs are ordered by the qualitative difficulty ranking (HFE 1B was not ranked since there was no empirical data for this HFE). Based on Figure 6-2 and the empirical findings, the assessment team observed that:

- There was significant variability among the team predictions of HEPs for each of the HFEs. Except for HFE 3A, the variability of the HEPs for each HFE given by the HRA teams is approximately 1 to 1-1/2 orders of magnitude. Interestingly, although all the HRA teams identified HFE 3A (the standard SGTR scenario with no further complications) to be the easiest HFE with the lowest HEP, this HFE showed the greatest degree of variability, with two predictions—Standardized Plant Analysis Risk-Human Reliability Analysis, (SPAR-H) Team 1 and A Technique for Human Event Analysis (ATHEANA) Team 1—significantly deviating from most of the others and leading to approximately 3 orders of magnitude difference among the HEPs.

- As noted above in the introduction, the consistentcy in difficulty ranking between the HRA teams' HEP estimates and the actual crew performance (by simulation observation and psot-simulation interview with the crew) of HFEs is rather important for HRA, since HRA is a tool not only to estimate probabilities of human errors for use in PRA, but also to identify important contributors and prioritize where fixes and improvements are needed in the plant. In most cases, the trend of HRA teams' HEP predictions is consistent with the crew performance (i.e., difficulty ranking based on the simulator observation and post-simulation crew interview). Inconsistences were observed in methods, for example, estimating higher HEPs for HFE 1C than HFE 2A, while HFE 2A was ranked more difficult than HFE 1C. For HFEs 1C, 1A, and 3A, the HEPs from each HRA team generally show good correlations with the difficulty ranking; in a couple cases the analysts did not produce different HEPs for HFEs 1A and 1C, which does not seem consistent with the empirical data. Note that HFEs 2A, 1C and 1A were all rather difficult events. While plausible, the events comprised far more difficulties for the crews than many of the HFEs in standard PRA scenarios and, therefore, may have taxed the ability of some methods to account for the differences in difficulty.

- It is recognized that this study cannot offer any conclusive evidence on the consistency or accuracy of the quantitative analysis from the methods since it is based on only two or three data points (HRA teams) per method. "Consistency" here refers to producing HEPs that (a) reflect the difficulty ranking of HFEs and (b) are close in values for a given HFE. "Accuracy" refers to producing values reflecting the empirical data in terms of failures observed. Nevertheless, on consistency, a review of Figure  suggests that ASEP (Accident Sequence Evaluation Program Human Reliability Analysis Procedure), ATHEANA and CBDT & HCR/ORE (cause-based decision tree and human cognitive reliability and operator reliability experiments (HCR/ORE) [9]) yielded somewhat more consistent quantitative results than SPAR-H across the analysis teams that used each of these methods. This is only true for CBDT & HCR/ORE if one takes into consideration that the low HEP of HFE 1C by CBDT & HCR/ORE produced by Team 2 was caused by the team's misunderstanding of the definition of the HFE. Regarding "accuracy," except ASEP, all other methods seemed to have underestimated HFE 2A and several underestimated the difficulty of 1C.

- HFE-1B is not much discussed in this study, because there were no observations in the empirical data for this HFE. This is because no crew waited until reactor tripped automatically in this study. All crews manually tripped the reactor before an automatic trip would have occurred. Nevertheless, comparing the predictions of HFE-1B shows that this HFE was even more uniform than the others. Eight of the nine HRA teams' predicted HEPs were well within 1 order of magnitude. The only exception was one of the ASEP teams, which was slightly lower than the rest. The two ATHEANA teams actually predicted the exact same mean HEP.

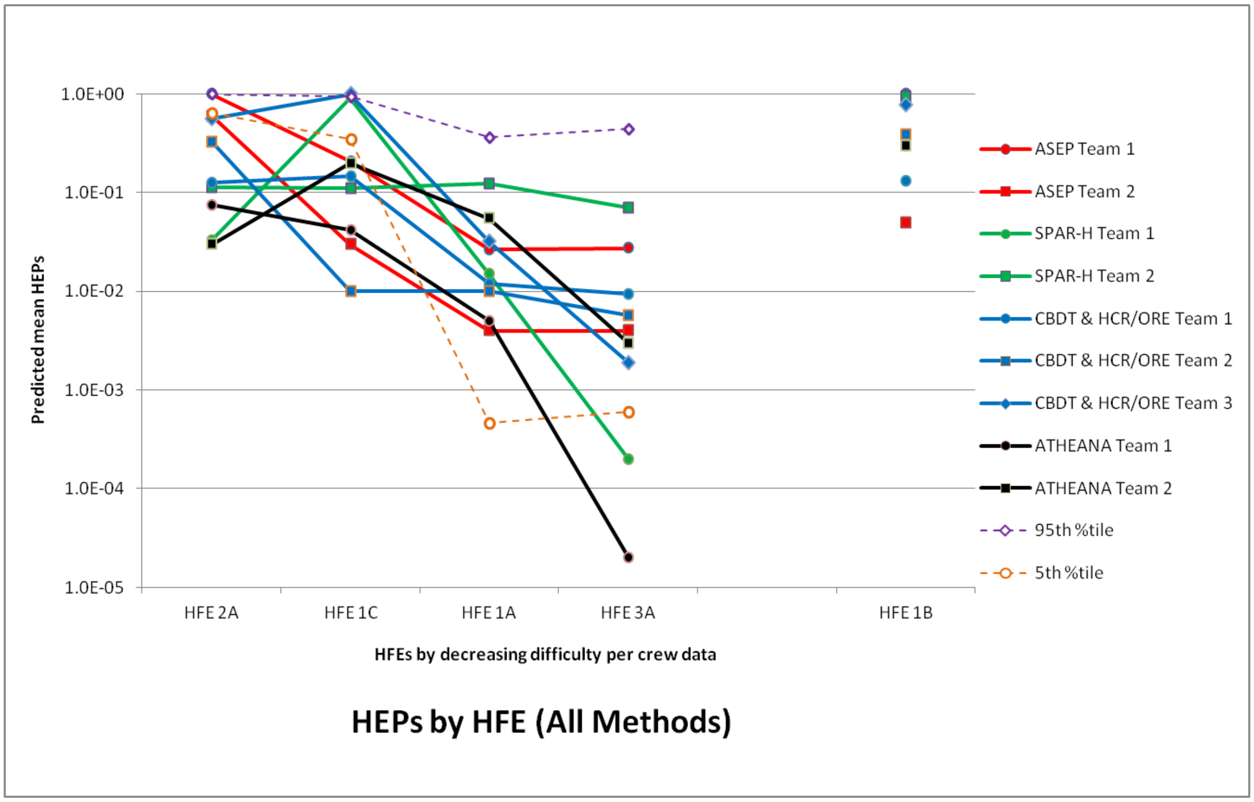



**Figure 6-2 Predicted mean HEPs by all HRA methods**

Turning to the differences in the diagnosis and execution HEPs predicted by each method, the results are plotted in Figure 6-3 and Figure 6-4, respectively, with the exception of ATHEANA which does not calculate a separate HEP for execution, but includes it in estimating the total HEP. As shown in Figure 6-3, the diagnosis HEPs of ASEP, SPAR-H and CBDT & HCR/ORE generally follow the HFE difficulty ranking. Although some teams produced relatively optimistic values for HFEs 2A and 1C, the HEPs differentiate difficult HFEs from easy ones fairly well. For HFEs 2A and 1C, the HEPs associated with diagnosis tend to dominate the total HEP value (i.e., diagnosis HEPs determine the trend/shape of the HEP curves); the same could be argued for ATHEANA as well based on the discussion of what was driving performance in the qualitative analysis. This result is in line with the empirical evidence and the study expectation that the complex scenarios would significantly increase the difficulty of diagnosis rather than execution for most HFEs. Similar to the total HEPs, the diagnosis HEPs exhibit some variability for HFE 3A. Regarding the CBDT & HCR/ORE results, Teams 1 and 2 produced a relatively conservative HEP for HFE 3A when compared with the results produced by Team 3. The difference is explained by the fact that the former two teams did not include recovery in their analysis, which, given the available time and conditions and the empirical results, would seem to be unnecessarily conservative. Thus, this would be one contributor to the variability and leaves only the SPAR-H Team 2 as an outlier with respect to the diagnosis HEPs for the teams presented.

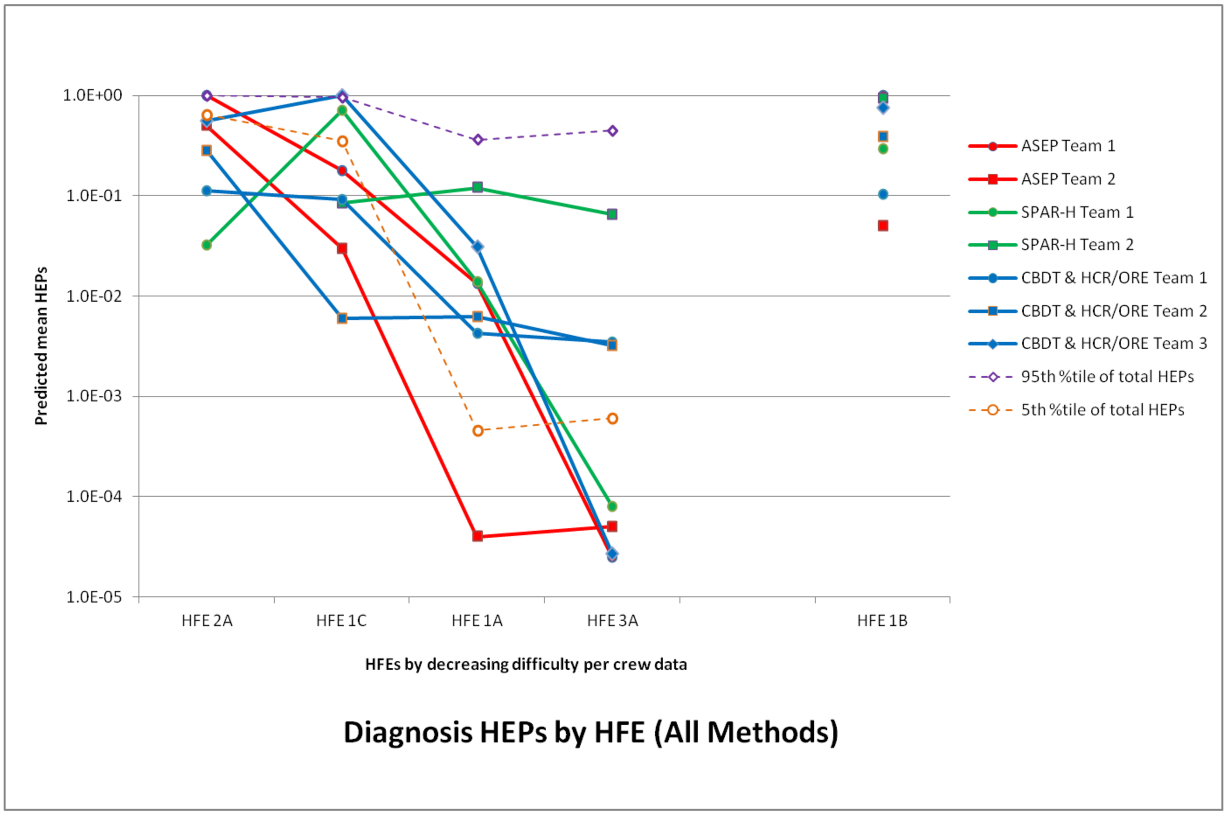



**Figure 6-3 Predicted diagnosis HEPs by HRA methods**
*(Note: ATHEANA Teams are not shown in this figure as ATHEANA does not calculate diagnosis and execution HEPs separately.)*

The values produced for the execution HEPs plotted in Figure 6.4 by ASEP, SPAR-H and CBDT & HCR/ORE do not show apparent differentiation across the HFEs as seen in the total HEPs, but this is a reasonable result for in-control room actions which are usually straightforward and typically accomplished quickly once the crew determines what is needed to be done. It is interesting to note that most of the teams obtained comparable execution HEPs for HFEs 1A and 3A. This may be partly because these teams used the same data source (e.g., Technique for Human Error Rate Prediction (THERP)) and the same or similar quantification approaches (e.g., THERP or ASEP). Nevertheless, for HFEs 2A and 1C, the execution HEPs show large variability. One contributor to variability is the high execution HEPs produced by ASEP Team 2 and SPAR-H Team 1 for HFEs 2A and 1C, respectively; as discussed below, these estimates seem to be unjustifiable. However, since the execution HEPs do not dominate the total HEP for the two HFEs, the variability does not affect the total HEPs.

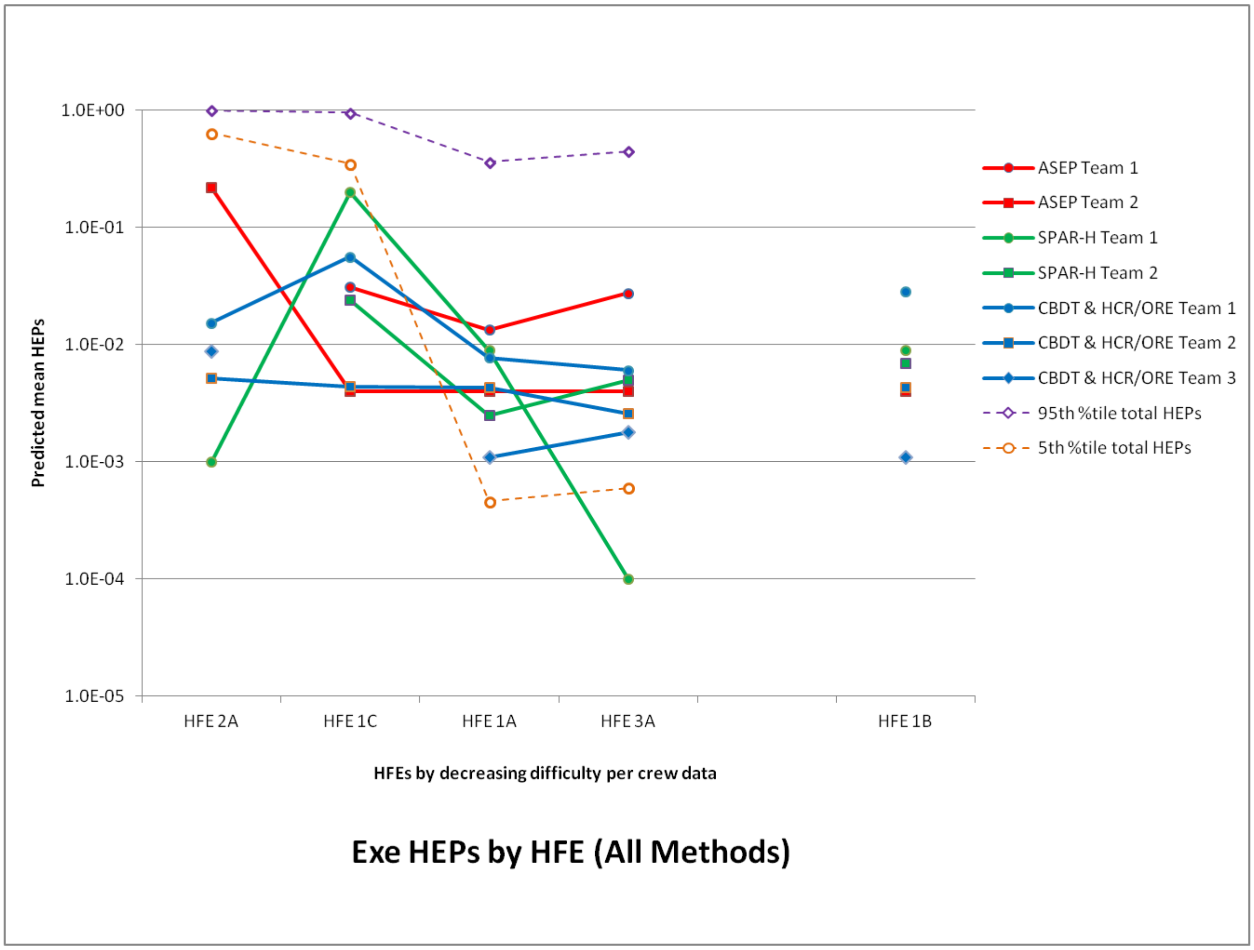



**Figure 6-4 Predicted execution HEPs by HRA methods**
*(Note: ATHEANA Teams are not shown in this figure as ATHEANA does not calculate diagnosis and execution HEPs separately.)*

It is interesting to note the following when examining Figures 6.2, 6.3 and 6.4 together. For methods that calculate final HEPs as the sum of diagnosis and execution HEPs, the final HEPs seem to be reasonable with respect to HFE difficulty rankings; however, this conclusion could be questionable if one considers the relative contribution of diagnosis and execution HEPs. For example, the diagnosis HEP predicted by ASEP Team 2 for HFE 1A is significantly lower than the execution HEP (i.e., execution HEP dominates the total HEP), which is not consistent with what would be expected for a challenging HFE. Similarly, the execution HEPs predicted by ASEP Team 2 for HFE 2A and SPAR-H Team 1 for HFE 1C, seem to be unjustifiably high. Addressing the question of what is the most important contributor to an HEP (diagnosis or execution) is an important aspect of HRA. Safety improvements to address diagnostic failures could be very different from execution failures.

Detailed discussions (HFE by HFE) about differences between the predictions of each method application and the crew data are presented in Appendix G and differences within the same or similar methods (intra-method comparisons) are presented in Appendix H. A summary of the results of the intra-method comparisons and the issues and factors identified as contributing to the variability in results are presented in Chapter 7. Detailed discussions of the contributors to variability are given in Chapter 8.

# 7. SUMMARY OF INTRA-METHOD COMPARISONS

The human error probabilities (HEPs) predicted by the same or similar methods are presented in Figure 7-1 alongside the Bayesian uncertainty bounds derived from the crew data. On the horizontal axis, the human failure events (HFEs) are ordered by their difficulty ranking. The following are brief summaries of the method performance (as shown in the figures below):

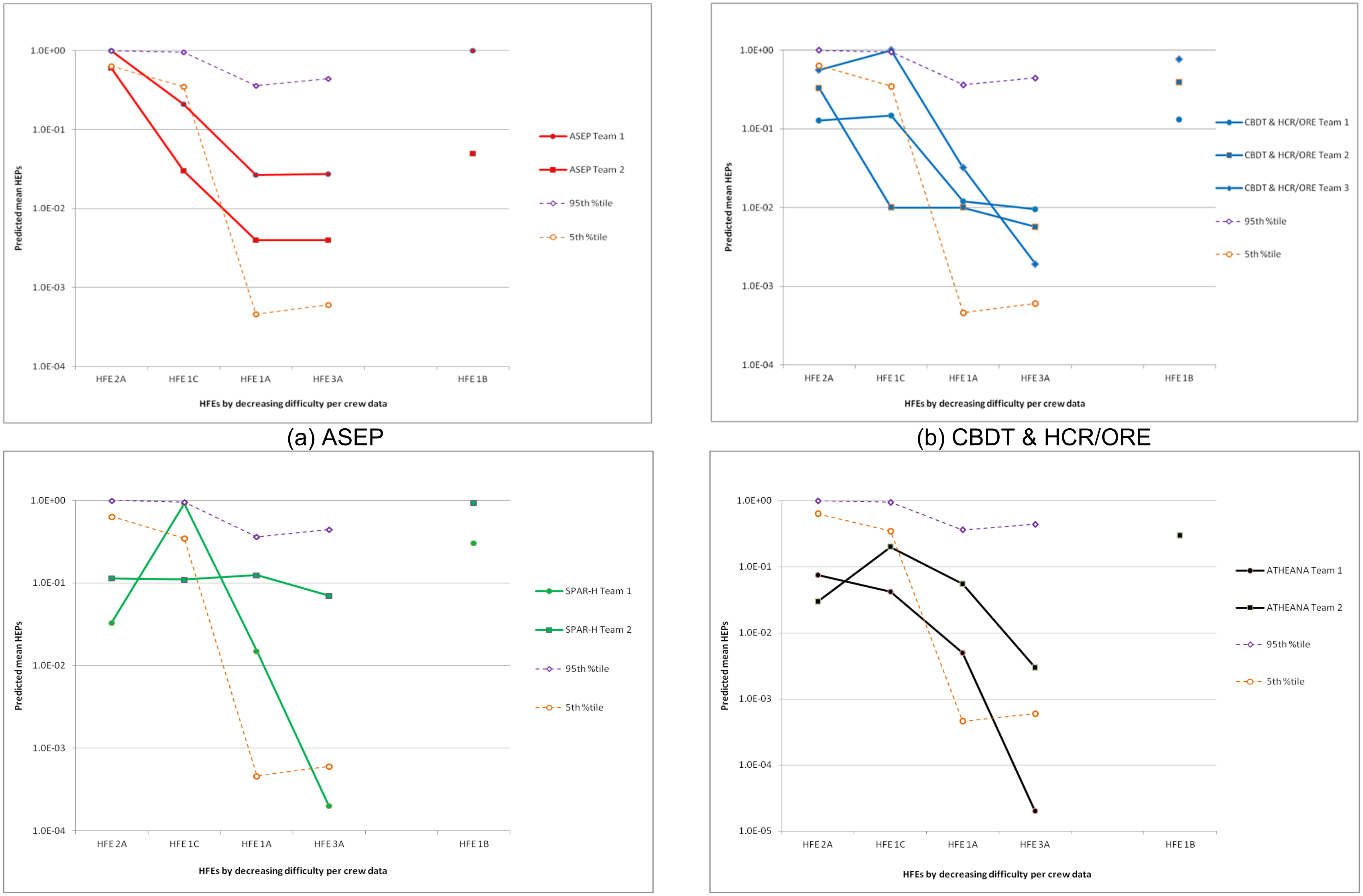
- For the ASEP (accident sequence evaluation program human reliability analysis (HRA) procedure) method, the HEPs from the ASEP teams show a consistent pattern across the teams, with one team consistently more optimistic than the other, but both reflecting similar differences in the difficulty of the HFEs.

- For the teams using the EPRI HRA Methodology implemented with the HRA Calculator with cause-based decision tree (CBDT) & HCR/ORE (human cognitive reliability and operator reliability experiments) [9]) (but see Section 7.2), with the exception of HFE 1C where Team 2 misunderstood the conditions related to quantifying the HFE, there is indication of reasonable consistency across the teams on the HEPs, which generally increased as HFE difficulty increases.

- For the Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H), the differences in the HEPs by one team are small across HFEs, i.e., they did not see much difference in the difficulty of the HFEs and, therefore, their HEPs do not differ much across the HFEs. The other SPAR-H team predicted relatively substantial differences in the HEPs, with their HEPs increasing as the difficulty of the HFEs increase, except for the HEP they provided for the most difficult HFE.

- For the teams using “A Technique for Human Event Analysis” (ATHEANA), the difference between the teams’ HEPs on HFE 3A are larger than those on other HFEs, but the HEPs for HFE 3A were consistent in the sense that they were the lowest for both teams. The main disagreement between the two ATHEANA teams was in the HEPs for HFEs 2A and 1C, but both teams appeared to underestimate the difficulty of 2A, as did several other teams using different methods.

Although the consistency of the quantitative results within methods is generally encouraging (with a few exceptions there is generally about 1 order of magnitude or less difference across teams within each method, see Figure 7-1), one question that arises is whether the qualitative predictions (performance issues, failure modes, dominant factors and failure “mechanisms”) from these analyses are consistent with the quantitative consistency. In other words, do the two ASEP, three CBDT & HCR\ORE, and two ATHEANA analysis teams obtain some consistency in their intra-method quantitative results <u>based on the same underlying qualitative and quantitative reasons</u>? The results show that although there was reasonable matching in some cases, the qualitative bases for the teams’ HEPs differed substantially in several instances. In addition, there were significant differences between teams using the same methods in the assessed contribution of the diagnosis and execution portions of the actions underlying the HEP in some cases. Thus, the similarities in the HEPs produced by teams using the same methods may have been somewhat serendipitous. This result raises potential questions about the reliability of the methods and their use for safety management, because they would be expected to lead to different safety insights and recommendations for potential improvements. The reasons for the differences are discussed in the intra-method comparisons below.




(a) ASEP

(b) CBDT & HCR/ORE

(c) SPAR-H

(d) ATHEANA

**Figure 7-1 Predicted mean HEPs by HRA methods with Bayesian uncertainty bound**

## 7.1 ASEP

Two teams (ASEP Team 1 and Team 2) performed analyses with ASEP [7]. Where a discrepancy between their qualitative analyses was identified, the qualitative analysis of Team 1 tended to be more consistent with the empirical data in terms of the performance drivers they identified and the operational stories they developed. It appears that this was the result of the more detailed qualitative analysis by Team 1, which seems to go beyond the guidance given in the ASEP method. In comparison, Team 2 tended to closely follow the specific guidance in ASEP. The value of Team 1's detailed qualitative analysis was particularly apparent in the team's ability to see how complications in the scenarios could influence operators' diagnoses as well as supporting their consideration of the role of operating procedures in operators' diagnoses. Such consideration is not explicitly addressed in ASEP, but enabled Team 1 to identify difficulties in the complicated scenarios that would prevent operators from making a timely diagnosis or delay them in reaching relevant procedure steps. In contrast, Team 2 only considered whether post-diagnosis actions were covered in procedures, per the ASEP guidance.

These observations to some degree relate above to one of the methodological features of ASEP. For estimation of the diagnosis HEP, ASEP relies on its Nominal Diagnosis Model (i.e., a time reliability correlation (TRC); ASEP Figure 8-1) with a few performance shaping factor (PSF) adjustments. This feature contributes to the model's strength of simplicity; however, with limited analysis of how PSFs and scenario conditions interact with operators' cognitive behavior (as with many other first generation HRA methods), such an approach is likely to limit the method's ability to identify important plant conditions or operational situations that would negatively affect operators' behavior. Consequently, ASEP may produce few insights for error reduction and potentially lead to unrealistic HEPs.

As shown in Figure 7-1a, the HEP curves from the two ASEP teams are close to parallel with one team consistently more optimistic than the other. On a general basis, they also reflect the HFE difficulty ranking. Unlike the qualitative analysis in which Team 1 was generally closer to the empirical findings, it is difficult to tell which team's HEP predictions tend to be more consistent with the crew data, because it is difficult to determine the true bounds because of the small sample of crew performance data. However, it does appear that Team 2 may have underestimated the true HEP for 1C since 3 out of 4 crews failed to complete the response within the time criterion.

The diagnosis and execution HEPs of the two teams are shown in Figure 7-2. Figure 7-3 shows the relative contributions of the diagnosis and execution HEPs to the total HEPs. For the easiest HFE (HFE 3A), the two teams obtained comparable diagnosis HEPs and agreed that the total HEPs were dominated by execution HEPs. For cognitively challenging HFEs (HFEs 1C and 2A), the diagnosis HEPs are larger than execution HEPs as expected. However, as can be seen in the pie charts presented in Figure 7-3, the teams did not always agree on what the relative contributions of these response stages were to the total HEPs.

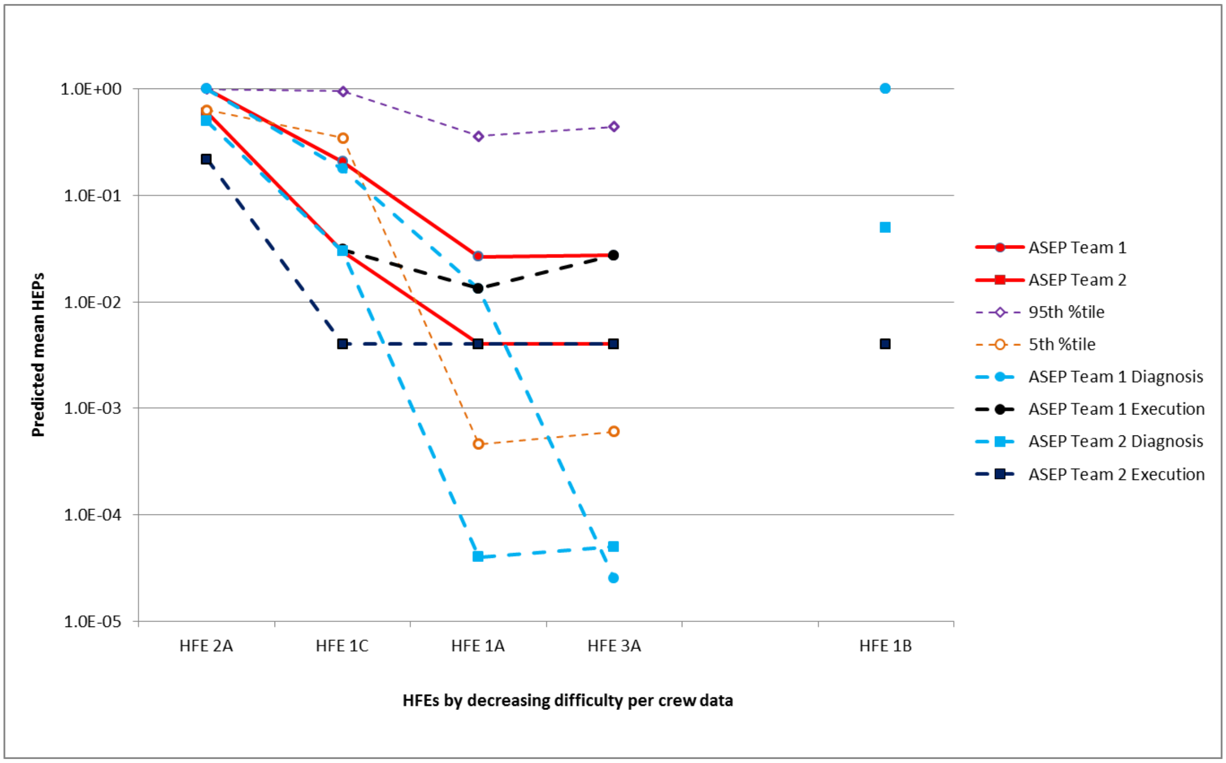



**Figure 7-2 Diagnosis and execution HEPs of ASEP teams**

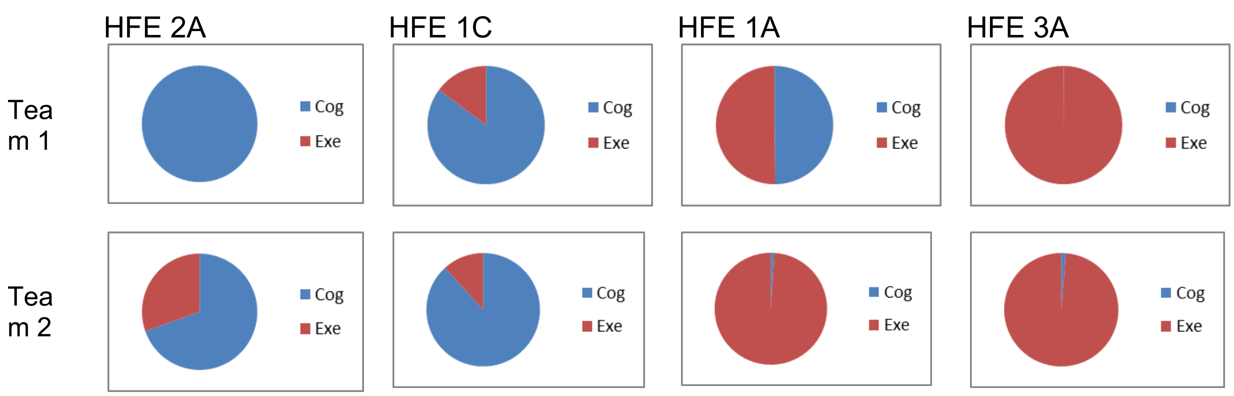



**Figure 7-3 Contributions of diagnosis and execution HEPs of ASEP teams**

As can be seen in Figure 7.3, where relatively large differences are found, Team 1's diagnosis HEPs were larger than Team 2's. Particularly for HFE 1A, Team 2's diagnosis HEP is quite small so that it can be neglected compared to the execution HEP (see Figure ). The differences seem to be largely caused by the differences between the two teams in their estimation of execution time, which affected the estimates of available time for diagnosis. Small differences in timing analysis results can lead to large differences in diagnosis HEPs because of the characteristics of the TRC. As discussed above, the timing differences can be partly attributable to Team 1's detailed qualitative analysis, in which they identified plant conditions and operational situations that would delay operators in reaching relevant procedural steps. As another example, Team 1 considered the time needed to finish some procedural steps that did not seem to be explicitly considered in the analysis of Team 2. Moreover, Team 1, in a couple of cases, showed more conservatism in estimating the time for some actions than did Team 2.

The discrepancies suggest that more specific guidance for execution time estimation is necessary to reduce variability in ASEP HRA results to obtain better HEPs.

Another important difference between the two teams is that they obtained different information/impressions from their interviews with trainers for the most difficult HFE (HFE 2A) with respect to crews' knowledge and experience. The information obtained by Team 2 led them to believe that crews' experience would help the crews' diagnosis and thus the team selected the nominal diagnosis HEP. In contrast, the information obtained by Team 1 led them to believe that crews' training and experience were not adequate and thus they selected the upper bound diagnosis HEP. Although the two teams' HEPs are not substantially different, this, to some extent, can illustrate how the information obtained from the interviews can affect analysts' judgments.

The HFE 1B analyses also showed how the HRA analysts' judgment about the information from interviews led to variability in their analyses. The two teams obtained the same information from interviews about crew training and experience, but their analyses and predictions differed. Team 1 decided to discount training and experience because of the high difficulty of the scenario and obtained a high diagnosis HEP, while Team 2 decided to credit crews' training and experience and obtained a small diagnosis HEP. The HEP difference (i.e., very high vs. relatively low) is large enough to lead to different conclusions. However, evaluating which prediction is more accurate was not possible because of the lack of empirical data for this HFE. A similar case occurred in the HFE 1A analyses. Despite that both teams acknowledged the crews' training and experience, Team 1 decided to choose the nominal diagnosis HEP to account for the difficulty in diagnosis, while Team 2 chose the lower bound of the diagnosis HEP without such a consideration. The conservatism of Team 1 in the two cases is another example of Team 1's relatively more detailed analysis, which enabled them to recognize some dynamics in scenarios that might delay or hinder the response.

In ASEP, the effect of crew experience and training is, to some degree, addressed by PSF adjustment; however, there does not seem to be explicit and/or adequate guidance to help analysts address knowledge-based behavior in identifying critical actions and procedural paths. Some operators took some actions based on their evaluation of the plant without procedure guidance, when, for example, the relevant procedures could not be reached. These knowledge-based actions led the operators through different procedural paths than the HRA analysts predicted. In one case, Team 1 considered procedures that the crews did not actually enter and as a result, they conservatively estimated the time required to complete post-diagnosis actions. As a result, they obtained a larger execution HEP compared to Team 2.

It is interesting to note that although the two teams arrived at the same conclusions for all HFEs about crew stress levels and whether post-diagnosis actions were dynamic or step-by-step, the execution HEPs of Team 1 are consistently about 1 order of magnitude larger than those of Team 2. One contributing factor to the numerical difference seems to be the different quantification approaches used by the two teams. Team 2 used ASEP rules, whereas the Team 1 followed the guidance in Item 2 in ASEP Table 8.5 to use THERP, as they decided that there was enough information to do task analysis. This finding may sound contradictory to the claim that ASEP generally yields more conservative HEPs than THERP, but it might be explained to some extent by the fact that Team 1 quantified more steps (i.e., decomposed the tasks differently).

## 7.2 CBDT and HCR/ORE

CBDT & HCR/ORE Team 1 did an analysis of the five HFEs using a hybrid CBDT & THERP & ASEP [7] method. Team 2 and Team 3 did the analysis using the EPRI HRA Methodology using the HRA Calculator version 4.1.1) [8]. The EPRI HRA Methodology divides an HEP into two elements: detection, diagnosis, and decisionmaking ($P_{cog}$), and action ($P_{exe}$). $P_{cog}$ is quantified using the larger HEP calculated by the CBDT and HCR/ORE [9], and $P_{exe}$ is quantified using THERP [10]. This intra-method comparison will compare the qualitative and quantitative performance of all three teams, while taking the method differences into account to the extent possible.

The CBDT & THERP & ASEP approach used by Team 1 differs somewhat from the approach recommended in the EPRI HRA Calculator (used by Teams 2 and 3), limiting the scope of the comparison in some cases. Although both approaches use CBDT [9] and a type of time reliability correlation to quantify the cognitive contribution to the HEP, the EPRI approach uses the maximum value between CBDT and the HCR/ORE, whereas Team 1's approach uses the sum of the two. Also, the two approaches use different TRCs – the EPRI method uses the HCR/ORE correlation, while Team 1 uses the TRC from THERP. In addition, the EPRI method uses THERP to quantify the execution contribution, while Team 1 used ASEP. Table 7-1 summarizes the methodology differences.

The CBDT values and contributing PSFs can be directly compared among the teams. The time reliability components cannot be compared directly because different correlations are used, however, at a more general level, the results of the timing analyses can be compared, as well as whether the obtained values were similar. The task decomposition for the execution components can also be compared.

There was significant variability in the amount and detail of documentation the three teams provided. In the following discussion, where comparisons for only two teams are discussed, the third team either did not provide sufficient documentation to allow for a comparison or the methods were different in that aspect precluding a direct comparison.

**Table 7-1 Summary of Method Differences Used by CBDT & HCR/ORE Teams**

| Team | Method | HEP |
|---|---|---|
| 1 | CBDT & THERP & ASEP | HEP = [CBDT & THERP(TRC)] & ASEP |
| 2 and 3 | EPRI HRA Methodology | HEP = Maximum[CBDT, HCR/ORE] & THERP |

Figure 7-4 gives the numerical results from each team plotted by the decreasing difficulty level of the HFEs, along with the uncertainty bounds for the empirical data. This figure includes decomposition of the HFEs into the cognitive and execution contributions. Figure 7-5 gives a summary of the overall results with commentary on certain HFEs. Overall, total HEPs match scenarios moderately well in rank and level of differentiation; however, the cognitive vs. execution contribution to the HEPs were not always consistent with what would be expected from the empirical data. In addition, as shown in the pie charts in Figure 7-6, the teams did not always agree on the relative contributions of these response stages to the total HEPs. These differences occurred for a variety of reasons including different assumptions about the presence of certain conditions and how the different portions of the HFEs were decomposed. Furthermore, in many instances there were inconsistencies in the teams' operational stories and the drivers they identified. However, although these various inconsistencies seem to even out in the final HEP, it appears that there were significant differences in the underlying analyses.

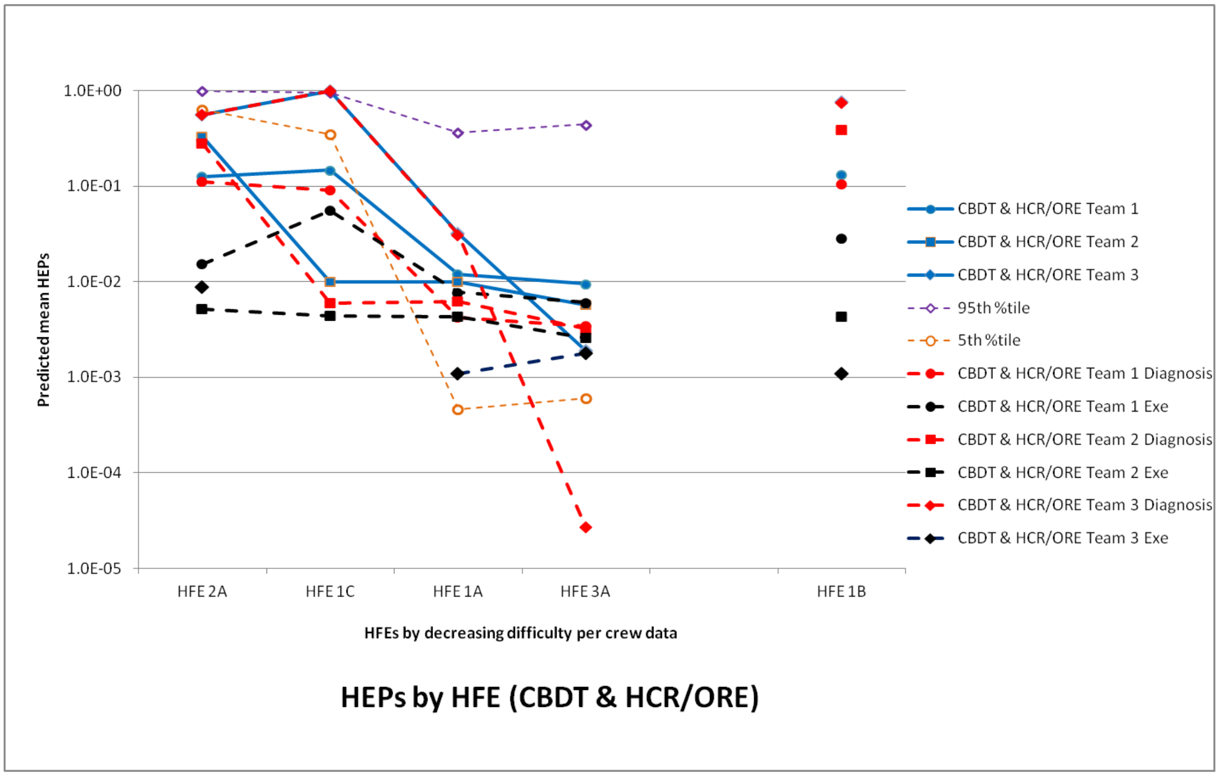



**Figure 7-4 Diagnosis and execution HEPs of CBDT & HCR/ORE teams**

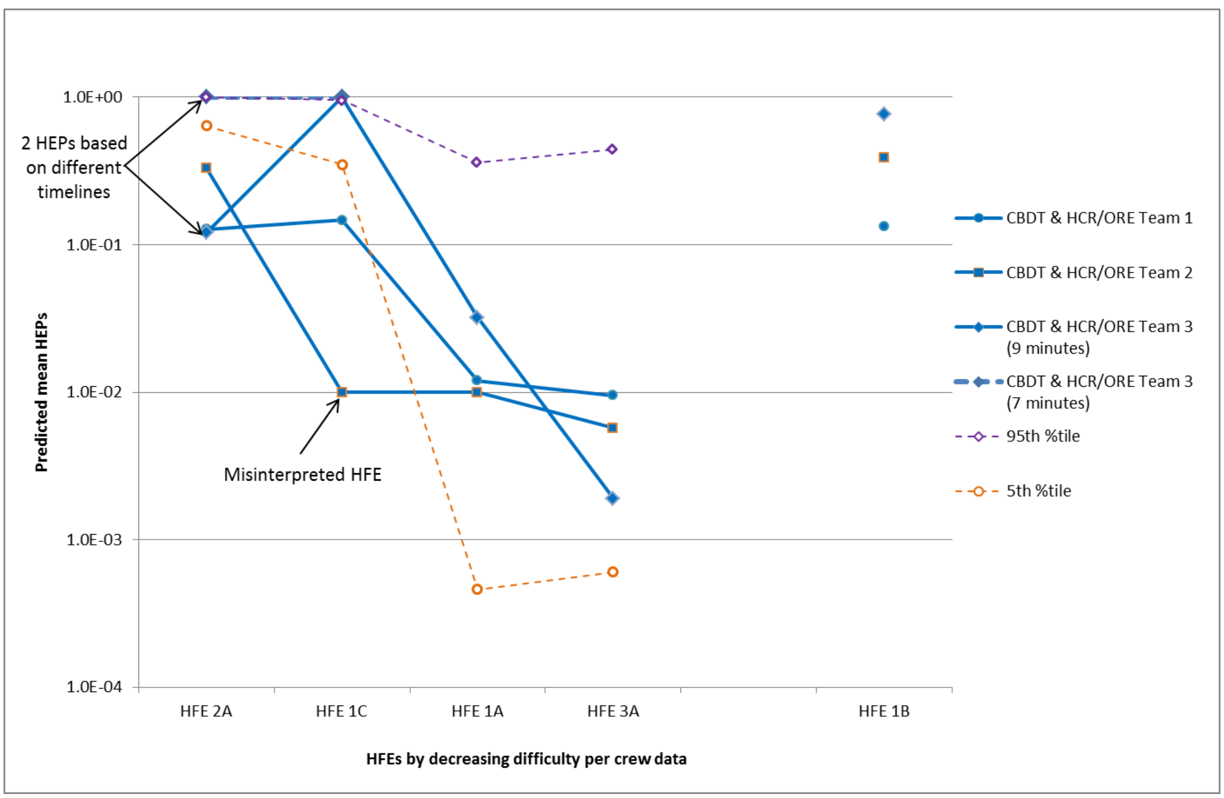



**Figure 7-5 Annotated HEP plot of CBDT & HCR/ORE teams**

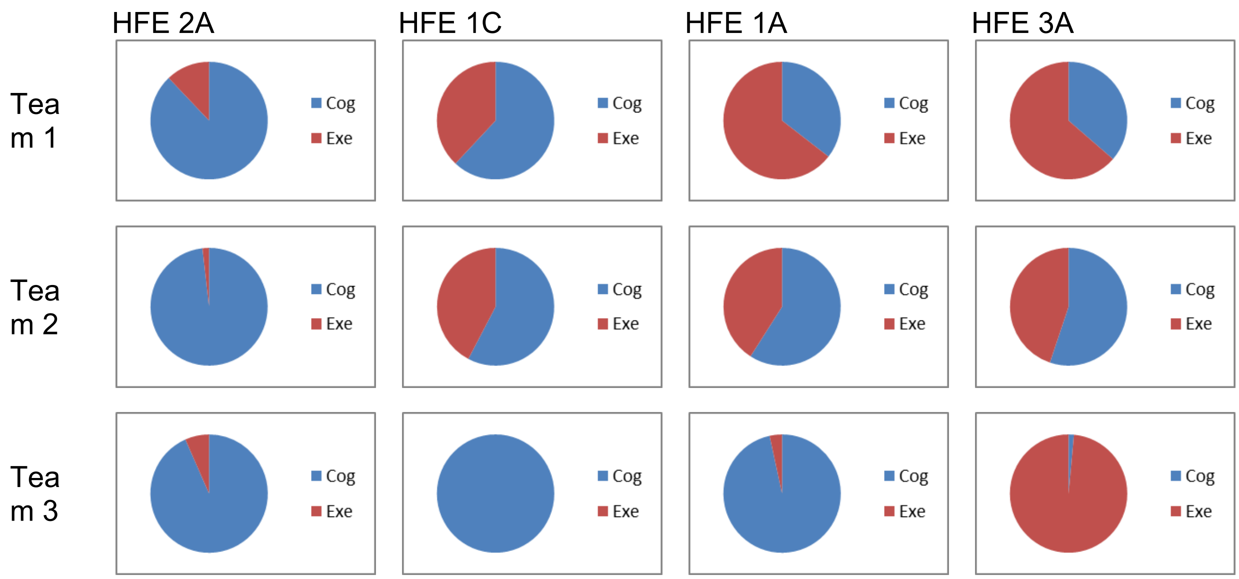



**Figure 7-6 Contributions of diagnosis and execution HEPs of CBDT & HCR/ORE teams**

As seen with some other methods, the more difficult HFEs produced less consistent results. HFE 1C was least consistent across the three teams because it had the least defined procedural path, as evident by the crew performance (most crew variability in the empirical data). In predicting the performance drivers, the drivers often matched the empirical data at a high level, but the reasoning did not consistently match what happened in the empirical data. For example, procedural guidance came up several times as a negative driver in the empirical data. In these cases, some of the analyses would also find procedural guidance as a negative driver. However, the empirical data refers to difficulty with the content and progression of the procedures, while the analyses found procedural guidance to be a negative factor based on the format of the procedure (e.g., existence of NOT or AND/OR statements), which was not supported by the data (at least as based on the trials in the study). Also, the magnitude of the effect of drivers was not always consistent between what was discussed in the qualitative analysis and what was manifest in the quantitative analysis.

Inconsistencies in the qualitative analysis drove some of the differences in the quantitative analysis. However, there was significant variability in the quality of the documentation available for the qualitative analysis, making it difficult to pinpoint the root cause of inconsistencies between teams in some cases. The operational stories were not always detailed enough to understand the expected progression of the scenarios (e.g., the expected procedural path, critical decision point, cues and other activity). Some teams gave a full description of the scenario, while others provided only the decision tree paths through CBDT, with minimal qualitative assessments. Overall, however, the operational stories were generally consistent across teams. The most variation was seen for HFEs where multiple procedural paths or knowledge-based success paths existed; the empirical data show much crew variability in these cases as well.

One major source of inconsistency was the assessed timing associated with each HFE. The timing estimate can be the most subjective portion of the EPRI HRA Calculator. Table 7-2 below gives a summary of the timelines from the analyses to the actual crew performance times or available time for successful response. Team 1's documentation did not include timelines for each HFE. There were large variations from team to team and between the teams' estimates and the empirical data, with the teams' assessed timing fairly consistently optimistic compared to the data. Differences in timing estimates came from differences in the teams' underlying assumptions and what was included in the estimate (e.g., steps, cues, factors that might elongate timing). Also, the analysts based their timing estimates on information from different sources; one team relied primarily on analyst judgment while the other team used input directly from the interviewed trainers/operators. It is important to note that, for time sensitive HFEs, small differences in timing can lead to large differences in HEP. For example, in HFE 2A, using a 7-minute timeframe vs. a 9-minute timeframe yields a difference between $1.0 \times 10^{-2}$ and 1.0 HEP (see Figure 7-5).

**Table 7-2 Comparison of Timing Estimates for Each HFE**

| HFE | Team 1 | Team 2 | Team 3 | Crew Response Times from Empirical Data |
|---|---|---|---|---|
| HFE 1A – Establish F&B | N/A | 21 min | 13 min | 10 – 17 min |
| HFE 1C – Isolate SG | N/A | 15 min | >40 min | 34 min to >1hr |
| HFE 2A – Trip RCP and start PDP | N/A | 6 min | 8.5 min | > 7.5 – 8.5 min* |
| HFE 3A – Isolate SG | N/A | 12 min | 10 min | 14 – 21 min |

* Because no crews started the PDPs in time, the timing here reflects the timing window for successful response, i.e., the time at which the RCP seal temperature exceeded 230 °F.

Although clear method guidance is given on the definitions of the relevant timing aspects that HRA analysts should consider (e.g., T_delay, etc.), analyst judgment is often needed to estimate these parameters. No specific guidance exists on how to obtain estimates of key values such as median response time when simulator data cannot be obtained. For many teams, the timing seemed to be based strictly on estimates of the time it would take to get cues and get through the procedure steps. The teams did not explicitly account for elongation of the time frame caused by distractions, parallel actions, etc. Cases with high complexity and limited time (e.g., HFEs 1C and 2A) do not seem to be well accounted for; the HCR/ORE method takes time into consideration, and other PSFs (e.g., complexity) are supposed to be reflected in the time estimates and sigma values, which can be difficult to make. The CBDT method accounts for complexity explicitly, with the prerequisite that time is sufficient.

The teams also varied significantly in how they conducted the operator interviews. The differences included variations in the questions asked, the depth and extent of their follow-up probes, the information they gave to the trainer/operator, and whether they asked for a step-by-step discussion of the relevant procedures. In general, the HEPs produced by teams that conducted more extensive interviews more closely matched the empirical data.

The EPRI HRA Calculator comprises three methods: CBDT branch points, HCR/ORE formula and THERP tables; dependency and recovery factors are then applied based on pre-defined rules. This structured approach offers a clear way to trace the relation between the input values and the final HEP. Although the EPRI HRA Calculator provides fields to capture some elements of the qualitative analysis, it does not give a specific format for the qualitative analysis other than the direct questions asked. Thus, the area of least traceability is the operational story, cues and timing analysis. Consequently, the analysts' bases for each input value (e.g., timing, branch points, recovery designations) is not always discernible – it depends on the quality of the analysts' documentation. For HCR/ORE, lack of traceability in developing the timeline translates directly into a lack of traceability in the final HEP.

Inconsistencies observed in the quantitative analyses stem from three sources: philosophy of method application/analyst judgment, division of the HFE and timing estimates (for HCR/ORE, discussed above).

In philosophy of method application there are two areas where differences creep in: conservatism vs. best estimate and cookbook application of method (verbatim) vs. analyst judgment. For example, Team 3 applied recovery to the cognitive and execution portions as per the method. Although Team 1 noted that the method allows for crediting cognitive recovery, they usually use the recovery factors for the cognitive part only in very specific situations, therefore, they only considered recovery for execution. Team 2 "conservatively" did not credit recovery. Similarly, in HFE 1C, both Teams 1 and 2 acknowledged the masking effect in the qualitative analysis. However, in the quantitative analysis, Team 2 chose the "accurate indications" branch (leading to a negligible contribution to the HEP) while Team 1 chose the "indications inaccurate" branch (leading to a $6X10^{-2}$ contribution to the HEP). Team 2 decided that the operators are monitoring the SG levels closely and, though delayed, the indications were technically correct. Team 1 stated they chose "indications inaccurate" to account for the masking effect and the fact that no alternate information sources were given in the procedures. Other less quantitatively significant versions of differences in analyst judgment were peppered throughout the analyses. For example, in the workload for HFE 1C, one team chose high, one team chose low, and the third team stretched the method and applied a medium workload. There were also several instances where there were differences in interpretation of the procedure logic category (e.g., one team applied the procedure logic questions to the entire procedure while the other team only applied those questions to the critical steps).

Another major source of inconsistency was in the division of the HFE for quantification. Nearly all of the HFEs were broken up differently between teams. For example, one team defined multiple cognitive failure mechanisms – each quantified separately – for each of HFEs 1A, 2A and 3A, while the other two teams quantified each of these HFEs as a single failure. Similarly, the teams quantified different execution steps. For HFE 3A, one team quantified 14 separate execution steps while another team quantified only two. In the quantification of the execution steps there was also a difference between the two teams which used THERP: for each execution step quantified, one team applied either the Error of Commission (EOC) or the Error of Omission (EOO), while the other team applied both. There is little guidance on how or when to break up HFE into subcomponents (e.g., different cognitive failure mechanisms).

Finally, harder HFEs yielded more variable HEPs which were less consistent with the empirical data, generally underpredicting the HEP. The basic assumption behind CBDT is that operators will follow procedures and generally trust their cues. That means that little to no credit is given for actions when there is a weak link between the procedure and the expected actions unless those actions are considered skill-of-craft. In cases where a weak procedural link existed (both HFE 1C and 2A exhibited these types of actions), significant analyst judgment had to be applied to quantify these cases and make the trees "fit" their understanding of the operational story. One example is HFE 1C, where the crew would not necessarily be in the appropriate procedure for SGTR in the right time frame (i.e., before they exit FR-H1FR-H1), but are allowed to use other procedures in parallel, if it does not distract from doing the higher priority procedure. Furthermore, for these harder HFEs, there are other factors that are not well accounted for by the method. For instance, in CBDT, procedural guidance only covers procedure format and whether or not procedure matches with the cues; the clarity of the workflow of the procedures is not addressed. Similarly, substantial complexity and/or teamwork are not dealt with crisply in CBDT and the mismatch between training and scenario (e.g., prioritization of starting the PDP in

HFE 2A) is not evaluated except in very specific circumstances (e.g., the tree for 'cue not as stated').

## 7.3 SPAR-H

The two SPAR-H [6] teams (SPAR-H Team 1 and Team 2) employed different approaches to the qualitative analysis. Team 1 used a holistic approach and treated HFEs as single tasks or failures without decomposing the HFEs; therefore, the PSFs and associated multipliers were evaluated at the HFE level. In contrast, Team 2 first used CRTs (Crew Response Trees) to determine the decision points and basic events in the scenarios based on the branching points and steps in the procedures. They then evaluated PSFs for individual subtasks and summed up the results. The CRTs are a qualitative analysis tool being developed for a hybrid HRA method called Integrated Human Event Analysis System (IDHEAS)[16]. The SPAR-H method focuses primarily on quantification and references other documents for guidance on doing qualitative analysis depending on the level of detail desired. As such, it does not provide much insight or detail on "conventional" modeling considerations, such as the level of decomposition of a scenario or how subtasks should be combined for analysis, as the method assumes that these are situation- and analyst-specific. The method states that the analyst may break the scenario into subtasks for analysis, and suggests that the ATHEANA decomposition and qualitative process could be used, but other approaches could be used as well. Therefore, the choices of both teams on how to do the qualitative analysis are in general accordance with the SPAR-H method.

Compared to Team 1, Team 2's use of CRTs allowed them to find detailed failure and success paths and provide greater qualitative insights into possible operational stories and performance drivers (e.g., limited time available in HFE 1C). Hence, Team 2's analysis would be expected to yield a better basis for error reduction. In addition, Team 2's qualitative analysis appeared to benefit from one of the team members' experience in nuclear operations, especially for the complex scenarios (HFEs 2A and 1C). However, this superior qualitative analysis was not reflected in their quantitative analysis. The main reason for this is not easy to determine when only given the final results of the analysis. Nonetheless, it seems that a better quantification could have been obtained with a thorough reasonableness check. Given the thoroughness of the qualitative analysis, such a check should have identified inconsistencies between the qualitative analysis and some of the choices of PSF multipliers. Especially for HFE 3A, which is a well-trained standard scenario, the justification for some choices of PSF multipliers do not seem to be reasonable.

The differences in the analysis of the procedures between the two teams can be illustrated by how the two teams addressed transitions in procedures. Team 1 did not discuss transitions in procedures and where exactly in the procedures the complications in the HFEs would have an effect. Although they gave some details about how the scenarios could evolve, the discussion seemed to be based on a good insight, partly a result of the team's HRA experience, about how the scenarios could progress. In contrast, transitions in the procedures were a significant focus of Team 2's analysis and served as a basis to decompose HFEs into basic events in CRTs.

Team 1 was optimistic on the most difficult HFE (HFE 2A); it appears that the optimism was based on their interpretations of the information (e.g., timing and complexity) they obtained from interviews with the trainers. However, it is apparent that they could have done a better qualitative analysis had they more thoroughly examined related factors including procedure branching and timing, as well as had they done a more detailed interview with the instructors. For example, Team 1 applied a nominal multiplier to the time PSF for HFE 2A as well as

HFEs 1C and 1B. Such an application is not well justified given the complexity of the HFEs and the time available; they could have obtained better results with a more careful timing analysis rather than just relying on trainers' opinions. These findings suggest that an HRA method should give guidance on analyzing scenarios at an appropriate level of detail and how to conduct interviews with plant personnel to support the identification of issues related to the complexity of the situation as well as timing issues associated with procedure following. Team 1 documented a good understanding of HFEs 1A and 1C, which contributed to good predictions of these HFEs. For the easy HFE, HFE 3A, Team 1's documentation indicated that they had done a rather high-level analysis. Nevertheless, this analysis seemed to be sufficient to make a fairly good assessment of the qualitative drivers. This finding indicates that a detailed qualitative analysis capturing the nature of the scenarios is more essential for the more difficult scenarios.

SPAR-H has a set of pre-defined PSFs and multipliers associated with different levels of effect. As with other methods, there is overlap in the dimensions used to characterize the PSFs. Lack of clear delineation of how to select a PSF as well as how to avoid double counting of a certain factor allows different analysts to account for a factor under different PSFs although their understandings of the factor may be quite similar. In the current study, Team 1 accounted for the misleading AFW flow in HFE 1A under complexity, while Team 2 accounted for the misleading AFW under human-machine interface (HMI): missing/misleading indicator. Both choices seem to be reasonable; however, the maximum HMI multiplier is 50 while the maximum complexity multiplier is 5. Thus, Team 2 used a much larger multiplier to account for misleading indicators than Team 1. Also although Team 1 construed the missing indication of the recirculation valve in the control room as an HMI issue, they only applied a multiplier of 10 to avoid double counting the misleading AFW flow, which was accounted for under complexity.

Furthermore, Team 2's analysis suggests that the range of PSFs offered by SPAR-H is not adequate to cover all scenario dynamics. For example, the team indicated that in HFE 1A, "[a] mitigating factor not specifically addressed by SPAR-H is that the cues for AFW flow are being looked at by multiple individuals. The STA typically makes the critical safety function status trees (CSFSTs) and the reactor operators monitor the main control board panels. Since there are two separate individuals with different backgrounds looking at the same data, it is more likely that the indication will be questioned. This 'back-up' or 'oversight' is not specifically credited by SPAR-H but does appear in other models." The mitigating factor can probably be included as part of the work processes PSF in SPAR-H though. This PSF is kind of a "catch-all" PSF in SPAR-H, and if these issues should be included, the SPAR-H guidance should be updated with more details. As noted elsewhere, mapping the qualitative analysis results to PSFs and choosing between PSF levels is a difficult process in all PSF-based methods. The guidance for how to do this in SPAR-H could be improved with additional guidance, examples and reference cases.

The quantification process of SPAR-H is very transparent and is highly traceable, given the PSF multipliers. However the traceability of the choice of a multiplier for a particular PSF depends on the documentation of the identified complications of HFEs in operational terms (especially for complexity issues). The HEP estimates by the two SPAR-H teams are presented in Figure 7.1c. In general, Team 1's HEP estimates for the scenarios reflect the empirical difficulty ranking, except for the most difficult HFE. In addition, there are substantial differences between the HEP estimates for the different HFEs. However, it is interesting to note that despite the detailed qualitative analysis, Team 2's HEP estimates do not really distinguish among the HFEs, and do not follow the difficulty ranking based on the empirical data. If we look at the diagnosis and execution HEPs separately (Figure 7-7), Team 1's diagnosis and execution HEPs are comparable for HFEs 1C, 1A and 3A. Both diagnosis and execution HEPs exhibit the same trend and level of differentiation as the total HEPs. Team 1's execution HEP for HFE 3A is relatively optimistic compared to most of the other teams. This result may be partly because the team applied a multiplier of 0.1 to the time PSF for both diagnosis and execution.

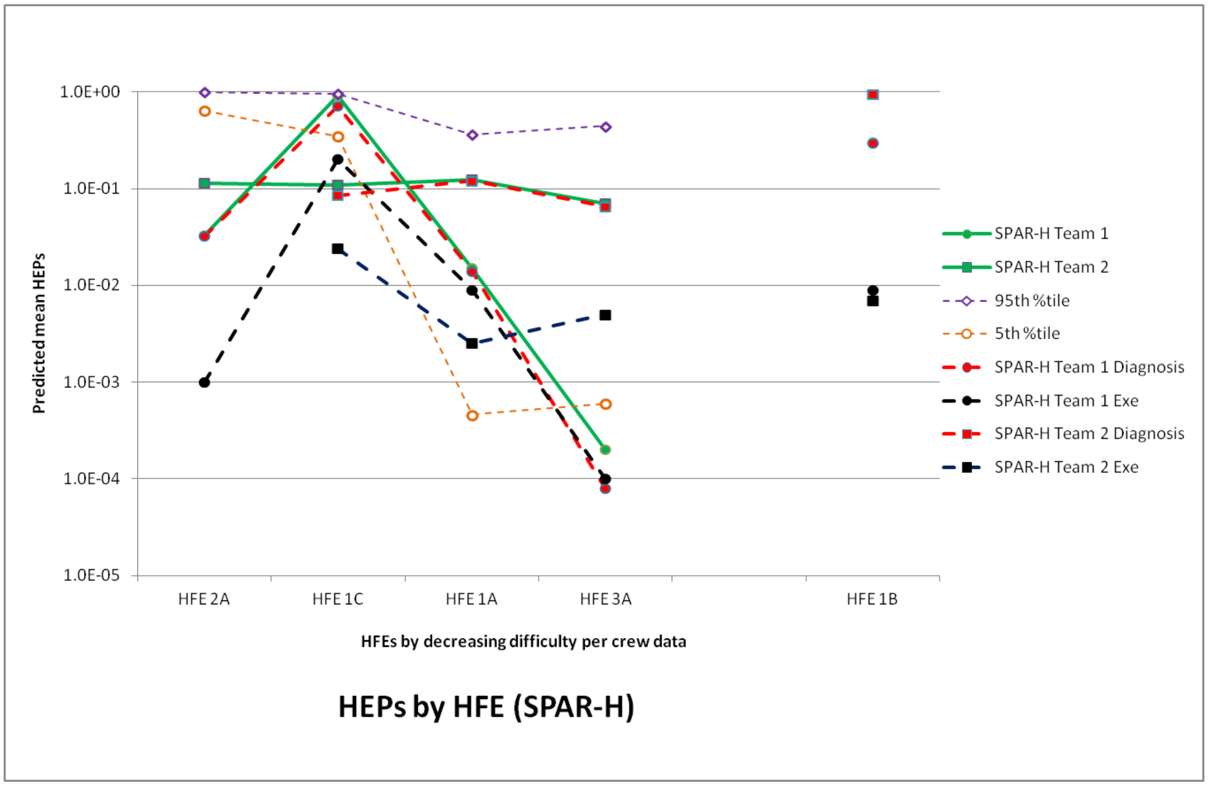



**Figure 7-7 Diagnosis and execution HEPs of SPAR-H teams**

Figure 7-8 shows the relative contributions of the diagnosis and execution HEPs to the total HEPs produced by the SPAR-H teams. As shown in this figure, Team 2's total HEPs are dominated by diagnosis HEPs for most of the HFEs (it is difficult to separate diagnosis and execution HEPs for HFE 2A because dependence between basic events was considered). Furthermore, Team 2's diagnosis HEPs are at least 1 order of magnitude larger than their execution HEPs for HFEs 1A and 3A.

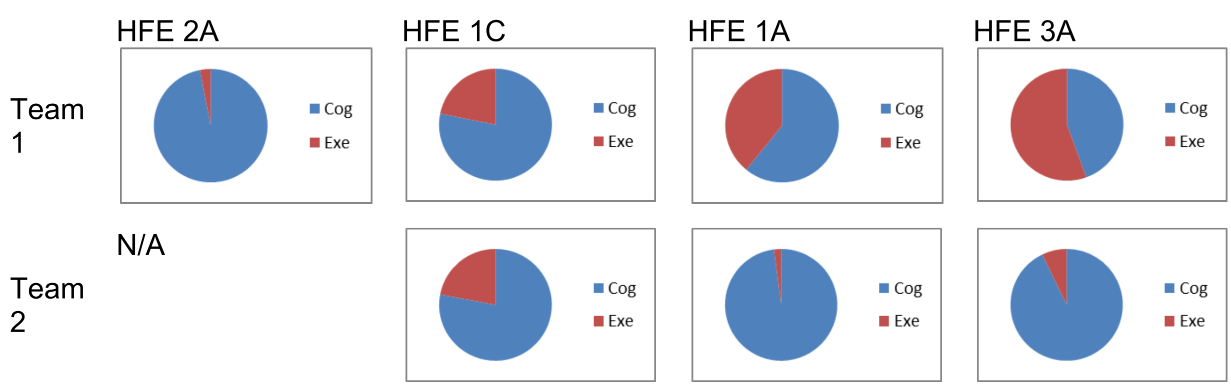



**Figure 7-8 Contributions of diagnosis and execution HEPs of SPAR-H teams**

In both HFEs 1C and 3A, Team 2 applied the highest multiplier for diagnosis complexity (a factor of 5). This treatment is well justified for HFE 1C given the complicating factors of the scenario. However, this treatment does not seem to be well justified for HFE 3A, because it an often trained HFE within a standard SGTR scenario. In contrast, Team 1 applied multipliers of 0.1 and 1 to the complexity PSF for diagnosis and execution, respectively, leading to a much lower HEP for HFE 3A. In addition, Team 2 selected nominal time for HFE 3A, although they recognized that there was sufficient time, which could have led to a lower multiplier. The selection of the nominal time for HFE 3A does not seem to be consistent with their evaluation of HFE 1A, where the team selected a lower multiplier for the time PSF to credit sufficient time. This inconsistency appears to be another reason for the high diagnosis HEP of HFE 3A. The use of high multipliers caused the diagnosis HEP for HFE 3A to be much larger than those obtained by other analyst teams for the same HFE. A reasonableness check of the relevant HEPs obtained, comparing the PSF multiplier choices and the explanations in the qualitative analysis, could have helped the analysts to realize that the estimates obtained did not reflect the level of difficulty of the HFEs that they identified during the qualitative analysis.

## 7.4 ATHEANA

Two teams used ATHEANA in this study (ATHEANA Team 1 and Team 2). As shown in Figure 7.1d, both teams' HEPs for the most difficult HFEs (HFEs 2A and 1C) appear to be underpredicted (optimistic) relative to the reference HEPs based on the crews' responses in the simulator scenarios. With the exception of Team 2's evaluation of HFE 2A, the HEPs of both teams are consistent with the difficulty ranking with significant differentiation among the HFEs. However, Team 1's estimates are consistently nearly an order of magnitude lower than Team 2's estimates for those HFEs. For HFEs 2A and 1C, Team 1's optimistic HEPs may be partly explained by their relatively optimistic timing estimates compared to the empirical data. For example, the team assumed that Procedure FR.P-1 in HFE 1C would guide the crews to respond to the SGTR without going through E-0 or E-30. Thus, they concluded that the 40-minute time window should be sufficient for most procedural paths once the crews entered FR.P-1. In addition, the team assumed that crews who failed to dispatch a plant operator to check the AFW recirculation valve would establish AFW to SGs per guidance in FR-H1. As a result, those crews would have relatively more time to respond to the SGTR because the crews would have completed more procedural steps in FR-H1 by the time the SGTR was initiated. However, this assumption does not seem to be supported by the crew data. The crew (Crew T) that did not check the AFW recirculation valve established AFW by cross-connection based on their knowledge.

Many discrepancies between the two teams seem to arise from differences in their application of the method. It appears that although much of the guidance was effectively used and closely followed by Team 2, they omitted or did some elements in less detail than the method guidance would suggest. This, to some extent, may be related to one weakness of the method. ATHEANA has a heavy emphasis on a good qualitative analysis and the full implementation of the method can be resource intensive. Team 2 simplified their analysis based on their experience of the tradeoffs identified in other applications of the method. Compared to Team 2, Team 1 took a more detailed approach at the analysis. Team 1 used 250 man-hours to perform the analysis relative to the 90 man-hours spent by Team 2. Nonetheless, it could be argued that Team 1 went beyond the basic ATHEANA guidance, while Team 2 still did a generally good and reasonable analysis.

The two teams' results differed in the following aspects:

- Scenario decomposition

The ATHEANA method requires analysts to define the expected scenario progression and then search for deviation scenarios and factors which could produce an Error Forcing Context (EFC) leading to Unsafe Actions (UAs) that can ultimately lead to failure in that scenario. For each scenario, Team 1 developed an expected response as well as numerous alternative procedural paths and sequences of actions in the performance of the HFE task. These alternatives are based on variations in the duration of operator subtasks or in the sequence of these subtasks, which, in turn, affect the plant response and the timing of cues. Team 2 did not explore alternative paths (which is not explicitly recommended in the ATHEANA guidance), which seemed to contribute to the team's failure to capture some of the main negative drivers for the most difficult HFE (HFE 2A). For this study, the EFCs were essentially pre-defined by the scenarios that were run.

Although Team 2 decomposed some of the HFEs into UAs, they did not attempt to quantify the UAs separately. Instead, they estimated probabilities for overall UAs corresponding to the HFEs themselves. As a result, it is more difficult to see where Team 2's estimates were derived from judgment and the bases for their judgments. However, it is not clear from the available information whether or to what extent the qualitative analysis and quantitative results might have been different if Team 2 had sufficient time to apply the full ATHEANA process described in the guidance.

- Quantification approach

  Team 1 developed a distribution from expert elicitation for each branch probability and each important time interval in scenario maps. The distributions were then used as inputs to Monte Carlo simulation to calculate branch probabilities, timing estimates, and associated uncertainty. This Monte Carlo treatment of the paths and their duration is to some degree an innovation relative to previously documented ATHEANA applications.

  Team 2's quantification process appeared to have been somewhat less formal, with each of three experts providing his or her inputs in a more holistic framework for the HFE. Moreover, Team 2's analysis stopped with the definition of the initial three points of a distribution (1st percentile, 99th percentile and most likely value), without following through further developing the distribution, which is not inconsistent with the ATHEANA guidance.

- PSF evaluation

  Team 1 considered a broad range of factors in estimating the distributions for the parameters such as timing and branch probabilities. The factors were not used as an explicit direct input to quantification. Team 2 closely followed the method guidance in PSF evaluation, using the PSFs proposed by the method.

- Timing analysis

  Both teams conducted a detailed timing analysis. Compared to Team 2, Team 1's results seem to be more sensitive to timing analysis or more time-oriented. Their analysis appeared much more focused on whether sufficient time was available given different conditions, and nearly all the dominant failures stemmed from finding a context where there was not enough time. For cases where operators are expected to carry out all anticipated actions in the specified time windows, HEPs are mainly determined by timing. In two instances, where the distribution for time required was always less than the time available, Team 1 developed a separate distribution to account for failures caused by "unexplained reasons." In contrast, Team 2's approach did not exhibit the same rigor in accounting for delays in the timeline, and only produced point-estimates in their timing estimations (vs. Team 1's distributions for both branch probabilities and action timing parameters). Moreover, it is difficult to tell how the Team 2's timing analyses influenced their judgment of the distribution of the HEP for the HFEs.

- Reasonableness check

  A "reasonableness check" is an important step in the HRA process. After all the HFEs were quantified, Team 1 made a reasonableness check to ensure the numerical ranking of the HFEs were consistent with the analysts' understanding of the relative degree of difficulty of the HFEs. It is interesting to note that the team had originally expected the HEP of HFE 1C to be higher than that of HFE 2A, but they changed their expectation when the HEP of HFE 1C that they estimated was actually lower than that of HFE 2A. This is understandable as the team was not quite sure which HFE was more difficult, given that HFEs 1C and 2A are both challenging. However, it suggests that analysts may tend to trust HEPs in a reasonableness check when the difficulty ranking is not obvious.

  In contrast, Team 2 did not make such a check because of time constraints. It is not possible to assess the effects that would have resulted from following the available guidance more explicitly.

Despite the simplifications mentioned above, Team 2 was generally quite effective in identifying many potential causes for the HFEs in terms of operational expressions. This was the case even for the difficult HFEs where the team was relatively less successful in finding important drivers.

Team 1's approach (using detailed scenario maps) and documentation gave a very clear and traceable link between the qualitative analysis and the resultant HEPs. In contrast, although it was relatively easy to understand Team 2's thought processes behind their qualitative analysis, the translation of the qualitative information into quantitative estimates was somewhat less transparent. ATHEANA, to a large extent, relies on expert judgment. It is challenging to give direct correlations of qualitative inputs to expert judgment estimates, as expert judgment depends in great part on the individual's background and experience. This challenge generally makes it inherently more difficult to reproduce quantitative results, and may have been magnified somewhat in Team 2's analysis because they were not able to follow the process as fully as the ATHEANA guidance called for. For the most difficult HFE (HFE 2A), Team 2 seemed to be biased by their experience and substantially underestimated the effect of the negative drivers. As discussed above, the bias may have been tempered had they considered possible deviation scenarios.

Another weakness of ATHEANA illustrated in the study was inadequate guidance to drive quantitative analysis with qualitative analysis. For Team 2, PSF rankings were used to guide their judgments, but they were not explicitly factored into quantitative results. For the most difficult HFE (HFE 2A), the team placed insufficient emphasis in the quantitative analysis on the qualitative factors identified as potential human error drivers.

## 7.5 Summary of Contributors to Predictive Differences Identified from Intra-Method Comparisons

The intra-method comparisons identified some factors and HRA-related issues that contributed to variability in the intra-method predictive results (and across methods in many cases) in terms of qualitative and quantitative analyses and in terms of several general HRA practices. The contributors included both method-driven factors and analyst-driven factors and their interaction, as follows:

- different approaches to qualitative analysis and different levels of qualitative analysis, including going beyond the method guidance (e.g., one ATHEANA team choosing to consider a range of failure paths—an activity that goes beyond the method)
- different approaches to task decomposition (including both task analysis breakdown and task execution elements)
- approach used for estimating the time required to perform the actions and treatment of the complexities associated with making the estimations
- difficulty in understanding and treating complexity (mainly a method effect or lack of guidance)
- use of different models to quantify execution as allowed by the method
- credit for recovery—use of different criteria, sometimes driven by desire for conservatism
- attempts to make up for inadequate range of PSFs
- differences in deciding which PSF to use to cover a particular effect or influence and what the weight or rating of the factors should be
- inadequate means to tie detailed qualitative analysis to the quantitative approach led to differences between applications of the same method
- compensating for a given method's poor treatment of diagnosis by some analysts
- different interpretations of what the results of interviews with trainers implied
- different approaches for collecting information in interviews
- appropriate use of a reasonableness check of obtained HEPs

# 8. DISCUSSION OF FACTORS CONTRIBUTING TO THE VARIABILITIES OF HRA RESULTS

The International HRA (human reliability analysis) Empirical Study identified that significant variability can occur in the results of applying different HRA methods for the same human failure event (HFE) because of the differences and limitations in the methods' technical and methodological bases [1–4] and in the associated guidance for applying the methods. In addition, although it appeared that the analysts themselves could contribute to differences in results (analyst-driven factors), the design of the International Study was not able to distinguish method effects from analyst effects because there was only one case where two teams used the same method. With at least two teams per method, the present study has information to probe the sources of variabilities in in the HRA results. The differences in the results from the HRA teams applying the same method observed in the present study underscore the need to enhance the guidance for applying methods, but more importantly they demonstrate the need to develop improved methodology and related guidance to address limitations in the methods. This section describes the reasons for the observed differences and the underlying methodological and guidance limitations that permitted the differences to arise.

In addition, a major source of variability within the same method was analyst decisions about how to apply various aspects of the method. Analysts are often called upon to make decisions in their analyses, and the guidance of the HRA methods are not sufficient or specific enough, so that analysts may, to some degree, have to rely on their own interpretation of the guidance. Moreover, the methods sometimes allow analysts to apply different analyses or modeling options without providing clear criteria for when to use the different options. The factors leading to analysts' subjectivity and thus contributing to HRA predictive differences are also discussed below.

## 8.1 <u>Factors Contributing to Differences in Qualitative Analysis</u>

### 8.1.1 Approaches, Scope and Depth of Qualitative Analysis

#### *8.1.1.1 Choosing Different Qualitative Analysis Approaches*

Some methods (e.g., SPAR-H) do not offer specific or complete guidance for doing the qualitative analysis, and thus HRA analysts are left to decide how they will perform the qualitative analysis and its level of detail. For example, standardized plant analysis risk-HRA (SPAR-H) guidance suggests using ATHEANA (a technique for human event analysis) if a detailed qualitative analysis is needed, but leaves analysts free to decide whether and to what extent they use it. Other methods, such as ATHEANA, give detailed guidance for qualitative analysis, but the guidance is somewhat open-ended and not always well-structured for translating the information into human error probabilities (HEPs) in a consistent manner; thus, to some extent leaving the level of detail up to the analysts. In addition, in the present study, one of the Accident Sequence Evaluation Program (ASEP) HRA procedure teams went beyond the method guidance in order to do what they thought was needed for an adequate analysis, suggesting an inadequacy in the ASEP's guidance.

Obviously, analysts may vary the scope and depth of their qualitative analyses. One reason for this variability in the analysis is the degree to which a specific method framework specifies or guides the qualitative analysis process. Some methods (e.g., ATHEANA) guide and support the qualitative analysis more strongly than others and/or are equipped with a relatively wider range of performance shaping factors (PSFs); such methods provide better support to analysts to identify the performance driving factors in a more comprehensive manner and with finer granularity.  Another factor is the analysts' level of effort devoted to their qualitative analyses.  In this study, some analysts took a detailed qualitative analysis because they saw it as a good HRA practice and considered a range of failure paths that went beyond method guidance (e.g., ASEP Team 1 noted above and ATHEANA Team 1).  Others decided to take a more holistic (less decomposition) and streamlined (e.g., SPAR-H Team 1 and ATHEANA Team 2) approach but staying within the specific method guidance.  In general, more detailed qualitative analyses tend to produce richer information about scenario underlying dynamics and crew failure mechanisms, along with a better understanding of what the crews are likely to do given the conditions.  However, it should be noted that in this study there were no clear indications that a more detailed qualitative analysis lead to better quantitative prediction results.

#### *8.1.1.2 Different Approaches to Task Decomposition (Including Both the Cognitive Task Analysis and Task Execution Breakdown*

Most HRA methods do not have a consistent approach or offer much guidance for task analysis and decomposition.  Insufficient task analysis and decomposition, both in terms of the various activities required to understand the scenario conditions and procedural interactions, and the key tasks required to execute the response after diagnosis has been made, may cause analysts to fail to understand the difficulty in a scenario, especially for complicated scenarios, and lead to different groupings of tasks which can affect dependency issues and quantification results.  It can also cause analysts to ignore cognitive activities involved in step-by-step actions (e.g., ASEP).  In addition, the level of task decomposition may affect the application and traceability of the quantitative analysis.  If the task is not decomposed appropriately or the decomposition of the task is not explicit and well-documented, the types of performance issues identified by the analyst for the HFE, the subtasks that they affect, and the quantitative impact of these factors may not be clear.

### 8.1.2 Estimations of Complexities and of Time Required

Variability in the timing analyses results of various HFEs was observed in the study.  This included variability across the teams using the same method and discrepancies between the analyses of the HRA teams and empirical data.

Timing is either explicitly or implicitly considered in various HRA methods; hence the uncertainties and variability in timing analysis can be propagated to the uncertainties and variability in HEPs.  For methods strongly based on time reliability correlations (TRCs) (e.g., ASEP and HCR/ORE (human cognitive reliability and operator reliability experiments)), HEPs are estimated as a function of the time available for crews to respond to accident scenarios and the HFEs of interest.  Considering the characteristics of TRCs, the HEPs are sensitive to timing analysis results, particularly for scenarios that are time critical, that is, small differences in time windows can lead to large differences in HEPs.  For example, for HFE 1A, ASEP Team 1 estimated the time required for execution to obtain 23 minutes as the time available for diagnosis, yielding a nominal diagnosis HEP of 0.0133.  In contrast, ASEP Team 2 obtained 36 minutes for the time available for diagnosis, which leads to a nominal diagnosis HEP of 0.0005.  Thus, a difference of 13 minutes in time window estimates led to a difference of 2 orders of magnitude in HEPs (using the same curve).  (Actually, ASEP Team 2 additionally credited the crews' training to select the lower bound TRC, obtaining a diagnosis HEP of 0.00004).  Although both HEPs can lead to the conclusion that diagnosis failures are unlikely, the differences suggest how HEPs can be sensitive to timing analysis and the importance of obtaining timing estimates that are as realistic as possible.

For some methods that are not TRC-based (e.g., SPAR-H), the availability of time is often considered as a PSF and differences in the rating of this PSF can lead to a variability in HEPs that is as large as that obtained with TRCs. For example, compared to the SPAR-H Team 1, the SPAR-H Team 2 tended to assign larger multipliers for the PSF of time (reflecting less time available) for diagnosis and/or execution HEPs in HFEs 1B, 1C, 2A and 3A. It should be noted that compared to the empirical data, the multipliers assigned to the PSF of time for HFEs 1B, 1C, and 2A by the Team 1 seemed to be optimistic (suggesting more time was available for diagnosis than there actually was), and the multiplier assigned to time for HFE 3A by the Team 2 seemed to be conservative.

The study suggests that variability in analysts' judgments about the time required to perform actions relevant to an HFE is caused, at least to some extent, by inadequate method guidance for estimating the time required. This can be a major source for the variability in timing analysis and the resulting HEPs. Issues related to making appropriate estimates of the time required to complete HFE related actions include the following:

- Analysts need to make decisions on which procedural steps to include in the timing analysis. For example, in some cases although the two teams assumed the same procedures, the ASEP Team 1 thought many of the steps would require operator attention and contribute to the time required, while ASEP Team 2 considered only a few key steps that they thought would influence performance time. As a result, the time available for diagnosis estimated by the ASEP Team 1 tended to be shorter than those estimated by the ASEP Team 2 because they thought reading through the procedures would require more time.

- Some methods do not give specific guidance on assessing the values of timing points, e.g., the median response time in Cause-Based Decision Tree (CBDT) when simulator runs cannot be made. Thus, analysts have to rely on their own judgment or operator opinions. For example, it seems that one factor leading to the differences in HRA Calculator teams' timing analyses is that the teams identified different plant indications as the relevant cues for the operator actions.

- To make accurate assumptions and/or judgments about the time required, factors such as the effect of distractions, delays, and competing task demands need to be considered. For example, in contrast to the ASEP Team 1, the ASEP Team 2 did not consider in HFE 1C the delay in working through FRH1 caused by HFE 1A; as a consequence, their timing analysis was optimistic compared to the empirical data. It can be argued that Team 2's analysis could have been improved with more detailed qualitative analysis, but the method does not provide such guidance to do so. ASEP Team 1 clearly went beyond the guidance offered by ASEP in delineating the context of HFE 1C in connection with HFE 1A in their qualitative analysis, which helped their assessment of the time required.

It also should be noted that the lack of traceability in timing analysis has also been observed. As mentioned above, analysts often rely on their own judgment and operator opinions (from interview) in their timing analysis. Without good documentation, it is difficult to understand the basis of the timing analysis.

### 8.1.3 Difficulty in Understanding and Treating Complexity (Method Effect or Lack of Guidance?)

Since the qualitative analysis of a given method is limited to the human performance issues addressed by the method, it can be difficult for analysts to characterize and represent the issues in complex scenarios that are beyond the scope of the method. In general, it appeared that the methods in the study had limitations for dealing with complexity and even though ATHEANA gives guidance for addressing complexity, there are some aspects that are still left to the analysts. While analysts can always go beyond the methods by doing a qualitative analysis for the HFE that is broader than the scope of the factors explicitly treated by a given method, such an extended analysis is subject to variability because the analysts will determine the appropriate scope of PSFs without a common basis for this judgment. Secondly, the results of this extended analysis may not be easily incorporated into the quantification model of a given method. Finally, an extended qualitative analysis in itself may not yield improved quantitative predictions, unless the qualitative and quantitative elements of the HRA method are closely coupled and the qualitative analysis is done appropriately.

In many methods, an HFE is basically decomposed into an assessment or decision component and an execution/implementation component. This approach may in some cases make it difficult to address complex HFEs with multiple assessment/decision and execution/implementation subtasks, taking place over a significant evolution of the scenario.

For complex HFEs, the narrative-based HRA methods such as ATHEANA do appear to present some advantages. Narrative-based methods usually do not try to address the HFE based on decomposition. Instead, they give attention to the unfolding of the scenario, to the evolving perspective of the operating crew, and to how the task is performed in operational terms, since these elements provide the basis for the context-based failure narratives on which their quantification is based. Other methods that examine how each HFE subtask may fail as a function of crew performance and the scenario evolution, such as CBDT, also have some capacity to capture and model context-related (in contrast to task-based) failure mechanisms and their consequences on the HFE failure probability. However, CBDT did not seem to address a broad enough set of factors that contribute to failure to adequately address complexity in this study and probably more generally (see discussion in Section 8.2.3).

For HRA methods in which decomposition is fundamental to the quantification model, the qualitative analysis could be strengthened by giving more attention to identifying how the operator actions (the set of tasks related to a given HFE) are embedded into the scenario operationally, including the interaction with other ongoing tasks. It is also important to identify the specific assessment/decision subtasks and execution/implementation subtasks over time, before attempting to decompose the operator action into its assessment/decision component(s) and its execution component(s).

## 8.2 <u>Factors Contributing to Variabilioty in Quantitative Results</u>

### 8.2.1 Use of Different Models to Quantify Execution

One source of variability in quantitative predictions is the differences in the models and data to quantify execution. For example, ASEP allows analysts to choose either ASEP or Technique for Human Error Rate Prediction (THERP) for execution quantification. The two methods have different models for execution. For instance, unlike THERP, ASEP treats all post-diagnosis actions (they have entered the appropriate procedure) the same regardless of whether they are following through procedures and assessing cues and making decisions or executing the action. As another example, of the three CBDT & HCR/ORE teams, Team 1 used ASEP to quantify execution (modelers choice apparently), while the other two teams used THERP per instructions of the HRA Calculator method. The choices of the quantification approaches can affect the consistency in HRA results across analysts using a given method, if the method includes optional approaches without clear criteria for when these should be used.

### 8.2.2 Judgments about When To Credit Recovery Influenced by Different Criteria and Sometimes Driven by Desire for Conservatism

Although some methods offer an option to consider recovery of an HFE, it is the analysts' decisions whether or not to consider it in an HRA application and whether the criteria are adequately met. The variability in these decisions may lead to variability in HRA results. For example, for the CBDT & HCR/ORE method, Team 3 considered recovery, while Teams 1 and 2 did not. As a result, the HEPs estimated by Team 3 tended to be relatively more optimistic than those by the other two teams. This may have reflected a preference toward a more conservative result on the part of Teams 1 and 2 rather than an attempt to obtain a more realistic assessment of the likelihood of recovery since the same criteria appeared to be adequate for Team 3. However, whether a preference for conservatism influenced the choice was not clear from the data. It may also have reflected differences in the timing estimates related to whether there was adequate time for recovery since guidelines for taking credit for recovery are relatively clear in CBDT. Alternatively, this finding may indicate that more guidance and better defined criteria are needed to help analysts make decisions with respect to crediting recovery to facilitate more consistent results.

More generally, different biases between analysts toward obtaining conservative results can obviously lead to variability in results. This type of variability is not necessarily a problem as long as the bases for the decisions are made clear and analysts have the opportunity to follow up on the estimates and do more detailed and realistic analysis for important contributors.

### 8.2.3 Attempting to Make Up for Inadequate Range of PSFs

It has been observed in the study that the PSF-based methods in some cases had difficulties in coping well with some HFE-specific performance issues and driving factors identified in the crew data simply because they were not addressed by the method. In some instances the correct issues were identified by the teams in their operational stories and in trying to compensate for such limitations, some analysts stretched the method (e.g., ASEP and CBDT) to fit the situation based on their experience, which led to quantitative differences in HRA results between the applications. This finding suggests that to be able to reliably predict performance, HRA methods need to cover an appropriate range of PSFs. Nevertheless, as was also shown and discussed in the International HRA Empirical Study [1-4]. This study shows that the PSF-based methods sometimes produced reasonable HEPs without identifying all relevant PSFs, particularly for the easy HFEs. It could not be determined whether this reflects an inherent characteristic of the methods or whether it was just a coincidental effect. In addition, both this study and the International Study seem to suggest that narrative-based methods (e.g., ATHEANA), which attempt to address a wide range of contextual factors, are relatively better equipped to address the full range of conditions associated with difficult HFEs and complex scenarios. While the present study was not able to resolve this issue, it does seem that it would be a good question to address in future HRA empirical studies: i.e., can methods using a key subset of factors, a corresponding qualitative analysis, and a dovetailing quantification process produce reliable and reasonable HEPs for most scenarios? Of course, a single method that adequately gives guidance for covering the full range of conditions in a relatively straightforward manner and consistently produces reasonable HEPs would be the ideal.

### 8.2.4 Deciding Which PSF To Use for Particular Effects and Rating the Factors

For some methods (e.g., SPAR-H, ASEP and CBDT), judgment about the relevance of a particular factor and the specific level of that factor in a given scenario must be made, and for others (e.g., ATHEANA) the analyst must determine what factors are present and characterize them, including the strength of their effects. Overlap in the definitions of the pre-defined PSFs and inadequate guidance on determining the level or strength of a PSF can cause observable variations in analysts' interpretation of the scope of the PSFs and in the ratings assigned to the PSF for a given issue or performance condition.

For example, there were a couple of cases in the application of CBDT where analysts seemed to make different judgments on whether the information/indications were accurate, when the information/indications were available and technically correct, but misleading. This led them to select different branches in CBDT decision tree (DT) Pca (availability of information) and DT Pcd (information misleading), where they were asked if the indication was accurate or all cues were as stated. For HFE 1A, both Team 2 and Team 3 used CBDT to estimate the diagnosis HEP (Team 1 was not compared here because of their misunderstanding on the effects of the auxiliary feedwater (AFW) flow indication), but Team 3's estimate of the diagnosis portion of the HEP is approximately 1 order of magnitude larger than that of Team 2 despite Team 3 having considered recovery (see Figure 8.3). This higher HEP seems to be partly caused by Team 3 accounting for the effects of the AFW flow indication and associated procedural inadequacy in both Pca and Pcd. In contrast, Team 2 accounted for the effects of the indication only in Pcd. It should be noted that Team 2 did not consider the inadequacy in procedures. The AFW flow indication functioned properly and was technically correct (i.e., the flow instrument correctly measured the flow at the location of the instrument). However, in the context of a misaligned recirculation valve, it was inappropriate to use the flow indication to reflect the condition downstream of the recirculation valve. Hence, it would be appropriate to say that the flow indication was misleading rather than incorrect. The observation seems to suggest the differences in analysts' interpretations of method guidance, which may cause conservatism in HEPs.

As another example, in HFE 1C, while Team 2 acknowledged that the masked SG and radiation levels (caused by the restored feedwater and isolated main steamlines in HFE 1A) would make diagnosis more difficult, they decided not to account for the masking effect in Pca and Pcd because they believed that the SG and radiation level indications were technically correct and would be closely monitored. In contrast, Team 1 accounted for the masking effect in both DTs.

Similar to CBDT & HRC/ORE teams, the SPAR-H teams made different judgments on the effects of masked SG and radiation level indications in HFE 1C. Team 1 assigned a factor of 10 to PSF ergonomics/human-machine interface (HMI), but Team 2 only assigned a nominal factor as they believed that the related instrumentation was functioning properly. For HFE 1A, Team 1 thought that the misleading AFW indication contributed to the scenario complexity, thus accounted for the misleading effect under PSF complexity rather than PSF ergonomics/HMI.

Obviously these types of effects can lead to significant differences in HRA results and points out the importance of identifying ways and guidance in the methods for avoiding such problems.

### 8.2.5 Inadequate Means to Tie Qualitative Analysis to Quantitative Approach Led to Differences between Applications of the Same Method and Mismatches with the Empirical Data

A broad qualitative analysis in evaluating likely crew performance does not necessarily generate HEP estimations that are consistent with the crew failure rates observed in the simulator experiments done in this study. For some methods the guidance on quantification of the effect of PSFs on crew performance is limited, and to varying degrees left to analysts' judgment, particularly when the analysts' qualitative analysis goes beyond the method guidance. In addition, it seems that not all HRA methods cover an adequate range of PSFs to predict operating crew performance for all circumstances; as a result, analysts may have to rely on their judgment to decide how to integrate the role of factors not explicitly covered by a method in HEP estimation, which can obviously lead to variability in results. A good tie or dovetailing between the qualitative analysis and the quantification approach is an important aspect for consistency in the results of HRA applications within methods.

### 8.2.6 Some Analysts Compensating for a Method's Poor Treatment of Diagnosis

Methods that strongly rely on TRCs to quantify diagnosis (e.g., ASEP) appear to be poorly equipped to address the difficulties in operators' cognitive activities. In addition, it does not appear that all the HRA methods in the study are well equipped to address the full scope of cognitive activities related to operators' overall response to the scenarios. This can lead analysts to attempt to compensate for the method's shortcomings by doing more qualitative analysis than is directed by the method and trying to incorporate the information into the quantification approach by adjusting the method based on their own experience, which can introduce variability in results. Moreover, when the methods do address diagnosis, they tend to focus on operators' cognitive activities to understand the plant situation and decide the appropriate response (i.e., initial diagnosis) without full consideration of the challenges with cognitive activities in making decisions during the execution of the selected response plan, which is important when the response plan is more complex than simple skill-of-the-craft.

The study results have shown that inadequate consideration of operators' cognitive activities and related potential failure mechanisms while they are working through the procedures can lead to a failure to identify important influencing factors and result in underestimating HEPs. Although analysts may compensate for methods' poor treatment of diagnosis based on their experience, it can lead to quantitative differences in HRA results.

## 8.3 HRA Practices (Interviews and Reasonableness check)

### 8.3.1 Different Impressions from Interviews with Trainers

Information from interviews with trainers was used for several purposes, such as estimating the time required for diagnosis and execution and for evaluation of training and experience. Most of the time, the information yielded valuable insights for analysts to understand the dynamics of scenarios. However, the analysts differed in the extent to which the interview information was used in their analyses. Some analysts tended to rely on input directly from the interviews while others tended to rely on their own analysis and judgment with interview information as a supplement. The study has shown that overreliance on trainer or operator opinions can sometimes negatively affect HRA results. Below is a list of such cases and it is interesting to note that the cases seemed to be dominant for the most difficult HFE (HFE 2A). This may suggest that the uncertainty in trainer or operator opinions may increase with the complexity of scenarios. Obtaining consensus among multiple trainers or operators and/or more detailed qualitative analysis may help reduce the negative effect caused by the uncertainty in trainer and operator opinions.

- SPAR-H Team 1 was apparently optimistic in their timing analysis for HFE 2A because a trainer underestimated the time needed to respond to the HFE and they did not confirm the estimate in their interview with a different trainer.

- The ASEP Teams 1 and 2 made different judgments on crew experience and training in HFE 2A because they seemed to obtain different information or at least a different impression or interpretation from their interviews with the trainers.

- Similar to the ASEP teams, the ATHEANA Team 2's analysis in HFE 2A on crew training and knowledge seemed to be influenced differently than the other ATHEANA team by information they obtained from their interviews.

### 8.3.2 Different Approaches for Collecting Information in Interviews

Different approaches for collecting information in interviews can cause differences in HRA results. Take HFE 2A as an example. One HRA team started out with a rather sparse qualitative scenario analysis. Then, the instructor in the interview claimed that the action would be easy and that the crews would manage it within the time available for the task. Following this, the HRA team judged the HFE to be rather easy, with a low HEP. Another HRA team had a quite detailed scenario analysis as a starting point but was also convinced by an instructor that the action would be easy. In contrast, a third HRA team also had a quite detailed analysis prepared, but instead of asking whether the crews would make the action or not, they asked the instructors to do a detailed talk-through together with them. The HRA team could then analyze how the crews would solve the scenario and dive into details of the procedures and analyze whether or not the crews would have time and ability to do the actions in the procedures, given the complexity of the situation. With this approach the team came away with a different impression of what would happen in the scenario and thought the crews would have more problems. The empirical results showed that the crews did have more problems and did not manage the action. After the complex situation led to a delayed start of the procedure, the crews simply did not have time to get through the procedure. (Note that the time constraints on these actions were rather tight.) It may not be easy for plant experts or instructors to understand the exact issues at hand in scenarios that they have not constructed themselves and that are difficult. The subtle details in a scenario that is discussed in an interview setting may not be so easy to grasp. In summary, a better interview technique and a self-standing detailed analysis for making a good qualitative scenario analysis that analyzes complexity and procedure progress in a better way may help here.

### 8.3.3 Appropriate Use of a Reasonableness Check of Obtained HEPs

The HEPs resulting from the application of each HRA method in the study were assessed with respect to their relative values (rank order of HFEs by failure probabilities) and the overall differentiation among these probabilities. Overall, the HEPs showed reasonable differentiation. However, in some cases, the HEPs for the two most difficult HFEs (2A and 1C) were not consistent with the difficulty ranking and/or fell into a narrow range. Particularly, the HEPs of SPAR-H Team 2 did not show the differentiation expected from their qualitative analysis. This suggested that the team did not check the reasonableness of the obtained probabilities or did an inadequate check in view of their qualitative findings. This is especially remarkable for their analysis of HFE 3A.

As was discussed in the International HRA Empirical Study report [1–4], although HEPs are often checked for reasonableness in external reviews of the probabilistic safety assessment (PSA)/HRAs, where each individual HEP cannot be reviewed in detail and emphasis is placed on the relative values of the HEPs, there appears to be little documented guidance on how to make reasonableness checks. A reasonableness check examines whether HEPs of comparable magnitudes are obtained for “similar” HFEs and whether the HEPs estimated for HFEs with more challenging performance conditions are indeed larger. Although inherently involving multiple dimensions, some of the factors to be considered in terms of similarity and levels of challenge include:

- time available (time window)
- decision complexity, basic vs. complex scenarios (number of issues, need to prioritize)
- task complexity, number of tasks, need for manual control, fine-tuning, adjustment
- number of issues, adverse performance shaping factors, and failure modes identified for the HFE

Comparing related HFEs (for the same tasks in different scenarios) or HFEs with similar performance conditions as represented by these factors typically leads analysts to review the contributions to the HEPs to determine whether they correspond to their expectations, based on their qualitative analysis. For the HRA teams that did make such checks, the identified discrepancies between HEP results and qualitative expectations would lead them to review the quantification and in some cases adjust the quantification of the HFEs.

In summary, the development of guidance for reasonableness checks that emphasizes the consistency between qualitative findings and quantification results would help to promote a structured review of HRA results.

# 9. INSIGHTS AND OVERALL CONCLUSIONS

This chapter summarizes the combined insights and conclusions obtained from the International and U.S. HRA Empirical studies about the four human reliability analysis (HRA) methods addressed in this study. In addition, it discusses the key findings on HRA in general; addresses areas where there is convergence between the two studies and where differences exist; identifies new findings, and summarizes other key achievements of the two studies taken together.

## 9.1 Findings on HRA Methods

### 9.1.1 Overall Assessment of ASEP

Accident Sequence Evaluation Program (ASEP) HRA Procedure is, as described in NUREG/CR-4772 [7], intended to be a less resource-intensive version of the THERP (Technique for Human Error Rate Prediction) method described in NUREG/CR-1278 (THERP Handbook) [10] and produces somewhat more conservative human error probabilities (HEPs) to account for the reduced analysis demands. ASEP also extends THERP in several ways, particularly with respect to the treatment of pre-initiators.

*Predictive Power*

In general, the two ASEP applications in the present study have illustrated ASEP to be a well-structured and easy-to-use method that produced fair to moderately good predictive power for the HFEs in the study, but that may not be adequate for reliably predicting HEPs for the full range of HFEs that may be expected in an internal events PSA. Despite its relatively good quantitative predictive power, the method was more limited in its qualitative predictive power because of some of its weaknesses, discussed in Section 9.1.1.2. However, as discussed in Chapter 7, while both ASEP teams in the present study followed the method's process, one team carried out the qualitative analysis at a more detailed level than suggested by the method guidance and this did seem to improve the predictive power for this team's application.

Variability between the teams' analyses and in their predictive power seemed to stem from the differences in analysts' assumptions and judgments in applying the method because of insufficient method guidance and the specific limitations of the method, which led analysts to attempt to compensate in some cases. As discussed in Chapter 7 and in more detail below, the analysis results can sometimes be sensitive to analysts' decisions (e.g., diagnosis HEPs can be sensitive to differences in timing analysis when the ASEP time reliability correlation (TRC) is used). Another source of variance is that ASEP allows analysts to decide whether to use ASEP's guidance or THERP to quantify the postdiagnosis actions; this choice will lead to different results.

*Traceability*

ASEP has good traceability because of its structured steps. It is possible for one to reproduce another's quantitative ASEP results given the same assumptions. Analysts may credit different indications of conditions and use different cues in timing analysis; nonetheless, once these items are clarified, the estimation of allowable diagnosis time and required postdiagnosis time is generally traceable. However, as with most methods, the basis for the analysts' decisions (e.g., the basis for selection of performance shaping factor (PSF) levels) needs to be documented to allow thorough traceability.

*Limited insight for error reduction*

Since ASEP relies heavily on its diagnosis curve with a few PSF adjustments to address diagnosis, it has a limited capacity for accounting for how PSFs and scenario conditions interact with operators' cognitive behavior. Correspondingly, the method guidance is not developed to identify the cognitive mechanisms that would lead to human failures, with the consequence that ASEP's ability to offer insights for error reduction is limited.

*9.1.1.1 Strengths*

In general, the two ASEP applications in the present study confirmed the following strengths identified in the International Study.

*Simplicity*

As mentioned above, ASEP seems to be a well-structured, easy-to-use, and straightforward method. It simplifies its human performance model by separating diagnosis from postdiagnosis actions, estimates the diagnosis HEP only with a diagnosis TRC with PSF adjustment, and focuses on the major procedural steps. On one hand, the simplifications make the method easy to use; on the other hand, they contribute to the weaknesses discussed in the following section. It should be noted that the simple analysis is justified in the method documentation by claiming that conservative HEPs will generally be obtained. However, this claim is challenged by the apparent optimism of the results seen in both the present study (e.g., final HEP of ASEP Team 2 for HFE 1C and the diagnosis HEP of ASEP Team 2 for HFE 1A) (see discussion below) and the International Study. The implication is that the tradeoffs between simplicity and a more comprehensive analysis need to be weighed before applications of the method.

*Traceability*

As discussed above, because of its structured steps, ASEP has good traceability and it can be considered a strength of the method.

*9.1.1.2 Weaknesses*

Evidence from both the present study and the International Study points to ASEP's limited guidance in the following aspects for doing detailed analysis.

*Insufficient guidance on when to include or exclude modeling of the diagnosis phase*

By segmenting total time available for coping with an abnormal event into two independent parts: diagnosis time and postdiagnosis time, ASEP provides an option to explicitly include and quantify diagnosis. However, since it appears that insufficient guidance is given as to when to include or exclude diagnosis, either additional guidance needs to be supplied or the option needs to be taken out of the methodology. As discussed above, such a weakness was evidenced in the International Study where analysts did not always quantify diagnosis (i.e., a decision component) for each HFE, where the task requirements and empirical observations suggested that one was involved. Although the two ASEP analyses in the present study explicitly modeled diagnosis, this weakness remains a potential problem, because, as mentioned above, the analysts in the present study might have learned from the International Study that not modeling diagnosis could lead to optimistic HEPs for many HFEs where the guidance would allow diagnosis not to be modeled.

*Limited guidance for estimating time requirements*

Because of the use of a TRC, the diagnosis HEP in ASEP is estimated as a function of the allowable diagnosis time, which is assessed by estimating the required postdiagnosis time (i.e., time to execute the response). Similar to the International Study, the present study has suggested that because of limited method guidance, analysts' judgments are often required to estimate these times and thus become a source of variability in time requirement estimation. In addition, there is also evidence in the present study that diagnosis HEPs for time-critical scenarios can be sensitive to analysts' judgments in postdiagnosis time estimation because of the characteristics of the TRC. Variability in analysts' judgments in the following aspects can lead to either conservative or optimistic diagnosis HEPs.

- Analysts need to make decisions on which procedural steps to include in estimating required postdiagnosis time. For the same HFE, analysts may assume different procedural paths. The effect of crew experience and training is, to some degree, addressed by PSF adjustment; however, there does not seem to be explicit or adequate guidance to help analysts address knowledge-based behavior in identifying critical actions and procedural paths. Even if the same procedure is assumed, analysts may consider different numbers of procedural steps.
- Analysts may need to make different assumptions and judgments on factors that could influence the time required, such as distraction, concurrent activities, and delay.

*Limited set of performance shaping factors (PSFs)*

The method shows inability to guide analysts to examine an adequate set of factors that would influence crew behavior for all circumstances. For example, the guidance to address diagnosis/cognitive tasks is minimal, and the method relies heavily on its diagnosis curve with adjustments for a few PSFs. Sometimes, analysts may anticipate a potential complication that operators will face, but the method does not give guidance for how to incorporate such additional information. Even with the guidance provided, it seems that the analysts would already have to have an idea of what they are looking for (i.e., they would need a good background in information needs for HRA in the context of PRA) in order to do an appropriate analysis. Both the present study and the International Study seem to suggest that better predictive power seemed to be partly because of analysts' experience and effort that went beyond ASEP guidance. In addition, it appears that it was more difficult for analysts to predict performance drivers as scenarios became more complicated in terms of cognitive demands. This suggests that improved methods and more guidance are needed to help analysts address the cognitive aspects of critical tasks, and it is necessary to include additional performance drivers to address complicated scenarios.

*Limited guidance for choosing PSF levels*

When addressing postdiagnosis actions, whether using ASEP or THERP (ASEP allows the use of THERP in quantification with respect to postdiagnosis actions), decisions need to be made regarding the specific levels of PSFs relative to a given scenario/HFE (e.g., stress levels and execution complexity). In both studies, differences across analysts were seen in the selection of PSF levels, and the differences led to observable variations in the HEPs for the same HFEs.

The guidance on those decisions appears to be limited for some situations, which may explain why analysts' decisions on those factors did not appear to correspond well to the factors and conditions observed from crew performance data.

*Limited guidance to examine low-level cognitive activities*

Another weakness of ASEP is its focus on procedural steps at a high level (e.g., identification of the initiating event and entry into the appropriate EOP), rather than the diagnosis and cognitive activities involved in following and responding to the steps in the EOPs. That is, in evaluating postdiagnosis tasks, lower-level cognitive activities, such as interpreting the plant status in the context of a step-by-step procedure and associated time-limiting conditions need more attention. As a consequence, HRA predictions are likely to be limited to the crew's interaction with the main procedural steps (e.g., entry conditions and key branching/transfer points within the procedure) and lead to optimistic HRA results by ignoring the difficulties that operators may face at the substep level.

### 9.1.2 Overall Assessment of SPAR-H

The Standardized Plant Analysis Risk-HRA (SPAR-H) method (NUREG/CR-6883) [6] was developed as a simplified HRA quantification technique based on eight pre-defined PSFs and separate nominal HEPs for diagnosis and action tasks. The results of the application is sensitive to the choice of levels for the PSFs and the PSFs selected to represent the condition imposed challenges on human performance, because different PSFs may weigh similar characteristics differently. In addition, both diagnosis and execution can make significant contributions to the total HEP obtained for an HFE using SPAR-H. As learned in the International Study, it is important to explicitly include diagnosis in evaluating most HFEs, even though the method allows diagnostic activities to be ignored once the overall event has been diagnosed and only quantify the execution portion of the action.

*Predictive Power*

The two SPAR-H applications in the study have illustrated that SPAR-H is easy to apply in order to arrive at the HEP, which may be seen as the method's greatest strength. Overall, the predictive power of the two applications was considered to be moderately poor to fair. SPAR-H is basically a quantitative method and analysts have freedom in choosing qualitative analysis approaches and deciding the scope and depth of the qualitative analysis, which can cause variability in HRA results and affect predictive power. Mostly, these shortcomings are seen as the byproduct of inadequate guidance on doing a successful and complete qualitative analysis in SPAR-H, but they may also reflect inherent limitations of the method. That is, extensions and modifications to the method will likely be required as opposed to simply adding more guidance. The current documentation does a good job of explaining the method, but it stops short in providing examples and guidance on deciding between competing levels of assignment for aspects such as PSFs.

It appeared that detailed qualitative analysis could help analysts understand scenario dynamics and improve qualitative predictive power, which seemed to be relatively more important for difficult HFEs than easy HFEs. However, as shown by SPAR-H Team 2's analysis, good qualitative analysis did not necessarily lead to good quantitative analysis because of the issues in translating qualitative analysis into the quantitative effect on HEPs.

*Traceability*

SPAR-H provides base probabilities (i.e., HEPs) for the diagnosis and execution portions of the crews' response and those HEPs are adjusted using multipliers to reflect the various levels of the PSFs identified from the analysis. Thus, the quantification process of SPAR-H is very transparent and is highly traceable. However the traceability of the basis for the choice of a multiplier for a particular PSF (i.e., the basis for the selection of the PSF level) will largely depend on the documentation of the scenario conditions in operational terms (especially for complex issues). The method does encourage appropriate documentation of why particular PSF levels are selected (i.e., what is it about the conditions that justify selection of a particular PSF level and the associated multiplier).

The traceability of the two applications in the present study differed fairly widely because of variability in the quality of the documentation. Thus, the traceability of quantitative analysis was considered relatively better than that of qualitative analysis for both applications. As discussed above, the selected PSF levels are clear for both diagnosis and execution, even though the accompanying justification from the underlying qualitative analysis could have been better in some cases.

*Insight for error reduction*

The PSFs included in SPAR-H can offer insights for error reduction through pinpointing error sources. However, the extent of insights largely depends on the quality of the qualitative analysis in understanding scenario dynamics and in identifying performance drivers, which is sometimes attributable to the analysts' skill since the method does not give detailed guidance for qualitative analysis (see discussion below).

*9.1.2.1 Strengths*

In general, the two SPAR-H applications in the present study confirmed the following strengths identified in the International Study.

*Simplicity*

SPAR-H is a simplified technique that uses easy-to-follow worksheets with base probabilities and PSF multipliers. Compared to other simplified techniques like ASEP, SPAR-H features few exceptions to the process flow, and the worksheets can be completed with minimal experience or training. However, the simplicity of the method also turns out to be a weakness of the method as is discussed below.

*Traceability*

As discussed above, traceability can also be seen as a strength of the method.

*9.1.2.2 Weaknesses*

*Simplicity*

While the ease of use of SPAR-H may be a benefit to experienced analysts, it also presents pitfalls for less experienced analysts because it belies the complexity in doing a thorough underlying qualitative analysis, which is not elucidated in the SPAR-H guidance and may be needed. In addition, if a good qualitative analysis is made, mapping the findings of the qualitative analysis to SPAR-H PSFs may be difficult. Moreover, because the method uses a checkbox approach, it is actually possible to complete an analysis without a thorough understanding of the HFE. One implication is that compared to standard or "vanilla" HFEs, analysts may need to devote relatively more effort to qualitative analysis to understand the difficulty and dynamics involved in complex HFEs. An implication of this may be that one should take care in using SPAR-H in a simplistic way, for example, as a screening or scoping tool, on complex HFEs. This may lead to optimistic HEPs.

The SPAR-H method is easy to apply, but it is also potentially easy to misapply. Additional guidance would help to prevent the unintentional misapplication of the method and potentially inappropriate results.

*Insufficient guidance for qualitative analysis (approach, scope and depth)*

SPAR-H focuses primarily on quantification and references other documents for guidance on doing qualitative analysis depending on the level of detail desired. As such, it does not give much insight or detail on "conventional" modeling considerations, such as the level of decomposition of a scenario or how subtasks should be combined for analysis. The method states that the analyst may break the scenario into subtasks for analysis, and suggests that the ATHEANA (A Technique for Human Event Analysis, NUREG-1624, Revision 1 [4], NUREG-1880 [5]) decomposition and qualitative process could be used, but other approaches could be used as well. For a given HFE, the level of decomposition will affect the quantitative result. Consequently, the freedom in choosing qualitative analysis approaches and deciding the scope and depth of qualitative analysis can lead to variability in HRA results.

*Insufficient set of PSFs*

Similar to some extent to other PSF-based methods examined in this study, there is evidence suggesting that the range of PSFs offered by SPAR-H is not adequate to cover all dynamics of complex scenarios. Analysts may compensate for such a limitation based on their judgment (e.g., by interpreting the scope of PSFs more broadly than suggested by the SPAR-H guidance), but the judgment involved in the compensation process can lead to variability in HRA results. Nevertheless, it can be argued that SPAR-H is a significant improvement compared to THERP and ASEP regarding the set of PSFs included and that the PSFs covered in SPAR-H is much closer to the recommended set of PSFs in the HRA Good Practices report by the NRC [12].

*Insufficient guidance for choosing PSFs and their levels*

Mapping the qualitative analysis results to PSFs and choosing between PSF levels is a difficult process in the PSF-based methods of this study. As with other methods, there is overlap in the dimensions used to characterize the PSFs in SPAR-H. Lack of clear delineation of how to select a PSF as well as how to avoid double-counting allows different analysts to account for a qualitative finding under different PSFs, although their understanding of the performance issue may be quite similar. All the issues can cause variability in translating qualitative analysis results to quantitative effects on HEPs.

### 9.1.3 Overall Assessment of CBDT & HCR/ORE

The method referred to as "CBDT & HCR/ORE (Cause-Based Decision Tree & Human Cognitive Reliability /Operator Reliability Experiments) in this study includes the Electric Power Research Institute (EPRI) HRA approach, which is applied using the EPRI HRA Calculator® [8] and refers to a combination of CBDT for diagnosis related actions and THERP for the execution portion, supplemented by the human cognitive reliability/operator reliability experiments (HCR/ORE) approach for time-critical actions. Two teams (CBDT & HCR/ORE Team 2 and Team 3) performed an analysis of the five HFEs using the EPRI HRA Methodology (implemented with the HRA Calculator software version 4.1.1). In addition, a third team (Team 1) performed the analysis using a hybrid CBDT & THERP & ASEP approach. A more detailed discussion of the Team 1 approach and its differences with respect to that used by the other teams (Team 2 and 3) can be found in Section 7.2. While the approach used by Team 1 is obviously different to some extent, based on the evaluation performed for this study, it is believed that at least with respect to CBDT and the general use of a TRC for time critical aspects, it has similar strengths and weaknesses and therefore can be discussed along with the EPRI HRA approach. However, there were some notable differences in the results and interested readers are encouraged to review Section 7.2 and Appendix G's Section 1.3 (particularly Section 1.3.6) for specifics. Section 9.1.1 above also gives specific conclusions about the ASEP method, which Team 1 used for quantifying post-diagnosis actions.

The findings with respect to the evaluation criteria used in the study are addressed first, followed by a summary of the strengths and weaknesses of the method. The adequacy of the analysis guidance is addressed more directly in the strengths and weaknesses section.

*Predictive Power*

Given the variability in the results across the different team applications, which was often because of how the different teams applied the methods, there was also variability in the predictive power of the method as applied in this study. In general the method showed fair to moderately good predictive power in terms of the quantitative predictions and in the identification of important drivers. However, the underlying reasons for the drivers did not always correspond to the empirical data and in some cases drivers were missed. In addition, in some cases drivers were identified as being important that did not appear to have an effect on performance, but there may not have been enough trials in this study to detect the effect. The operational stories produced by the method generally showed the weakest qualitative predictive power relative to the drivers in that the analysts did not always develop a good understanding of what would occur. This aspect was often a function of the extent to which the analysts made a detailed qualitative analysis and since there may not have been adequate guidance for doing this analysis (unless the analysts went to other sources); it may be related to the HRA skills of the different team members. In addition, the use of the TRC in some cases and team conclusions that there simply was not enough time available for the crew actions in others precluded the need for analysts to perform additional analysis that might have provided a more detailed discussion of what the crews would experience during the scenarios.

While the success of the applications of the (CBDT & HCR/ORE) method in this study can be seen as varying with the choices of the analysts to some extent (i.e., teams made different assumptions that led some teams to be more successful in particular cases than others), all of the applications generally had better predictive power on the easier or simpler HFEs than the more difficult, more cognitively complex HFEs. In addition, while the results suggest that a reasonable analysis can be done and reasonable results obtained with the method in some cases, the results from the comparison of the methods with the data and the intra-method comparisons still identified some important weaknesses or limitations of the method that could hinder obtaining appropriate results and limit consistency in HRA results across teams. Although some new findings emerged, these aspects are consistent with the results from the International Study [1–4] and are discussed below in Section 9.1.3.2, along with some of the strengths of the method in Section 9.1.3.1.

*Traceability*

A strength of the CBDT part of the method for evaluating diagnosis and the THERP approach for quantifying execution of the actions was their traceability (at least in one aspect of traceability). The derivation of the HEPs within the CBDT part of the method and the identification of which factors contributed to the HEPs are generally traceable. How the various factors are weighted in determining the final HEP can be determined by examining the contributions of various factors from the decision trees. However, the ability to trace the basis for the judgments about the branch points in the trees will rely on the analysts' documentation. Similarly, if analysts attempt to incorporate factors or expected effects not directly addressed by the decision trees, good documentation of the rationale will be necessary for traceability. The basis for the execution HEPs derived using THERP can also be traced (i.e., the weight of the factors considered, but again, to understand why those judgments were made will rely on the documentation).

However, since HCR/ORE estimates HEPs based on estimates of the median crew response time and associated variability, the traceability of the reasons for the HEPs using the TRC will depend on the extent of the documentation of the factors considered in estimating the time required for the action and the rationale for choosing parameters such as median response time and sigma. The results of the applications indicated that how the teams obtain these estimates (what factors are considered in evaluating timing issues) can be a significant source of variability in the results. Nevertheless, even HCR/ORE is traceable in the sense that the associated calculations used to obtain the HEPs once the conditions are determined is based on a set of clear rules to be applied based on well-defined criteria.

*Potential for Error Reduction*

For certain types of errors, this method can be useful in providing insights for error reduction. There are generally two categories of error reduction: 1) improving human factors to reduce errors (e.g., improve location and clarity of indications, improved format of procedures, etc.) and 2) reducing the difficulty of cognitively challenging actions (e.g., improved training, clearer procedures, or improved systems to address the scenario). This method clearly provides insight into the former category when CBDT and THERP are used, at least to the extent that the factors are addressed by these methods and in some cases may give information relevant to the second category. However, this method offers a more limited set of insights for reducing error caused by complex scenarios. The primary source of insights for error reduction of these scenarios comes from the qualitative analysis, which is supposed to include operator interviews and/or simulator observations, which can be difficult to do thoroughly in some instances. In addition, there is limited guidance in the method beyond the direct, but limited questions asked in CBDT and THERP, for how to do this analysis.

When HCR/ORE is used, the only insight to error reduction that can be gleaned is to find a way (either through training, revised procedures, etc.) to improve the timeline if a good understanding of the factors affecting timing is obtained. However, additional guidance for estimating the time required for actions is needed when adequate simulator observations cannot be observed and analyzed and such guidance may help find ways to reduce errors.

9.1.3.1 *Strengths*

*Cause-based approach of CBDT can identify failure mechanisms and factors important to performance and provide reasonable predictive power for some types of events*

The cause-based approach underlying the CBDT method did in many cases lead analysts to identify relevant failure mechanisms (i.e., how the crews might fail) and factors that were important contributors to the crews' performance. However, in some instances the factors were identified as being important for reasons other than those identified in the actual crew data (see the discussion in the Weaknesses section below). In addition, while the method did not appear to cover an adequate range of important factors to address the range of potential scenarios (discussed below), it did appear to reliably find some of the driving factors, that for at least some HFEs (e.g., non-challenging events), led to moderately good predictive power.

*Traceability and structured approach*

As discussed above, another strength of the method was its traceability. It gives a structured and generally traceable approach to estimate the HEPs and, as long as the basis for the selection of the paths through the decision trees in CBDT, the timing assumption made in HCR/ORE, and the tables and PSFs used in THERP are provided, good traceability is possible.

*9.1.3.2 Weaknesses*

*Inadequate set of factors covered*

An important CBDT & HCR/ORE limitation or weakness identified in the study is that the factors addressed or covered by the CBDT model (and, more generally, the HRA Calculator) may not always be adequate to find important driving factors that influence crew performance (i.e., the model did not always lead the analysis teams to address significant aspects of the scenario and detect important factors). This seemed to be particularly true for cases where there is moderate to high complexity. The problem can be magnified when complexity is paired with moderate to low time availability because HCR/ORE will be used to cover time-limited scenarios and the complexity may not be adequately addressed. Furthermore, in some cases analysts identified factors as being important that were consistent with the data, but the underlying reason for the effect did not match the data. For example, the analyses sometimes found procedural guidance to be a negative factor based on the format of the procedure (e.g., existence of NOT or AND/OR statements), when the problem that appeared had more to do with the content and progression of the procedures relative to the scenario.

Even if analysts identify operational conditions in the scenario that could be a problem, the model may not offer a direct means to incorporate this information. This was evidenced to some extent by the fact that a good operational story developed by the analysts and consistent with the data did not always translate into HEPs consistent with the data. It also appears that in some cases, the approach may select some PSFs as important contributors that lead to higher HEPs. In other words, the PSFs are judged to be at a level that significantly increases the HEP, but no effect is seen on the crews. However, it is difficult to get any confirmation of this conjecture about particular PSFs since it may be that there were simply not enough trials run in the study to detect the effect. Nevertheless, taken together, these potential issues with the method may have contributed to the lack of differentiation (in the International Study) that was seen between some of the HEPs where significant differences in error rates were obtained for the crew data and where there was also an underestimation of HEPs for some HFEs in the present (U.S.) study.  However, the method did show some sensitivity to the relative difficulty of HFEs, particularly in the U.S. Study.

*Use of HCR/ORE time reliability correlation and inadequate guidance for estimating the time required for HFE actions*

In the third phase of the International Study, analysts chose to use both the CBDT quantification approach and the TRC from HCR/ORE to quantify diagnosis for the scenarios (summing of the results from each model).  This was done because the CBDT approach is known to produce relatively lower HEPs than HCR/ORE when time pressure is a key driver, and because the analysts wanted to make sure that diagnosis was not ignored as was the case in some instances in the steam generator tube rupture (SGTR) scenarios (phase 2).  However, this solution led to overly conservative HEPs in several cases, as suggested by the crew data.  That is, the use of the HCR/ORE model for all HFEs caused the time available to become a driving factor in some cases where it did not have a detectable effect on the crew data.  In addition, the summing of the results from CBDT and HCR/ORE (which is not the normal approach used the EPRI method) may also have contributed to the apparent conservatism).

On the other hand, in the present study, while HCR/ORE was generally used appropriately, the estimation of the time required for operators to work through procedures and respond to scenario conditions was a variability contributor.  This was consistent with the observation of ASEP application in this study that used TRC to calculate cognitive error probability.  Ideally, HCR/ORE would rely on simulator observations to estimate the time required for crews to initiate and complete the actions, but in many cases this is not feasible and (as in the present study) analysts must go through a process to realistically estimate the time required.  However, the empirical studies have shown that there are multiple aspects that need to be considered (e.g., concurrent activities, use of multiple procedures, and workload) in estimating the time required and adequate guidance for how to obtain the correct information is currently not given by the methods.

*Potential for inadequate treatment of diagnosis*

In the International Study, the CBDT method was generally used to quantify the diagnosis portion of the HFE, and the THERP method was used to quantify the execution portion; however, for several HFEs, the analysts assumed that, based on the identified conditions, there would not be any additional diagnosis needed.  Similar to arguments in ASEP and THERP, they argued that after the initial diagnosis of the event and the presence of straightforward cues for the actions, the crews would simply follow the procedures, and limited diagnosis would be involved.  For these HFEs, the HEP was quantified solely on the assessment of response execution using THERP.  This decision, which in general appears to have been a modeler's choice and not a function of the software tool (EPRI HRA Calculator) or the associated methods (specifically the CDBT and HCR/ORE approach described in EPRI TR-100259 [9]), meant that for some HFEs the analysts did not investigate potential negative diagnosis factors that could influence performances based on the CBDT decision trees.  In some cases, disregarding the crews' cognitive activities and related failure mechanisms while they were following procedures apparently led to a failure to find some important negative drivers, which in turn led to apparent underestimations of HEPs.

The above approach was not taken by the CBDT & HCR/ORE analysis teams in the present study in the sense that the diagnosis portion of each HFE was explicitly addressed and resulted in better consideration of cognitive activities while working through the procedures and to overall better results. Thus, inadequate treatment of diagnosis in the sense above cannot be seen as an inherent weakness of the method, but the findings from the two studies point out the importance of adequate consideration of diagnosis activities as operators work through procedures. This conclusion is supported with a similar pattern of results found in the application of ASEP (i.e., addressing diagnosis activities after entering procedures led to better method performance).

*Inconsistency in estimating execution HEPs*

While total HEPs estimated by the CBDT & HCR/ORE teams often matched the scenarios moderately well in rank and level of differentiation, the cognitive vs. execution contribution to the HEPs were not always consistent with what would be expected from the empirical data. In addition, as can be seen in the pie charts presented in Figure 7-6 in Section 7.2 above, the analyses did not always agree on what the relative contributions of these response stages were to the total HEPs. These differences occurred for a variety of reasons including different assumptions about the presence of certain conditions and how the different portions of the HFEs were decomposed. Furthermore, in many instances there are inconsistencies in operational story and drivers relative to the response stages. While these various inconsistencies seem to even out in the rolled up HEP, it appears that there are significant differences in the underlying analyses. Apparently better guidance is needed for how to appropriately and consistently model these diagnosis and execution portions of the response. This is a new finding from the present study that was not detected in the International Study and it should be noted that the others methods in the U.S. Study showed similar problems. With respect to the decomposition of actions, HRA in general needs better guidance and approaches for how this should be accomplished.

*Improved guidance for operator interviews is needed*

The operator interviews across the three CBDT & HCR/ORE applications varied significantly in style, questions asked, depth, and even information received by the trainer/operator. For these assessments it was found that the better the operator interviews (i.e., how well the analysts understood the scenario going into the interview, the greater the depth of the interviews, a step-by-step discussion of procedure, the more formal the process), the closer the analysis matched the empirical data. This is a finding that is also relevant to all HRA methods. General HRA guidance needs to be developed to address this factor.

*Several areas where additional guidance is needed to support quantitative analysis judgments for CBDT & HCR/ORE*

There were three main sources of inconsistency in the quantitative analysis: philosophy of method application or analyst judgment, division of the HFE (task decomposition) and timing estimates (for HCR/ORE). Additional discussion on these issues is provided in the intra-method comparisons in Section 7.2.

### 9.1.4 Overall Assessment of ATHEANA

The ATHEANA method was applied by two HRA analysis teams in this study.

*Predictive Power*

The ATHEANA applications in the present study suggest that the method has moderately good to good predictive power for the HFEs analyzed in the study. The applications by both teams identified many of the performance difficulties or issues that were consistent with the observations of the plant crews in these scenarios. With respect to the quantitative results, the ranking of the HFEs by both teams was consistent with the empirical difficulty ranking, with the exception of one team's (Team 2) estimated HEP for the HFE that posed the most difficulties for the operating crews (HFE 2A). In terms of the absolute magnitudes of the HEPs, both teams underestimated the HEPs for the two more difficult HFEs (HFEs 2A and 1C), where several crews failed to meet the HFE success criteria.

The detailed examinations of the HRA analyses by both teams and the obtained predictions provide some evidence that the ATHEANA method's strength lies in the qualitative analysis, as discussed further below. In contrast, the quantification of the qualitative findings may be somewhat problematic. The results suggest that after having identified the potential performance issues, the experts may still find it difficult to quantify the effect of these issues or the capability of positive factors such as training and procedures to compensate for these issues.

*Traceability*

The traceability of the ATHEANA applications was good to very good. This is notable in light of the significant (operations) expertise required as well as the expert judgment involved in carrying out an ATHEANA analysis. The applications of the ATHEANA method by the two teams both showed consistency with the extensive guidance that the method gives for the qualitative analysis. Although there were significant differences in the level of detail of these analyses, the ATHEANA teams developed detailed descriptions of the expected crew response or responses and associated timelines. These path descriptions and timelines constitute an HFE-specific "model" of the crew response. In terms of level of detail, one team (Team 2) produced an overall timeline for the success path while the other team (Team 1) produced a tree of potential success and failure path, referred to as a scenario map.

In terms of quantification, the two teams applied rather different approaches. Both approaches were moderately traceable. Nevertheless, they are discussed separately because of the differences in the quantification approach and the implications of these differences for traceability.

The team that produced an overall success path and timeline, ATHEANA Team 2, noted that its overall ATHEANA application was streamlined. The success path was analyzed in terms of the major challenges (or performance issues) identified for the crews. It should be stressed that Team 2 did not find multiple error-forcing contexts or Unsafe Actions (UAs) and quantify these in terms of separate failure scenarios. Instead, the identified performance issues along the success path were used to assess an overall difficulty level for the HFE. This difficulty level and the arguments used to support this qualitative assessment were then used in conjunction with the calibration table suggested in the ATHEANA User's Guide (NUREG-1880 Table 3.8-2). The individual analysts separately estimated the HEP distributions and convened to derive a consensus distribution for the HEP.

The second ATHEANA analysis team, ATHEANA Team 1, produced a scenario map. Team 1 derived consensus distributions for the (sub)task durations and for the branch probabilities. The scenario map was then simulated using Monte Carlo sampling (i.e., obtaining response histories by sampling from the duration and branch probability distributions). The time to complete the response was compared to the available time (i.e., the time window for the HFE). The HEP is then the fraction of response histories exceeding the available time. For HFEs with a high expected level of performance reliability, in other words, actions without an error-forcing context or negative performance drivers, ATHEANA Team 1 estimated a "residual" HEP to represent "unexplainable" failure causes. The traceability of the numerical result obtained for the HEP is very good: the response paths that are considered and the underlying assumptions concerning durations and branching probabilities are clearly documented and yield, after Monte Carlo simulation, the HEP.

*Insights for error reduction*

The development of a detailed description of the expected response paths of the crews provide a strong basis for identifying specific potential performance issues for the tasks associated with an HFE. The embedding of these performance issues within a detailed response path supports insights for error reduction that are equally specific and targeted.

The qualitative predictions of both ATHEANA analysis teams could be clearly associated with such insights. As mentioned, these identified issues were in many cases subsequently observed in the crew performance data. Moreover, those issues that were not consistent with the simulator performance data nevertheless offer insights for potential error reduction.

#### *9.1.4.1 Strengths*

*Qualitative analysis and identification of potential performance issues*

The qualitative analysis guidance clearly constitutes one of the major strengths of the ATHEANA method. Both analysis teams applied the guidance to successfully find many of the performance issues observed in the plant crew performances in the simulator. The potential performance issues that were identified are highly detailed and specific; they specify where in the crew response to the scenario the performance issue may arise and why it arises.

*Development of a detailed performance timeline and narrative*

One of the reasons for the strength of the qualitative analysis is the need for the HRA analysis team to develop a detailed performance timeline and narrative. Rather than examining a set of characteristics of the HFE (e.g., assessing procedural guidance and other PSFs at an overall level), the development of the timeline and response path or paths requires close interaction with plant experts and a thorough understanding of the scenario, the required tasks, how these are done, and the contexts for these tasks. In an ATHEANA application, the understanding of the HFE is not taken for granted as a basis for the quantification but is instead developed explicitly in the analysis process. This explicit documentation of the understanding of the HFE increases traceability.

#### *9.1.4.2 Weaknesses*

*Quantification of qualitative findings*

The ATHEANA applications examined in this study show that both ATHEANA analysis teams had difficulties in quantifying the effect of the potential performance difficulties that they identified. It should be noted that neither ATHEANA team appeared to apply the quantification methodology as described in the method guidance. As mentioned, one team streamlined the process and did not attempt to specify multiple failure paths, deviant scenarios, or unsafe actions. The second team can be said to have formalized the notion of multiple failure or response paths by developing scenario maps for each HFE, extending this notion to the degree of considering the detailed variability (e.g., of subtask durations). The quantification by a Monte Carlo simulation of stochastic distributions of subtask durations and scenario branches may to some degree be seen as the logical "next step." However, this highly detailed simulation may miss holistic considerations with respect to the failure scenario related to an unsafe action.

In the case of both ATHEANA analysis teams, the quantitative predictive power was not as good as one might hope, given that the teams identified specific performance difficulties that were consistent with the empirical evidence. In summary, the teams were unable to assess the quantitative effect of these issues and/or the capability of positive performance factors to compensate for these difficulties. It should be noted that these quantitative aspects would be expected to be highly plant- and scenario-specific; consequently, such quantification will probably have to continue to rely on plant expertise.

*Lack of a formal quantitative approach for translating qualitative into quantitative findings*

Both analysis teams applied many of the recommendations of the ATHEANA guidance concerning the qualitative process. For instance, they developed distributions rather than point estimates, they used scaling guidance to support consistency, and they developed their estimates individually before deriving a consensus value. Nevertheless, the striking differences in the quantification approach applied by the two teams suggest that the quantitative approach may not be sufficiently formalized. In any case, it seems that the benefits of the documented ATHEANA approach were not sufficient to encourage both teams to apply the approach as described.

*Extent of resources needed for ATHEANA application*

The resources needed for the ATHEANA application are extensive. The highly detailed qualitative results and the value of the qualitative results obtained suggest that this cost does have its benefits: quality of the qualitative results, traceability (of the qualitative aspects of the HFE), and insights for error reduction. However, the quantitative stage of the analysis is also highly resource-intensive. Having established their own understanding of the HFE, collected information from plant experts to confirm or correct this understanding, ATHEANA analysis teams need to develop a set of failure scenarios and to quantify each by eliciting probabilities distributions for the elements of these failure scenarios. When the variability in the timing of the scenario evolution or the crew response is systematically considered, as the second analysis team addressed, more failure scenarios need to be quantified for a single HFE.

## 9.2 Findings on HRA General Issues

This section summarizes some of the main findings and conclusions obtained from both the International and U.S. HRA empirical studies. Section 9.3 below discusses the general convergence of the two studies and areas where differences were found.

Consideration of cognitive activities

The U.S. Study echoes the International Study that qualitative analysis done to support HRA quantification and the consideration of diagnosis is an important contributor to the adequacy of HRA predictions, especially the benefits of detailed qualitative analysis of cognitive challenges in complex scenarios. Compared to the International Study, one major improvement made in the U.S. Study is that all teams explicitly considered the effect of diagnosis and related activities on crew performance and did not assume that no diagnosis was required once the crews entered symptom-based procedures. The consideration of diagnosis contributed to improved qualitative analyses by the HRA teams in the U.S. Study, by helping the analysts understand the difficulties in crews' assessing the situations and/or making new response plans while the scenarios progressed. However, the U.S. Study has revealed that even when diagnosis is explicitly considered, the methods still show limitations in the ability to assess crews' cognitive activities in order to adequately support understanding and identification of failure mechanisms and quantification of HFEs, particularly for the more difficult HFEs.

Structured and consistent qualitative analysis with explicit guidance and framework

While a good qualitative analysis is a relative strength of some methods (e.g., ATHEANA), one conclusion from both studies is that qualitative analysis is a shared weakness across all methods (i.e., they all can be improved) and has a significant effect on the robustness of HRA applications. The HRA applications in the studies varied in qualitative analysis approach, both in scope and in depth. The variability is not unexpected given the differences in the technical bases and approaches of the methods. However, it seems that the variability also has its root in the fact that the methods do not give sufficient guidance or an explicit framework for analysts to conduct a structured and consistent qualitative analysis. This is clearly evidenced by the variability in timing analysis, HFE decomposition, the range of factors considered, and the treatment of complexity observed in the intra-method comparisons in the U.S. Study.

Treatment of complexity with method improvement and extended qualitative analysis

The two studies suggest that HRA methods differ in their abilities to deal with complexity. Compared to methods that rely on a defined (and usually limited) set of PSFs to represent the conditions driving performance (e.g., SPAR-H and ASEP), methods that develop a narrative or operational story that allows consideration of a broader range of potential PSFs in conjunction with the scenario and plant conditions (e.g., ATHEANA and CBDT to a more limited extent) tend to produce richer insights into scenario dynamics because of their focus on failure mechanisms and contextual factors. However, it appears that even for the more narrative and cause-based methods, there are still some aspects left to the analysts when attempting to adequately cover complex scenarios. For some of these analyses, it was observed that performance drivers were sometimes identified based on incorrect rationale, and in some cases, operational stories were generally consistent with the empirical data at a high level, but disagreed on details. These findings indicate the need for method improvement to cover a broader scope of performance drivers in all methods. Given that complex scenarios normally involve relatively more cognitive challenges compared to easy scenarios, one priority of method improvement should focus on providing means and frameworks for analysts to identify and characterize contextual factors and mechanisms that can cause failures at the cognitive level and provide a structured and systematic way to incorporate the information into the quantification process.

However, even with method improvement, it will still be difficult to cover all scenario-specific issues and performance drivers to the extent that all subjectivity in HRA will be eliminated. This means that HRA applications will always rely to some extent on analysts' experience and expertise, but the goal should be to give as much structure and guidance as possible to support analysts at differing levels of expertise.

Coherent coupling between qualitative analysis and quantitative model

As discussed above, extended qualitative analysis can help analysts understand the context and dynamics of complex scenarios and uncover scenario-specific performance drivers. However, it may not necessarily lead to appropriate HEPs in all cases because of the difficulties in translating qualitative analysis into HEP effect, especially for complex scenarios. The difficulties appear to be caused by two reasons. First, when scenario-specific issues, particularly cognitive challenges, identified in the qualitative analysis are beyond those addressed by a method (e.g., go beyond the conditions represented by the PSFs or failure mechanisms addressed in a given method), analysts have to compensate for method limitations by stretching the method to fit the situation based on their experience. Second, for methods that rely on selecting among a limited set of PSFs and estimating their level among several possibilities, inadequate guidance for choosing the PSFs and making the associated ratings is another cause for the breakdown of the interface between qualitative analysis and quantitative analysis, which results in variability in quantitative results.

Adherence to good practices with improved guidance

Most of the teams in the U.S. Study benefited from their multi-disciplinary composition. In addition, most of them did walkdowns of the main control room, observed a simulator exercise, and interviewed trainers, which analysts were not able to do in the International Study. These activities may have been a contributor to somewhat better HRA predictions in the U.S. Study than in the International Study, but also caused problems contributing to variability. It appears that the benefits of plant visits were maximized and trainer interviews went efficiently when analysts conducted a careful pre-analysis before such activities. Differences in the following aspects were observed, suggesting that improved guidance is needed for doing plant visits and personnel interviews.

- differences in interview skills and styles
- differences in the extent to which analysts relied on information from interviews
- differences in interpreting information from interviews

Additionally, the results of the U.S. Study supports the idea that reasonableness checks on obtained HEPs and associated qualitative analysis are an important HRA practice; however, there is little documented guidance on how to make such checks. Besides checking the relative values of the HEPs, a structured review of HRA results also needs to focus on the consistency between qualitative findings and quantitative results in terms of performance conditions.

## 9.3 Convergence of Findings between the U.S. HRA Empirical Study and the International Study and New Findings

Design Differences

There were several notable differences in the design of the International Study and the present study. First, there was only one case in the International Study where the same HRA method was applied by different teams, but at least two teams applied the same method in the U.S.

Study. This was because the goal of the U.S. Study was to find aspects of the methods that were susceptible to different applications or usages by different analysts that would lead to differences in results. The design of the International Study made it difficult to clearly separate method-specific effects from variability created by the analysts' differences in applying a method. Although it is not unreasonable to expect different analysts to make different decisions if the guidance is incomplete or subject to different interpretations and while it is unrealistic to expect that there will never be any subjectivity involved in HRA results, it is important to find areas in HRA methods where additional guidance or modeling approaches could improve the robustness of HRA performance.

International Study

Another difference between the two studies was that in the International Study, the HRA teams were unable to visit the Halden simulator to collect HRA related information (e.g., interviewing the plant operators and trainers and observing simulator exercises). Simulator observation and crew interviews are typical information collection activities for doing an HRA. This type of information was given to the HRA teams to the extent possible by the study team in the International Study. In addition, the HRA teams were allowed to submit written questions that were answered by the study team and plant personnel as needed. Nevertheless, some of the HRA teams in the International Study felt that this significantly limited their ability to do an adequate HRA and it has been acknowledged that good HRA practice involves visiting the plant, observing crew performance, interviewing operators, trainers, and other plant personnel about specific aspects of the scenarios being addressed, and obtaining other plant-specific information as needed. The HRA teams in the U.S. Study were able to visit the reference plant to collect information for their analyses.

Finally, there was a concern that because the International Study was based on the results of simulator runs using European crews at the Halden Reactor Project, the results might not be directly generalizable to what would occur with U.S. nuclear power plant crews. Some of the HRA teams in the International Study thought that their expertise was more geared to understanding what U.S. crews would do and that their U.S. bias might have influenced HRA teams' decision-making in applying their HRA methods. Thus, although the design of the U.S. Study would not allow an assessment of whether or to what extent the results of the International Study were in fact influenced by this situation, the U.S. Study (using a U.S. plant simulator and U.S. crews) provided information to evaluate this potential issue. If the International and U.S. Studies obtained the same or similar HRA results, the conclusions of the International Study are reinforced in the U.S. Study.

New Findings and Convergence of Results

As described in the sections above, there was significant agreement in the findings and conclusions between the International and U.S. Studies in terms of the strengths and weaknesses of the methods evaluated in both studies and in the overall findings about HRA and the identified needed improvements. In addition, while examination of the inter-analyst effects did not reveal any contradictory findings, some new issues were identified in the U.S. Study and additional information on others was obtained. For example, additional information on the factors that should be considered for estimating the time required for operator actions was obtained (e.g., crew procedure usage) and the importance of additional guidance on how to obtain the information was strongly supported in the present study. In addition, inconsistencies in modeling execution parts of the HFEs were seen and the importance of systematic interviews with plant personnel and the need for guidance for conducting the interviews was identified.

Other examples are discussed in Chapters 7 and 8 and in Section 9.1 above. However, there were also some interesting differences observed in the results of the two studies related to the performance of the HRA methods.

For one thing, based on the overall quantitative results obtained in this study, it could be argued that with the exception of HFE 3A (discussed further below), there was substantially less variability in the predicted HEPs across the methods in the U.S. Study compared to the International Study. In addition, the difference in the intra-method HEP predictions was less than might have been expected based on general HRA performance in the International Study and on the couple cases where similar methods were used. In other words, one might argue that the analysts using the same methods in the U.S. Study did a relatively good job in many cases and often corresponded relatively well in their predicted difficulty rankings of the HFEs.

Potential reasons for the better performance of the U.S. Study participants in predicting results compared to the International Study are the following:

- There may have been some learning effects between the International Study and the U.S. Study. Some of individuals on the HRA teams participated in both studies and most participants in the U.S. Study were familiar with the results of the International Study. Thus, the lessons learned may have improved the HRA teams' applications in that they had a better idea of what they needed to do to perform a better analysis with the method they were using. Some improvements were also seen between the SGTR and loss-of-feedwater (LOFW) phases of the International Study that appeared to be due, at least in part, to learning effects (but the nature of the LOFW scenarios may also have contributed).

- It is also possible that the HRA teams were, in fact, better at predicting the performance of U.S. crews on a U.S. simulator than the performance of foreign crews. That is, their previous experience with how U.S. crews work and interact with procedures etc. may have facilitated their performance.

- Similarly, the ability of the participants in the U.S. Study to visit the plant, interact with trainers, and obtain the usual HRA related information may also have contributed to improved performance.

It is certainly possible that all of these factors played a role and it was at least somewhat encouraging to see the level of consistency in the predictions both within and across the different methods. However, as discussed in Chapters 7 and 8 and Section 9.1 above, there were still important differences in the results of the applications (both quantitative and qualitative) within and between the methods, particularly in the predictions for the difficult HEPs. Furthermore, some of the consistency that was seen appeared to be based on different underlying reasons (i.e., they obtained the same results for different reasons) and in spite of differences in detailed analysis results (e.g., the differences in relative contributions of the diagnosis and execution components). Thus the consistency in results was potentially coincidental to some extent. And where differences did occur between the applications, it appeared that the differences between analysts' decisions and judgments might be improved with better methodology and guidance.

It should also be noted that some of the differences between the intra-method and across-method application might also be reduced with improved training. It is certainly possible that skill differences may have contributed to differences in decisions and assumptions made by the analysts, but it is common for less than expert HRA analysts to apply HRA methods and it should be possible to temper such effects and support such analysts.

One finding worth noting as being a significant surprise was the large variability in the predicted HEPs for HFE 3A, which was clearly identified by most everyone as the easiest HFE. Since it was the easiest, one might have expected this HFE to show the least variability in HEPs. Although much of the variability seen between the HEPs might be attributed to a couple of outliers that were much lower than the other HEPs for 3A, it still suggests that there may not be much consensus in HRA in terms of what the baseline HEP for generally good conditions should be. It would seem important to try to resolve this issue for future HRA modeling.

In any case, it is clear that while there was significant convergence in the findings from the two studies, the ability to compare the intra-method applications expanded the findings from the International Study and provided additional information on how to improve the consistency and robustness of HRA.

## 9.4 Achievements and Overall Conclusions

Taken together, the U.S. HRA Empirical Study and the International HRA Empirical Study [1–4], represent a large systematic data collection effort, with appropriate controls, to allow clear conclusions on the strengths and weaknesses of several HRA methods currently being used in the industry and on HRA practices in general.

Using a small number of operating crews and simulated accident scenarios, but having at least two HRA teams applying each method to predict crew performance with the use of standard HRA practices to the extent possible, the present study produced a set of findings on HRA methodology and process and on how multiple analysts can apply the same method differently to generate different results. In particular, the intra-method comparisons conducted in the study provided valuable insights on method effects and analyst effects in causing variability in HRA results. The information allowed us to identify methodological enhancements to obtain consistent, reliable and robust HRA results. It could be argued that the differences in the applications of both the same and different methods were, to some extent, caused by analyst experience levels. As discussed earlier in this chapter and in Chapter 8, analysts did rely on their experience to compensate for methods' weaknesses, leading to variability in analysis results. The study team tried to control the effect of experience levels to the extent possible with two measures in teaming up HRA analysts. First, the study team tried to ensure that each HRA analyst team in the study had a multi-disciplinary composition with expertise in HRA methods, operating experience, and human behavior, but there were varying degrees of expertise in human behavior. Second, we tried to ensure that there was at least one HRA analyst on each team with significant experience in applying HRA. However, it was the case that one or two teams may not have had as much applied experience as the others and of course it is very difficult to control for overall expertise. It is important to note however, that the objective of the study was to show where and how HRA method guidance could lead analysts to obtain different HRA results. That is, where the methods were susceptible to different applications and where the guidance appeared to be weak. The study certainly afforded us the opportunity to do this and important results were obtained. We recognize that if method guidance were perfect, there would not necessarily be variability in HRA results regardless of analysts' experience. Therefore, while we acknowledge that the analysts' experience levels could affect the analysis, we did not specifically investigate this effect and do not believe that it diminishes the findings from the study.

In addition to the important information on HRA methodology, related achievements from the study, in conjunction with the International Study, are summarized below.

- The U.S. Study and the International Study demonstrated the feasibility of using simulator data in HRA studies. The methodological tools developed in the studies, such as: (1) the development of the experimental design with focus on evaluating HRA methods; (2) the method for collecting crew data; (3) the method for analyzing crew performance data in terms of tasks corresponding to PRA-type HFEs and in terms of the corresponding HFE boundary conditions; and (4) the method for data-to-method comparisons were tailored to HRA needs and are proving to be very useful achievements. They have also demonstrated that important information on HRA and HRA methods can be obtained without using impractically large numbers of operating crews and scenarios, which is another important achievement.

- The studies have shown that simulator data are highly useful for HRA studies. Although simulator data was used as the empirical basis against HRA predictions, the promising results from this study encourage and promote the use of simulator data in the future, as well as encouraging analysts to use it in different ways. The studies also show the potential of using and aggregating empirical simulator results from multiple studies to strengthen the empirical basis for both method assessment and extending the scope of methods to address some of the identified shortcomings. In summary, while there are other sources of HRA data, this study reinforced the relevance of simulator data for HRA in general.

- The scenarios developed in the studies are similar to those modelled in PRA and represent difficulty levels from basic to highly complex. They can be used as standard scenarios for other HRA benchmarking studies. Complex scenarios can be used to determine whether an HRA method would underestimate HEPs in highly complex scenarios. Basic scenarios can be used to establish a baseline performance. The difficulty levels can be used to test whether a method can produce HEPs with appropriate differentiation.

- The study also yielded valuable empirical evidence on how crews will perform and why, and documented the variability in actual crew performance. Although these findings on crew performance variability are not typically considered in HRA, they can be important under some circumstances. For example, the findings offer significant value to the participating plant in helping them improve plant procedures and training programs.

# 10. REFERENCES


[1] Lois, E., Dang, V.N., Forester, J., Broberg, H., Massaiu, S., Hildebrandt, M., Braarud, P.Ø., Parry, G., Julius, J., Boring, R., Männistö, I., Bye, A. (2009). "International HRA Empirical Study—Phase 1 Report: Description of Overall Approach and Pilot Phase Results from Comparing HRA Methods to Simulator Data." NUREG/IA-0216, Vol. 1. U.S. Nuclear Regulatory Commission, Washington, DC.

[2] Bye, A., Lois, E., Dang, V.N., Parry, G., Forester, J., Massaiu, S., Boring, R., Braarud, P.Ø., Broberg, H., Julius, J., Männistö, I., Nelson, P. (2011). "International HRA Empirical Study—Phase 2 Report: Results from Comparing HRA Method Predictions to Simulator Data from SGTR Scenarios." NUREG/IA-0216, Vol. 2. U.S. Nuclear Regulatory Commission, Washington, DC.

[3] Dang, V.N., Forester, J., Boring, R., Broberg, H., Massaiu, S., Julius, J., Männistö, I., Liao, H., Nelson, P., Lois, E., & Bye, A. (2011). "The International HRA Empirical Study—Phase 3 Report: Results from Comparing HRA Methods Predictions to HAMMLAB Simulator Data on LOFW Scenarios." NUREG/IA-0216, Vol. 3, U.S. Nuclear Regulatory Commission, Washington, DC.

[4] Forester, J., Boring, Dang, V.N., Bye, A., Lois, E., Massaiu, S., Broberg, H., Braarud, P., Boring, R., Männistö, I., Liao, H., Julius, J., Parry, G., & Nelson, P. (2013). "The International HRA Empirical Study—Final Report: Lessons Learned from Comparing HRA Methods Predictions to HAMMLAB Simulator Data." NUREG-2127. U.S. Nuclear Regulatory Commission, Washington, DC.

[5] "Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA)," (May 2000). NUREG-1624, Revision 1, U.S. Nuclear Regulatory Commission, Washington, DC.

[6] Forester, J., Kolaczkowski, A., Cooper, S., Bley, D., and Lois, E. (2007). "ATHEANA User's Guide," NUREG-1880. U.S. Nuclear Regulatory Commission, Washington, DC.

[7] Gertman, D., Blackman, H., Marble, J., Byers, J., Haney, L., & Smith, C. (2005): "The SPAR-H Human Reliability Analysis Method," NUREG/CR-6883. U.S. Nuclear Regulatory Commission, Washington, DC.

[8] Swain, A.D. (1987). "Accident Sequence Evaluation Program Human Reliability Analysis Procedure" (1996). NUREG/CR-4772/SAND86-1996, Sandia National Laboratories for the U.S. Nuclear Regulatory Commission, Washington, DC.

[9] The EPRI HRA Calculator® Software Users Manual, Version 4.1.1, EPRI, Palo Alto, CA, and Scientech, a Curtiss-Wright Flow Control company, Tukwila, WA: 2009. Software Product ID #: 1020436

[10] Parry, G. et al. (1992). "An Approach to the Analysis of Operator Actions in PRA," EPRI TR-100259, Electric Power Research Institute, Palo Alto, CA.

[11] Swain, A. D & Guttman, H. E. (1983). "Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications." NUREG/CR-1278-F, U.S. Nuclear Regulatory Commission, Washington, DC.

[12] Guenther, S. (2004). "Operator Licensing Examination Standards for Power Reactors." NUREG-1021, Revision 9, U.S. Nuclear Regulatory Commission, Washington, DC.

[13] Kolaczkowski, A., Forester, J., Lois, E., & Cooper, S. (2005). "Good Practices for Implementing Human Reliability Analysis (HRA)." NUREG-1792. U.S. Nuclear Regulatory Commission, Washington, DC.

[14] Brown, L.D., Cai, T.T., DasGupta, A. (2001). "Interval Estimation for a Binomial Proportion." Statistical Science, 16(2): 101-133.

[15] Siu, N.O., Kelly D.L. (1998). "Bayesian parameter estimation in probabilistic risk assessment." Reliability Engineering & System Safety, 62: 89-116.

[16] U.S. Nuclear Regulatory Commission. (2012). NRC/EPRI Draft Report, "An Integrated Decision-Tree Human Event Analysis System (IDHEAS) *Method for NPP Internal at-Power Operation*." Washington, DC: U.S. Nuclear Regulatory Commission.

# APPENDIX A INFORMATION PACKAGE

This appendix provides most of the information package that was sent to the HRA teams participating in the study. Specific items not included are the Agreement Forms for participation in the study (Item 2 of the package) and the actual plant procedures (Item 5 of the package). The items included are presented as they were sent to the HRA teams. Some minor aspects may have changed along the away. The actual final design aspects are presented in the body of the report. When the package was sent out, some aspects had not yet been specified and the HRA teams were notified as the study progressed.

## A.1 Item 1 of Information Package

### - Overview and Instructions to HRA Teams

**Objective of the Study.**
A diversity of older, established HRA methods and more recent methods are applied in the Probabilistic Safety Assessments of nuclear power plants and other complex human-technical systems. There are differences in the scope, approach, and models underlying these methods. Consequently, there is significant interest in assessing the performance of HRA methods, their consistency/reliability, and, to the extent possible, their validity. In the present study, an evaluation of HRA methods in light of simulator-based data is being performed, aiming to develop an empirically-based understanding of their performance, consistency, strengths, and weaknesses. The objective of this empirical study is to evaluate a specific set of HRA methods used in regulatory applications through a comparison of HRA predictions to crew performance in simulator experiments performed in a US nuclear power plant.

It is expected that the results of this work will provide the technical basis for the development of improved HRA guidance and, if necessary, improved HRA methods. The study is a follow-on to the recent international empirical study conducted at the HAMMLAB simulator at the Halden Reactor Project.

**HRA Team Information package**

This package consists of the instructions, background information, and information on the three specific scenarios and 5 HFEs addressed in this study.

Item 1 (this item) consists of:

- descriptions of the contents of the information package (items 1-6)
- instructions to the HRA Teams
- schedule and deadlines for this study phase
- agenda and logistical information for the plant visit (PENDING)

**Contents of the Information Package**

| Info package item | |
|---|---|
| 1. Overview and instructions to the HRA teams | - overview of the info package<br>- instructions to the HRA teams. Additional, specific instructions appear on the response forms.<br>- schedule and deadlines<br>- site visit logistical information (PENDING) |
| 2. Administrative (e.g. protecting data, procedures) | - agreement forms for the study participants |
| 3. Study outline | - outline of the overall study, including the aspects related to the measures used in the simulator experiments (experimental measures), data analysis, and the assessment of the HRA team responses |
| 4. Scenario description and HFEs | - description of the three simulated scenarios and the associated HFEs to be analyzed by the HRA teams |
| 5. Procedures | - The procedures relevant to each simulated scenario |
| 6. Forms/guidance for the responses of the HRA teams | There are three types of documentation required from the HRAs performed by the HRA teams<br>- fill-out Form A to document key aspects of the HRA<br>- fill-out Form B to document the general analysis process and qualitative analysis *<br>- provide the usual method guided documentation for an HRA analysis |

* Participants in the (earlier) Int. HRA Empirical Study, note that Form B in this study has a completely different purpose and content from that in the earlier study.

**Instructions to the HRA Teams**

1. Please read, sign, and return the included agreement forms. One person from each participating organization (except the NRC participants) should sign the forms.
2. In referring to this study and in all publications related to this study, please do not refer by name to the plant and plant organization.

   The following expressions (or equivalent expressions) are recommended:

   – "the PWR simulator at a US reference plant" to refer to the study reference plant

   – "licensed PWR operators operating the PWR simulator at the US reference plant"

   – "US reference plant EOPs" etc.
3. Review the information package and prepare questions, if desired.

   There is a period for clarification questions to the study coordinators until May 24th, as noted in the study schedule (following these instructions). These questions may address the study plan and instructions as well as the scenario and HFE information that has been provided. Important: questions about the understanding of the scenarios and HFEs should be asked at this stage. The plant visit in Step 4 is intended to provide you with the opportunity to ask plant staff about how they handle certain situations, etc. The plant staff will not be briefed in detail concerning the scenario and HFEs of interest.

Answers will be provided as fast as possible. Please note, only clarification questions will be answered at this time; teams will have the opportunity to visit the plant and simulator to ask technical and information gathering questions related to the plant, crews, scenarios and HFEs. The answers provided during the plant interviews will not be provided to all HRA teams. That is, each HRA team will have their own plant team to interview. For more general clarification questions on the materials provided in the information package, some answers may be shared with all teams if relevant and appropriate as decided by the assessment team.

Please do not wait until the end of the question period to submit your questions. Submit these as you go along. The study organization will be compiling the questions from the different teams and merge these if appropriate. By not waiting until the end of the question period, you create the opportunity for follow-up questions.

4. Perform site visit to gather data necessary for analysis. Each team will have an opportunity to visit the simulator, observe a training simulation and, independently, interview plant operator(s) and/or trainers. Specific dates for each teams' plant visit have already been assigned. An agenda and logistical information for this visit will be provided as soon as possible. Note that multiple team members may attend if desired, but it may not be possible for all team members to be able to attend the training run. We will try to include everyone, but space may be limited.
5. Perform your HRA analyses of the 5 Human Failure Events (HFEs) in the three scenarios. These are defined and described in item 4 of this information package.
6. Prepare your response package, which should consist of the following
    A. responses to Form A – key aspects of the results of the HRA
    B. responses to Form B – documentation of the general qualitative analysis process compiled during the analysis
    C. documentation of the HRA analysis and quantification (the usual method guided documentation for an HRA analysis).

    The responses of the HRA teams will initially be reviewed by the assessment group concerning its clarity and conformity with instructions. This step is to ensure that the instructions have been sufficiently clear and that there is consistency in the way the HRA teams have interpreted the questions and provided their answers. The assessment group may have questions to the HRA teams following this review and request additions or amendments of the submitted response.
7. After the HRA teams have submitted their final responses, the assessment group will assess the teams' predictions against the experimental outcomes. (See the study schedule for more information and the study outline for details.) This draft assessment will be initially provided in draft form and the HRA teams will be asked to review any summaries of their analyses and the draft comparisons.

**Schedule/Timeline**

**(Deadlines for HRA teams are shown in bold)**

**2010**

| | |
|---|---|
| April 30 | Information package sent to HRA teams by E-mail. (Beginning of 3-week period for questions from HRA teams.) |
| **May 3-May 24** | **Questions from HRA teams**. Please submit questions as early as possible. You will get answers as fast as we can answer them. |
| May 28 | Last responses to HRA team questions. |
| June 2 and 3 OR | |
| June 23 and 24 | **Plant visit** |
| **Aug 31** | **Deadline for HRA teams to submit responses** (forms and documentation) |
| Sept 30 | Review of responses by assessment group and clarification questions to HRA teams |
| **Oct 15** | **Deadline for HRA teams to submit revised responses, if requested.** |
| TBD | Summary of HRA team results for comparison with crew/experimental data |
| TBD | First results of experimental data analysis (experimental outcomes) by Halden, for review by assessment group. |
| TBD | Release of experimental outcomes to HRA teams |
| TBD | Release of comparison results (draft) from assessment group |
| TBD | Workshop on Empirical Study Results and Outlook, ??? USA |
| | Finalization of comparison results |

**Site Visit Agenda and Logistics**

Pending at the time

## A.2 Item 2 of Information Package -Agreement Forms

Not included in this appendix

## A.3 Item 3 of Information Package -Study Outline Study Outline

1. Objectives of the study
2. Study design and protocol (how the study is organized)
3. Experimental protocol, measures, and analysis (how the simulator sessions are organized, raw data and data analysis)
4. Participants, roles and tasks

5. Comparison of experimental findings with HRA method outcomes

6. Comparison Between HRA Team Outputs

7. Specification of HRA Team information package

## 1. Objectives of the study

The overall objective of the study relates to the validation of HRA methods. The specific objective of this study is to assess HRA methods against a specific set of data. The study will focus on the following methods:

- EPRI HRA Calculator
- ATHEANA
- SPAR-H
- ASEP/THERP

However, other methods may be included for special testing.

- The focus of the study is the comparison of experimental findings with HRA method outcomes. A comparison of the methods will be performed, addressing the strengths, weaknesses, predictive power and areas for improvement identified in the course of the analyses and the comparisons.

- In addition to comparing experimental data with the HRA outcomes, another major objective of the study is to test the consistency of HRA predictions a) among analyst teams using the same methods and b) across different methods. A particular area of interest in this comparison is examination of the qualitative analysis performed by different methods and teams to identify shortcomings that contribute to inconsistencies.

- The third objective is to compare conclusions of this study against those of the Halden experimental study (Int. HRA Empirical Study) to confirm results using US crews.

## 2. Study design and protocol (how the study is organized)

The overall design of this study is shown in Figure 1. The activities of the teams applying the HRA methods ("HRA Teams") are shown on the left. The collection of the simulator data will be performed by the experimental team, using measures that have been used in previous work. The HRA teams will have the opportunity to visit the plant and collect the data necessary to perform their analyses (e.g., interview of plant operators). The simulator data analysis and the comparative analysis will be performed primarily by experimental team, with the HRA teams providing input in the form of completed HRA assessments and associated documentation for their assigned method (see Item 6 of the information package).

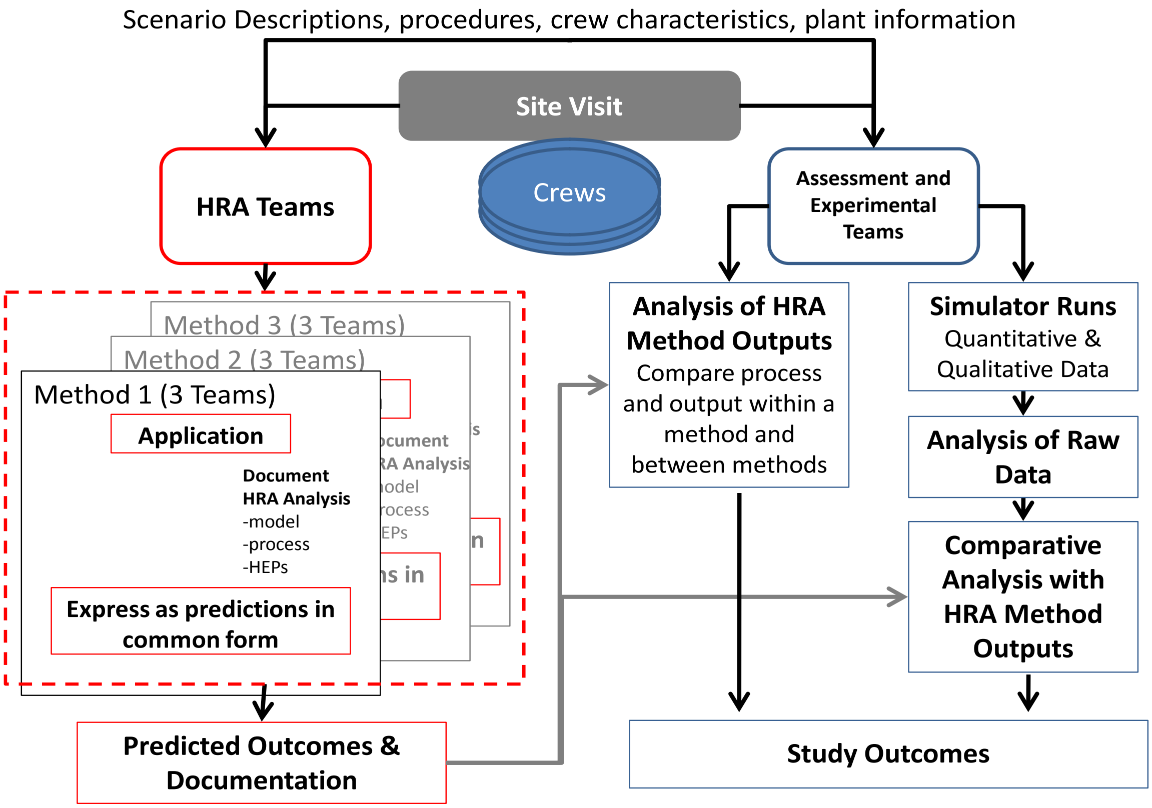



Figure 1.  Overview of the study.

# 3. Experimental protocol, measures, and analysis

## 3.1 Experimental Measures

The raw experimental measures include:

- occurrence (or non-occurrence) of human failure events of interest and performance of human actions related to these HFEs
- timing data on when human actions related to the HFEs were performed
- debriefing of crews in terms of important PSFs and drivers of behavior, areas of difficulties
- impressions of observers, including perceived basis of decisions, apparent basis for difficulties, evidence of communication breakdown, crew characteristics that seem to have contributed to the crew response (such as difference in rate of working through procedures, patterns of interaction, etc)

### 3.2 Analysis of experimental results

The analysis of experimental results to obtain the experimental outcomes will include:

- identification of influencing factors for the performance of the HFEs
- form of the difficulties experienced by the crews for these HFEs

### 3.3 Scenarios

The reference plant is a PWR 4-loop Westinghouse simulator.

Four crews of licensed operators are participating in the simulator runs; each crew was comprised of 5 members in this study. Normally, there is a 6 member crew at this plant: shift manager, unit supervisor, shift technical advisor and three reactor operators. However, it is not unusual for only 5 members to be in the control room at a given time. See the "Conduct of Operations' procedure in item 5B of the information package for additional information.

Each crew responded to 3 scenarios: LOFW and SGTR with complications; loss of CCW and RCP Sealwater; SGTR without complications. The crews did not have knowledge of the types of scenarios they would face. Across the crews, the order of the scenarios was varied.

## 4. Participants, roles and tasks

### 4.1 Experimental Team (Study Organization)

The roles of the experimental/assessment group include:

- design of the overall study
- preparation of the information package for the HRA teams
- answering programmatic and clarification questions from the HRA teams
- gather data during site visit:
  - notes from simulation observation (control room and simulator booth)
  - simulator logs
  - interviews with simulation crews, including a PSF questionnaire
  - audio/visual recording of simulations
  - audio recording of interviews between HRA teams and plant personnel
- analysis of raw data from the simulation
  - determination of driving factors for each HFE
  - operational story/description of crew response with respect to the HFE
  - assessment of difficulty of each HFE for comparisons with HRA team HEPs

- review/summary of the HRA team responses

- assessment (comparison) of the predicted outcomes vs. experimental outcomes

- assessment (comparison) of the consistency of HRA predictions of different HRA teams across the same method and different methods

### 4.2 HRA Teams

Each HRA team will apply a specified HRA method. The number of members on a team and the composition of the team are not restricted. The HRA team leads were all selected due to their extensive HRA experience and the teams have experience or training in the methods they are applying.

In order to allow teams to perform a realistic HRA, all the teams will have the opportunity to observe a regular training simulation, tour the facility and independently interview plant personnel (a different plant SRO and instructor team for each HRA team) with an opportunity to follow-up with a different instructor to ask additional questions.

After the site visit, the HRA teams will then perform an assessment of the defined HFEs using their assigned method. For each method, the response should include:

- Form A, where the HEP is reported and two open-form questions need to be answered (one form A for each of the 5 HFEs being considered over the three scenarios)

- Form B, a guided documentation of the HRA analysis (a Diary); in particular information about the qualitative analysis done to support the quantification.

- The normal documentation associated with applying the HRA method and quantifying the HFEs

## 5. Comparison of experimental findings with HRA method outcomes

Following the experimental data analysis to derive the experimental outcomes, the assessment team will compare the HRA predicted outcomes to the experimental outcomes. This comparison will be performed using three measures:

1) HEPs. Given that the experiments will not yield HEPs, a surrogate for experimental HEPs will be used. The ranking of the HFEs based on the predicted HEPs will be compared against measures of crew performance. (It is expected that one of these measures will be the “difficulty ranking” or (qualitative) relative failure likelihood used earlier.)

2) Influencing factors predicted for each HFE. (Form A item 2)

3) Narratives or operational stories describing difficulties/problems the crews might have with the actions associated with each HFE and why; or why things might go well in the scenario and with performing the HFE related actions (Form A item 3).

## 6. Comparison Between HRA Team Outputs

In addition to comparing the HRA analyses against experimental outcomes, the experimental team will also compare the HRA analyses against eachother to some extent.  Three teams are assigned to each of the HRA methods addressed in this study (EPRI HRA Calculator, ATHEANA, SPAR-H and ASEP/THERP), allowing the experimental team to compare the consistency of the outputs both within a method (is there a large team effect?) and between methods. This comparison will be performed by examining the qualitative and quantitative (Forms A and B and the HRA application documenattion) analyses using the metrics described in section 5 above,  as well as examining the actual process each team went through in order to develop their qualitative analysis and output.

## 7. Specification of HRA TEAM information package

This package of material includes

- overview and instructions
- agreement forms
- study outline
- specific scenarios and HFE definitions being analyzed in the study
- forms/guidance for the responses of the HRA teams

The package supports the analysis.

## A.4.1 Item 4a of Information Package –Scenario

## Item 4A.  Scenarios

## US HRA simulator study, Scenario 1: LOFW and SGTR

### Plant technical information

There are three main feedwater pumps: 11, 12 and 13.
There are four auxiliary feedwater pumps: 11, 12, 13 and 14. AFW pump 14 is turbine-driven and the other three motor-driven.

### Situation from start

- The Shift Technical Advisor is not in the control room. He or she will arrive 5 minutes after being called. The other participating crew members are in the control room (SM, US, 2 ROs)
- The plant is operating at 100%
- Core burnup is 19,000 MWD/MTU (End of life)

**Total loss of feedwater**

2 min Loss of main feedwater pump 11, and subsequent trip of feedwater pump 12 and 13 within the next 10 seconds.

All main feedwater pumps are tripped, and if the crew doesn't trip manually the reactor will trip on low SG level (20%). (The start up feedpump cannot be started.)

At autostart, Auxiliary feedwater (AFW) pump 14 will overspeed and cause damage that cannot be repaired. AFW pump 11 will have a seized shaft and trip and will not be available. AFW pump 13 will start but the shaft will shear and no flow will be indicated.

AFW pump 12 will start automatically and indicate full flow, but this flow will not reach (feed) the steam generator because a recirculation valve is mis-positioned (it is open). There is no indication of the valve's position in the control room.

There is no AFW flow to the SGs, and the SG levels go down. In reality, criteria to start FR-H1FR-H1 are met. Because of the indicated flow from AFW pump 12, the plant computer will not show a red path on the heatsink status tree.

According to procedure FR-H1FR-H1, Bleed and Feed (F&B) shall be established when the WR level on any two SGs are less than 50%. (In this case F&B refers to *primary F&B,* with feed from safety injection and bleed through the PZR PORVs*, and not F&B through the SGs.*

**Establish AFW to SGs**

To establish AFW flow to the SGs, the crews can (in principle):

- dispatch a plant operator (PO) to check and close the open recirculation valve (feed SG B). In the simulation, one of the instructors acted as plant operator (and other plant personnel that are not in the control room) and could be reached on the phone.
- cross-connect AFW flow from pump 12 to SG A, C or D. The cross-connection can be done from the central control room.

However, if the crew sends a PO to the recirculation valve before start of Bleed and Feed, the PO will delay closing the valve until F&B is established.

Similarly, if the crew tries cross-connecting before F&B, the breaker for the power to the valve would open (as part of the study) and the valve would remain closed.

After F&B, the valve breaker would be reclosed by a PO. If the crew were to cross-connect after F&B, the valve would open.

**Steam generator tube rupture (SGTR)**

After F&B has been established, the crew will be able to establish AFW flow to one or several SGs. A tube rupture occurs in the first SG that is fed. The crew will want to fill a SG to be able to exit FR-H1FR-H1, and the tube rupture may be masked by AFW flow to the SG, as long as it is being fed. The leak size of the ruptured tube is about 500 GPM at 100% power, but the flow will depend on the differential pressure between the RCS and the ruptured SG.

There is initially no secondary radiation because there is a minimum steam flow. The blow down (BD) and sampling is secured because of the SI.

The crew is working in FR-H1FR-H1, and may have criteria for FR-P1 as a consequence of the F&B.

**Procedures that may be used**

- 0POP04-FW-0002 “Steam Generator Feed Pump Trip”
- 0POP05-EO-EO00 “Reactor Trip Or Safety Injection”
- 0POP05-EO-ES01 “Reactor Trip Response”
- 0POP05-EO-F003 “Heat Sink Critical Safety Function Status Tree”
- 0POP05-EO-FRH1 “Response to Loss of Secondary Heat Sink”
- 0POP05-EO-ES11 “SI Termination”
- 0POP05-EO-EO10 “Loss of Reactor or Secondary Coolant”
- 0POP05-EO-F004 “Integrity Critical Safety Function Status Tree”
- 0POP05-EO-FRP1 “Response to Imminent Pressurized Thermal Shock Condition”
- 0POP05-EO-EO30 “Steam Generator Tube Rupture”
- 0POP05-EO-EC31 “SGTR with loss of reactor coolant – subcooled recovery desired”

## US HRA simulator study, Scenario 2: Loss of CCW and RCP sealwater

**Plant technical information**

Component Cooling Water (CCW)
CCW pump 1A, powered by E1A
CCW pump 1B, powered by E1B
CCW pump 1C, powered by E1C

RCP sealwater
Charging pump 1A, powered by E1C
Charging pump 1B, powered by E1A
Positive Displacement Pump (PDP), powered by 1G8-bus (remains energized), cooled by air (doesn’t use CCW)

**Situation from start**

- All participating crew members in control room (Shift Manager, Unit Supervisor, Shift Technical Advisor and two Reactor Operators)
- The plant is operating at 100%
- Core burnup is 19,000 MWD/MTU (End of life)
- CCW pumps 1A and 1C are in service. Charging pump 1A is in service.
- B train out of service for CCW pump 1B and ECW pump 1B planned maintenance. The following equipment is unavailable:
    - CCWP 1B
    - ECWP 1B
    - Diesel Generator 12
    - AFWP 12

**Failure of distribution panel**

2 min Failure of Distribution Panel 1201, 120V AC Class Vital Distribution.

Consequences include failure of the controlling channels for:

- A and B SGs.
- PZR level control
- Rod control
- Nuclear Instrumentation (NIS)
- PZR pressure control

The crew needs to take the equipment above in manual control, in particular they need to take manual control of the feedwater flow to SG A and B.

**Reactor trip on high SG level**

The Feedwater regulation valve on SG A cannot be operated manually and remains fully open, feeding the SG. If the crew does not trip the reactor, there will be an automatic turbine trip on high SG level (87%), which causes a reactor trip.

**Loss of CCW and sealwater**

On R-trip Bus E1C will have bus lockout due to a bus fault. (The busbar is de-energized and the DG breaker cannot be closed.)

On R-trip CCW pump 1A breaker will trip due to failed, seized shaft.

There are no CCW pumps in service (B pump out of service, A pump tripped, C pump de-energized), and no charging pump running (A pump de-energized). If charging pump 1B is started, it will trip 2 minutes after reactor trip.

**Procedures that may be used**

- 0POP04-VA-0001 “Loss Of 120 VAC Class Vital Distribution”
- 0POP04-RC-0002 “Reactor Coolant Pump Off Normal”
- 0POP05-EO-EO00 “Reactor Trip Or Safety Injection”
- 0POP05-EO-ES01 “Reactor Trip Response”
- 0POP09-AN-02M3 “Annunciator Lampbox 2M03 Response Instructions” (page 10, CCW PUMP 1A(2A) TRIP)
- 0POP09-AN-04M7 “Annunciator lampbox 4M07 Response Instructions” (page 3, RCP 1A(2A) SEAL WTR INJ FLOW LO)

## US HRA simulator study, Scenario 3: Steam Generator Tube Rupture

**Situation from start**

- All participating crew members in control room (Shift Manager, Unit Supervisor, Shift Technical Advisor and two Reactor Operators)
- The plant is operating at 100%

- Core burnup is 19,000 MWD/MTU (EOL)

**Steam generator tube rupture**

About 1 min after the start of the scenario, a tube rupture occurs in steam generator C. The leak size is about 500 GPM at 100% power.

**Procedures**

- 0POP05-EO-E000 “Reactor Trip Or Safety Injection”
- 0POP05-EO-EO30 “Steam Generator Tube Rupture”

## A.4.2 Item 4b of Information Package –HFEs for Simulator Study

**Item 4B. HFEs for Simulator Study**

**Overview.** There are 5 HFEs in 3 scenarios. Scenario 1: HFEs 1A, 1B, 1C; Scenario 2: HFE 2A; Scenario 3: HFE 3A. (For detailed scenario descriptions, see item 4A.)

**Scenario 1**: Total Loss of Feed Water (LOFW), followed by Steam Generator Tube Rupture (SGTR)

In this scenario, the time at which the reactor is tripped will impact the time available to initiate Bleed&Feed (F&B) before core damage (CD). If the crews trip within approximately 30 - 45 seconds of the loss of feed water, they will have approximately 45 minutes before CD. If they fail to manually trip, the plant will trip automatically on low-low SG NR level (20%) approximately 50-60 seconds after the loss of feed water. If the plant trips on low-low SG level, the crews will have about 13 minutes to initiate F&B to avoid CD. According to FR-H1FR-H1, F&B shall be initiated when WR level on any two SGs are less than 50%. This criterion should be reached approximately 2 to 2.5 minutes after the LOFW. Please estimate the following probabilities:

**HFE 1A**: The probability of failing to establish F&B within 45 minutes of the reactor trip, given that the crews initiate a manual reactor trip before an automatic reactor trip.

**HFE 1B**: The probability of failing to establish F&B within 13 minutes of the reactor trip, given that the crews do not manually trip the reactor before an automatic reactor trip occurs.

The actions to start Bleed & Feed include:

- Actuate Safety Injection
- Open both of the PZR PORVS

Per the description of Scenario 1, if the crews successfully initiate F&B, they will be able to establish AFW to one or several SGs. However, if they do so, an SGTR will occur in the first SG that is fed. For the next HFE, assume that crews are successful in establishing feed-andbleed (F&B). Please estimate the following probability.

**HFE 1C**: Failure of crew to isolate the ruptured steam generator and control pressure below the SG PORV setpoint to avoid SG PORV opening.

The time window to perform the required actions is estimated to be approximately 40 minutes.

The actions include:

- Isolate the ruptured SG (feedwater and main stream isolation valves closed)
- Maintain SG pressure below the setpoint by cooling down the RCS (cooling the secondary by dumping steam and depressurizing the RCS).

----------------

**Scenario 2**: Loss of RCP sealwater (loss of CCW due to failure of distribution panel DP1201 and CCW pump 1B out of service, along with reactor trip)

**HFE 2A**: Failure of the crews to trip the RCPs and start the Positive Displacement Pump (PDP) to prevent RCP seal LOCA

Success requires that the crew:

Trip the RCPs after the loss of CCW and start the PDP to provide seal injection before sealwater inlet or lower sealwater bearing temperatures are greater than 230 degrees (per ES01 Step 6 or 0POP04-RC-0002: Reactor coolant pump off-normal) to avoid potential (not necessarily immediate) RCP seal LOCA. Time to reach 230 degrees is about 7-9 minutes from loss of CCW.

--------------------------------

**Scenario 3**: Basic SGTR

**HFE 3A**: Failure of crew to isolate the ruptured steam generator and control pressure below the SG PORV setpoint before SG PORV opening

The time window to perform the required actions is estimated to be 2 to 3 hours.

The actions include:

- Isolate the ruptured SG (feedwater and main stream isolation valves closed)
- Maintain RCS pressure below the setpoint by cooling down the RCS (cooling the secondary by dumping steam and depressurizing the RCS).

## A.5 Item 5 of Information Package – Plant Procedures

Not included in this appendix

## A.6.1 Item 6a of Information Package –Form A for HRA Team Responses

**Form A for HRA Team Responses**

**Instructions.** For each HFE, fill out a Form A. There are 5 HFEs in the 3 scenarios: HFE 1A, 1B, 1C, 2A, and 3A (see item 5 for the definitions). ***Please make sure that each page shows your team ID, method name, the HFE ID (in the header), and has a page number.***

item 1) HEP = _________(provide a mean and uncertainty measures)

item 2) Provide a summary of the most influencing factors on the crews' behavior with respect to this HFE and why they are important. The description should be in terms of the PSFs, causal factors, and other influence characterizations explicitly identified through the HRA method being used.

Factors relevant to the success and/or failure of the HFE should be described, but of particular interest are the factors that may drive the crews to fail. The discussion should reflect the basis for the HEP obtained for the HFE.

This discussion should be in terms of the "factors" or characterizations explicitly identified as important from the application of the HRA method. Do use the terminology of the HRA method.

item 3) Provide a qualitative assessment discussing (a) the perceived difficulty or ease that you predict the crew will have in performing the action of interest and (b) why the action should be easy or difficult, based on insights from using the HRA method.

Explicitly discuss the difficulties associated with the HFE in operational or scenario-specific terms. How will the driving factors be manifested in the crews' performances? If there are specific observations you expect to see in the crews' performances, such observations should be stated here.

The statements you make here may be conditional. Crews with characteristic ___ would be expected to ______.

You may additionally include in your assessment other predictions that are not directly based on the application of the HRA method. Please identify clearly these parts of your assessment and discuss.

### A.6.2 Item 6b of Information Package –Form B for HRA Team Responses

## Form B for HRA Team Responses

**Form 6B:** **HRA Diary**

> One goal of the 2010 HRA study is to understand how variability in HRA results is related to the interaction between method factors and analyst factors. To understand the strength and weaknesses of HRA methods, we need to consider both the analysis *results* and the analysis *process* (how the method was applied).
>
> In particular, we are interested in how the qualitative analysis is performed and how it is used as the basis for the quantification process. In some cases this may not be as explicitly specified as the quantification process and we are interested in the process used, whether it is specifically part of the method or not.

**Instructions.** This document contains forms for documenting the analysis process. Please use them throughout your analysis to create a continuous record of your work.

Form 6B-1, **Analysis Plan**
Complete this at the start of your work. Describe the team composition, major phases of the analysis, and time allocations.

Form 6B-2, **HRA Diary**
Complete this form for each analysis session. What was the goal of the session, what was done, what information was used and how, what difficulties were encountered?

Form 6B-3, **Site Visit**
Complete this form before and after the site visit. What are your goals for the visit? What information are you seeking, what information did you obtain?

***Please make sure that each page shows your team ID and method name.***
**Form 6B-1: Analysis Plan**

ANALYSIS TEAM

*Please provide a short description of the analysis team*, including names, roles and professional experience (professional background and work experience, experience as HRA analyst, experience with the specific method used).

| |
|---|
| Jane Smith (JS). Analysis leader and Human Factors expert. Has been working in HRA for 12 years, mainly using <method>. Besides HRA, I also work on xxx, mainly focusing on yyy. [...] |
| John Doe (JD). Process expert. John has experience in [...] |
| John Hancock (JH). [...] |

ANALYSIS PLAN AND SCHEDULE

*Please describe the major phases of your analysis*. What will be done and by whom? Approximately how much time is budgeted for each phase?

| |
|---|
| Phase 1: JS and JD will work on [...]. This will involve [...] and will probably require half a day of [...] and some telecons to clarify [...]. Total time for this phase (JS+JD) approx xx hrs. |
| Phase 2: JH will [...] |
| ... |

OTHER COMMENTS

*Please note anything else that might be relevant for understanding your work.*
Is this a 'textbook' application of the method? Have you modified or adapted the method (e.g. due to constraints, or due to the context of this study)? Is information from other guidance documents used to support the qualitative analysis? Are there any constraints on your work (e.g. the analysis team is distributed geographically; limitations on the number of meetings)? Do you foresee any challenges?

| |
|---|
| Add comments here [...] |

**Form 6B-2: HRA Diary**

*Complete one form for each analysis session*, i.e. every time you work on this project individually or as a group. Please keep all forms in one file in chronological order, and append any working documents generated during the session

| | | |
|---|---|---|
| Session# and date | #1, May 15 | (*increment session# for each work session)* |
| Analysis phase | _2__ | (*corresponds to analysis plan in form 6B-1)* |
| HFE | _1A&1B_ | (*if applicable)* |

PARTICIPANTS AND DURATION

JS (2hrs), JD on the phone for half an hour.

GOALS

*What was the purpose of the session*, what did you aim to achieve?

PSF identification. This was the second session on PSFs. We specifically wanted to look at xxx and yyy for HFE-1A.

SUMMARY & COMMENTS

*Summary of what happened in the session*. What critical decisions (e.g., PSFs, failure mechanisms) were made and how did the method guide/fail to guide these decisions? What information was used and how? Was expert judgment used, and how? Was there disagreement among the team members? Did the method provide adequate support and guidance? Did you achieve the goals for this session (if not, why)? Please note any challenges (e.g. problems understanding or modeling the scenario; lack of data or process expertise; difficulties using the method).

JS went over the notes from the site visit to identify xxx. Clarification was needed on yyy, so called JD to discuss zzz. Eventually resolved this by [...]. Took longer than expected, could have saved time if [...]. Still need more input on [...] to verify that [...]

Please remember to append any working documents generated during the session, e.g. notes, tables, diagrams

**Form 6B-3: Site visit**

PARTICIPANTS AND GOALS

*Please complete this section before the site visit*. Who is involved? What information are you planning to collect during the visit, and why (e.g. clarification of the scenario; work practices; etc)? Please describe any method-specific guidance for data collection during the site visit.

| |
|---|
| JS and JD will go on the site visit. We need to get information on the following issues: [...] |

INFORMATION OBTAINED

*Please complete after your visit*. What have you learned? Did you obtain all the necessary information? Would you consider this a 'typical' site visit, or did you obtain more / less information than normal. What worked well and why? Did you encounter any difficulties or surprises?

| |
|---|
| [...] |

COMMENTS AND INSIGHTS

*Please note any comments regarding the process of on-site information gathering*, e.g. expertise necessary to conduct the interviews; the structure and content of an effective interview; and guidance provided by the method (i.e. was it easy to identify which information was required based on the method, or did you have to rely mainly on your professional expertise)?

| |
|---|
| [...] |

Please append any working documents that were generated during the site visit, including notes (scan handwritten notes if necessary)

# APPENDIX B EXPERIMENTAL PROCEDURE FOR EXPERIMENTERS

This appendix describes the procedures under which the data was collected in the simulator. This guidance was discussed with the experimenters prior to the data collection in the simulator.

**General rules**

- Please, *no talking* in the *control room* or in the *simulator booth* during the scenario.
- Mobile phones must be switched off
- Observers that do not have one of the roles described below may observe from the observer room on the second floor.
- Observers in the control room must avoid disturbing the crew members
    - Avoid eye contact
    - Avoid standing close to crew members
    - Avoid visibly checking areas of the control room where important indications are present
    - Avoid making facial expressions

**Start and stop of scenario procedure**

Before start of scenario, Experiment leader will:

1. Verify with booth instructor:
    a. Correct scenario is loaded according to schedule
    b. Process data is being logged
    c. Trainer 1 is ready for start
2. Verify Halden interviewer has started recording of CSF in simulator booth
3. Verify with floor-instructor:
    a. Control room prepared for scenario (procedures clean and switch check done)
    b. Sign on simulator(s) door to avoid that someone enters
    c. Crew members present (Scenario 1: STA outside simulator)
    d. Turnover sheet handed to crew (including STA in sc 1)
4. Verifiy trainers and observers have taken their places
5. Remind experimental staff and crew members that mobiles must be turned off
6. Check that experimental staff in control room is ready for start
7. Check that crew members are ready to take the watch
8. Announce that start is coming

<u>Start of scenario:</u>

1. Experimental leader asks trainer 1 to start simulation
   - *Trainer 1 starts the simulator*
   - *Trainer 1 resets the clock*
   - *Trainer 1 announciates that the the crew has the watch*

<u>End of scenario:</u>

1. Experiment leader decides when to stop scenario and communicates this to Trainer 1
   - *Trainer 1 announces stop of scenario and stop simulation*
   - *Trainer 1 checks against schedule which crew is present, and saves data on memory sticks (Process parameters log, Annunciator Windows Logger and Simulator Action Monitor Log)*
2. Experiment leader tells the crew members at what time the interviews will start, and who will go where (LOR classroom/ Critique room)

**Roles and responsibilities**

| ROLE | RESPONSITILITIES |
|---|---|
| Experiment leader | • Lead the start and stop of scenario according to "Start and stop of scenario procedure" |
| Booth-instructor | • Load scenario in simulator booth<br>• Record simulator session on video<br>• Reset simulator time = 0 at start of scenario<br>• Announce start of the scenario after sign from Experimental leader<br>• Run simulator and initiate malfunctions<br>• Respond to phone calls from crew members and act as called out personnel<br>• Scenario 1: Send STA to control room 5 minutes after he or she is called<br>• Scenario 1: Note the initial breakflow (RCS-SG) and the current RCS and ruptured SG pressures (one time)<br>• Collect the following data during the scenario<br>  o Communications over phone<br>  o Timeline with actions<br>• After Experiment leader has decided to stop scenario, announce that the scenario has ended<br>• Save the following data after each scenario to two memory sticks (DATA_1 and DATA_2):<br>  o Process parameter logs<br>  o Annunciator Windows Logger<br>  o Simulator Action Monitor Log<br>• After saving data, participate in Halden interview<br>• After interview, hand over notes (timeline, communicaitons outside CR and FR-P condition) to Halden interviewer |

| | |
|---|---|
| Floor-instructor | • Put sign on simulator door(s) to avoid that someone enters<br>• Prepare control room for scenario, switch check (do not wipe procedures until interviews finished)<br>• Fetch the crew members.<br>• Hand out turnover sheet to the crew before each scenario<br>• Ask STA to wait outside simulator in scenario 1<br>• Collect the following data during the scenario<br>  o Pre-defined actions (template)<br>  o Start and stop of procedures<br>  o Timeline/ Events<br>• After scenario, participate in Halden interview<br>• After interview hands over notes (timeline, predefined actions and procedures) to Halden interviewer<br>• Wipe procedures clean after each scenario, after both interviews are completed.<br>• Participate in the crew feedback after all scenarios |
| INL interviewer(s) | • Lead interview with additional operator, STA and reactor operators after the scenario |
| Halden interviewer(s) | • Record FR-P condition (Sc. 1) and RCP seal temperature (Sc. 2) as well as the experimental clock time on DVD in booth<br>• Collect the following data during the scenario<br>  o Time line/ Events<br>  o Mark events for questions in interview<br>• Lead interview with Trainer 1, Trainer 2, Shift Manager and Unit supervisor after the scenario<br>• Mark notes from interview with Crew (A, B, C, D) and Scenario (1, 2, 3) and save them on computer and memory stick (DATA_3). |
| NRC observer | • Note general observations and collect data on FR-P entry condition (latter only in scenario 1). |

# APPENDIX C NRC CREW PERFORMANCE ASSESSMENT FORM

Adapted from NUREG-SR1020r9, Form ES-604-2

**Diagnosis of Events and Conditions Based on Signals or Readings**

Did the crew—

(a) recognize off-normal trends and status?

| 3 | 2 | 1 |
|---|---|---|
| Recognized status and trends quickly and accurately | Recognized status and trends at the time of, but not before, exceeding established limits. | Did not recognize adverse status and trends, even after alarms and annunciators sounded. |

(b) use information and reference material (prints, books, charts, emergency plan implementation procedures) to aid in diagnosing and classifying events and conditions?

| 3 | 2 | 1 |
|---|---|---|
| Made accurate diagnosis by using information and reference material correctly and in a timely manner. | Committed minor errors in using or interpreting information and reference material. | Failed to use, or misused, or misinterpreted information or reference material that resulted in improper diagnosis. |

(c) correctly diagnose plant conditions based on control room indications?

| 3 | 2 | 1 |
|---|---|---|
| Performed timely and accurate diagnosis. | Committed minor errors or had minor difficulties in making diagnosis. | Made incorrect diagnosis, which resulted in incorrect manipulation of any safety control. |

**Understanding of Plant and System Responses**

Did the crew—

(a) locate and interpret control room indicators correctly and efficiently to ascertain and verify the status/operation of plant systems?

| 3 | 2 | 1 |
|---|---|---|
| Each crew member located and interpreted instruments accurately and efficiently. | Some crew members committed minor errors in locating or interpreting instruments or displays. Some crew members required assistance. | The crew members made serious omissions, delays, or errors in interpreting safety-related parameters. |

(b) demonstrate an understanding of the manner in which the plant, systems, and components operate, including setpoints, interlocks, and automatic actions?

| 3 | 2 | 1 |
|---|---|---|
| Crew members demonstrated thorough understanding of how systems and components operate. | The crew committed minor errors because of incomplete knowledge of the operation of the system or component. Some crew members required assistance. | Inadequate knowledge of safety system or component operation resulted in serious mistakes or plant degradation. |

(c) demonstrate an understanding of how their actions (or inaction) affected systems and plant conditions?

| 3 | 2 | 1 |
|---|---|---|
| All members understood the effect that actions or directives had on the plant and systems. | Actions or directives indicated minor inaccuracies in individuals' understanding, but the crew corrected the actions. | The crew appeared to act without knowledge of or with disregard for the effects on plant safety. |

**Adherence to and Use of Procedures**

Did the crew—

(a) refer to the appropriate procedures in a timely manner?

| 3 | 2 | 1 |
|---|---|---|
| The crew used procedures as required and knew what conditions were covered by procedures and where to find them. | The crew committed minor failures to refer to procedures without prompting, which affected the plant's status. | The crew failed to correctly refer to procedure(s) when required, resulting in faulty safety system operation. |

(b) correctly implement procedures, including following procedural steps in correct sequence, abiding by cautions and limitations, selecting correct paths on decision blocks, and transitioning between procedures when required?

| 3 | 2 | 1 |
|---|---|---|
| The crew followed the procedural steps accurately and in a timely manner, demonstrating a thorough understanding of the procedural purposes and bases. | The crew misapplied procedures in minor instances, but made corrections in sufficient time to avoid adverse effects. | The crew failed to follow procedures correctly, which impeded recovery from events or caused unnecessary degradation in the safety of the plant. |

(c) recognize EOP entry conditions and perform appropriate actions without the aid of references or other forms of assistance?

| 3 | 2 | 1 |
|---|---|---|
| The crew recognized plant conditions and implemented EOPs consistently, accurately, and in a timely manner. | The crew had minor lapses or errors. Individual crew members needed assistance from others to implement procedures. | The crew failed to accurately recognize degraded plant condition(s) or execute efficient mitigating action(s), even with the use of aids |

**Control Board Operations**

Did the crew—

(a) locate controls efficiently and accurately?

| 3 | 2 | 1 |
|---|---|---|
| Individual operators located controls and indicators without hesitation. | One or more operators hesitated or had difficulty in locating controls. | The crew failed to locate control(s), which jeopardized system(s) important to safety |

(b) manipulate controls in an accurate and timely manner?

| 3 | 2 | 1 |
|---|---|---|
| The crew manipulated plant controls smoothly and maintained parameters within specified bounds. | The crew demonstrated minor shortcomings in manipulating controls, but recovered from errors without causing problems. | The crew made mistakes manipulating control(s) that caused safety system transients and related problems |

(c) take manual control of automatic functions, when appropriate?

| 3 | 2 | 1 |
|---|---|---|
| All operators took control and smoothly operated automatic systems manually, without assistance, thereby averting adverse events. | Some operators delayed or required prompting before overriding or operating automatic functions, but avoided plant transients where possible. | The crew failed to manually control automatic systems important to safety, even when ample time and indications existed. |

**Crew Operations**

Did the crew members—

(a) maintain a command role?

| 3 | 2 | 1 |
|---|---|---|
| The crew took early remedial action when necessary. | In minor instances, the crew failed to take action within a reasonable period of time. | The crew failed to take timely action, which resulted in the deterioration of plant conditions. |

(b) provide timely, well-planned directions to each other that facilitated their performance and demonstrated appropriate concern for the safety of the plant, staff, and public?

| 3 | 2 | 1 |
|---|---|---|
| Supervisor's directives allowed for safe and integrated performance by all crew members. | In minor instances, the supervisors gave orders that were incorrect, trivial, or difficult to implement. | The supervisor's directive(s) inhibited safe crew performance. Crew members had to explain why order(s) could not or should not be followed. |

(c) maintain control during the scenario with an appropriate amount of direction and guidance from the crew's supervisors?

| 3 | 2 | 1 |
|---|---|---|
| Crew members stayed involved without creating a distraction, the crew members anticipated each other's needs, and the supervisors provided guidance when necessary. | Crew members had to solicit assistance from supervisors or each other, interfering with their ability to carry out critical action(s). | Crew members had to repeatedly request guidance. The crew failed to verify successful accomplishment of orders. |

**Crew Operations (Continued)**

Did the crew members—

(d) use a team approach to problem solving and decision making by soliciting and incorporating relevant information from all crew members?

| 3 | 2 | 1 |
|---|---|---|
| Crew members were involved in the problem solving and decision making processes for effective team decision making. | At times, crew members failed to get involved in the decision making process when they should have, detracting from the team-oriented approach. | The crew was not involved in making decision(s). The crew was divided over the scenario's progress, and this behavior was counter-productive. |

**Communications**

Did the crew—

(a) exchange complete and relevant information in a clear, accurate, and attentive manner?

| 3 | 2 | 1 |
|---|---|---|
| Crew members provided relevant and accurate information to each other. | Crew communications were generally complete and accurate, but sometimes needed prompting, or the crew failed to acknowledge the completion of evolutions, or to respond to information from others. | Crew members did not inform each other of abnormal indication(s) or action(s). Crew members were inattentive when important information was requested. |

(b) keep key personnel outside the control room informed of plant status?

| 3 | 2 | 1 |
|---|---|---|
| Crew members provided key personnel outside the control room with accurate, relevant information throughout the scenarios. | In minor instances, the crew needed to be prompted for information and/or provided some incomplete/inaccurate information. | The crew failed to provide needed information. |

(c) ensure receipt of clear, easily understood communications from the crew and others?

| 3 | 2 | 1 |
| --- | --- | --- |
| The crew requested information/clarification when necessary and understood communications from others. | In minor instances, the crew failed to request or acknowledge information from others. | The crew failed to request needed information, or was inattentive when information was provided; serious misunderstandings occurred among crew members. |

# APPENDIX D SCENARIO 1 CREW BY CREW

## D.1 Scenario 1, Crew Q

### *Performance of Human Failure Events*

HFE1A: success. The crew manually tripped the reactor when they lost all feedwater, and had established bleed and feed 17 minutes after the reactor trip.

HFE1C: failed to meet 40 minute time criterion but PORV did not open. This crew did isolate the ruptured SG and control pressure (secure the HH safety injection pump) before the PORV opened, but did not complete the actions within 40 minutes of the SGTR. The level in the ruptured SG reached 100%, but the SG pressure was kept below the SG PORV setpoint. (Procedure E-30 was not entered.)

### *Scenario development and crew responses*

Normal text is based on observations and simulator logs.
*Texts in italics are based on interviews.*

**Reactor trip**
The crew manually tripped the reactor 37 seconds after total loss of heatsink.

*The Unit Supervisor (US) ordered the reactor trip when all feed pumps were tripped. The US knew that they couldn't recover.*

**Identification of total loss of feedwater and start Bleed and Feed**
The secondary reactor operators reported to the US that the SG level was going down in the only SG that has an indicated AFW-flow. The US ordered the RO to check the valve. The STA added that the WR-level was going down and the US ordered the STA to check criteria for FR-H. The crew transferred to FR-H1 (response to loss of secondary heat sink) 12 minutes after the reactor trip.

*The US made the decision to go to FR-H1 based on the level in SG B, verified by the level in the AFW storage tank. The crew was helped to identify the LOFW situation by the RO who reported not getting expected response (level) in SG B. The US immediately suspected the recirculation valve and used the STA and the RO to verify that they had no AFW flow.*

The US opened the CIP in FR-H1 but did not act from the F&B criteria. The US returned to the procedure. Three minutes after having started the procedure the US asked the primary RO if WR levels were above 50%, and ordered the RO to start SI when the answer was negative. The crew had established bleed and feed 17 minutes after the reactor trip, 20 minutes into the scenario. At that time the second lowest SG WR level was at 41% (figure 1). The water level level is a limiting condition to start F&B.

*When starting FR-H1 the US was not aware that the levels were already below 50%. The US read 50% on the CIP but didn't say that they were there. The US wanted to do step 1 in the procedure. The US ordered RCP trip and start of F&B from directions in FR-H1 step 2.*

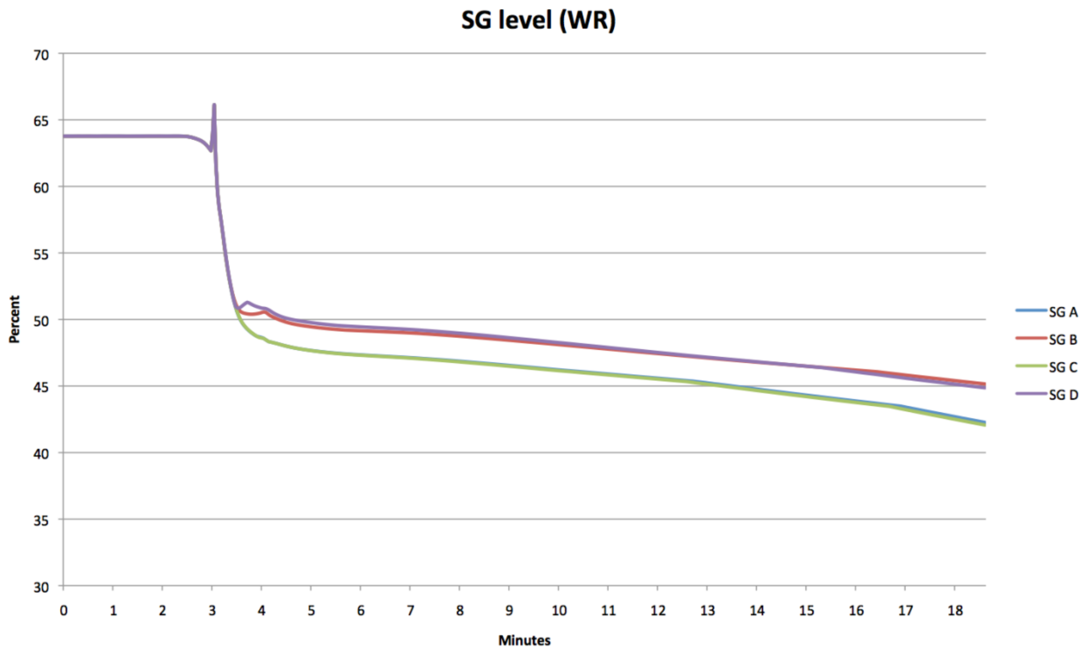



*Figure 1. SG level (Wide Range) from start of scenario to the time when the crew started Bleed and Feed*

**Reestablishment of AFW and start of SGTR**

The US directly suspected that the recirculation valve for AFW pump was open when an RO reported that the level was going down in the SG that had indicated flow. The RO tried to cross-connect the AFW, but it did not work (simulated, to force them to FR-H). At 20 minutes into the scenario, the plant operator was ordered to close the recirculation valve. When the recirculation valve was closed AFW started filling SG B at a flow rate of above 600 gpm (figure 2).

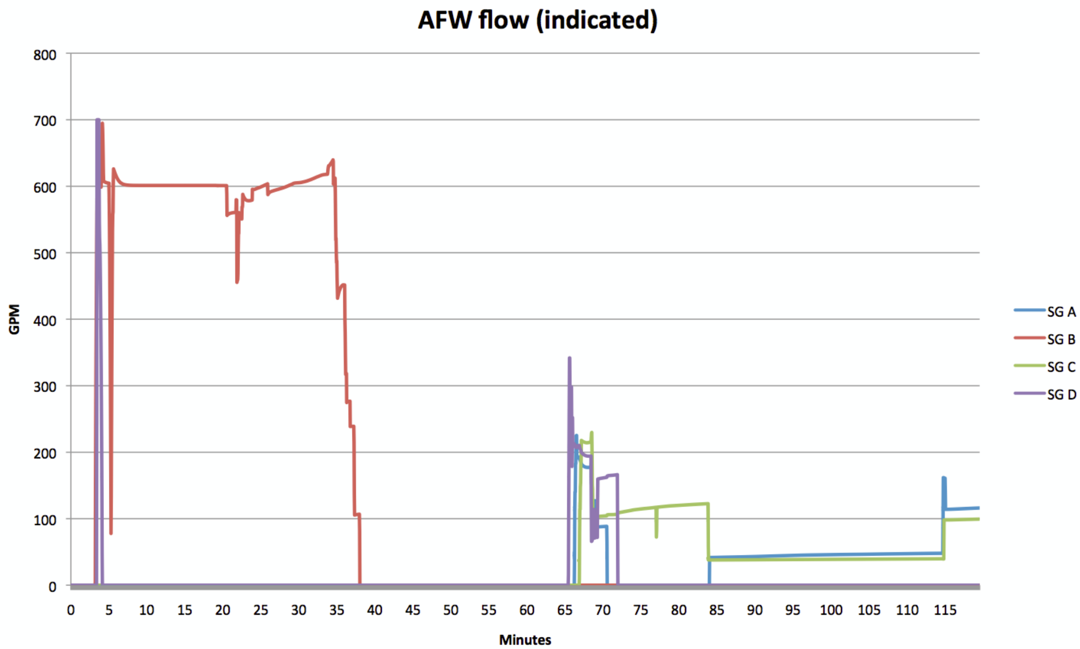



*Figure 2. Indicated AFW flow to the SGs. Before the recirculation valve on AFW pump 12 is closed (at 22 minutes) the indication of flow to SG B is false.*

At 22 minutes into the scenario, just a minute after having established an AFW flow to SG B, there was a tube rupture in the SG that was fed by AFW (B). When the tube rupture started the WR level in SG B was 36% and rising, and the levels in the other SGs were 29-31% and slowly going down.

About 30 minutes into the scenario, the STA told the US that the NR-level in SG B was 33% and rising. The US ordered a reactor operator to control the level. At 35 minutes the RO reduced and then stopped the AFW-flow to SGB.

It took a long time after AFW was stopped to SGB until it was established to the other SGs (figure 2) [THIS SENTENCE IS NOT CLEAR.]. After more than an hour into the scenario the US ordered a reactor operator to cross connect to feed the other SGs.

**Identification and isolation of SGTR**

When the tube rupture started at 22 minutes into the scenario, the crew was working in procedure FR-H1, verifying that they had AFW flow to SGB. They were filling SG B in order to stop F&B.

About 35 minutes into the scenario the RO stopped AFW to SG B, and a minute later the STA told the US that they had a red path on integrity, meaning that they would have to go to procedure FR-P if they were not in FR-H.

The US noticed the rising level in SGB after having ordered the secondary RO to control level in SGB. At 42 minutes, the US ordered the RO to evaluate if they had a tube leak. At the same time the US was working in FR-H1 with the STA and the primary RO to stop the high head safety injection (HHSI). At 44 minutes into the scenario, the secondary RO reported believing that they had a tube leak. The NR-level in SG B was above 90% and rising. The STA called for sampling. When the secondary RO reported that the level was 100% the US checked again with the RO that AFW was isolated. The US ordered the third RO to make a plant announcement about the tube rupture at 59 minutes.

*The crew identified the tube rupture based on the rising level in SG B after AFW flow was isolated, backed up later by radiation in the steam line. The crew managed to identify SGTR from checking levels, board awareness. The third RO and the STA pointed out the steam line rad monitor{THIS SENTENCE IS INCOMPLETE}.*

The US was aware of the tube rupture, and was concerned about increasing the SG pressure. According to procedure FR-H1, they needed to improve their subcooling to stop the last HHSI-pump. The crew stopped AFW to the ruptured SG. They did not enter E-30 as they were held up in FR-H1 and FR-P1 (respond to imminent reactor pressurized thermal shock condition). Even though the crew reached 100% level in the ruptured SG (figure 3 and 4) the pressure was below the SG PORV setpoint and they did not release radioactive water to the environment (figure 5) because they had secured all HHSI-pumps.

*When the crew identified the tube rupture, the priority of the US was to get through FR-H1 and to stop F&B. The US wanted to balance the RCS pressure without raising the SG pressure too high, to prevent the SG PORV from lifting. The SG was already isolated, the feed reg valve was closed, and the US had an RO verify steamline isolation (SLI). The US thought about adjusting the setpoint but it wasn't in any of the procedures they were in and since they were far from the setpoint it wasn't a concern. The concern was to control the RCS pressure to keep the margin from the lifting point.*

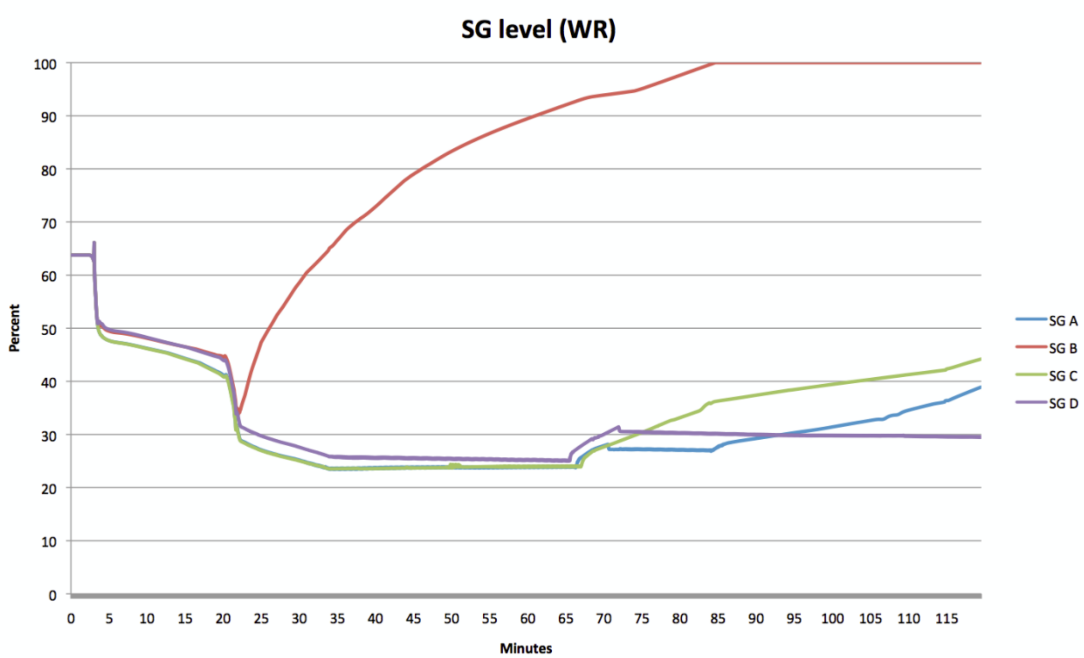



*Figure 3. Wide Range levels in the SGs.*

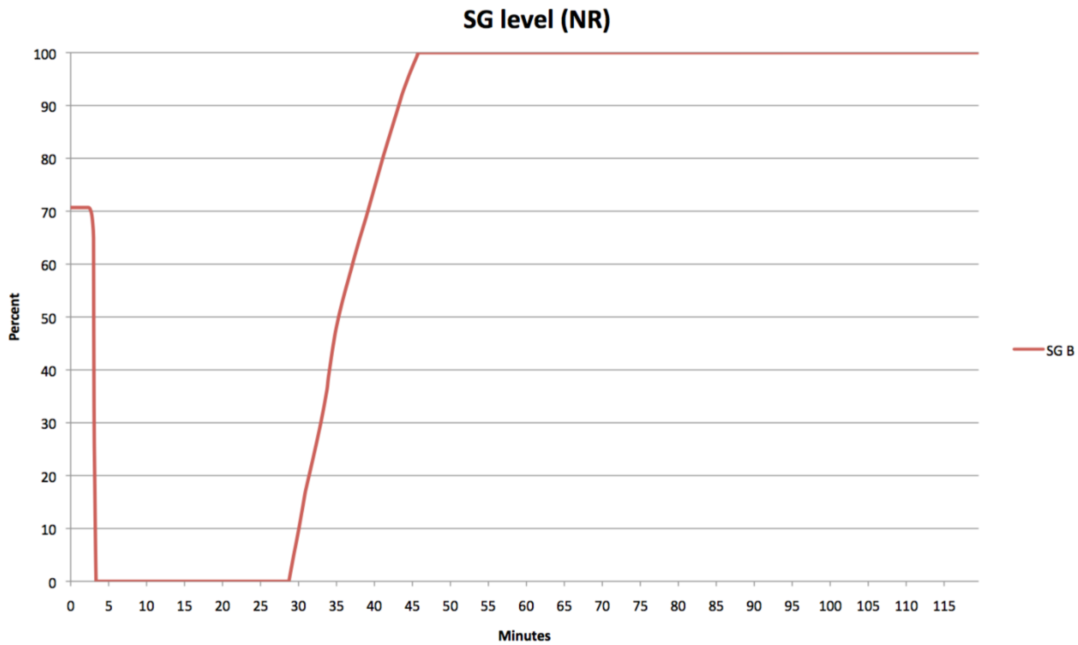



*Figure 4. Narrow range level in the ruptured SG.*

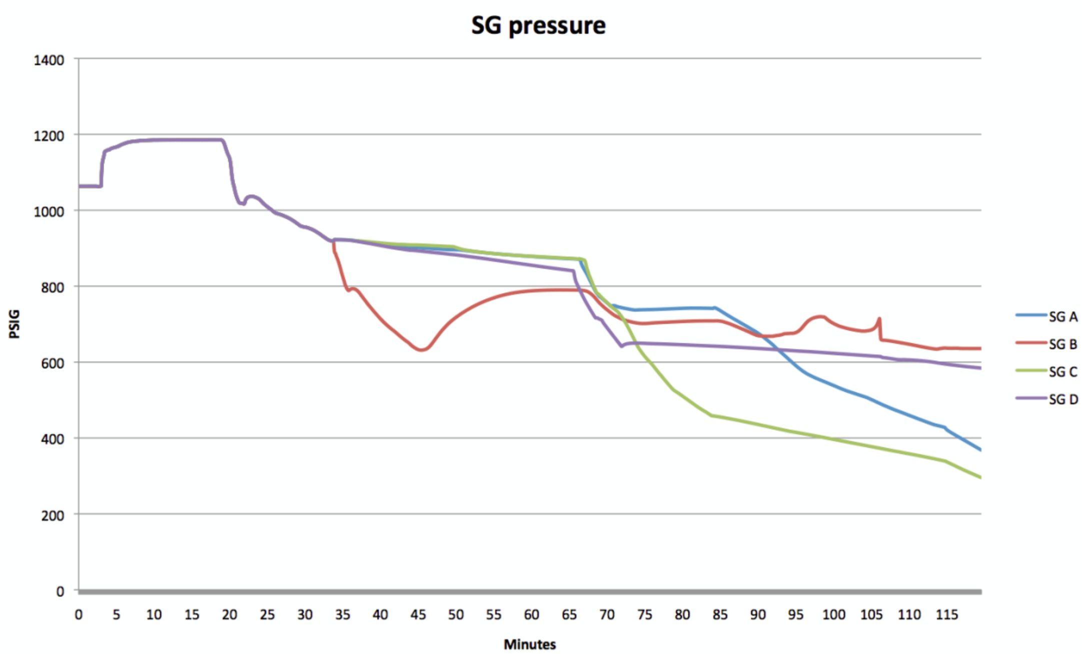



*Figure 5. Pressure in the SGs.*

At 1 hour 38 minutes, the US held a crew brief and explained that the goal was to have subcooling and stop the HHSI pump to be able to continue in FR-H1. At 1 hour 45 minutes the crew stopped the last HHSI-pump. They closed the PORVs and could exit FR-H1. When exiting they had an orange path on integrity and they started FR-P1 at 1 hour 50 minutes. They were still working in this procedure when the scenario was stopped at 2 hours 18 minutes.

## D.2 Scenario 1, Crew R

### Performance of Human Failure Events

HFE1A: success. The crew manually tripped the reactor when they lost all feedwater, and had established bleed and feed 16 minutes after the reactor trip.

HFE1C: failure. The crew isolated feedwater to the ruptured SG but the level in the ruptured SG reached 100% and the SG pressure reached the SG PORV setpoint. The SG PORV on the ruptured SG opened. (Procedure E-30 was entered.)

### Scenario development and crew responses

Normal text is based on observations and simulator logs.
*Texts in italics are based on interviews.*

**Reactor trip**
The crew manually tripped the reactor 29 seconds after the total loss of heatsink.

*The crew tripped the reactor when they had lost all feedwater because they knew that they couldn't maintain SG level. The US said that they had to trip the reactor and the shift manager (SM) concurred.*

**Identification of total loss of feedwater and start Bleed and Feed**
Seven minutes into the scenario the secondary RO said that SG B was not going up. The US and the SM were busy trying to restart AFW pumps. The RO continued to monitor and report the WR levels in the SGs. At 9 minutes, the STA arrived in the control room and was ordered by the US to check the critical safety function staus tree (CSFST) for heatsink, because they had indications of AFW flow but suspected that they had no flow. The US ordered the SM to take out FR-H1 to look at it. At 11 minutes into the scenario, the US ordered the secondary RO to check the recirculation valve. Three minutes later the plant operator called to inform the crew that the valve was open. When the plant operator (PO) could not open the valve the STA said that they had a red path on heatsink. The crew transferred to FR-H1, 13 minutes after the reactor trip.

*The crew decided to move on in ES-01 because they had indicated flow but they were monitoring the level in the B SG because it was going down. They sent a PO to check the recirculation valve and determined that they didn't have heat sink based on the information from the PO that the valve was open. When the PO couldn't open the valve the US decided to go to FR-H1. Previous training helped the identification of loss of feedwater (LOFW). The crews are trained to use the CSFST and to look at both flow and level indications to see if they meet guidelines. The crew did not transfer to FR-H1 earlier because they had indication of AFW flow, and they wanted to make some evaluation before starting a bleed-and feed (F&B). They were also aware of the SG levels and thought that 48% would be okay.*

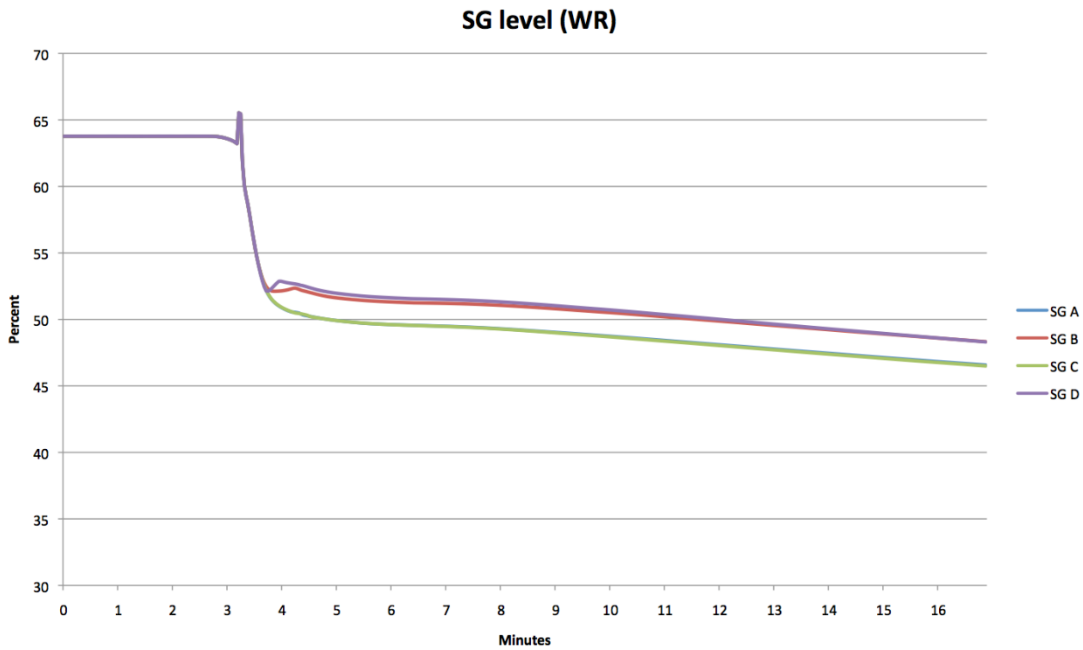



*Figure 1. SG level (Wide Range) from start of scenario to the time when the crew started Bleed and Feed*

The US read FR-H1 step 1 and then read the CIP. The STA said that they had less then 50% WR in all SGs. The crew had established F&B 16 minutes after the reactor trip, 18 minutes into the scenario. At that time, the second lowest SG WR level (limiting in the start F&B criterion [DO NOT KNOW WHAT THIS MEANT]) was 46% (figure 1).

*The crew started F&B as directed from FR-H1 CIP. When starting FR-H1 they already knew that the level in SG B was below 50% but the US wasn't sure of the other levels until checking them according to the CIP.*

**Reestablishment of AFW and start of SGTR**
19 minutes into the scenario, just after the crew had started F&B, the plant operator closed the recirculation valve over AFW to SG B (as ordered before F&B). AFW started filling SG B at a flow rate greater than 600 gpm (figure 2).

At 20 minutes into the scenario, 1 minute after having established an AFW flow to SG B, there was a tube rupture in the SG that was fed by AFW (B). When the tube rupture started, the WR level in SG B was 42% and rising, and the levels in the other SGs were 39% and decreasing slowly.

At 26 minutes into the scenario, the primary RO reported to the US that the level in SG B was more than 14%, allowing the crew to continue in FR-H1 to stop F&B. The crew started stopping HHSI-pumps, and at 35 minutes they cross-connected AFW to SG A, C and D, and stopped the AFW to SG B. The crew stopped two HHSI-pumps but did not stop the last one.

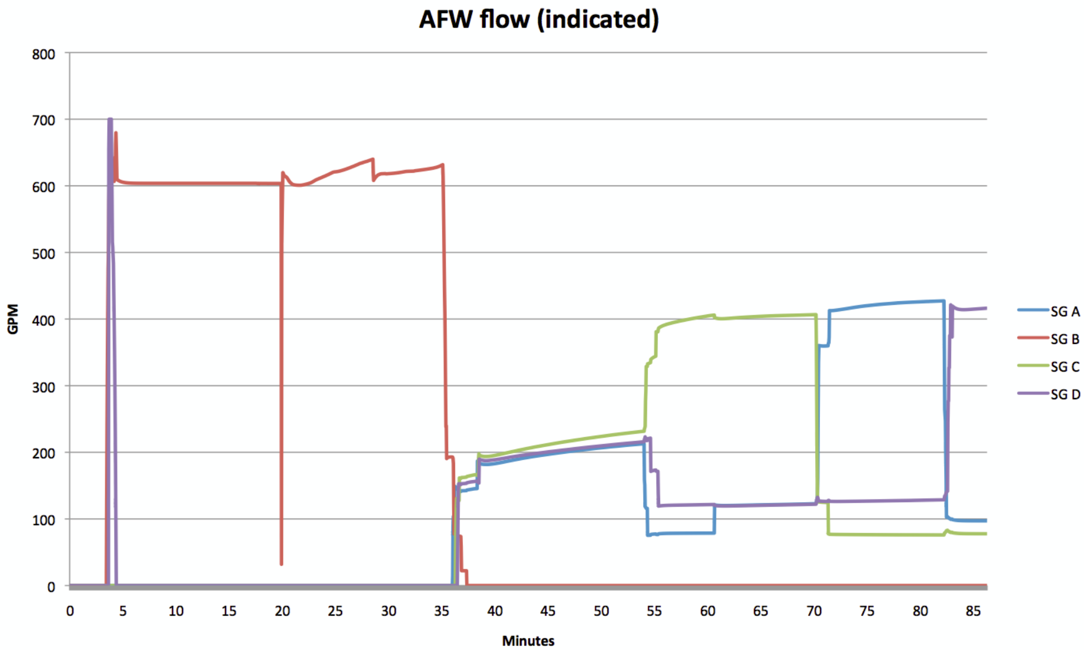



*Figure 2. Indicated AFW flow to the SGs. Before the recirculation valve on AFW pump 12 is closed (at 19 minutes) the indication of flow to SG B is false.*

**Identification and isolation of SGTR**

When the tube rupture started at 20 minutes, the crew was working in procedure FR-H1 with reestablishing AFW to be able to stop F&B.

At 37 minutes into the scenario, the crew stopped AFW to SG B. At that time they had 75% WR level and 87% NR level in the ruptured SG. The US and SM worked in FR-H1, preparing to stop the last HHSI-pump and discussing the procedure with the STA. The secondary RO noticed that the level in SG B continued to rise and consulted the primary RO, two minutes after the AFW to SG B was stopped.

At 41 minutes the secondary, RO told the US that the level in SG B was still rising, and repeated it since the US, SM and STA were busy in the FR-H1 steps. The STA checked radiation and said that they only had activity in the containment (expected after F&B, and meaning no indication of SGTR).

The US and the SM decided not to stop the last HHSI-pump in FR-H1 step 28 based on having an active loop hot leg temperature greater then 405 °F. The mistake they made was thinking that they had an active loop. They did not verify this, which normally is the STA's task. Natural circulation is verified by five criteria of which one is core exit temperature stable or lowering. The core exit temperature was rising when they were in step 28 at about 43 minutes (Figure 3). As a result they continued to the next step instead of stopping the last HHSI-pump. Leaving one HHSI-pump running had serious consequences for the RCS pressure and the pressure in the ruptured SG. At this time the US and the SM were not aware that they had a tube rupture, and they were following procedures as they interpreted the situation.

(For comparisons sake: When the other three crews reached FR-H1 step 28 right column the hot legs temperatures were below 405 °F for one crew, for another crew only the ruptured SG loop had a hot leg temperature above 405 °F, and for the third crew two not ruptured loops were above 405 °F. This means that two of the other crews also had to evaluate if they had an active loop.)

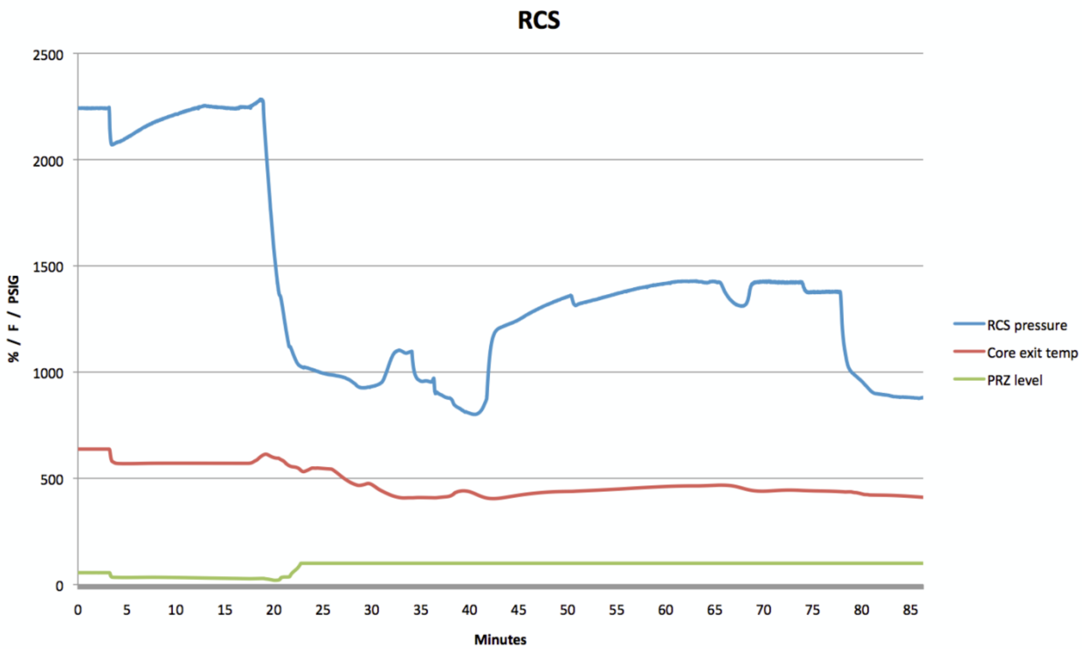



*Figure 3. RCS pressure, Core exit temperature and PZR level.*

At 44 minutes, they transferred from FR-H1 step 29a right column to E-10 (LOCA procedure). The secondary RO reported that the B SG was full.

When exiting FR-H1 the crew had a yellow path on FR-P, which was different from the other three crews that had to start FR-P1. All crews had the FR-P entry criterion switching between yellow, orange and red in this part of the scenario except the crew R. At the time the crew R left FR-H1 the CSFST indicatin on FR-P was okay.

The US held a brief and concluded that the plan was to go through E-10 and cool down. At 46 minutes, they opened the CIP, and then took a crew conditional brief. The SM said that they needed to watch SG B and the primary RO reported that they had steamline radiation on SG B. When the brief was over three minutes later the US decided to go to E-30 (SGTR procedure). They didn't start step 1 in E-10 but went directly to E-30 from the CIP.

*The crew became aware of the tube rupture during the brief when they had left FR-H1. They had already discussed the rising level and the secondary RO said the radiation in the steam lines was rising. Good board awareness from the RO helped the crew with the identification. The identification could have been faster if the main steam isolation valves (MSIVs) weren't closed preventing radiation. (The instructor adds that throtteling AFW earlier could also have helped, giving the crew more time to react to the rising SG level.)*

The crew started procedure E-30 49 minutes into the scenario and did the isolation steps, including adjusting the SG power operated relief valve (PORV) setpoint. They still had one HHSI pump running, and at 58 minutes the secondary RO said that the pressure in the ruptured SG was 1230 psig and rising. At 63 minutes into the scenario, the SG PORV on SG B started to open. At the same time the crew started the depressurization step in E-30 and established auxiliary spray. The RCS pressure started to go down at 65 minutes, just after the US had ordered maximizing pressurizer spray. At the same time the SG PORV was releasing (figure 7), decreasing the SG pressure and the RCS pressure.

At 67 minutes, just after the crew had closed the auxiliary spray valves, the primary RO detected steam flow from SG B and showed the rest of the crew. The US, SM and STA were at the same time in E-30 step 21 to evaluate if they could stop SI. They had decreasing pressure in the reactor coolant system (RCS), which led them to transfer to EC-31 (SGTR with loss of reactor coolant). They did not however wait to see if the pressure came back. Nor did they do the connection that the open SG PORV caused the RCS pressure to drop. The primary RO called health physics to inform them about the steam flow from the ruptured SG, and at the same time the US pulled out procedure EC-31. At 68 minutes, the STA said that the SG pressure was going up again and the crew saw that the SG PORV opened.

EC-31 is the procedure for tube rupture in combination with a primary leak. In this situation the RCS pressure was going down because of the open SG PORV and not because of a primary leak, and if the crew would have waited a little longer they would have seen the pressure go up again. If the crew had continued in E-30 the next step would have been to stop the HHSI-pump, which would have reduced the RCS pressure. The crew literally followed the procedure and went to EC-31 and continued to release through the SG PORV for another 10 minutes. At 78 minutes, the crew started RCS cooldown with intact SGs and the leak through the SG PORV stopped.

*The crew incorrectly assumed that they had active loop (natural circulation) without verifying it with the STA. In the interview after the scenario the trainer pointed out to them that the correct thing would have been to judge the loop not active and to continue in FR-H1 step 28 to stop the last HHSI-pump. If they would have stopped the pump they would not have lifted the SG PORV later.*

*The crew isolated the SG B according to procedure E-30, but were kicked out to EC31 before turning off the high head SI pumps. The RCS pressure was going down and the crew didn't know if the SG PORV was lifting at that time. The crew couldn't depressurize because everything was solid.*

*A trainer reviewing the scenario said that they went to EC-31 because of lack of understanding of step 21 in E-30. The ROs had noticed the steam flow and the STA should have been able to explain to the crew that the decreasing RCS pressure was caused by the SG PORV and not a primary leakage. The crew appeared to be following procedure literally without understanding the intention of the step.*

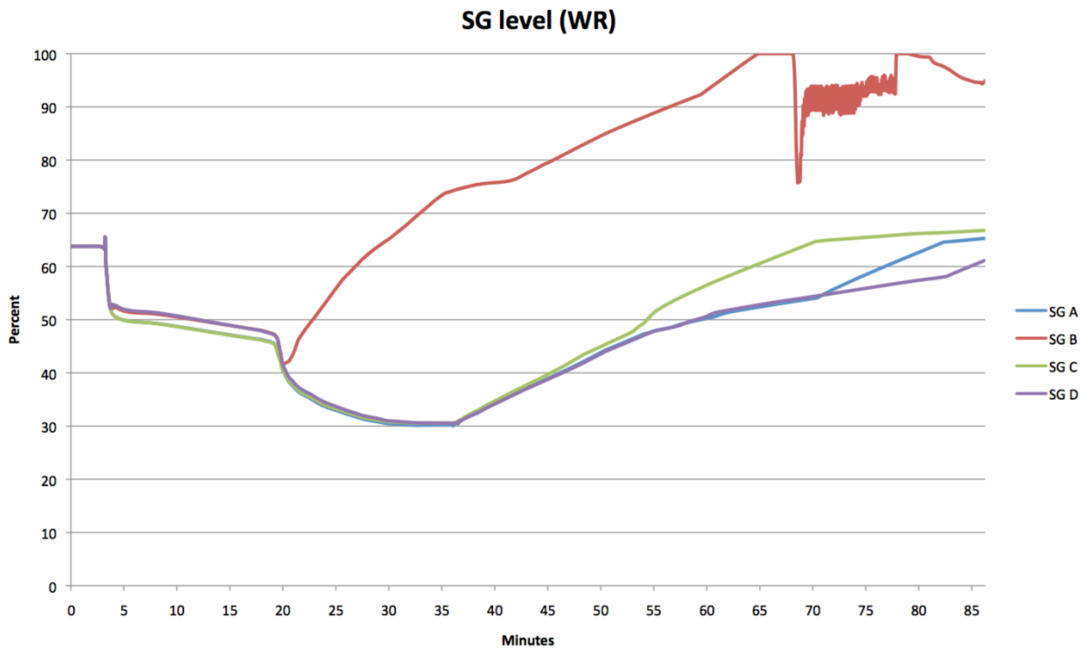



*Figure 4. Wide Range levels in the SGs.*

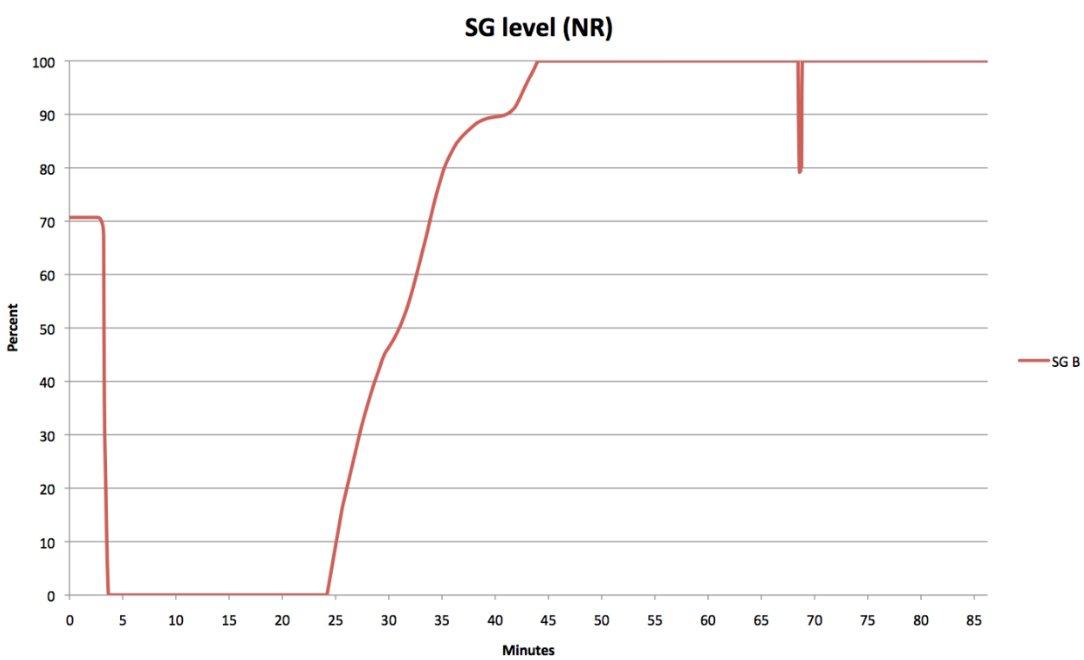



*Figure 5. Narrow range level in the ruptured SG.*

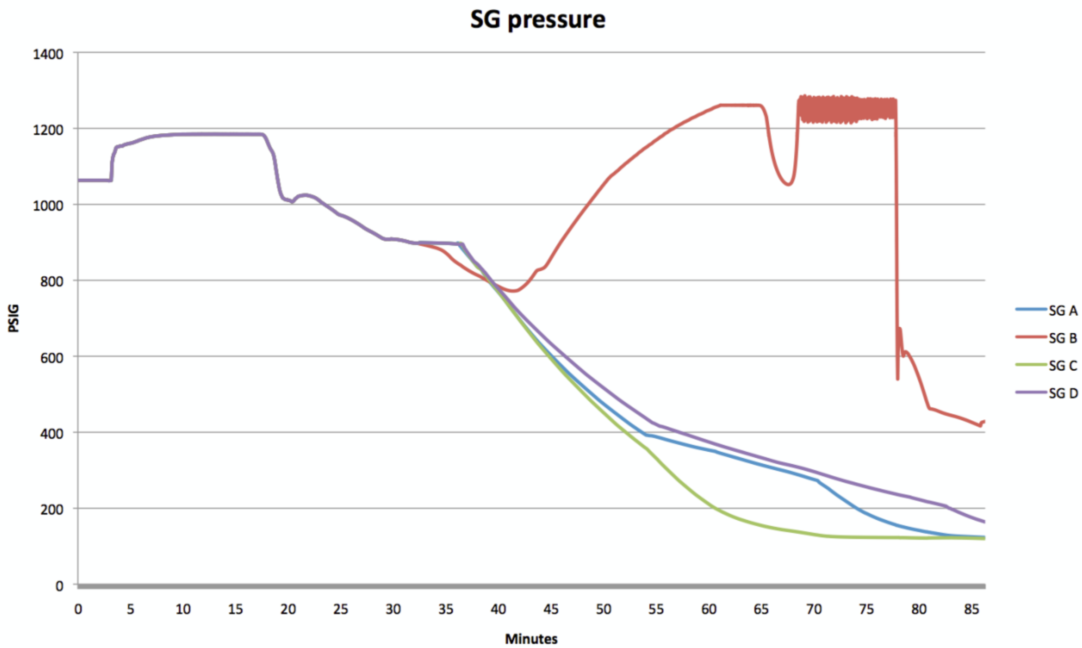



*Figure 6. Pressure in the SGs.*

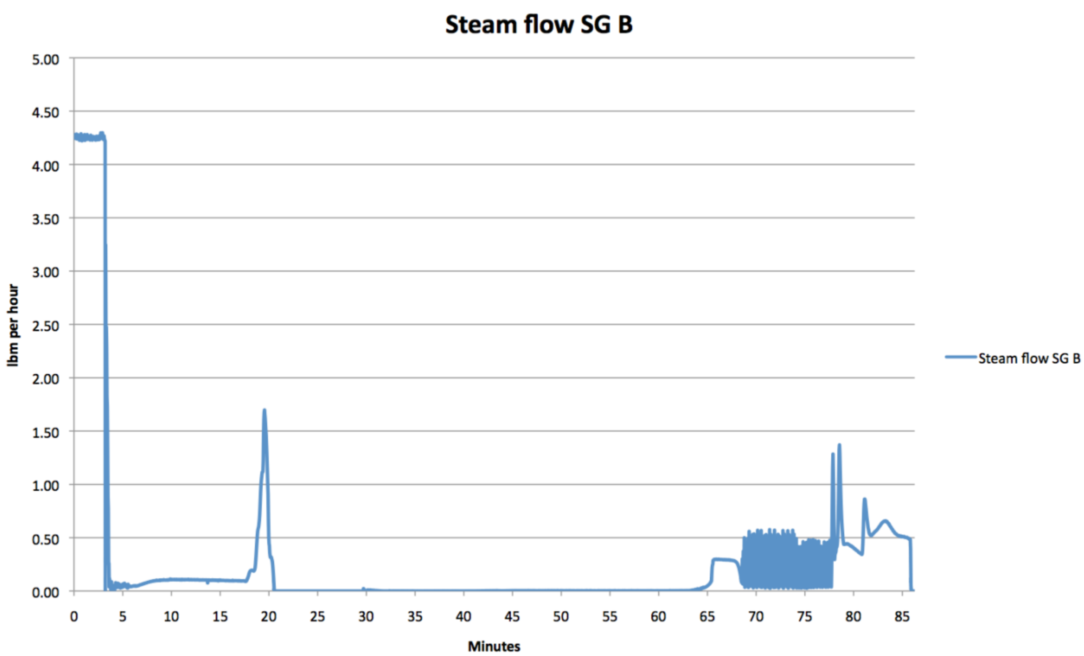



*Figure 7. Steam flow from the ruptured steam generator.*

## D.3 Scenario 1, Crew S

### *Performance of Human Failure Events*

HFE1A: success. The crew manually tripped the reactor when they lost all feedwater, and had established bleed and feed 12 minutes after the reactor trip.

HFE1C: success. The crew isolated feedwater to the ruptured SG and kept the SG pressure below the SG PORV setpoint and accomplished the needed actions (e.g., secure HH SI pumps within 40 minutes). (Procedure E-30 was not entered.)

### *Scenario development and crew responses*

Normal text is based on observations and simulator logs.
*Texts in italics are based on interviews.*

**Reactor trip**
The crew manually tripped the reactor 37 seconds after total loss of heatsink.

*When the second feed pump tripped the US knew that there was no recovery and decided to trip the reactor. The US was also aware that they were going to trip automatically if they did nothing.*

**Identification of total loss of feedwater and start Bleed and Feed**
At 8 minutes into the scenario, the secondary reactor operator said that they don't have any AFW because there is no level in SGB (which had indicated AFW flow). The primary RO immediately suspected that the recirculation valve was open and sent a plant operator to check. The US started ES-01 (reactor trip response) and made a crew update and told the crew to monitor critical safety functions. At 12 minutes into the scenario, the US and SM decided to go to FR-H1 based on the level in SG B. The crew transferred to FR-H1 9 minutes after the reactor trip.

*After suspecting that the recirculation valve was open the crew started a monitoring phase. They wanted to validate that the valve was open. Finally, they made the decision to go to FR-H1 based on the low level in the SG. They were helped by previous training to make the decision. They have had scenarios before when the flow is indicated but not getting to the SG. They also have experience from the plant that a recirculation valve was open some years ago. The crew believes that if the STA had arrived earlier (the STA arrived only a minute before they decided to go to FR-H1) they might have taken the decision even earlier.*

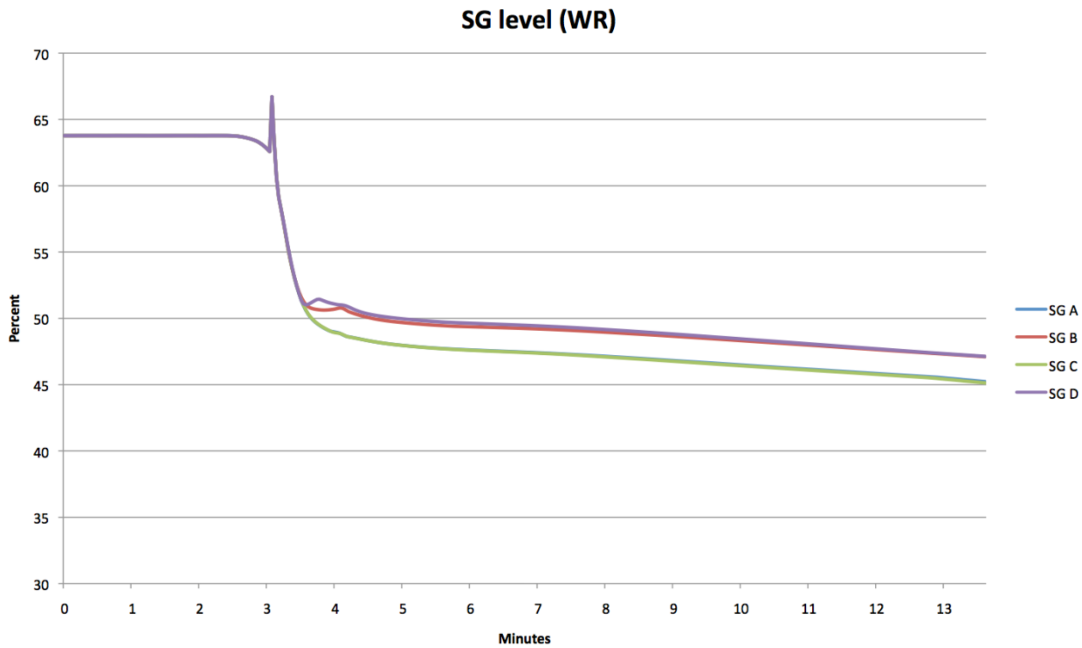



*Figure 1. SG level (Wide Range) from start of scenario to the time when the crew started Bleed and Feed*

One minute after starting FR-H1, the US said that they have F&B criteria and ordered the primary RO to actuate SI. The crew had established F&B 12 minutes after the reactor trip, 15 minutes into the scenario. At that time the second lowest SG WR level (limiting water level to start F&B) was at 45% (figure 1). The crew didn't stop the reactor coolant pumps (RCPs) before starting F&B, as directed by procedure FR-H1, but the STA reminded them and they stopped the RCPs within a minute after having established F&B.

*When entering FR-H1 the crew already knew that they had to start F&B. After having completed step 1 they went to the CIP and then directly to step 10 which they recognize was an error because they missed stopping the RCPs first. The STA asked why the RCPs were not tripped and said that they should be tripped. The US thinks that the STA detected that they had missed the RCP trip because the STA always reviews the CIP notes.*

**Reestablishment of AFW and start of SGTR**

The crew identified that the recirculation valve for AFW pump 12 was open and sent a plant operator to close it. The valve was closed at 21 minutes into the scenario, 18 minutes after the reactor trip, and 6 minutes after F&B was established. When the recirculation valve was closed AFW started filling SG B at a flow rate of greater than 600 gpm (figure 2).

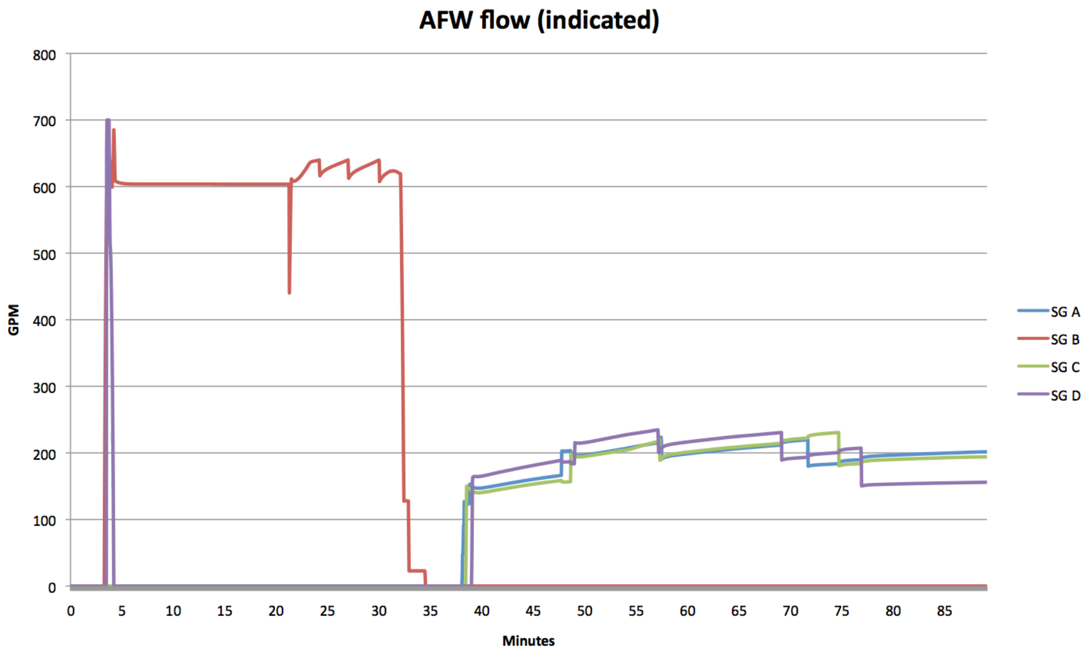



*Figure 2. Indicated AFW flow to the SGs. Before the recirculation valve on AFW pump 12 is closed (at 21 minutes) the indication of flow to SG B is false.*

At 24 minutes into the scenario, a few minutes after having established an AFW flow to SG B, there was a tube rupture in the SG that was fed by AFW (B). When the tube rupture started the WR level in SG B was 50% and rising, and the levels in the other SGs were 41-44% and constant.

At 30 minutes into the scenario, 6 minutes after the start of the tube rupture, the STA said that they had 47% NR-level in SG B and needed to stop feeding 600 gallons to it. They stopped AFW to SG B and the unit supervisor ordered one of the reactor operators to cross connect. At 37 minutes they started feeding the other SGs.

**Identification and isolation of SGTR**

When the tube rupture started, the crew was working in procedure FR-H1 with reestablishing AFW to be able to stop F&B.

At 40 minutes into the scenario, the secondary reactor operator told the US that the level in SG B was going up even though there was no AFW flow into it. While the US and the SM discussed the SG-level, suspecting a tube leak, the crew also detected a red path on FR-P. The STA added that there was slightly elevated radiation in SG B. This small change was visible on the plant computer. The SM decided that procedure FR-H1 was still applicable. The US ordered the secondary RO to isolate the ruptured SG, but to continue feeding the others. The SM and US decided to do FR-P1 while still in FR-H1, and informed the crew in a crew update. The US took about 9 minutes to go through FR-P1 before returning to FR-H1. Before exiting FR-H1 they stopped all HHSI-pumps. In step 28 the STA evaluated the loops as not active, which led them to stop all HHSI pumps.

*The crew identified the tube rupture when the SG level rose in an uncontrolled manner after having secured AFW. The US and SM first discussed if the level came from swelling, which is expected to some degree, or from a tube leak. They were prompted by the secondary RO who kept them informed about the rising level in SG B. They were helped to identify the tube rupture by previous training, because they knew what is a normal amount of swell after stopping feed to a SG.*

Even though the crew identified a tube leak and ordered a reactor operator to isolate AFW to it, they didn't use E-30 to do all isolation steps including adjusting the SG setpoint, and later in the scenario the SGTR was not discussed.

*The US and SM discussed what guidance they had in the situation and decided that FR-H was the highest priority. They could not transition. ZA18, the EOP users guide, has guidance on what you can do when you identify SGTR. They didn't open the procedure but used it from knowledge. The normal action is to pull the AFW pump to stop, but they did not want to do that. The crew verified that the AFW reg vavle was closed. They could have adjusted the SG PORV setpoint, but were aware that the pressure was so low that they were not at risk of lifting the PORV. The MSIVs were shut and the US thought that the most actions to fully isolate the SG were essentially done.*

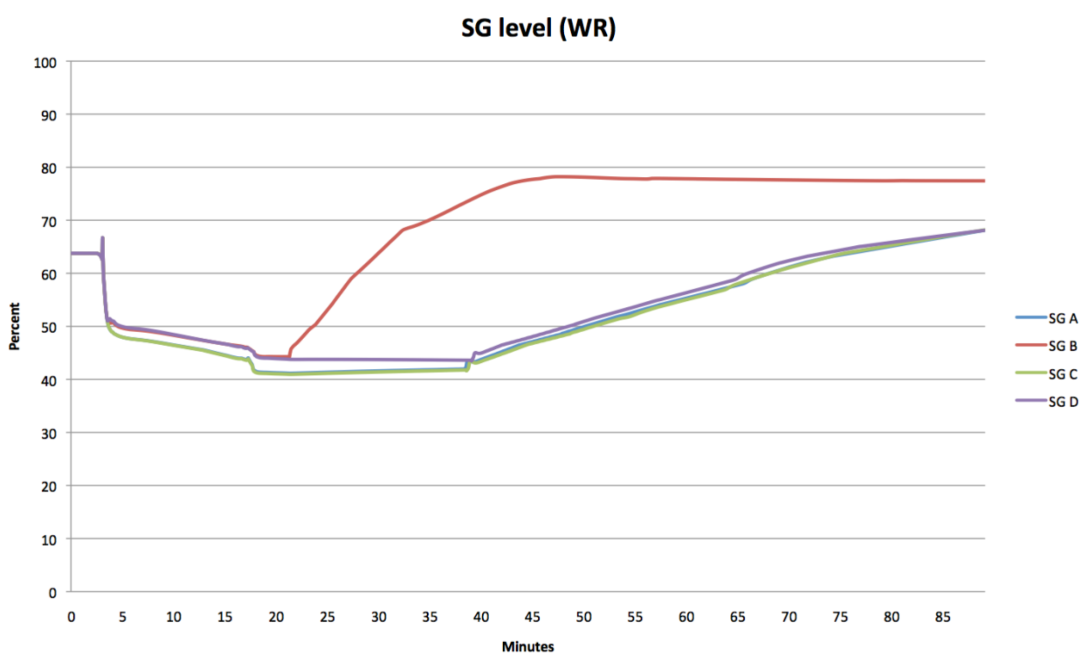



*Figure 3. Wide Range levels in the SGs.*

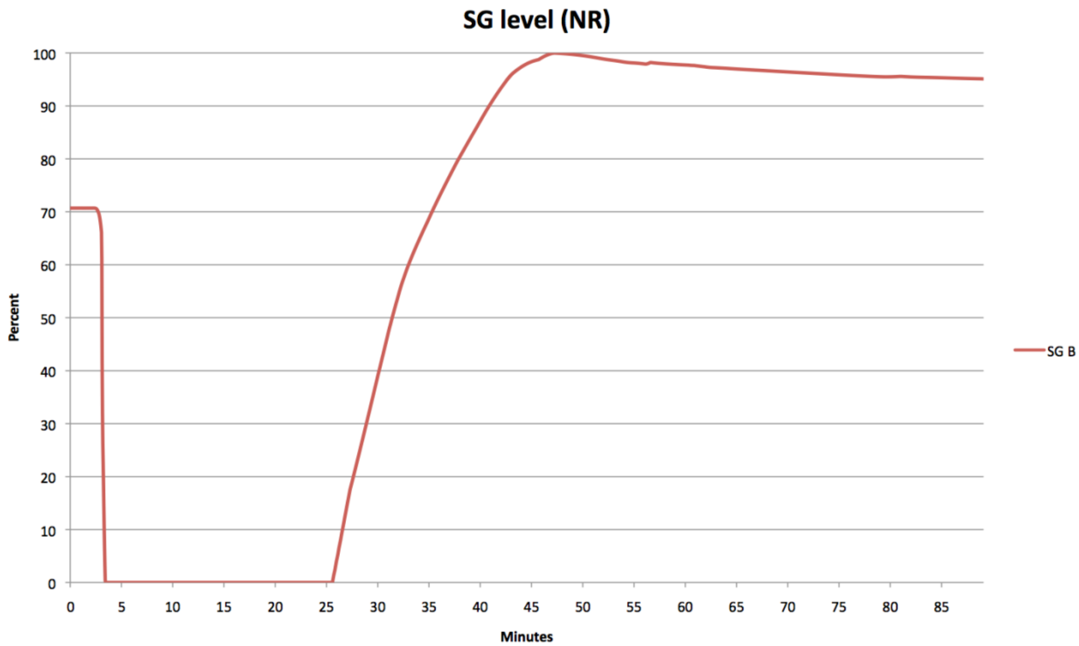



*Figure 4. Narrow range level in the ruptured SG.*

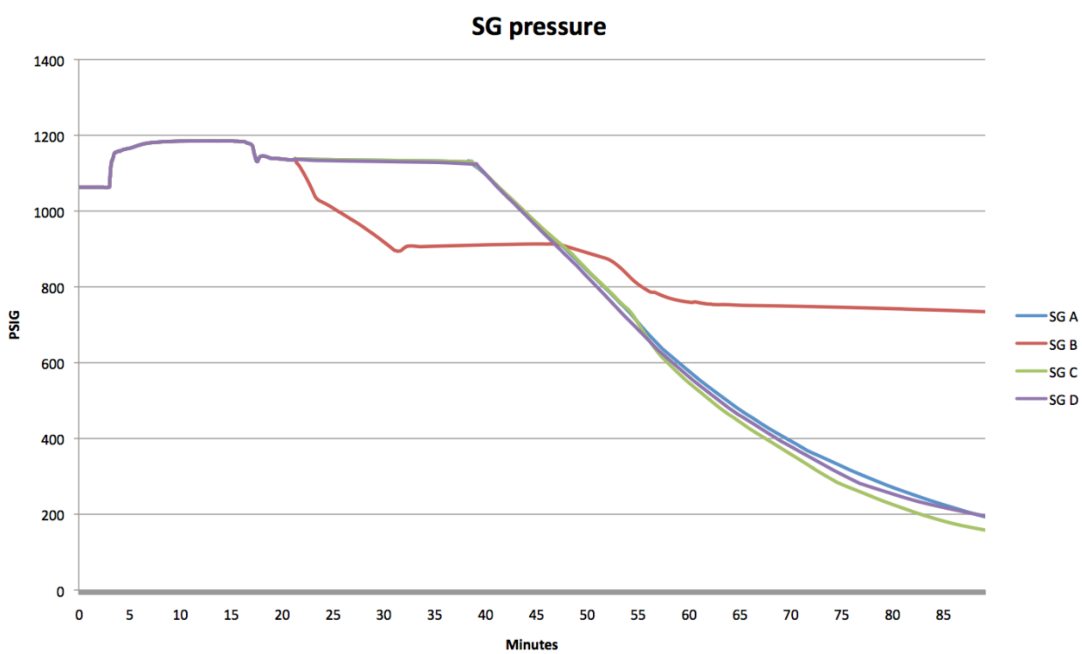



*Figure 5. Pressure in the SGs.*

At about one hour into the scenario, the crew had stopped all HHSI-pumps, the LHSI pumps and closed the PZR PORVs. They had stopped F&B and left procedure FR-H1 from step 33 and transitions to ES-11 (SI Termination). They had a red path on integrity and went to procedure FR-P1 (for the second time), at 1 hour and 4 minutes.

At 1 hour 25 minutes, the US made a crew update. The US informed the crew that they were in a loop to minimize subcooling and that the goal was to take letdown in service and establish a bubble. Then they must soak for an hour before leaving the procedure. They didn’t mention the tube rupture. At 1 hour and 29 minutes, the scenario was stopped.

Even though procedure E-30 is never entered, the crew isolated the ruptured SG and kept the SG pressure below the SG PORV setpoint (figure 5).

## D.4 Scenario 1, Crew T

### *Performance of Human Failure Events*

HFE1A: success. The crew manually tripped the reactor when they lost all feedwater, and had established F&B 10 minutes after the reactor trip.

HFE1C: failed to meet 40 minute time criteria but PORV did not open). This crew did isolate the ruptured SG and control pressure (by securing the HHSI pump) before the PORV opened, but did not complete the actions within 40 minutes after the SGTR. (Procedure E-30 was entered.)

### *Scenario development and crew responses*

Normal text is based on observations and simulator logs.
*Texts in italics are based on interviews.*

**Reactor trip**
The crew manually tripped the reactor 51 seconds after the total loss of heatsink.

*The US ordered reactor trip after loss of all feed-pumps because there was no win-scenario. The US knew from training that they would meet the trip criteria within seconds and wanted to maintain as much inventory as possible.*

**Identification of total loss of feedwater and start Bleed and Feed**
At 4 minutes into the scenario, the secondary RO noticed that they had indication of a red path on heatsink. The US acknowledged it but they did not yet monitor the critical safety functions at that point. At 8 minutes into the scenario, the US noticed that the level was going down in SG B and asked the secondary RO if they had feed flow. The RO answered affirmatively but the US was still suspicious. No one came to think of checking the recirculation valve. The STA arrived and was informed by the SM about the decreasing level in SG B. The STA started verifying the AFW status and the US decided to transfer to FR-H1. The crew transferred to FR-H1 at 8 minutes after the reactor trip.

*The crew identified that they had little or no feedwater based on the decreasing level in SG B. They knew from training that AFW should raise the SG levels after reactor trip. They were helped by the STA who pulled ut the fault tree for heatsink. They also knew from training that they must go to FR-H1 if they don’t have any feedwater.*

When the US started reading FR-H1 step 1 the SM pointed to the CIP and the F&B criteria. The US ordered stop of RCPs and start of F&B from the CIP. The crew had established bleed and feed 10 minutes after the reactor trip, 13 minutes into the scenario. At that time the second lowest SG WR level (the limiting SG level for starting F&B) was at 37% (figure 1).

*The US was aware of the low SG levels when starting FR-H1. After checking that a heatsink was needed in step 1 they started F&B as directed from criteria on the CIP.*

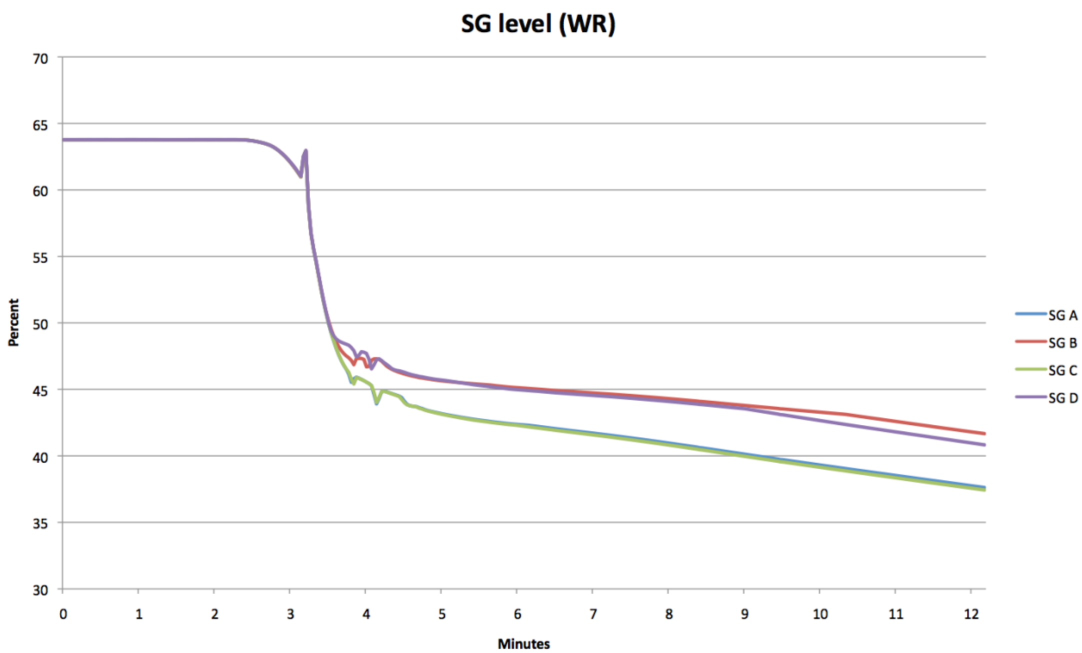



*Figure 1. SG level (Wide Range) from start of scenario to the time when the crew started Bleed and Feed*

**Reestablishment of AFW and start of SGTR**
The crew did not identify the open recirculation valve, it was never mentioned as a possibility.

At 20 minutes into the scenario, the US wanted to cross connect but said that they only had two reactor operators available. Five minutes later the US ordered the secondary RO to connect AFW to SG C. The US expressed that the goal was to have 14% NR to have heat sink. They started feeding about 600 gpm to SG C. The reason why this crew had the tube rupture in SG C is that they didn't close the recirculation valve as the other crew did, thereby starting feeding SG B. The tube rupture occurred in the first SG that was fed, which was C in this case.

At 28 minutes into the scenario, 2 minutes after having established an AFW flow to SG C, there was a tube rupture in the SG that was fed by AFW (C). When the tube rupture started the WR level in SG C was 26% and rising, and the levels in the other SGs were 23-26% and constant.

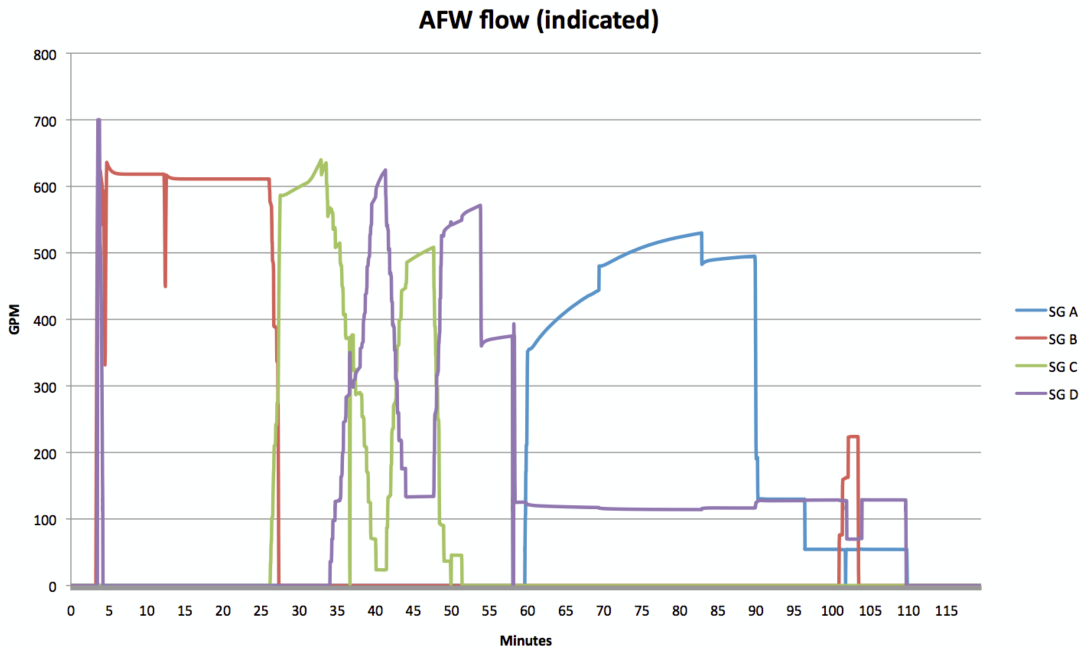



*Figure 2. Indicated AFW flow to the SGs. The indication of flow to SG B is false in the beginning of the scenario (first 25 minutes) because of an open recirculation valve.*

When they reached 37% WR in SG C (over hot empty SG according to CIP, meaning the SG cannot be fed if they reach adverse containment) the US ordered the secondary RO to establish AFW flow also to SG D to get as many SGs available as possible. When the WR level reached 37% in SG D, the US chose to only feed SG C again, to get the SI reduction and get out of FR-H1. At this time, they have two percent NR level in SG C and trending up. At 48 minutes, they had adequate heatsink in SG C and started filling SG D again. At 57 minutes they started to fil SG A.

**Identification and isolation of SGTR**

At 51 minutes into the scenario, AFW was stopped to the ruptured SG (C). The secondary RO reported that the level continued to rise even without feed. The US said that it could be a leak and told the RO to continue monitoring. The US also informed the SM and the STA about the rising level. The STA and the ROs continued to watch the level and the US continued working in FR-H1. One hour into the scenario, the STA told the US that they had a leak in SG C, but the US replied that it didn't make any sense and said that it could be heat up. A couple of minutes later the US made a crew update to inform everyone that they were waiting to stabilize the RCS pressure to get out of FR-H1 and start FR-P1. The US asked the STA about the tube leak, and the STA still thought they had one. The US said that they had to go to FR-P.

*The secondary RO was the first to detect the SGTR. The RO said they were getting .1% level increase every 10 seconds with no feed. The US was busy with FR-H1 and thought: "Let's keep looking. We can have tube rupture, we can have heating, or a leak pass the regvalve." The teamwork, that they all started looking at the SG level, helped the crew identify the tube rupture. Even though they thought they probably had a tube rupture they couldn't go to E-30.*

The crew had problems stabilizing RCS pressure because of the leak, but at 1 hour and 11 minutes they transferred from FR-H1 to ES-11, and then to FR-P1. In FR-P1 the crew had to open a PZR PORV. The STA said that they had to make the RCS pressure less than the pressure in SG C and the US agreed. When working in FR-P the US looked ahead in procedures and at 1 hour and 30 minutes into the scenario the US said that once they got out of FR-P to ES-11 they had criteria to go to E-30. Two hours into the scenario they transferred to E-30 and did the isolation steps, including adjusting the SG PORV setpoint. They had already secured the AFW reg valve and had SLI.

*In FR-P1, the crew's goal was to get out of the procedure, since they knew they had been stable for some time. The decision to go to E-30 was procedure driven, from the ES-11 CIP once they left FR-P1. Even though the isolation in E-30 was some time after the tube rupture occurred they had already isolated the main steamline (MSL) and secured the AFW regvalve.*

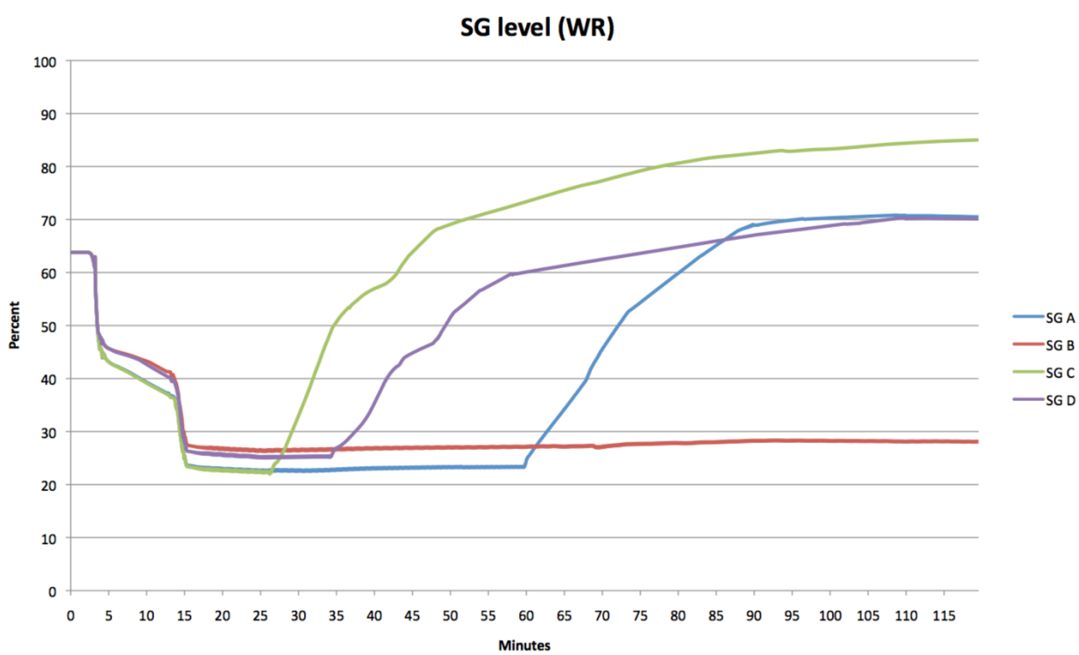



*Figure 3. Wide Range levels in the SGs.*

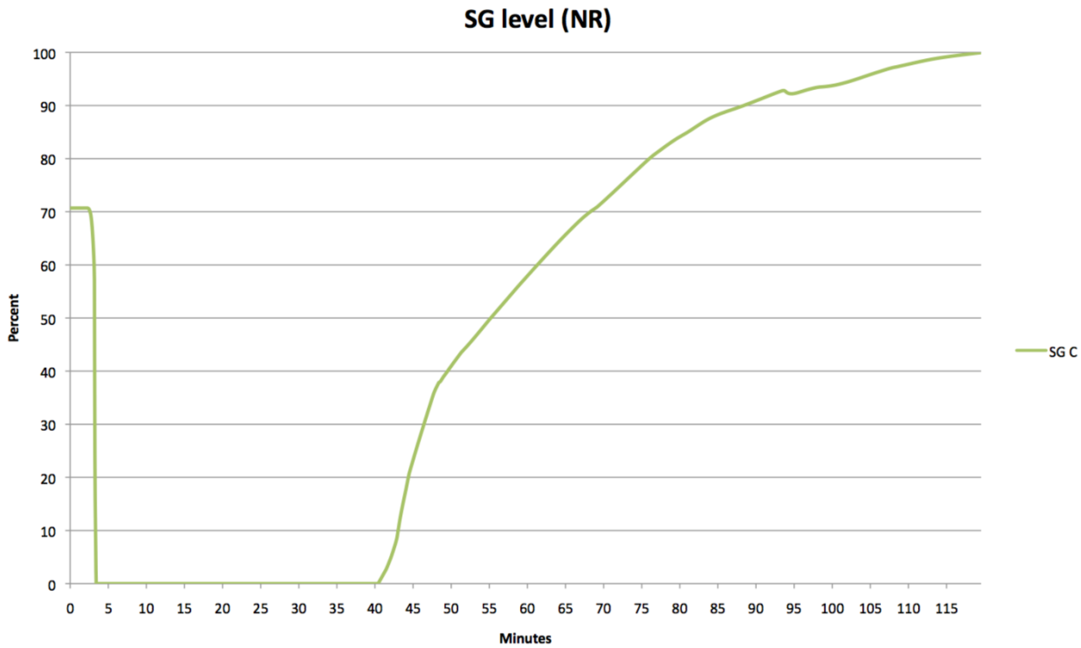



*Figure 4. Narrow range level in the ruptured SG.*

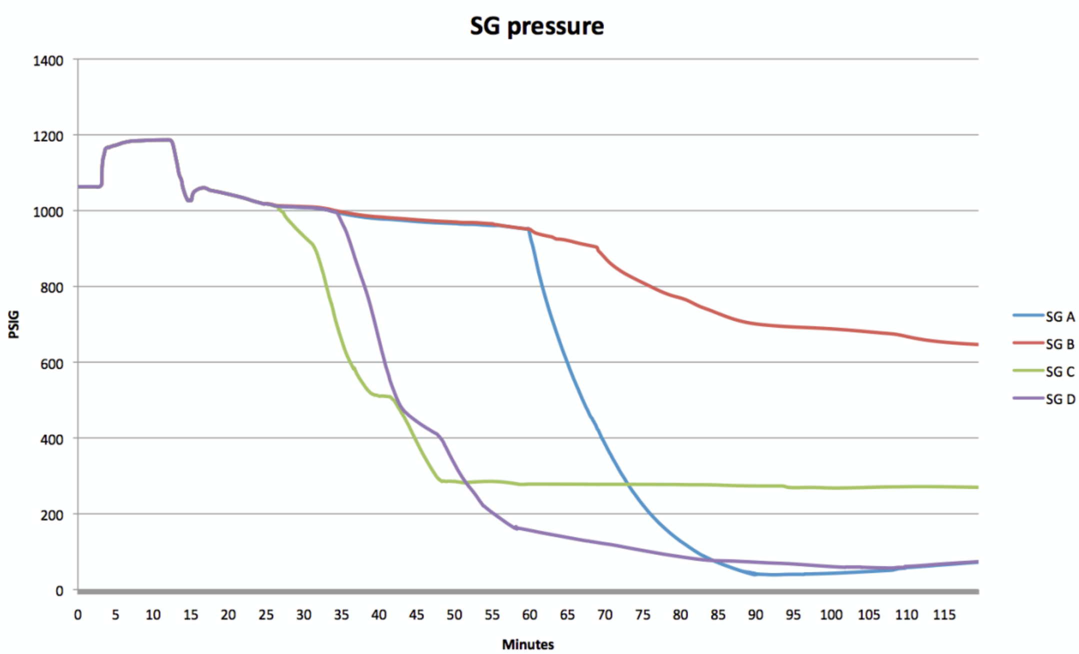



*Figure 5. Pressure in the SGs.*

At 2 hours and 2 minutes, the C SG was isolated and the US ordered a transfer to procedure EC-31 (SGTR with loss of reactor coolant) on low pressure.

# APPENDIX E SCENARIO 2 CREW BY CREW

## E.1 Scenario 2, Crew Q

### *Performance of Human Failure Event*

HFE 2A: failure. The crew tripped the Reactor Coolant Pumps (RCPs) 9 minutes after the loss of Component Cooling Water (CCW) and RCP sealwater. They reached 230 °F 11 minutes after the loss of CCW and RCP sealwater, and did not start the Positive Displacement Pump (PDP).

### *Scenario development and crew responses*

Normal text is based on observations and simulator logs.
*Texts in italics are based on interviews.*

| Action/ Event | Scenario time | Source* |
|---|---|---|
| Start of scenario | 0:00:00 | |
| Reactor trip | 0:02:54 | SAML |
| Loss of CCW and sealwater | 0:02:57 | SAML |
| Start procedure E-0 | 0:03:05 | OBS |
| Start procedure ES-01 | 0:08:05 | OBS |
| Detect no CCW or sealw. | 0:09:15 | OBS |
| Trip all RCPs | 0:11:36 | SAML |
| RCP seal temp. > 230 °F | 0:10:30 | FILM |
| Start “RCP-procedure” | 0:10:00 | OBS |
| Start PDP | - | |
| * SAML = Simulator Action Monitor Log, OBS = Observer's notes, FILM = video camera of screens | | |

Note: The make up of this crew was different compared with the other crews. In this crew, the SM was not present and they had 3 ROs instead of 2 ROs.

**Failure of distribution panel and reactor trip**
At 2 minutes into the scenario, the distribution panel was lost. The US identified that the distribution panel 1201 was lost and ordered the operators to check the electrical status and the levels in Steam Generators (SGs) A and B. The secondary RO reported that SG A feed regulation valve did not go to manual, and the US ordered a reactor trip.

*At the identification of the loss of the distribution panel 1201 the US stepped through immediate actions for all that was lost. The US knew that there is a procedure but it was not opened. The US prioritized watching the ROs to see that the plant was stable. The US decided to trip the reactor because of the rising SG level and that there was no success path.*

**Loss of CCW and sealwater**
When the reactor was tripped 3 minutes into the scenario, the ESF bus C was lost and the CCW pump A stopped. The running charging pump was fed by the C-bus and consequently all CCW and sealwater was lost to the RCPs. The reactor trip was also complicated by the failed distribution panel.

When the reactor tripped the crew started procedure E-0 immediate actions. The primary RO detected that the C-bus was lost less than 30 seconds after the reactor trip. The US ordered the

RO to pull to lock the C diesel generator (DG). At 4 minutes into the scenario, the US took a crew update, informing about the loss of 1201, the reactor trip and the C DG pulled to lock (stop). At 6 minutes into the scenario, the US ordered the secondary RO to start the Auxiliary Feedwater Pump (AFP) in train A. At 8 minutes into the scenario, the US said that SI was not required and started ES-01 (reactor trip response).

At 9 minutes into the scenario, 6 minutes after the loss of CCW and RCP sealwater, the primary RO reported that they had no charging pump, no seal injection to RCP seals, and added that they had no CCW. The US ordered the third RO to take the POP 4 procedure “RCP off-normal”, while the US continued in ES-01. The RCP trip criteria was not recognized from the loss of CCW and RCP sealwater.

At 11 minutes into the scenario, 8 minutes after the loss of CCW and RCP sealwater, the third RO showed the STA and the US the RCP trip criteria on the POP4 Conditional Information Page (CIP). The US ordered stopping all RCPs. The RCPs were stopped 9 minutes after the loss of CCW and RCP sealwater, and over 2 minutes after the crew had detected the loss of CCW and sealwater. When the crew tripped the RCPs they already exceeded 230 °F, and the time limit allowed which is 1 minute. According to procedure 0POP04-RC-0002 “Reactor Coolant Pump Off Normal”, any RCP that experiences a simultaneous loss of seal injection flow and loss of CCW flow to thermal barrier shall be stopped within 1 minute. The risk of a seal failure increases after 1 minute.

*The primary RO detected the loss of CCW and sealwater. The ROs walked past the indications that they did not have CCW a couple of times before finally the primary RO detected the CCW alarm when attempting to start a charging pump. The multiple failures and the complication of the scenario delayed the detection. The US handed the POP 4 procedure to the third RO to implement so that US can continue to work on the higher hierarchy procedure. The US expected the RO to come back if they experienced any problems. They stopped the RCPs from the POP4 CIP, as directed by the procedure.*

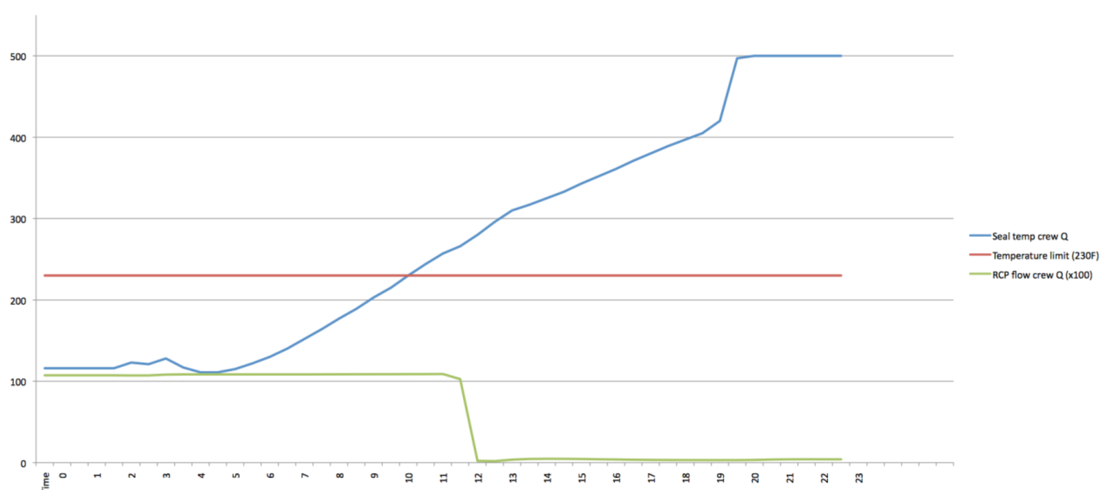



*Figure 1. RCP flow in loop D and RCP seal temperature*

The STA and the third RO (who was working in POP 4) checked the seal temperature which was already greater than 230 °F. They did not start the PDP. At 19 minutes into the scenario, the US ordered the primary RO to start a charging pump (not recognizing that it did not have cooling). The charging pump was not started. Less than a minute later, 16 minutes after the loss of CCW´ and RCP sealwater, they had a seal LOCA (instructor action).

*The US was not aware that they had high RCP seal temperatures. The US did not hear or know where the third RO and the STA were in the POP 4 procedure. The US ordered the primary RO to start charging pump B from ES-01 without recognizing that it had no CCW. The US thinks that the crew verbalized the need for a charging pump to cool the seals, but the US did not internalize [initialize(?)]it.*

*The STA pulled up trends for the RCPs before going into the POP 4. The POP 5 adresses seals (ES-01 step 6). The third RO told the US that when they got into the evaluation step the temperature was already above 400 °F at the lower bearing, but the third RO was not sure that the US had a chance to repeat it back. The STA said that the POP5 takes care of everything they need and that they can pull actions out of the POP4. The STA knew that there was a goal to get charging back to cool the seals. The RO that handled the procedure thought that the procedure didn't give any inclination to the procedur path that teh operators want to get to. They thought that it was difficult to recognize the urgency with all the alarms.*

## E.2 Scenario 2, Crew R

### *Performance of Human Failure Event*

HFE 2A: failure. The crew tripped the Reactor Coolant Pumps (RCPs) 7 minutes after the loss of Component Cooling Water (CCW) and RCP sealwater. They reached 230 °F 10 minutes after the loss of CCW and RCP sealwater, and did not start the Positive Displacement Pump (PDP).

### *Scenario development and crew responses*

Normal text is based on observations and simulator logs.
*Texts in italics are based on interviews.*

| Action/ Event | Scenario time | Source* |
|---|---|---|
| Start of scenario | 0:00:00 | |
| Reactor trip | 0:02:45 | SAML |
| Loss of CCW and sealwater | 0:02:47 | SAML |
| Start procedure E-0 | 0:02:50 | OBS |
| Start procedure ES-01 | 0:08:30 | OBS |
| Detect no CCW or sealw. | 0:08:50 | OBS |
| Trip all RCPs | 0:09:32 | SAML |
| RCP seal temp. > 230 °F | 0:10:10 | FILM |
| Start "RCP-procedure" | 0:13:00 | OBS |
| Start PDP | - | |
| * SAML = Simulator Action Monitor Log, OBS = Observer's notes, FILM = video camera of screens | | |

**Failure of distribution panel and reactor trip**
At 2 minutes into the scenario the distribution panel 1201 was lost. The primary RO and the STA identified that 1201 was lost. The secondary RO reported that the regulation valve to Steam Generator (SG) A did not respond and that the level was rising. The US ordered a reactor trip.

*The crew identified the loss of DP1201 from the indications in the control room. The operators knew what they needed to do in that situation from training. There is a procedure that guides them but they didn't use it.*

*The crew decided to manually trip the reactor because the RO reported that the feedwater reg valve could not be controlled and the level was rising. They realized that they had no chance of controlling the SG level. They knew that they would have had an automatic trip and should take manual action before an automatic setpoint.*

**Loss of CCW and sealwater**
When the reactor was tripped 3 minutes into the scenario, the ESF bus C was lost and the CCW pump A stopped. The running charging pump was fed by the C-bus and consequently all CCW and sealwater was lost to the RCPs. The reactor trip was also complicated by the failed distribution panel.

When the reactor was tripped the crew started procedure E-0 immediate actions. The primary RO detected that the C-bus was lost within a minute after the reactor trip. The US ordered the RO to pull to lock the C diesel generator (DG). The secondary RO reported that AFW 11 failed to start. At 5 minutes into the scenario, the immediate actions were done and the US asked if there was any actions to do before reading. The US restarted reading procedure E-0.

At 8 minutes into the scenario, the US checked if SI was required to transfer to ES-01, the primary RO said that charging was not in service. They also detected no CCW in service. The US started ES-01 and ordered the RO to start a CCW pump. When they couldn't start the CCW pump, the US ordered the primary RO to stop the RCPs. All RCPs were stopped 10 minutes into the scenario, 7 minutes after the loss of CCW and RCP sealwater, which was before reaching 230 °F, but after the time limit allowed which is 1 minute. According to procedure 0POP04-RC-0002 "Reactor Coolant Pump Off Normal", any RCP that experiences a simultaneous loss of seal injection flow and loss of CCW flow to thermal barrier shall be stopped within 1 minute. The risk of a seal failure increases after 1 minute.

*One of the reactor operators detected the loss of CCW and seal injection. The number of alarms and the other ongoing action with the loss of 1201 and the loss of power delayed the detection. They think that if they would have had the third RO present he could have looked at the alarms.*

*The crew decided to stop the RCPs based on having no cooling flow. They knew from training that they needed to stop the RCPs and the decision was knowledge based, not procedure driven. The US was not aware of the limiting time, but the SM knew that they had only a minute. The trip of the RCPs was delayed because they recognized the loss of CCW late in the scenario.*

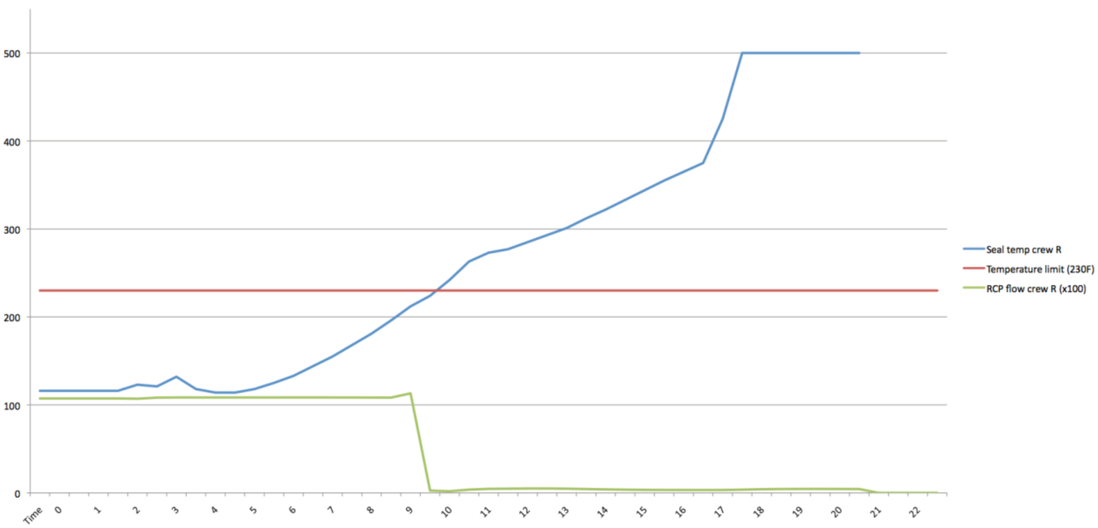



*Figure 1. RCP flow in loop D and RCP seal temperature*

After stopping the RCPs, the primary RO asked the US if they should pursue the PDP and the US said yes. The RO was working from the POP 9 annunciator response procedure.

At 12 minutes into the scenario, the crew pulled up the RCP seal temperatures. At this time the RCP seal temperature was already more than 230 °F. The crew talked about the possibility to start the PDP. They talked about the temperature limit and started the POP 4 RCP off-normal procedure.

At 15 minutes into the scenario, 12 minutes after the loss of CCW and RCP sealwater, the SM, the US and the STA had a discussion about the 230 °F temperature limit (that was exceeded). The crew did not start the PDP. At 17 minutes into the scenario, 14 minutes after the loss of CCW and RCP sealwater, they had a seal LOCA (instructor action).

*They did not start the PDP because there were other things going on. They don't know if there is anything they could have done differently. They would still have to go through the same process to check the temperatures. The crew normally trains on starting the PDP to protect the seals in a loss of outside power (LOOP) scenario. If they loose all AC they have the seals in mind, but in this situation they thought that the seals would be protected.*

## E.3 Scenario 2, Crew S

### *Performance of Human Failure Event*

HFE 2A: failure. The crew tripped the Reactor Coolant Pumps (RCPs) 5 minutes after the loss of Component Cooling Water (CCW) and RCP sealwater. They reached 230 °F 8 minutes after the loss of CCW and RCP sealwater, and did not start the Positive Displacement Pump (PDP)

### *Scenario development and crew responses*

Normal text is based on observations and simulator logs.
*Texts in italics are based on interviews.*

| Action/ Event | Scenario time | Source* |
|---|---|---|
| Start of scenario | 0:00:00 | |
| Reactor trip | 0:02:51 | SAML |
| Loss of CCW and sealwater | 0:02:53 | SAML |
| Start procedure E-0 | 0:02:55 | OBS |
| Start procedure ES-01 | 0:09:40 | OBS |
| Detect no CCW or sealw. | 0:07:05 | OBS |
| Trip all RCPs | 0:07:42 | SAML |
| RCP seal temp. > 230 | 0:10:45 | FILM |
| Start “RCP-procedure” | 0:11:50 | OBS |
| Start PDP | - | |
| * SAML = Simulator Action Monitor Log, OBS = Observer's notes, FILM = video camera of screens | | |

**Failure of distribution panel and reactor trip**
At 2 minutes into the scenario, the distribution panel was lost. The US recognized the channel failure and ordered the crew to take the plant to stable and take rods to manual. The primary RO said that feed regulation valve A was not responding and that the Steam Generator (SG) level was rising. The US ordered a reactor trip.

*Both the US and the SM had experienced failure of a distribution panel in the plant. They decided to trip the reactor because they had no control of the rising level and they were approaching a trip setpoint on high SG level (87.5% NR).*

**Loss of CCW and sealwater**
When the reactor was tripped 3 minutes into the scenario the ESF bus C was lost and the CCW pump A stopped. The running charging pump was fed by the C-bus and consequently all CCW and sealwater was lost to the RCPs. The reactor trip was also complicated by the failed distribution panel.

When the reactor was tripped the crew started procedure E-0 immediate actions. The primary RO detected that the C-bus was lost in less than 30 seconds after the reactor trip. The US ordered the RO to pull to lock the C diesel generator. After immediate actions the US asked the crew about the status, and they talked about the C-bus and missing AFW on train A. The crew manually started the A AFW pump. At 6 minutes into the scenario the primary RO reported that they had no charging pump running. The US repeated the info and said that they would go back and restart E-0.

At 7 minutes into the scenario, 4 minutes after the loss of CCW and sealwater to the RCPs, the SM said that they had no CCW and no sealwater. The US ordered the secondary RO to trip the RCPs. All RCPs were tripped 5 minutes after the loss of CCW and sealwater, which was before reaching 230 °F, but exceeded the allowable time (1 minute) per procedure. According to procedure 0POP04-RC-0002 “Reactor Coolant Pump Off Normal”, any RCP that experiences a simultaneous loss of seal injection flow and loss of CCW flow to thermal barrier shall be stopped within 1 minute. The risk of a seal failure increases after 1 minute.

*The information that they had no charging came from the RO during the "grand pause", which is a moment after immediate actions to see where they are. When the RO reported that they didn't have a charging pump, the priority of the US was to continue in E-0. When the SM heard "no charging" the SM looked to make sure that they had CCW, which they didn't. The SM checked if they had CCW from knowledge that if you loose charging, as long as you have CCW you are fine. The crew tripped the RCPs based on knowledge when they realized that they did not have CCW or seal injection.*

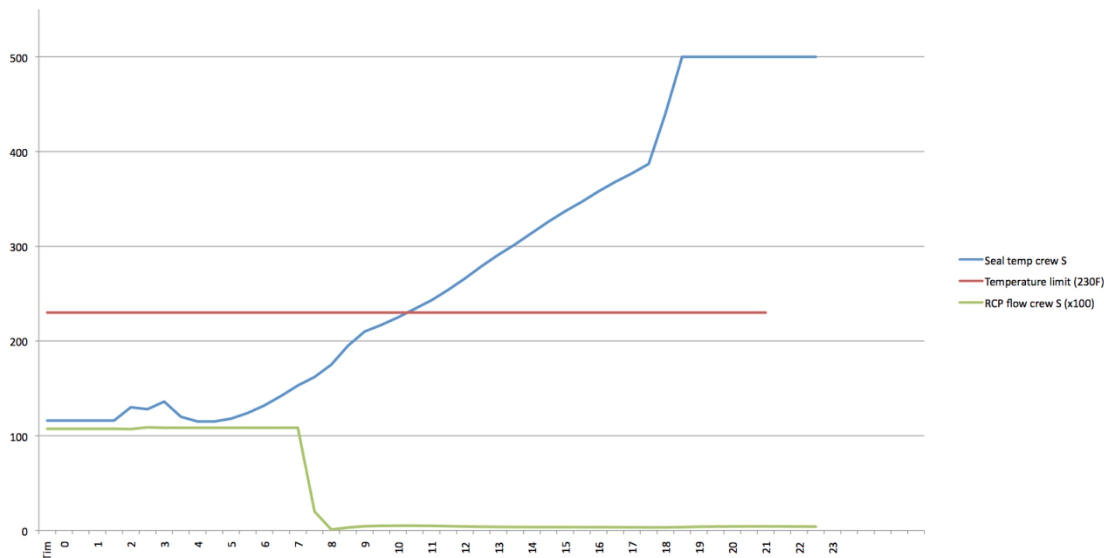



*Figure 1. RCP flow in loop D and RCP seal temperature*

At 9 minutes into the scenario, the STA brought up the possibility to start the PDP to save the seals, if they could get there quickly. (At this time the RCP seal temperature was 195 °F.) The SM and US decided that the US should go to ES01 (they had just verified that SI was not needed), but that an RO should start the RCP-off-normal procedure (POP4-RC-0002). At about 10 minutes the STA noticed that the seal temperature was under 200 °F. The US handed the RCP-off-normal procedure to the secondary RO, who started the procedure 12 minutes into the scenario, when the RCP sealwater temperature was already 255 °F. At 16 minutes, the RO consulted the STA about the RCP seal temperature. The RO informed the US that they had exceeded 230 °F and the procedure said not to attempt to restore the real cooling. At 17 minutes the RO put down the POP4 procedure. At 18 minutes into the scenario, 15 minutes after the loss of CCW and RCP sealwater, they had a seal LOCA (instructor action).

*The hierarchy procedure was the POP 5 (ES-01), meaning that it was the driving procedure and had to be addressed first, but they wanted to get someone in POP4 also. It was difficult when they didn't have a third RO, and one RO was in the emergency procedures. The US handed the POP 4 to a RO with the purpose of reestablish seal cooling with PDP. They knew that there was a short time before the seals heats up and you cannot start the PDP to cool a dry hot seal. The STA had the temperature display up on computer screen. Before the RO got to that point in the procedure it was too late (the seal temperature was greater than 230 °F).*

*There is no other help to start the PDP except the POP 4 procedure. The crew thinks that an EOP users guide would be helpful, or that it would be included in the POP 9 annunciator response procedures on low charging flow. The crew knew what they wanted to do but they were looking for procedure guidance. They could not step through fast enough.*

## E.4 Scenario 2, Crew T

### *Performance of Human Failure Event*

HFE 2A: failure. The crew tripped the Reactor Coolant Pumps (RCPs) 7 minutes after the loss of Component Cooling Water (CCW) and RCP sealwater. They reached 230 °F at 12 minutes after the loss of CCW and RCP sealwater, and did not start the Positive Displacement Pump (PDP)

### *Scenario development and crew responses*

Normal text is based on observations and simulator logs.
*Texts in italics are based on interviews.*

| Action/ Event | Scenario time | Source* |
|---|---|---|
| Start of scenario | 0:00:00 | |
| Reactor trip | 0:03:03 | SAML |
| Loss of CCW and sealwater | 0:03:05 | SAML |
| Start procedure E-0 | 0:03:16 | OBS |
| Start procedure ES-01 | 0:07:15 | OBS |
| Detect no CCW or sealw. | 0:09:20 | OBS |
| Trip all RCPs | 0:10:34 | SAML |
| RCP seal temp. > 230 | 0:11:35 | FILM |
| Start “RCP-procedure” | - | OBS |
| Start PDP | - | |
| * SAML = Simulator Action Monitor Log, OBS = Observer's notes, FILM = video camera of screens | | |

**Failure of distribution panel and reactor trip**
At 2 minutes into the scenario, the distribution panel was lost. The primary Reactor Operator (RO) took the rods in manual and the secondary RO tried to take manual control of feedwater to the A an B Steam Generator (SG). The RO reported that manual control of SG A failed. The US checked the SG level and ordered a reactor trip, but they already had an automatic reactor trip.

*One of the crew members said “Channel 1 bystables in”, and then the US saw the 1201 and called out 1201 failure and take your appropriate manual actions. The US tried to de-select the channel and select the second channel, which is an improbable but possible failure. The secondary RO reported the increasing SG level. The crew tried to preclude an automatic trip, but the order to trip the reactor was too late.*

**Loss of CCW and sealwater**
When the reactor tripped (automatically) 3 minutes into the scenario, the ESF bus C was lost and the CCW pump A stopped. The running charging pump was fed by the C-bus and

consequently all CCW and sealwater was lost to the RCPs. The reactor trip was also complicated by the failed distribution panel.

When the reactor tripped the crew started procedure E-0 immediate actions. The crew detected the loss of the C-bus, and the US ordered the primary RO to pull the Diesel Generator (DG) to lock 1 minute after the reactor trip. The STA and secondary RO worked on AFW to SG A. At 7 minutes into the scenario, the US started ES-01.

At 9 minutes into the scenario, 6 minutes after the loss of CCW and RCP sealwater, the primary RO said that they had no seal injection and no CCW to RCPs. The STA walked the board to check. When the CCW pump couldn't be started, the US ordered to trip the RCPs. All RCPs were stopped 7 minutes after the loss of CCW and RCP sealwater which was before reaching 230 °F, but after the time limit allowed which is 1 minute. According to procedure 0POP04-RC-0002 "Reactor Coolant Pump Off Normal", any RCP that experiences a simultaneous loss of seal injection flow and loss of CCW flow to thermal barrier shall be stopped within 1 minute. The risk of a seal failure increases after 1 minute.

*The crew detected the loss of seal injection attributed to direct observations of the board indications. The RO saw that the seal injection flow indicated zero (0), saw that no charging pump was running, and then looked up and saw the CCW alarm. They could have detected it earlier if they had not had the 1201 failure and such a complicated scenario.*

*The US made the decision to stop the RCP based on the knowledge that with loss of all CCW the motors would overheat. The US had this knowledge attributed to training. There are procedures but there was no time to pull them. The US tripped the RCPs to save the motors.*

*The US became aware of the loss of all seal injection when they secured the RCPs. Then they talked about having no charging pump. The primary RO said that there was no seal injection.*

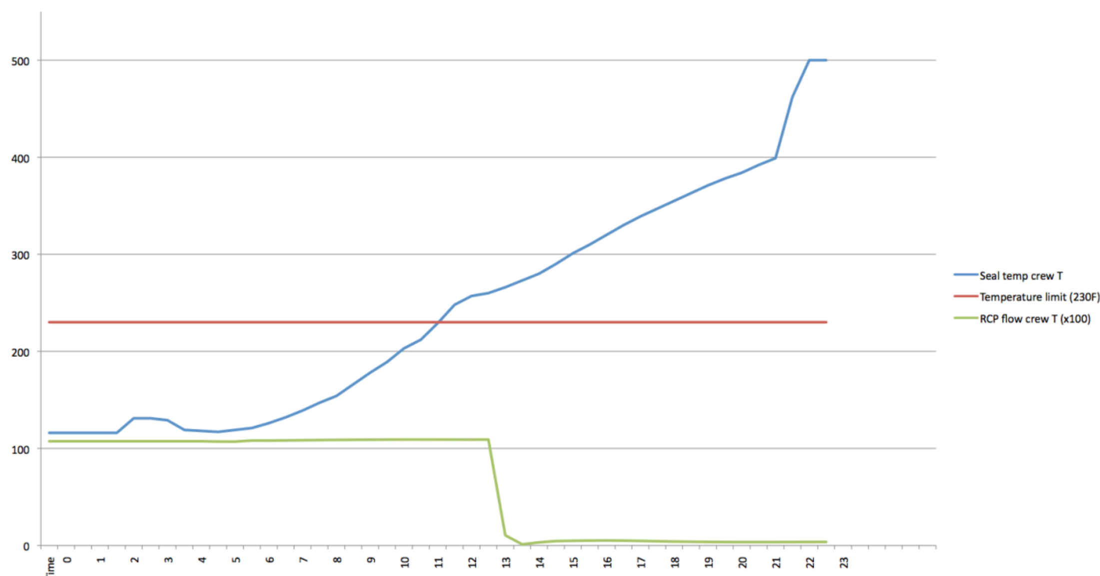



*Figure 1. RCP flow in loop D and RCP seal temperature*

At 13 minutes into the scenario, while working in ES-01 step 6, but before getting to the PDP start later in the step, the US decided to start SI and made a crew update informing everybody that they would go back to E-0. (They did not yet have a seal LOCA.) They made the decision to start SI based on loss on subcooling which is on the CIP in ES-01. The SM pointed the US to this fact. Crew T was the only crew that lost subcooling, the other crews were only close to loosing it.

The crew never started the POP4 procedure for RCP off-normal. The STA took it out but didn't start it. The crew didn't start the PDP. At 22 minutes into the scenario, 18 minutes after the loss of CCW and RCP sealwater, they had a seal LOCA (instructor action).

*The crew did not start seal injection because they knew that you are not to establish seal injection if you are above 230* °F*. They couldn't start charging pumps without CCW, which was pointed out by the SM and the STA. The US didn't think about starting the PDP.*

*The priority of the US was to stabilize the reactor and stopping the cooldown. The RCP seals were not on the list. Nothing in this scenario could have made them start the PDP. They didn't have time with the reactor trip, the AFW work, stepping through the procedure, recognizing the loss and tripping the RCPs.*

*The scenario was challenging and the request for PDP did not match training expectation. The US wasn't worried about the seals because they don't get much leak from them.*

# APPENDIX F SCENARIO 3 CREW BY CREW

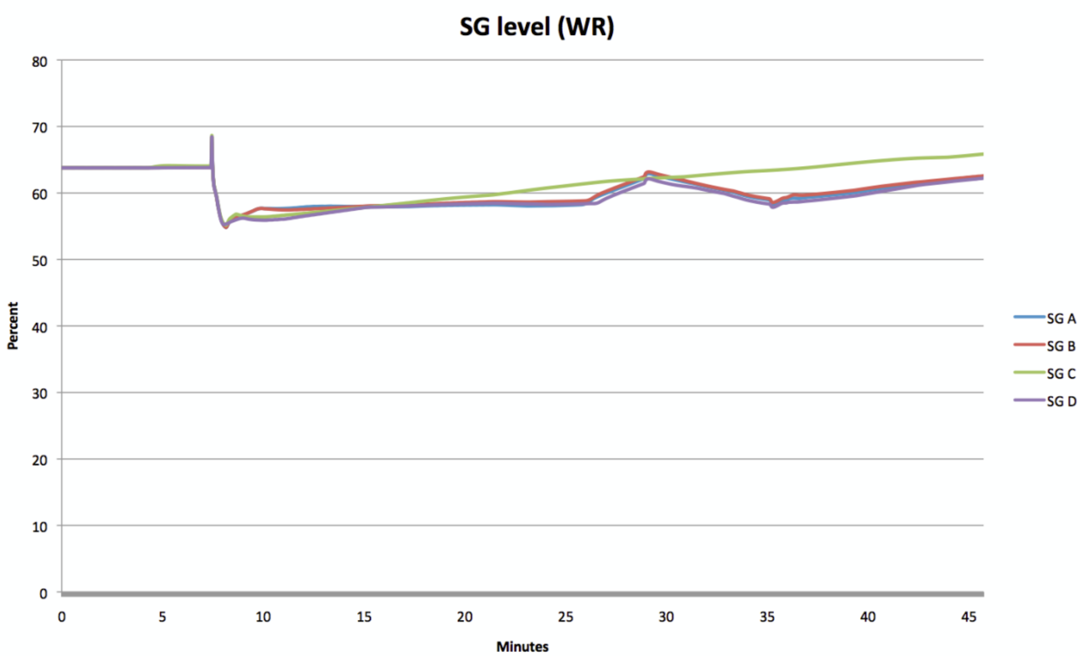
## F.1 Scenario 3, Crew Q

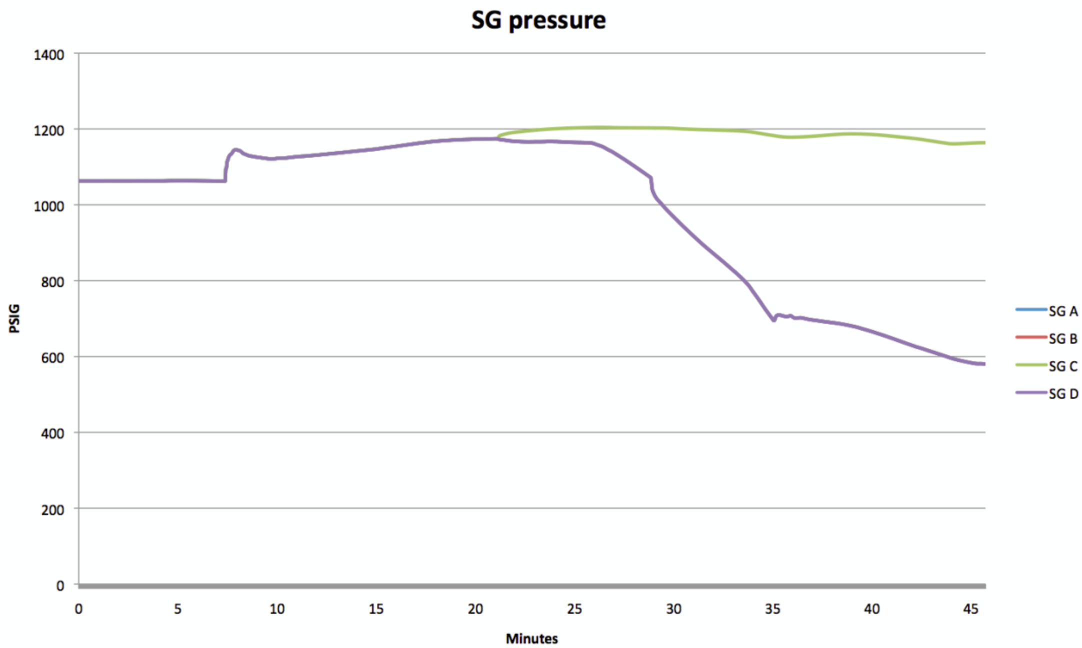
The crew received the instrumentation and control system (ICS) radiation alarms at 00:50 (mm:ss). The crew entered POP4-RC4 (steam generator tube leakage) at 01:29. They identified an SGTR in SG “C”. At 03:45, the US ordered a reactor trip.

*The ICS computer alarm helped the crew identify the tube leak. The US assessed that SG “C” had higher radiation than the others and concluded that they only had one SGTR in “C”. The RO raised charging flow in order to calculate the leak rate. The N16 alarm, the steam line radiation and blowdown radiation monitor helped the crew identify the SGTR.*

At 13:06, the crew entered E-30. At 16:29 the RO adjusts the PORV setpoint to 1264 psig. They close the main tseam isolation valve (MSIV) and the main steam bypass valve (MSBV) at 17:32. At 19:38 they close the AFW OCIV (Outside Containment Isolation Valve). The crew starts cooldown at 24:55. At 31:39 the crew stops cooldown. During cooldown, the SG levels dropped and re-actuated AFW and also the isolation valve. However no flow was initiated because the pump was pulled to lock earlier (at 05:00). They commence RCS cooldown and depressurization at 34:06 and stop it at 40:10.

*The crew used maximum steam dump for the RCS cooldown. The US used time available during this phase to look ahead in the procedure. The crew used maximum spray for depressurization. They checked the criteria for transition to EC-31 but determined that the met both requirements for staying in E-30. Procedure support helped them during this phase.*




SG pressure
1400
1200
1000
800
600
400
200
0
PSIG
0
5
10
15
20
25
30
35
40
45
Minutes
SG A
SG B
SG C
SG D

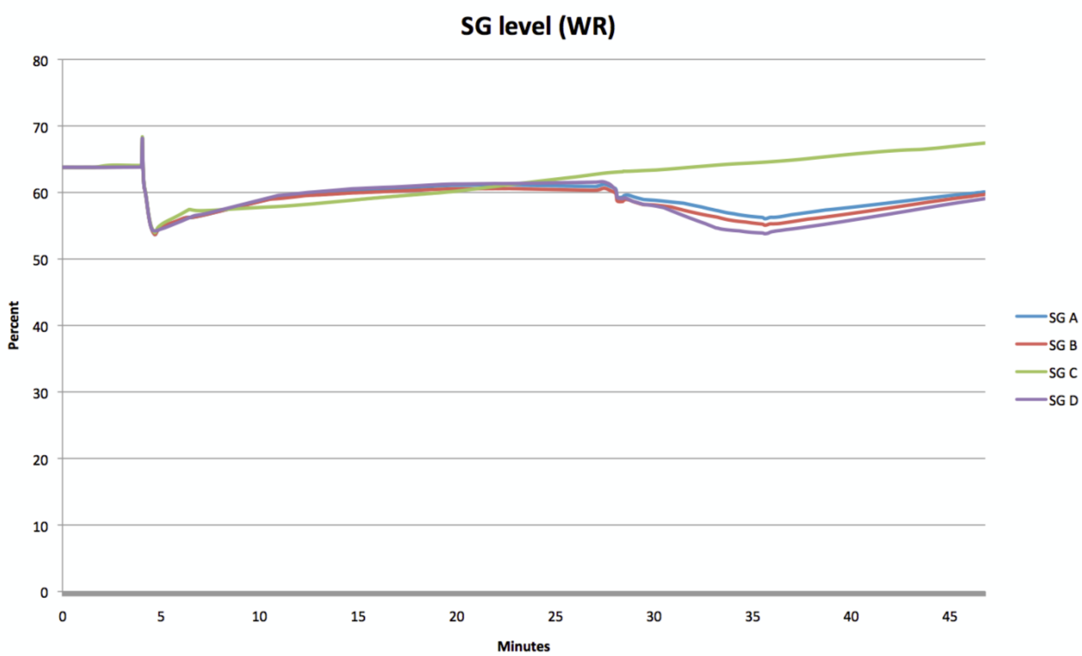
## F.2 Scenario 3, Crew R

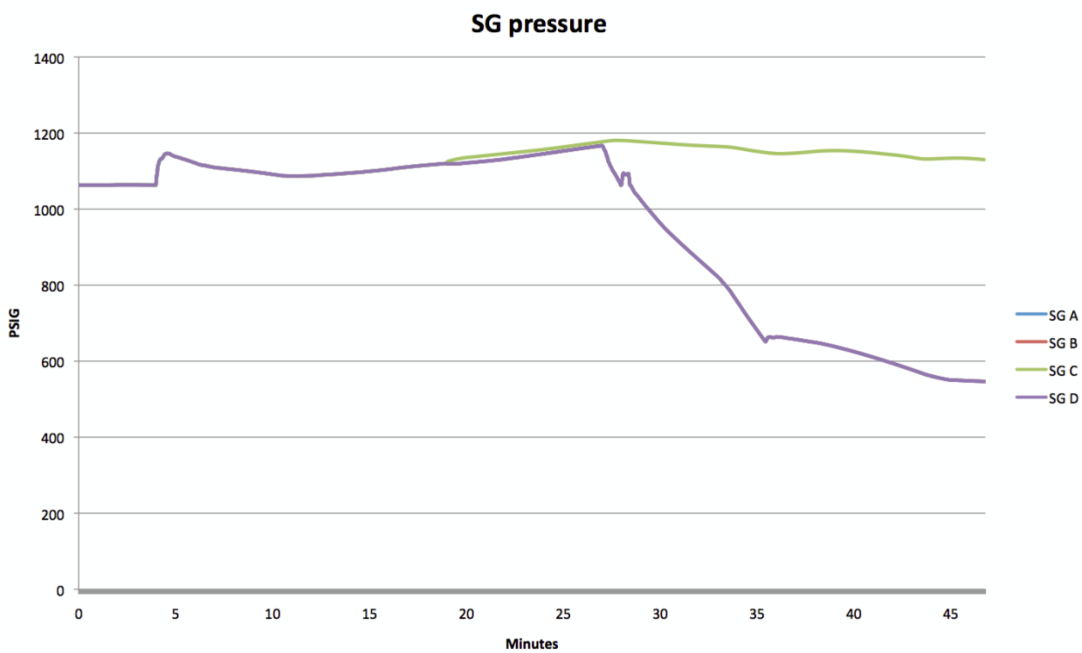
The crew received N16 alarms at 01:25. At 03:38, the US authorized reactor trip and safety injection. Immediate actions were completed at 05:00. At 06:02, AFW pump 13 was pulled to lock. At 15:30 the US decides to go to E-30.

*The computer alarms triggered the crew to monitor primary parameters and SGs. The US noticed a red monitor alarm from SG C on main steam line and later blowdown (BD). The crew then tried to establish the leak size and increased charging flow to maximum. They established that the leak was bigger than 200 gpm, which led to the decision to trip the reactor and start SI. Training, procedure and teamwork (input from board operators and STA) supported the identification. They transitioned to E-30 based on indications on C SG.*

At 17:52, the crew adjusted C PORV to 1264 psig. At 18:42, the MSIV and MSBV were closed. At 22:10, the AFW OCIV is closed. The crew starts cooldown at 26:35 until 34:56. They depressurize from 37:00 until 43:10.

*Procedure, training and teamwork helped the crew in the isolation and cooldown phase.*




SG pressure
1400
1200
1000
800
600
400
200
0
PSIG
0
5
10
15
20
25
30
35
40
45
Minutes
SG A
SG B
SG C
SG D

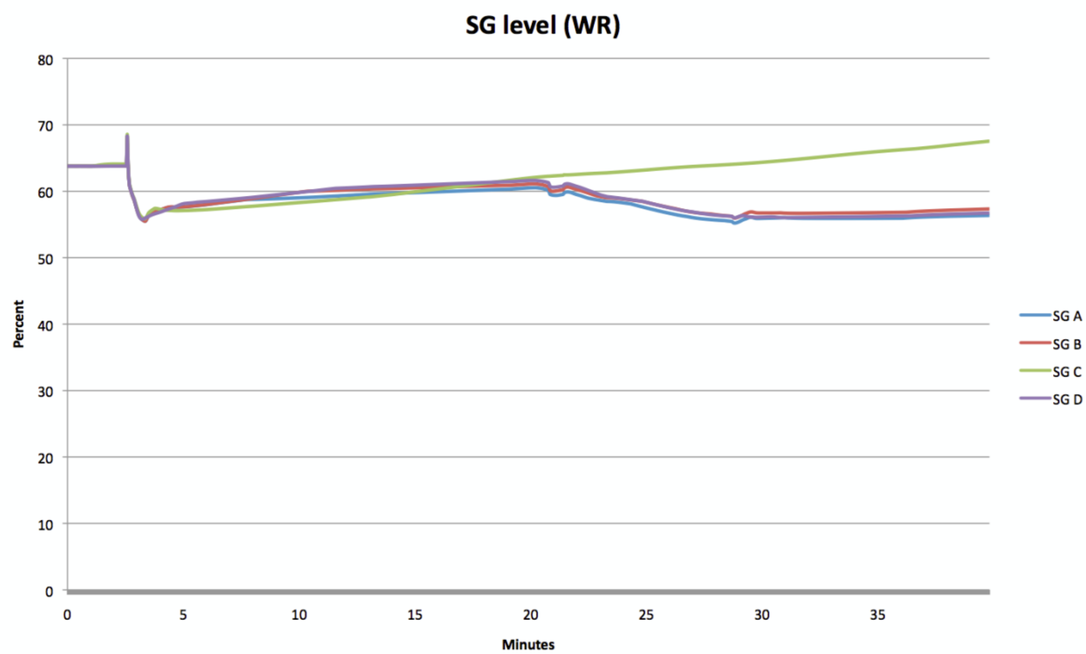
## F.3 Scenario 3, Crew T

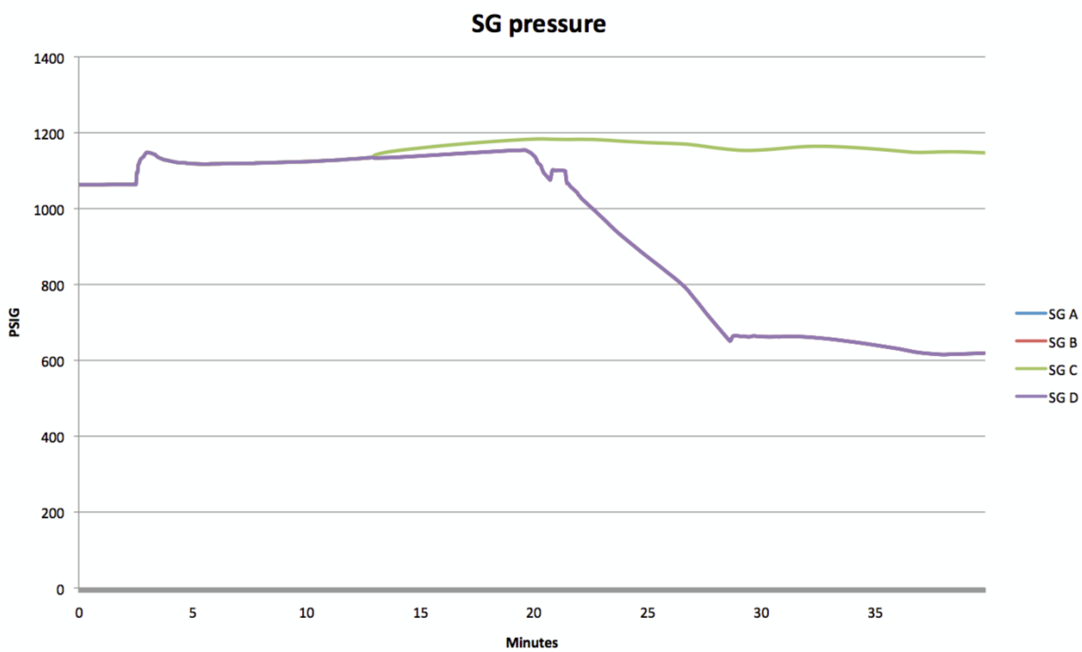
The crew received radiation alarms at 01:08. At 01:49, the US enters POP4-RC4. At 02:21, the US orders reactor trip and safety injection. They go to E-30 at 10:10.

*The computer alarm, annunciator response and training helped the crew to identify the SGTR.*

At 12:02, the crew adjusted the PORV setpoint to 1264 psig, and closed the MSIV and MSBV at 12:54. The AFW OCIV is closed at 15:18. Cooldown was commenced at 19:38 and was terminated at 28:30. The crew started depressurization at 30:09 and terminated it at 36:20.

*Procedures and training helped the crew during the isolation and cooldown phase.*





SG pressure
1400
1200
1000
800
600
400
200
0
PSIG
0
5
10
15
20
25
30
35
Minutes
SG A
SG B
SG C
SG D

# APPENDIX G COMPARISON OF RESULTS FROM HRA METHOD PREDICTIONS TO EMPIRICAL DATA AND OVERALL ASSESSMENTS PER METHOD

The comparisons in this appendix are for each method application and for each HFE. Note that in the summary tables of driving factors presented below, the following scale was used by the assessors to rate the impact of the various factors:

- MND = Main negative driver
- ND = Negative drivers
- 0 = Not a driver (effect could not be determined)
- N/P = Generally positive effect and contribution to the total HEP being small (Note that some methods use the term “Nominal” to denote a default set of positive circumstances and our use of the N/P rating is consistent with that terminology)
- N/A = Not addressed by the method

Also, note that the summaries of the different applications were done by different assessors and although they were used as a basis for the final conclusions on the methods presented in Chapter 9 in the body of this report, the assessment team worked together and consulted with the HRA teams in some cases to obtain a consensus on the final conclusions based on the empirical evidence. In addition, the findings from the intra-method comparisons and the results of the International Study also contributed to our final summary/conclusions about each method. Thus, the final summaries may not in all cases match exactly to the conclusions of the single analyst who performed the initial summary in this appendix.

## G.1 <u>ASEP (Team 1)</u>

### G1.1 HFE 1A

*G1.1.1 Summary of Qualitative Findings*

Time available for this HFE is reduced by initial procedure layout and system status assessment, as well as a complex sequence of procedure transfers or concurrent execution involving the following procedures:

- EO00 “reactor trip or safety injection”
- ES01 “reactor trip response”
- F003 “heat sink critical safety function status tree”
- FRH5 “containment critical safety function status tree”
- FRH1 “response to loss of secondary heat sink”

This time reduction coupled with nominal diagnosis curve were used to calculated the diagnosis HEP.

Symptom-based EOPs do not allow for prompt identification of the scenario under evaluation. Operators would not readily recognize the event because they are trained to rely on the qualified display processing system (QDPS) computer displays to assess critical safety function status, which are affected by wrong AFW flow indication. It is expected that, upon the identification of flow deviation in AFW Pump 12, some discussion or hesitation will take place (perhaps as a team briefing) until the proper way to transfer to the right procedure (FRH1) is found through procedure F003 status trees. The delay in these discussions will depend on how well training has helped support these challenging decisions; this is an uncertainty in the analysis. In summary, it can be concluded that the cognitive aspect of behavior is very important and thus diagnosis can be considered knowledge-based.

Control panel indications needed for diagnosis are simple and easily found. Time trends are provided through electronic recorders for each SG; no calculations or trend annotation/memorization is needed. However, once SG level falls below narrow range, wide range instrumentation is only available through plant computers, and commonly used displays only provide point values. This may obscure required SG trend information.

Once correct diagnosis is made the scenario becomes a classical event and well represented in the EOPs. From this point, the crew is expected to easily follow the correct path through FRH1 to reach required action steps.

Required post-diagnosis actions are straight forward, and frequently practiced in training scenarios. None of the actions are exercised during normal plant operations. Control panel indications and controls for post-diagnosis actions are simple and easily found. No readings or actions take place outside of the main control room.

Deviations from the correct mitigation path may occur randomly, with no adverse driving factors identified. There are recovery opportunities at a later stage, and available time is enough for return to the correct path.

ASEP methodology considers diagnostic HEP as one joint probability for the crew team. It could be claimed that the STA works independently from the team, representing thus an error recovery opportunity. However, considering the STA works mainly with plant computer information (ICS or ODPS), he or she would also be misled by the incorrect safety function status shown on the displays in this scenario.

The qualitative summary included the following:

1. Scenario: LOFW followed by SGTR.

2. Actions and localization of manipulations

   - Required equipment and/or instrumentation is available.
   - No readings or actions are performed outside of the main control room.
   - Since the execution actions are clearly specified in EOPs and constitute step-by-step tasks, the human errors of interest are EOOs (error of omissions) in relevant EOPs.

3. Characterization

   - Diagnosis is knowledge-based because symptom-based EOPs do not allow for prompt identification of the scenario under evaluation.
   - The post-diagnosis actions are considered step-by-step as they are clearly specified in EOP action steps and frequently practiced in training scenarios. Deviations from the correct mitigation path may occur randomly without adverse driving factors identified.
   - Stress level is moderately high, which is the lowest level considered in ASEP, as the event has slow evolution and there is only one safety system failure.
   - Although procedure transfers reduce the time available for this HFE, time is sufficient to perform actions. However, extra time does not allow for recovery to be performed, unless immediate steps provide indications for error recovery.

*G1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

1. Nominal rather than lower bound diagnosis HEP is selected according to ASEP Table 8-3 based on the following factors.

   - This event with the identified complexities (flow diversion and misleading safety function status) is covered in training.
   - Symptom-based EOPs do not allow for prompt identification of the scenario and operators are likely to be misled by wrong AFW flow indication. Hence, the lower bound diagnosis HEP was not selected.
   - Instructor interviews showed that the specific knowledge required to deal with this event is reinforced during simulator training.

   Assuming no additional abnormal events occur, the mean diagnosis HEP is estimated to be 0.0133 (Error Factor (EF) = 10, Median HEP = 0.005) based on ASEP Figure 8-1 and Table 8-2.

2. Post-diagnosis actions are clearly specified in EOPs and constitute step-by-step tasks under moderately high stress, thus all nominal HEPs are multiplied by a factor of 2 according to THERP Table 20-16 Item (4). Since sufficient information can be obtained per task analysis for the assessment, nominal execution HEPs are obtained from THERP Table 20-7 Item (2). Error recovery is not credited. By adding the failure points in the HRA event tree, the total execution HEP is estimated to be 0.0134 (EF = 3.2).

Therefore, the final HEP for HFE 1A is the sum of the diagnosis and execution HEPs, which is 0.0267 (EF= 6.8).

*G1.1.3 Summary Table of Driving Factors*
The following scale was used by the assessors to rate the impact of the various factors:

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Time is considered sufficient, but the ASEP nominal diagnosis curve (given the time available and required) shows time is a contributor to diagnosis HEP. Extra time does not allow for recovery to be performed, unless immediate steps provide for indications. It appeared that adequacy of time was considered at least a minor negative driver in the ASEP analysis and did contribute to some degree to the HEP. | ND |
| **Time Pressure** | Covered under stress | N/A |
| **Stress** | Moderately high stress, which is the lowest level of stress in ASEP. | N/P |
| **Scenario Complexity** | ASEP does not explicitly consider scenario complexity. However, the HRA team pointed out that high complexity in the diagnosis stage due to AFW flow diversion and misleading safety function status, and knowledge-based diagnosis occurs as symptom-based EOPs do not allow for prompt identification of the scenario under evaluation. | N/A |
| **Indications of Conditions** | Although control panel indications needed for diagnosis are simple and easily found, there is wrong AFW flow diversion and misleading safety function status at the initial stage. In addition, once SG level falls below narrow range, wide range instrumentation is only available through plant computers, and commonly used displays only provide point values; this may obscure required SG trend information. | MND |
| **Execution Complexity** | The execution actions are step-by-step and straightforward. | N/P |
| **Training** | Adequate training is provided to deal with the complexities (AFW flow diversion and misleading safety function status) associated with the scenario. | N/P |
| **Experience** | | N/A |
| **Procedural Guidance** | The design of symptom based EOPs does not allow for prompt identification of the scenario under evaluation. However, once correct diagnosis is made, the scenario becomes a classical event and well represented in the EOPs. | ND |
| **Human-Machine Interface** | | N/A |
| **Work Processes** | | N/A |
| **Communication** | | N/A |
| **Team Dynamics** | | N/A |
| **Other** | | N/A |

*G1.1.4 Comparison of Drivers to Empirical Data*
Comparison between predicted drivers and empirical crew data reveals the following differences.

- Adequacy of time: The ASEP assessment Team 1 predicted adequacy of time to be at least a minor negative driver, as they were not optimistic that the crews would discover a flow diversion at the first procedural opportunity. Instead, they suspected that the crews would likely enter FRH5 before FRH1. Hence, the team concluded that time available for HFE 1A would be reduced by a complex sequence of procedure transfers or concurrent execution (FW-0002[3], EO00, ES01, F003, FRH5, FRH1), which, coupled with nominal diagnosis curve, would lead to a contributing diagnosis HEP. According to the crew data, one crew did not diagnose flow diversion like the assessment team suspected. Nevertheless, all crews entered procedure FRH1 without entering FRH5 or FW-0002 to initiate F&B in conjunction with the SG WR levels and based on their experience. This enabled the crews to establish bleed and feed well within the time limit (45 minutes). As such, adequacy of time can be considered a nominal/positive factor and not a negative driver as predicted by the HRA team.

- Stress: ASEP Team 1 predicted stress to be a nominal/positive factor, because stress was expected to be at the lowest level considered in ASEP. This is consistent in that the crew data also indicated that stress was not a negative driver.

- Scenario complexity and indications of conditions: The two drivers were compounded in this scenario in that the scenario was complicated by misleading indications of conditions. The ASEP assessment Team 1 team predicted these two factors to be main negative drivers as symptom-based EOPs would not allow for prompt identification of the complicated scenario. Although the assessment of the crew data did not identify these factors as MNDs, they were considered to be negative drivers; therefore there is a general agreement that the factors would have some impact.

In summary, while the HRA team was not correct about adequacy of time being a negative driver, they were generally correct that scenario complexity and indications of conditions would not be ideal. Nevertheless, they did obtain a relatively low HEP, which appeared to be consistent with the data.

*G1.1.5 Comparison of Qualitative Analysis to Empirical Data*
ASEP Team 1 expected that the crews would not readily recognize the false flow indication, as such a significant amount of time would be spent on diagnosis and procedure transfers. The team also expected that the crews would likely transfer from FW-0002 and FRH5 to FRH1 to initiate F&B. In contrast, as discussed in the above section, the crews started F&B without going through FRH5 or FW-0002, and most crews quickly recognized the misleading flow indication. Additionally, the crews did not feel stress under the familiar scenario, which was consistent with the HRA team’s prediction.

[3] 0POP04-FW-0002 “Steam Generator Feed Pump Trip” procedure.

*G1.1.6 Impact on HEP*
The factors identified as potentially affecting performance of this HFE had a direct impact, but the contributions to diagnosis HEP appeared to be small as a relatively low diagnosis HEP was obtained.

## G1.2 HFE 1B

*G1.2.1 Summary of Qualitative Findings*
Time available for this HFE is spent with the initial procedure layout and system status assessment, following failure to perform an early reactor trip, as well as a complex sequence of procedure transfers or concurrent execution (EO00, ES01, F003, FRH5, and FRH1). Much of this time would not be advantageous anyway, since there is a wait time for the SG level indication to come back on scale.

Symptom-based EOPs do not allow for prompt identification of the scenario under evaluation. Operators may not readily recognize the event because they are trained to rely on process computer (QDPS) displays to assess critical safety function status, which are affected by wrong AFW flow indication. It is expected that, upon identification of flow deviation in AFW Pump 12, some discussion or hesitation will take place (perhaps as a team briefing) until the proper way to transfer to the right procedure FRH1 is found through procedure F003 trees.

Even if correct diagnosis were made from the onset of the scenario, and procedure FRH1 was entered early, the time available is scarce for completion of all the necessary steps, and the crew would have to rush to perform F&B within 13 minutes. According to instructor interviews, there is not a clear feel of the timing for SG dryout, and estimates appear to be optimistic. Also, once SG level falls below narrow range, as expected in this scenario, wide range instrumentation is only available through plant computers, and commonly used displays only provide point values. This may hinder appreciation of SGs level trends. Thus, it is expected that crews will not perceive the need for a fast implementation of FRH1.

ASEP methodology considers diagnostic HEP as one joint probability for the crew team. It could be claimed that the STA works independently from the team, representing thus a recovery possibility. However, considering the STA works mainly with plant computer information (ICS or ODPS), he or she would also be misled by the incorrect safety function status shown on the displays in this scenario, especially when the STA arrived 5 minutes into the event.

The qualitative summary included that following:

1. Scenario: LOFW followed by SGTR.

2. Actions and localization of manipulations

    - Required equipment and/or instrumentation is available.
    - No readings or actions are performed outside of the main control room.
    - Since the execution actions are clearly specified in EOPs, the human errors of interest would be EOOs (error of omissions) in relevant EOPs if there were sufficient time for successful diagnosis.

3. Characterization

    - Diagnosis is knowledge-based due to the complex nature of the event.

- The post-diagnosis actions are considered step-by-step as they are clearly specified in EOP action steps and frequently practiced in training scenarios.
- Stress level is initially moderately high, which is the lowest level considered in ASEP, as there is only one safety system failure and the event, which actually has a fast evolution, will be perceived to have a slow evolution until RCS parameters start to degrade later.
- There is not enough time for successful diagnosis.

*G1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

1. Since there is not enough time for successful diagnosis, there is no possibility of successful diagnosis in this scenario. This inadequacy of time holds even if correct heat sink status was initially determined. Based on ASEP Figure 8-1 and Table 8-2, probability of diagnosis failure is 1.0 (EF: N/A).

2. Since there is no possibility for successful diagnosis, execution HEP is not calculated for this scenario.

Therefore, the final HEP for HFE 1B is 1.0 (EF: N/A).

*G1.2.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Time is not sufficient for successful diagnosis. Even if correct diagnosis were made from the onset of the scenario and the correct procedure (EO-FRH1) were entered early, the time available would still be too scarce to complete all the necessary steps. | MND |
| **Time Pressure** | Covered under stress | N/A |
| **Stress** | Moderately high stress, which is the lowest level of stress in ASEP. | N/P |
| **Scenario Complexity** | ASEP does not explicitly consider scenario complexity. However, the HRA team pointed out that high complexity in the diagnosis stage due to wrong AFW flow diversion and misleading safety function status, and knowledge-based diagnosis occurs as symptom-based EOPs did not allow for prompt identification of the scenario under evaluation. | N/A |
| **Indications of Conditions** | Although control panel indications needed for diagnosis are simple and easily found, there is AFW flow diversion and misleading safety function status at the initial stage. In addition, once SG level falls below narrow range, wide range instrumentation is only available through plant computers, and commonly used displays only provide point values; this may obscure required SG trend information. | MND |
| **Execution Complexity** | Time is not sufficient to execute post-diagnosis actions. If the time were sufficient, the actions would be step-by-step and straightforward. | 0 |
| **Training** | Adequate training is provided to deal with the complexities (AFW flow diversion and misleading safety function status) associated with the scenario. However, the operators are trained to rely on process computer (QDPS) display to assess critical safety function status, which in turn are affected by wrong AFW flow indication. Hence, the operators are expected to make correct initial diagnosis, but they would take some time to evaluate their diagnosis. | N/P |
| **Experience** | | N/A |

| | | |
|---|---|---|
| **Procedural Guidance** | Assuming procedures are being followed. The scenario is covered by procedures. | N/P |
| **Human-Machine Interface** | | N/A |
| **Work Processes** | | N/A |
| **Communication** | | N/A |
| **Team Dynamics** | | N/A |
| **Other** | | N/A |
| [redacted] | | |

Note that comparisons with the crew data were not possible for this HFE since all crews performed a manual trip before the automatic trip.

### G1.3 HFE 1C

*G1.3.1 Summary of Qualitative Findings*

The specific diagnosis and actions for mitigating the SGTR event can be considered easy to recognize and perform, as symptom-based EOPs are in place and well designed for the scenario under evaluation. However, this scenario starts from a condition that delays the appearance of cues. Such delay, in turn, may be affected both by how specific actions are performed and how the plant responds to those actions. This includes how early AFW was recovered, the way to exit the previous stage of the scenario (F&B), the approach taken to refill one or several SGs, and the choice to enter other procedures. While it is not expected that the crew will have difficulties making this diagnosis once they acquire sufficient stimuli, additional delays as suggested above will result in the crew not meeting the specified time window. In summary, rather than observing failures to diagnose, it is expected to see some crews failing to complete the scenario on time or barely meeting the time. Therefore, although the required inputs are easy to find and interpret and operators are familiar with expected plant responses, the diagnosis HEP is expected to be high due to short diagnosis time.

Execution time estimation is important in this calculation, and results are very sensitive to small changes. The estimates were based on simple and conservative rules mainly from ASEP. Some conservatism could be removed based on simulator observations or operator talk through at the panels.

An alternate mitigation path is possible through procedure EO-EO10, transferring either from FRH1 or ES11. This trajectory is equivalent to the one delineated here, since it contains the same critical actions. However, the procedure layout is more suitable for the purpose of identifying an SGTR, and contains explicit secondary radiation monitoring. It was concluded that EO-EO10 is less likely to be accessed, and the HFE probability for this path might be slightly lower than the one selected. For the end result, there is no impact from this choice, as diagnosis is based on alarms and the only instruction credited is CIP transferring to EO-EO30, which exists in both procedures.

The possibility of entering higher priority functional restoration procedure FRP1 to focus PTS concerns has not been considered. This entry highly depends on the specific thermal hydraulic conditions attained during F&B in the simulation. From the HFE results, it can be concluded that entry into FRP1 will result in the inability to timely diagnose SGTR, because FRP1 includes several steps that will delay exit from FRH1.

The qualitative summary included that following:

1. Scenario: LOFW followed by SGTR.

2. Actions and localization of manipulations.

   - Required equipment and/or instrumentation is available.
   - No readings or actions are performed outside of the main control room.
   - Execution actions consist of both step-by-step and dynamic actions, thus the human errors of interest would be both EOMs and ECOMs.

3. Characterization

   - Stress level is moderately high, which is the lowest level considered in ASEP, as the only safety system failure has been recovered.
   - Time available for diagnosis is not sufficient enough. Delay of the appearance of cues for SGTR may be a challenge for operators to make correct diagnosis in the specified time window.
   - The execution actions consist of both step-by-step and dynamic actions.

*G1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

1. Lower bound diagnosis HEP is selected according to ASEP Table 8-3 based on the following factors.

   - The scenario is a well recognized event, covered in training, practiced in simulator requalification exercises, such that all operators know the pattern of stimuli associated with the event.
   - Symptom-based EOPs are in place and well designed for the scenario under evaluation; operators recognize the event and are familiar with applicable procedures and actions.

   Assuming no additional abnormal events occur at the SGTR stage of the scenario, the mean diagnosis HEP is estimated to be 0.178 (Median HEP = 0.129) based on ASEP Figure 8-1 and Table 8-2. Since median value is close to 1, a maximum entropy distribution is assigned with range (0.0129, 1) with parameters $\lambda_1 = 2.828$ and $\lambda_2 = -5.345$.

2. Post-diagnosis actions are clearly specified in EOPs and most of them are step-by-step tasks under moderately high stress, thus all nominal HEPs are multiplied by a factor of 2 according to THERP Table 20-16 Item (4). Since sufficient information can be obtained per task analysis for the assessment, execution HEPs are obtained from THERP Table 20-7 Item (2) and Table 20-12 Item (3) (for memorized and skill-based actions, the HEPs are obtained from ASEP Table 8-5 Item (10)). By adding the failure points in the HRA event tree, the total execution HEP is estimated to be 0.0310 (EF = 3.3).

Therefore, the final HEP for HFE 1C is the sum of the diagnosis and execution HEPs, which is 0.209. Since median value is close to 1, a maximum entropy distribution is assigned with range (0.0273, 1) with parameters $\lambda_1 = 2.828$ and $\lambda_2 = -5.345$.

*G1.3.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Delay of the appearance of cues for SGTR may cause operators to fail to complete the scenario on time or barely meet the time window. Although lower bound diagnosis HEP is selected, the HEP is high due to short diagnosis time. | MND |
| **Time Pressure** | Covered under stress | N/A |
| **Stress** | Moderately high stress, which is the lowest level of stress in ASEP. | N/P |
| **Scenario Complexity** | ASEP does not explicitly consider scenario complexity. However, the HRA team pointed out that the SGTR scenario was a well recognized event, which was covered in training and practiced in simulator requalification exercises, hence diagnosis and actions for mitigating SGTR could be considered simple. | N/A |
| **Indications of Conditions** | The scenario starts from a condition that delays the appearance of SGTR cues. | ND |
| **Execution Complexity** | Time is sufficient to execute post-diagnosis actions, which are clearly specified in EOPs and consist of both step-by-step and dynamic actions. | N/P |
| **Training** | Adequate training and practice is provided. | N/P |
| **Experience** | | N/A |
| **Procedural Guidance** | Assuming procedures are being followed. The scenario is covered by procedures. | N/P |
| **Human-Machine Interface** | | N/A |
| **Work Processes** | | N/A |
| **Communication** | | N/A |
| **Team Dynamics** | | N/A |
| **Other** | | N/A |

*G1.3.4 Comparison of Drivers to Empirical Data*

Comparison between predicted drivers and empirical crew data reveals the following differences.

- Adequacy of time: When the 40-minute time window defined in HFE 1C is considered, it is concluded from the crew data that adequacy of time was a negative driver. Similarly, the ASEP Team 1 predicted that adequacy of time was a main negative driver.

  The HRA team predicted that the time available for diagnosis would be shortened by the delay of SGTR cues, which in turn resulted in a high diagnosis HEP even though the HEP lower bound was selected due to favorable performance shaping factors (e.g., training and experience). This is consistent with the crews' report that the diagnosis might have been done earlier if AFW to the ruptured SG had been stopped sooner. In addition, as the HRA team expected, the crews' performance was affected by their actions earlier in the scenario. For example, the crews had to first follow FRH1 steps to stop SI and close PZR PORVs as a function of RCS subcooling, which was a crucial step to maintain SG pressure below the setpoint. As a result, although three of the four crews prevented SG PORVs from opening, only one crew isolated the ruptured SG and stopped the HH SI pumps barely within the 40-minute time limit. This observation is consistent with the assessment team's prediction that although several procedure instructions exist, they would not be helpful because of insufficient time.

- Scenario complexity: The ASEP assessment Team 1 considered scenario complexity to be a nominal/positive factor, because although SGTR was initially masked by AFW flow it was not difficult to be identified once sufficient SGTR cues were provided. However, the crew data indicated that scenario complexity was a main negative driver due to the initial masking by the AFW flow to the ruptured SG and the low steam flow prevented the radiation alarms, particularly under the context of the limited time available for performing the task.

- Procedural guidance: ASEP Team 1 considered procedural guidance to be a nominal/positive factor as symptom based EOPs are in place and well designed for the SGTR scenario. However, as the crew data suggested, most crews were not able to get through the procedures to E-30 quickly since they were held up in FR-H1 and FR-P1, and thus two crews isolated the ruptured SG based on knowledge. Hence, it was concluded that procedural guidance was a negative driver, as there were no specific instructions for mitigating SGTR when operators had to take care of a task with a higher priority. On the other hand, the fact that the SGs were isolated on the containment isolation signal may have led the crews not sensing an urgency to entering EO30. In addition, the concern in this scenario was to control the RCS pressure, and most of the crews achieved this through implementing FR-H1 to throttle SI without having to enter EO30 to adjust the SG PORV setpoint. Thus, it may not have been a procedural limitation per se.

In summary, the assessment team's analysis and the crew data were consistent on adequacy of time, but disagreed on scenario complexity and procedural guidance. The discrepancy between predicted drivers and crew data may be attributed to the fact that the SGTR scenario was complicated by the LOFW scenario that occurred earlier. ASEP Team1 took crews' training and experience into consideration, but did not fully consider the context in which some crews would be held up in FR-H1 and FR-P1. Nevertheless, the HRA team did predict a relatively high HEP (0.209), which was consistent with the data (3 out of 4 crews failed to meet the time criterion for the actions).

#### *G1.3.5 Comparison of Qualitative Analysis to Empirical Data*

As discussed in Section 0, the SGTR scenario was complicated by the LOFW scenario that occurred earlier. Due to the containment isolation in the LOFW scenario, the SGs were all isolated automatically. Hence, a key aspect of maintaining SG pressure was to control RCS pressure by reducing SI. This was particularly critical for the crews who had difficulty to reach the EO30's procedure steps inadjusting the SG PORV setpoint. ASEP Team 1 recognized the importance of completing FRH1 to stop HH SI pumps to prevent SG PORVs from opening. Consistent with the assessment team's analysis, it was difficult for the crews to isolate the ruptured SG and stop HH SI pumps within the 40-minute time window. The assessment team took crews' training and experience into consideration, but did not fully consider the context in which some crews would be held up in FR-H1 and FR-P1.

As predicted by the assessment team, two crews transferred to EO30 from the CIPs of ES11 and EO10, but the assessment team did not expect some crews to isolate the ruptured SG from knowledge without entering EO30. It was assumed by the assessment team that FR-P1 would not be entered because entry into FR-P1 would result in the inability to timely diagnose the SGTR; however, three crews actually entered this procedure and one crew met the 40-minute time criterion.

The assessment team determined the cross connection of AFW pump 12 to other SGs to be knowledge-based dynamic actions as there are no written instructions. The crew data suggested that the operators did not have trouble performing this action probably due to good training and experience.

### G1.4 HFE 2A

*G1.4.1 Summary of Qualitative Findings*

The crew has been preconditioned in the initial phase of Scenario 2, where control panel indicators are considered unreliable. However, upon a reactor trip, the scenario is significantly complicated by several high impact malfunctions (e.g., failures of Vital AC PDP 1201 and ESF 4160V Bus E1C), leading to a total loss of CCW and CCP. According to interviews, the scenario of total loss of CCW is not frequently trained, and moreover, operators are not readily aware of its full impact and the immediate actions they may take.

Successful mitigation requires following a combination of alarm response, abnormal operation and emergency operation procedures, as well as transfers among them and intricate jumps between them. The failures of three support systems (Vital AC PDP 1201, ESF 4160V Bus E1C, and CCW) will lead to a large number of alarms competing for crew's attention, and requires manual operation of important control systems. The primary operator is overloaded because of the recent reactor trip and manual PZR pressure and level control. In addition, the primary operator is in principle responsible for all the cues, alarm procedures, and control panel operations associated with this HFE, including a time-consuming RCP parameter verification from plant computer displays. Alarm response prioritization is needed in this scenario and involves important cognitive elements. As such, it can be concluded that the complexity and stress levels of the scenario are significantly increased by the context where the event has fast evolution and there are several and diverse safety system failures.

There are generally two diagnosis approaches. The first one is a cognitive immediate response to a major functional loss, which leads to a knowledge-based fast track to trip RCPs and align PDP under concurrent loss of CCW and CCP based on the procedure FW-0002. The complementary measure of starting the PDP will take place if FW-0002 instructions are followed to Addendum 1 Step 11. Interviews suggested that this fast track response could not be credited as an immediate and memorized post-initiator action with HEP less than one due to inadequate training and experience. Instead interviews showed that training emphasizes on monitoring RCP operation parameters and the use of alarm response procedures to follow on their deviations, but no fast discrimination of all pumps affected by a high-level cause. It should be noted that it is expected that most crews will be able to reach either RCP trip through a fast track to RC-0002 or PDP alignment in ES01, but most of them will not complete both assignments on time unless the primary operator puts aside other tasks or gets support to perform them. With this approach, time does not allow for recovery to be performed, unless immediate steps provide for indications.

The second approach is a proceduralized step-by-step response to existing alarms without explicit identification of the problem. The execution time for all steps substantially exceeds the time available. Therefore, it is concluded that there is not sufficient time for successful diagnosis even with a non-conservative estimate of execution time.

For the main HFE operator actions, RCP pump trip is very simple and PDP alignment is considered a dynamic task because it requires controlling flow to achieve a 1°F/min temperature decrease in the seals. Therefore, in addition to implementation of procedure where jumps (or discretional applicability) are allowed, PDP alignment and flow control are also likely to result in execution errors.

Execution time estimates are determinant for the results obtained. However, a more realistic task analysis will require accounting responses to other simultaneous operator tasks, such as loss of PDP 1201, Bus E1C, and CCW, and their effects, including PZR pressure and level control, loss of indications, loss of HVAC, etc. The conservative estimate of execution time covers to a certain extent these unquantified times.

The qualitative summary included that following:

1. Scenario: Loss of CCW and RCP sealwater.

2. Actions and localization of manipulations.

    - Required equipment and/or instrumentation is available.
    - No readings or actions are performed outside of the main control room.
    - Execution actions consist of both step-by-step and dynamic tasks, thus the human errors of interest would be both EOMs and ECOMs.

3. Characterization

    - Stress level is extremely high, which is the top level considered in ASEP, as the event has fast evolution and there are several and diverse safety system failures, including some affecting control room I&C and others having a widespread support system effect.
    - Inadequate training is provided. Operators are not readily aware of the full impact of a total loss of CCW and CCP and the immediate actions they may take.
    - Time is not sufficient for proceduralized step-by-step diagnosis. Fast-track diagnosis is knowledge based and involves a cognitive immediate response. In addition, the primary operator needs to put aside other tasks or gets support to perform the fast-track actions.
    - A large number of alarms compete for crew's attention and the primary operator is overloaded. The need for prioritization of these responses further increases the complexity level of the scenario and has important cognitive elements.

*G1.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

1. For fast-track diagnosis mentioned above, an upper bound diagnosis HEP is selected, based on the following factors according to ASEP Table 8-3.

    - Training and experience is assessed as low.

- Instructor interviews showed that not all the operators know the pattern of stimuli associated with the fast track being postulated.

Based on ASEP Figure 8-1 and Table 8-2, the mean fast-track diagnosis HEP is estimated to be 1.0 (Median HEP = 1.0, EF: N/A).

For proceduralized step-by-step diagnosis, the mean diagnosis HEP is estimated to be 1.0 (Median HEP = 1.0, EF: N/A) as there is not sufficient time.

It should be noted that the failures of three support systems may be sorted out to a large extent by implementing procedure ES01, as SI is not required. However, prioritization of alarm responses has important cognitive elements for process deviations not directly focused in ES01. Since an RCP trip to avoid overheating falls in this situation, the argument to neglect diagnosis HEP cannot be sustained. As such, the diagnosis HEP is 1.0.

2. Since there is no possibility for successful diagnosis, execution HEP is not calculated for this scenario.

Therefore, the final HEP for HFE 2A is 1.0 (EF: N/A).

*G1.4.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Time is not sufficient for successful proceduralized step-by-step diagnosis. The fast-track diagnosis requires the primary operator to put aside other tasks or get support to perform fast-track actions. With fast-track diagnosis, time does not allow for recovery to be performed, unless immediate steps provide for indications. | MND |
| **Time Pressure** | Covered under stress | N/A |
| **Stress** | Stress is extremely high as the event has fast evolution and there are several and diverse safety system failures, including some affecting control room I&C and others having a widespread support system effect. | MND |
| **Scenario Complexity** | ASEP does not explicitly consider scenario complexity. However, the HRA team pointed out that the initial phase of Scenario 2 was a well recognized and frequently practiced event. However, upon a reactor trip, the scenario was significantly complicated by several high impact malfunctions, leading to a total loss of CCW and CCP. A large number of alarms would compete for crew's attention and the primary operator is overloaded. The need for prioritization of these responses would further increase the difficulty in diagnosis. | N/A |
| **Indications of Conditions** | Control panel indicators are considered unreliable. A large number of alarms compete for crew's attention and prioritization is needed for operators' responses. | ND |

| | | |
|---|---|---|
| **Execution Complexity** | The execution HEP was not calculated; hence it did not drive the total HEP. If the execution HEP had been considered, it might have been a negative driver. The HRA team pointed out that successful mitigation would require following a combination of alarm response, abnormal operation and emergency operation procedures, with several transfers among them and intricate jumps within them. With proceduralized step-by-step diagnosis, time would not be sufficient to execute post-diagnosis actions. With fast-track diagnosis, PDP alignment and flow control would be considered dynamic tasks, and require the primary operator to put aside other tasks or get support to perform the actions. | 0 |
| **Training** | Inadequate training and practice is provided. Operators are not readily aware of the full impact of a total loss of CCW on RCPs and the immediate actions they may take. As a result, upper bound diagnosis HEP is selected for fast-track diagnosis. | MND |
| **Experience** | | N/A |
| **Procedural Guidance** | Successful mitigation requires following a combination of alarm response, abnormal operation and emergency operation procedures, as well as transfers among them and intricate jumps within them, which increases the complexity level of the scenario. Fast-track diagnosis is knowledge based and requires a cognitive immediate response to a major functional loss. Procedures with high priorities do not provide instructions for some alarm responses (e.g., ES01 does not provide instructions to stop RCPs). | ND |
| **Human-Machine Interface** | | N/A |
| **Work Processes** | | N/A |
| **Communication** | | N/A |
| **Team Dynamics** | | N/A |
| **Other** | | N/A |
| [redacted] | | |

*G1.4.4 Comparison of Drivers to Empirical Data*

Compared to crew data, ASEP Team 1 accurately predicted most important negative drivers, including adequacy of time, scenario complexity, indications of conditions, training, and procedural guidance. It should be noted that both the assessment team and the crew data agreed that the factor of adequacy of time would have a negative impact on HEP, but they differ in the extent to which the factor would impact the HEP. The crew data suggested that if the crews had had more training or experience on this scenario, it is likely that there would have been enough time. Hence, adequacy of time was not considered as a main negative driver. However, in this scenario, given their current level of training, it is clear that it was an important factor. In addition to training or experience, the ASEP assessment Team 1 also considered that operators would not be able to trip RCPs and start PDP on time unless the operators put aside other tasks or got support to perform them. This, to some degree, is consistent with the crews' comments that if they had had one more reactor operator, they might have detected loss of CCW and sealwater earlier. Hence, the assessment team expected the factor to be a main negative driver.

The assessment team's analysis and crew data differed on the following two factors.

- Stress. As mentioned in Section 0, ASEP assessment Team 1 expected the stress level in this scenario to be extremely high, as the event has fast evolution and there are several and diverse safety system failures. In contrast, the crew data suggested that stress was probably not an important driver, as there were only some observations of stress. According to the comments from the crews, it was difficult for them to recognize the urgency of this scenario with all the alarms.

- Execution complexity. ASEP Team 1 expected that the successful mitigation would require following a combination of alarm response, abnormal operation and emergency operation procedures, with several transfers among them and intricate jumps within them. Furthermore, dynamic and knowledge-based actions, such as PDP alignment and flow control, were expected. Hence, the assessment team concluded that the execution complexity would be a negative driver. Since none of the crews went far enough to start PDP injection, there was no evidence to show that execution complexity was a negative driver. It is certainly possible that there may have some execution complexity. The plant participants involved in the study thought that: it would "depend on what procedure guidance they used, and who read the procedure (A RO may struggle with steps that a US has no problem with.).

In summary, the ASEP team's assessment and the crew data agreed on most drivers, but were not consistent on stress. The impact of execution complexity could not be evaluated. If the crews had started PDP injection, the ASEP assessment team's prediction on execution complexity might have been consistent with actual crew data.

*G1.4.5 Comparison of Qualitative Analysis to Empirical Data*
As mentioned in previous sections, ASEP Team 1 expected that the operators would go through a combination of alarm response, abnormal operation and emergency operation procedures. Specifically, they expected that operators would reach RCP trip through RC-0002 or PDP alignment through ES01. They concluded based on their interviews that RCP trip and PDP alignment should not be considered as an immediate memorized operator response to a total loss of CCW, even when charging flow was consequentially lost.

The crew data showed that the crews did not go through as many procedures as the assessment team predicted. Rather, the crews took several critical actions without entering procedures. For example, all crews quickly identified and responded to the failure of Distribution Panel 1201 without entering 0POP04-VA-0001 as the assessment team predicted. And three crews tripped RCPs from knowledge without entering 0POP04-RC-0002. While this provided some evidence that RCP trip could be considered as an immediate memorized operator response to a total loss of CCW, there was no evidence that that the PDP alignment would happen in this way, which was consistent with the crew data.

As predicted by the assessment team, the crews were overloaded partly due to poor procedural guidance and insufficient experience. While they worked in prioritized procedures E000 and ES01 to stabilize the reactor, the crews also had to respond to alarms that were not directly focused in these two procedures (e.g., ES01 does not provide instructions to stop RCPs). Consistent with the prediction of the assessment team, no crew was able to complete PDP alignment on time.

*G1.4.6 Impact on HEP*
Except stress and execution complexity, other factors identified as potentially affecting performance of this HFE had a direct impact.

## G1.5 HFE 3A

*G1.5.1 Summary of Qualitative Findings*
Time available is abundant for this HFE. The SGTR scenario is a well recognized event, which is covered in training and practiced in simulator requalification exercises. Symptom-based EOPs are in place and well designed for the scenario under evaluation. The crew is expected to easily follow the correct path through EOPs to reach required action steps. Hence, it can be concluded that the cognitive aspect of behavior for this event is very small, and diagnosis could be either neglected or converted from knowledge-based into rule-based behavior in a more realistic assessment. Low diagnosis HEP is expected.

Post-diagnosis actions are straight forward and well specified in EOPs. They are frequently practiced for the same or similar training scenarios. Some actions are exercised during normal plant operations. Thus, the actions can be considered step-by-step. Time is sufficient for recovery when subsequent steps provide for indications. Low execution HEP is also expected.

Control panel indications are simple and easily found. When time trends are needed, these are provided through electronic recorders or plant computer displays; no calculations or trend annotation or memorization is needed. Negative trends can be established by cross comparison among similar, adjacent instruments. No readings nor actions take place outside main control room.

Deviations from the correct mitigation path may occur randomly, with no adverse driving factors identified; in most of these cases there are recovery opportunities at a later stage, and available time is plenty for return to the correct path.

The qualitative summary included that following:

1. Scenario: SGTR.

2. Actions and localization of manipulations.

    - Required equipment and/or instrumentation is available.
    - No readings nor actions are performed outside of the main control room.
    - Parameters trends are easily available and no calculations or trend annotation or memorization is needed.
    - Execution actions are considered step-by-step, and the human errors of interest are both EOMs and ECOMs.

3. Characterization

    - Stress level is moderately high, which is the lowest level considered in ASEP, as the operators are quite familiar with the scenario.
    - The post-diagnosis actions are considered step-by-step as they are clearly specified in EOP action steps and frequently practiced in training scenarios.

- Time is abundant for both diagnosis and execution, and is also sufficient for recovery actions.

*G1.5.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

1. Lower bound diagnosis HEP is selected based on the following factors from ASEP Table 8-3.

   - The scenario is a well recognized classic event, covered in training, and practiced in simulator requalification exercises, such that all operators know the pattern of stimuli associated with the event.
   - Symptom based EOPs are in place and well designed for the scenario under evaluation; operators recognize the event and are familiar with applicable procedures and actions.

   Assuming no additional abnormal events occur, the mean diagnosis HEP is estimated to be 2.5E-05 (EF = 30, Median HEP = 3E-06) based on ASEP Figure 8-1 and Table 8-2.

2. Post-diagnosis actions are clearly specified in EOPs and constitute step-by-step tasks under moderately high stress, thus all nominal HEPs are multiplied by a factor of 2 according to THERP Table 20-16 Item (4). Since sufficient information can be obtained per task analysis for the assessment, execution HEPs are obtained from THERP Chapter 20 according to Item (2) in ASEP Table 8-5. By adding the failure points in the HRA event tree, the total execution HEP is estimated to be 0.0275 (EF = 2.8).

Therefore, the final HEP for HFE 3A is the sum of the diagnosis and execution HEPs, which is 0.0275 (EF: 2.8).

*G1.5.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Time is abundant for both diagnosis and execution, and is also sufficient for recovery when subsequent steps provide for indications. This contributes to low total HEP. | N/P |
| **Time Pressure** | Covered under stress | N/A |
| **Stress** | Moderately high stress, which is the lowest level of stress in ASEP. | N/P |
| **Scenario Complexity** | ASEP does not explicitly consider scenario complexity. However, the HRA team pointed out that the scenario was a well recognized event, which was covered in training and practiced in simulator requalification exercises; hence the cognitive aspect of behavior for this event was very small, and diagnosis could be either neglected or converted from knowledge-based into rule-based behavior in a more realistic assessment. | N/P |
| **Indications of Conditions** | Control panel indications needed for diagnosis are simple and easily found. | N/P |
| **Execution Complexity** | Time is sufficient to execute post-diagnosis actions, which are clearly specified in EOPs and are considered step-by-step. This contributes to low execution HEP. | N/P |

| | | |
|---|---|---|
| **Training** | Adequate training is provided and operators are quite familiar with the pattern of stimuli associated with the event. Thus diagnosis can be considered rule based and execution can be considered step-by-step. This contributes to low total HEP. | N/P |
| **Experience** | | N/A |
| **Procedural Guidance** | System based EOPs are in place and well designed for the scenario. This contributes to low total HEP. | N/P |
| **Human-Machine Interface** | | N/A |
| **Work Processes** | | N/A |
| **Communication** | | N/A |
| **Team Dynamics** | | N/A |
| **Other** | | N/A |

#### *G1.5.4 Comparison of Drivers to Empirical Data*

Consistent with the actual crew data, ASEP assessment Team 1 did not predict any negative drivers. The assessment team expected stress, adequacy of time, scenario complexity, and execution complexity to be nominal drivers and contribute to a low HEP, while the crew data suggested that these factors were not a driver. Considering the assessment team indicated that the combination of training and symptom-based EOP was the major reason for a low HEP, the assessment team's analysis was considered to be consistent with crew data.

#### *G1.5.5 Comparison of Qualitative Analysis to Empirical Data*

The qualitative analysis of ASEP Team 1 was in agreement with crew data. As predicted, the crews were able to detect tube rupture in SG C based on radiation alarms and other cues. After quick diagnosis, the crews tripped the reactor, adjusted the SG PORV setpoint, isolated the ruptured SG, and achieved cooldown and depressurization according to procedure steps. It should be noted that the assessment team did not expect any crew to enter Procedure POP4-RC4. This may be because the procedure was not provided to them. The crews who entered this procedure only stayed momentarily in the procedure because the procedure was not relevant to the scenario.

#### *G1.5.6 Impact on HEP*

The factors identified as potentially affecting performance of this HFE had a direct impact. Execution HEP drives the total HEP.

### G1.6 Comparison Summary

#### *G1.6.1 Predictive Power*

Overall, the predictive power of the Team 1 ASEP analysis in this study is considered to be moderately good. Qualitatively, the predicted performance drivers do not correspond exactly with the factors identified from the crew data, but the analysis did generally identify the most important factors, particularly for HFE 2A. The drivers seemed to have the biggest disconnect with HFE 1C, in that the analysis failed to discover some important factors that affected crew behavior. Quantitatively, the final HEPs reflect the impact of the performance drivers that were accurately identified in the analysis, and seem to be consistent with crews' failure rates and HFE difficulty ranking.

G1.6.1.1 Qualitative Predictive Power in Terms of Drivers

The ability of the Team 1 ASEP analysis to predict important performance drivers varied across HFEs, but in general the assessment team did a moderately good job. Notably, the method accurately predicted most important drivers for the most difficult scenario – HFE 2A. Also the rationale for the accurately predicted important drivers was in agreement with the crew data.

It appears that the Team 1 ASEP analysis is sensitive to the time window defined in HFEs. This may be because time estimation is required to estimate diagnosis HEP. On one hand, this makes it easy for the method to identify if time is sufficient to complete a certain task. For example, the method successfully identified the time issue in HFE 1C. On the other hand, adequacy of time is likely to be falsely identified as a negative driver due to conservatism or incorrect assumptions in estimating execution time ($T_a$) and maximum allowable time ($T_m$) defined in ASEP Figure 6-3. Take HFE 1A as an example, the ASEP assessment Team 1 overestimated $T_a$ by assuming that the crews would go through more procedures than they actually did, as a result, the method predicted adequacy of time to be a negative driver.

It seems that there are two reasons for the method to make inconsistent prediction with empirical data. The first reason may be that the method did not lead the assessment team to fully understand the nature of the scenarios that the crews would face. For instance, the assessment team failed to consider the SGTR scenario under a time pressured context and concluded that scenario complexity was not a negative driver. The other reason may be that crews fail before evidence can be observed to evaluate assessment team's prediction on a certain driver. For example, although both the HRA team and the crew data agreed that execution complexity was not a driver for HFE 2A, the HRA team suggested that the factor might have been a negative driver if the diagnosis HEP had not been 1.0 and the execution HEP had been considered. In a scenario where evidence cannot be observed but predictions are made, discrepancies between predictions and actual crew data may occur.

G1.6.1.2 Qualitative Predictive Power in Terms of Operational Expressions

In general, the qualitative analysis provided a moderately good description of what would be going on in the scenario. Particularly for HFE 1C, the analysis accurately predicted that rather than observing failures to diagnose, it is expected to see some crews failing to complete the scenario on time based on the time limit.

Partly due to the need for execution time estimation, the analysis focused on stepping through procedures and associated response execution at a high level. Possible procedural paths and procedural diagnosis opportunities were identified with reasonable assumptions made. The procedure-related prediction was logical and reasonable, but it showed optimism as well as conservatism. On the optimistic side, the analysis assumed diagnosis success once the crews entered a procedure step without examining the sub-steps within the step. For instance, stopping HHSI pumps was a critical procedure step in HFE 1C. One crew failed the step due to a cognitive error during a sub-step. On the conservative side, the analysis did not consider the possibility that the crews would take shortcuts. The crew data showed that the crews did not enter some procedures as expected due to their knowledge and experience.

Influence of performance shaping factors (PSFs) (e.g., training, experience, and stress level, symptom-based procedures) on crews' behavior was assessed. The ability of the method seemed to be limited in addressing the plant conditions (operational situation) that could cause the crews to have problems with understanding the situation and appropriately completing the action. Take HFE 1C as an example, although the method recognized that time would not be sufficient, it did not address the possibility that the crews would be held up in FRH1 and were not able to reach EO-30.

G1.6.1.3 Quantitative Predictive Power

The Team 1 ASEP analysis provided a good HEP prediction for all HFEs. As shown in Figure 0.1, the HEPs are consistent with the crews' failure rates and HFE difficulty ranking. One may argue that the analysis did not discriminate HFE 1A and HFE 3A, because HFE 1A was ranked as more difficult than HFE 1C, but they had approximately the same HEP. On the other hand, the HEPs were consistent with the fact that the failure rates of the two HFEs were both zero. In addition, the fact that the error factor for HFE 3A was smaller than that for HFE 1A may suggest that the difference in difficulty levels is factored into the difference in error factors.

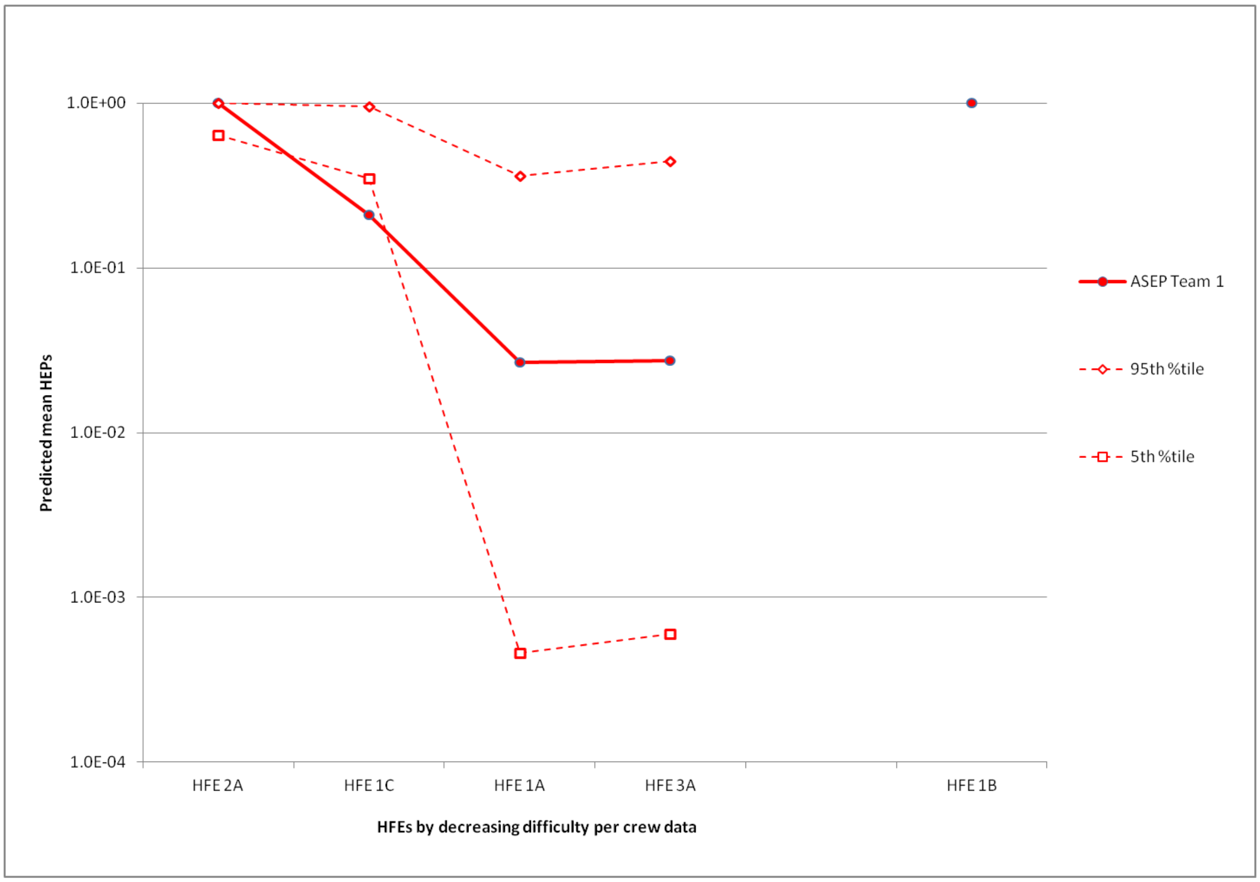


Figure 0.1 Predicted Mean HEPs with Uncertainty Bounds by HRA Team 1 with ASEP

*G1.6.2 Assessment of Guidance and Traceability*

As discussed in Section 0, the guidance of the method is limited in identifying important performance drivers and addressing plant situations that would affect crews' performance. The limitations seem to have their roots in the method's deficiencies in addressing crews' diagnosis/cognitive activities. Firstly, the focus on crews' interaction only with the main procedure steps is likely to cause analysts to ignore diagnosis activities required within a particular procedure step, and thus lead to optimistic results by failing to identify important driving factors for cognitive errors.

Secondly, the impact of fast diagnosis on diagnosis HEP due to PSFs such as experience and training is, to some degree, taken into account in the form of PSF adjustment to the HEP; however, there is not adequate guidance on how to address the impact of fast diagnosis on execution HEP estimation. The inadequacy in guidance will likely to lead to conservative execution time estimation by ignoring the possibility of fast diagnosis, which will in turn result in conservative total HEP.

Thirdly, diagnosis HEP is estimated only with ASEP diagnosis curve with PSF adjustment. Simplicity is one of the advantages of this approach, however, by ignoring the details of how PSFs interact with crews' cognitive behavior, this will probably limit the method's ability to identify important plant conditions or operational situations that will negatively impact crews' behavior. Therefore, it is concluded that although there may be situations where this approach would be appropriate, it seems clear that additional explicit guidance is necessary, especially for critical tasks at the more cognitive level.

The derivation of the HEP within the method and what is important to performance given the factors considered is generally traceable, and how the various factors are weighted in determining the final HEP can be determined. However, how analysts might bias the rating the factors considered, based on other information identified that is not covered by the method would be difficult to trace. In summary, traceability is considered to be moderately good.

*G1.6.3 Insights for Error Reduction*

One conclusion from the ASEP analysis of the study is that experience and training can enhance crews' performance. Given that the insight of the conclusion is quite broad and already widely recognized, it is concluded that the ASEP analysis did not appear to provide good insights for error reduction. As discussed in Section 0, this may be partly due to the simplicity of the method, which limits the method's ability to discover error mechanisms. Without a clear understanding of error mechanisms, it is not surprising that the method cannot offer many insights for error reduction.

*G1.6.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses*

Based on the discussion in the previous sections, it is concluded that the predictive power of the ASEP Team 1 application in this study was considered to be moderately good. Admittedly, in addition to the method's guidance, the assessment team should also take some credit for their detailed analysis and careful thinking.

Simplicity, ease of use, and traceability are the method's strengths. Interestingly, the method's weaknesses seem to stem from its strength – simplicity. Diagnosis HEP is estimated only with ASEP diagnosis curve with PSF adjustment, and diagnosis success is assumed once a major procedure step is entered. This simplified approach limits the method's ability to discover error mechanisms for cognitive failures, and leads to deficiencies in identifying important plant conditions or operational situations that will negatively impact crews' behavior.

Compared to diagnosis, the method appears to focus relatively more on post-diagnosis actions. A set of execution specific PSFs is evaluated, however, the simple consideration of stress level and whether the actions are step-by-step or dynamic would not appear to be adequate in some cases.  In addition, the focus on procedures tends to lead to the method's deficiency in addressing the impact of fast diagnosis on HEP estimation.

As pointed out in Section 0, more assessment guidance appears to be necessary to improve the method's ability to discover useful information that could affect crew performance, especially for critical tasks at the more cognitive level.

## G.2 ASEP/THERP (Team 2)

### G2.1 HFE 1A

*G2.1.1Summary of Qualitative Findings*

The HRA team noted that, for this HFE, diagnosing the need for F&B (F&B) should be easy as the operator should enter the correct procedure based on loss of feedwater and will be monitoring steam generator water level. The HRA team did recognize, however, that initial efforts related to closing the operating AFW recirculation valve may inhibit the operator from deciding to initiate F&B to avoid releasing primary fluid into the containment. Nonetheless, the HRA team concluded that the actual initiation of F&B should be relatively easy and quick to accomplish.

*G2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP= .004, (lognormal EF= 10)

Diagnosis HEP:

- $T_m$= 45 minutes – 1 or 2 minutes (for cue) = 43 or 44 minutes. 44 minutes is chosen for maximum time to diagnose and complete actions.
- Assumes 5-minute delay due to reading written procedures (ASEP step 5 a) and 1 minute per control room action, totaling 2 minutes for actions. This leads to $T_a$= 7 minutes.
  - $T_d$= $T_m$-$T_a$, thus $T_d$ (time for diagnosis) = 36 minutes.
    - Analyst's note: The minimum Tm of 43 minutes (= 45-2) was used in evaluating HFE-1a because MAPP run provided that 50% WR level was reached 80 seconds after a manual reactor trip. In reality, both the minimum and maximum Tms (i.e., 1 and 2 minute cue times) were used in the evaluation, and the minimum was selected because the results were insensitive to $T_m$ (the HEP is dominated by failure to perform the actions to initiate F&B, and not by diagnosis). (Figure 8-1 cannot be read with great accuracy and thus using an exact value of Tm is not critical).
- Using Figure 8-1 in ASEP manual and Td, the Diagnosis HEPs are as follows:
  - Nominal: 5E-4
  - Lower bound: 4E-5 (*This is selected based on operator interviews)
  - Upper bound: 7E-3

Post-Diagnosis Actions HEP:

- As post-diagnosis actions are determined to be "step-by-step" with "moderately high stress", analyst uses ASEP table 8-5 to determine that the probability of failure to initiate F&B is 0.02 per action (2, completely dependent actions) with recovery being 0.2
- This leads to Post-Diagnosis Action HEP = (0.02)(1.0)(0.2) = 4E-3

Total-Failure Probability (approximation): $P_d + (1 - P_d)P_a$ = 4E-5 + 4E-3 = **4E-3**

*G2.1.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | The HEP for time for diagnosis was chosen from the lower bound curve. Also, ASEP specifies at least moderately high stress in the first two hours, and all of these actions need to be done in two hours. | N/P |
| **Time Pressure** | Reflected in the base HEP selected based on the adequacy of time and the actions subsequent to an initial human failure event | N/A |
| **Stress** | ASEP only considers two degrees of stress: moderately high and extremely high stress. Either option is then combined with execution complexity. The analyst chose moderately high stress for this case. | N/P |
| **Scenario Complexity** | ASEP does not explicitly consider scenario complexity, but it does include explicit consideration of training, experience, procedure guidance, indications of conditions, and familiarity with the pattern of stimuli, which may get at complexity related issues (and these are discussed under the appropriate criteria below). | N/A |
| **Indications of Conditions** | The operator's cue to initiate F&B per FRH1 when wide range levels on 2 SGs are less than 50%-- this analysis assumes this occurs within 1 to 2 minutes of the scenario. The analyst noted that this should be a straightforward diagnosis as operators should be looking at the steam generator level indication and correctly enter procedure | 0 |
| **Execution Complexity** | ASEP only considers two options for execution complexity: step-by-step or dynamic. This scenario involves step-by–step actions, which imply simple execution (actions could be performed without procedure guidance). | N/P |
| **Training & Experience** | Analyst chose the lower bound for diagnosis HEP as operators indicated during their interviews that they are aware of the need for F&B under these circumstances. Note: ASEP assumes that a novice operator is quickly replaced with an experienced operator for the purposes of assigning the HEP. Also, the analyst selected step-by-step execution complexity which implies "a routine, procedurally guided set of steps performed one step at a time without a requirement to divide ones attention between the task in question and other tasks. With high levels of skill and practice a step-by-step task may be performed reliability without recourse to written procedures." This could mean that the operators are well-trained in such routine tasks. | N/P |
| **Procedural Guidance** | Assuming procedures are being followed. Analyst chose "step-by-step" for execution complexity, so this implies that there is procedural guidance. (Again, the definition of step-by-step is "a routine, procedurally guided set of steps," e.g. procedures exist for such tasks.) | N/P |

| | | |
|---|---|---|
| **Human- Machine Interface** | Covered by “adequacy of time” and “indications of conditions”-specific rules are assigned depending on control panel layout. For example, in Table 8-1, step 5b states “Assess 1 minute as the required travel and manipulation time combined for each control room (CR) control action taken on the primary operating panels which are normally in visual access of the CR operator.” In this way, HMI considerations in ASEP are deferred to adequacy of time calculations. | N/A |
| **Work Processes** | ASEP does not explicitly consider work processes—in this table, its influences are considered under team dynamics. | N/A |
| **Communication** | ASEP does not explicitly consider communication—in this table, its influences are considered under team dynamics. | N/A |
| **Team Dynamics** | | N/A |
| **Other** | Consideration of multiple events; subsequent events given higher HEP. In other words, dependency and recovery taken into account. In other words, the default for ASEP is that the HEP covers the entire crew (see step 9g), but there is some provision for breaking out individuals if justified. The one exception can be found in Table 8-5, where recovery credit can be assigned for a “second person who checks the performance of the original performed”—the analyst gives credit for this recovery in this analysis. | N/A |

*G2.1.4 Basic comparison of qualitative predictions—operational stories*

In general, the analyst’s qualitative assessment was consistent with the crews’ operational stories. The analyst predicted that the diagnosis of the need for F&B would be easy in this scenario as the operators would enter the correct procedure (FR-H1) based on loss of feedwater and that the crews will be monitoring steam generator water levels. In addition, the analyst assessed that the actual initiation of F&B in this situation is also relatively easy and quick to accomplish. The relative ease of both the diagnosis for and initiation of F&B, as predicted by the analyst, proved to be true based on crew data. All four crews manually tripped the reactor and were able to start F&B within the 45-minute time limit. All crews detected that the level in SG B continued to decrease even though there was indication of AFW flow, and the crews questioned if they actually had flow to the SG. The STAs in all four crews verified that they had entry criteria to start FR-H1, which all crews did within 11 to 15 minutes of the scenario. When the crews started FR-H1 the wide range levels in the SGs were at 50%, and they, therefore, met the criteria to start F&B, which they all did.

The analyst did identify that the only potential complicating factor would involve initial efforts related to closing the operating AFW recirculation valve. It was observed that the difficulty in the scenario was, in fact, identifying that all AFW was lost due to a mispositioned recirculation valve’s remaining open. Three of the four crews suspected this. Nonetheless, all crews successfully initiated F&B within the time criterion.

*G2.1.5 Basic comparison of qualitative predictions—drivers*

Though not explicitly called out in the qualitative assessment, the analyst seemed to identify both negative and positive driving factors in this scenario. The analyst seemed to distinguish Adequacy of Time, Execution Complexity, Training and Experience, Stress, Team Dynamics, and Procedural Guidance as either positive or nominal. The empirical data was consistent with all of these indications except the last. Both the analysis and the data suggested that 45 minutes was enough time to both diagnose and initiate F&B. The analyst chose the lower diagnosis HEP, and also designated post-diagnosis actions as the less conservative "moderately high stress" (as actions needed to be performed in less than 2 hours). Both of these decisions indicated a confidence in the crews' time adequacy. Both the analyst and the data suggested a relatively straightforward Execution Complexity—the analyst implied this by selecting step-by-step complexity versus dynamic complexity. In addition, both the data and the analyst stated that the crews should be well-trained and do have experience (indicated during interviews and actual plant events) on the components of this scenario, thus a positive factor. Another nominal/positive factor that analyst indicated is stress—as ASEP only considers two degrees of stress, the analyst chose the less conservative, moderately high stress level. This nominal effect seems consistent with the data's zero effect of stress, as is also the case with Team Dynamics. The analyst's last N/P factor, Procedural Guidance, was based on his decision again to choose step-by-step execution complexity. This implied that there is procedural guidance for this scenario, and thus N/P effect would be appropriate. The data uncovered, however, that the Critical Safety Functions status trees do not address misaligned valves.

The data indicated Scenario Complexity, Indications of Conditions, and Procedural Guidance as negative drivers. Although the analyst did not explicitly identify any negative drivers, their qualitative explanations seemed nonetheless consistent with the negative drivers' effects, besides the Procedural Guidance explanation. For instance, Scenario Complexity and Indications of Conditions were identified as negative drivers based on the loss of AFW's being masked by an open recirculation valve—a difficulty the analyst qualitatively identified but that did not cause him to select the more conservative options and thus did not drive the HFE. In addition, ASEP does not explicitly consider Scenario Complexity, so this needed to be evaluated within other categories, such as training and experience, procedural guidance, and indications of conditions.

*G2.1.6 Quantitative Comparison*

For this analysis, the HEP is dominated by failure to perform the actions to initiate F&B (post-diagnosis actions rather than by diagnosis). In this respect, both the minimum and maximum times for diagnosis ($T_d$, for this scenario-1 and 2 minute cue times) were used in the evaluation, and the analyst chose the minimum time because the results were insensitive to $T_d$. Figure 8-1 cannot be read with great accuracy and thus using an exact value of $T_d$ is not critical.

The HEP of 0.004, which the analyst calculated, loosely aligned with the empirical data. In reference to all of the scenario HEPs, this HFE ranked third (tied with HFE 3A) and was, in fact, third in the observed difficulty scale. The HEP was relatively low and is consistent with the fact that no crews failed this scenario.

## G2.2 HFE 1B

*G2.2.1 Summary of Qualitative Findings*

The HRA team noted that, for this HFE, diagnosing the need for F&B should be easy as the operator should enter the correct procedure based on loss of feedwater and will be monitoring steam generator water level. The HRA team did recognize, however, that initial efforts related to closing the operating AFW recirculation valve may inhibit the operator from deciding to initiate F&B to avoid releasing primary fluid into the containment. Nonetheless, the HRA team concluded that the actual initiation of F&B should be relatively easy and quick to accomplish. (same qualitative findings as for HFE 1A)

*G2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
HEP= .05 (lognormal EF= 10)

Diagnosis HEP:

- $T_m$= 13 minutes (given in scenario) – 1 or 2 minutes (for cue) = 11 or 12 minutes. 12 minutes is chosen as maximum time to diagnose and complete actions.
- Assumes 5-minute delay due to reading written procedures (ASEP step 5 a) and 2 minutes to actually perform the necessary actions in the control room. This leads to $T_a$ = 5 + 2 = 7 minutes.
  - $T_d$= $T_m$ – $T_a$, Thus $T_d$ (time for diagnosis) = 5 minutes.
    - Analyst's note: If the reactor is not manually tripped, per a simulator run, an automatic reactor trip occurs at 96 seconds and the 50% SG WR level is reached 8 seconds later (the PRA used 75 seconds). The minimum $T_m$ of 43 minutes (45 minus 2) was used in evaluating HFE-1a based on the 80 8seconds time from the MAPP run. In reality, both the minimum and maximum $T_m$'s (i.e., 1 and 2 minute cue times) were used in the evaluation and the minimum was selected because the results were insensitive to Tm (the HEP is dominated by failure to perform the actions to initiate F&B, and not by diagnosis). The maximum $T_m$ of 12 minutes (13-1) was used in evaluating HFE-1b based on the 75 second cue time from the MAPP run. Again, both the minimum and maximum $T_m$ were used in the evaluation and the diagnosis error probability was found to be relatively insensitive to which value was used. (Figure 8-1 cannot be read with great accuracy and thus using an exact value of $T_m$ is not critical). However, for the HFE-1b case, the diagnosis dominates the HEP and thus the maximum Tm was chosen as most representative based on the provided information.
- Using Figure 8-1 in ASEP manual and $T_d$ of 5, the Diagnosis HEPs are as follows:
  - Nominal HEP= 0.2
  - Lower bound HEP= .05 (*This is selected based on operator interviews)
  - Upper bound HEP= 1.0

Post-Diagnosis Actions HEP:

- As post-diagnosis actions are determined to be "step-by-step" (because actions could be performed without procedural guidance) with "moderately high stress" (since multiple systems failed and the actions must be performed in less than 2 hours), analyst uses ASEP table 8-5 to determine that the probability of failure to initiate F&B is 0.02 per action (2 actions required to initiate F&B, which are completely dependent) with recovery being 0.2.
- This leads to Post-Diagnosis Action HEP = (0.02)(1.0)(.2)= 4E-3

Total-Failure probability (approximation): $P_d + (1-P_d)P_a$ = 5E-2 +4E-3= **5E-2**

*G2.2.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Although the lower bound for diagnosis HEP was chosen, it was a significantly higher HEP than HFE 1A, which was the same task in a longer time. Also, the fact that actions needed to be performed in less than 2 hrs allowed moderate stress to be designated for post-diagnosis actions. | MND |
| **Time Pressure** | Reflected in the base HEP selected based on the adequacy of time and the actions subsequent to an initial human failure event | N/A |
| **Stress** | ASEP only considers two degrees of stress: moderately high and extremely high stress. Either option is then combined with execution complexity. The analyst chose moderately high stress for this case. | N/P |
| **Scenario Complexity** | ASEP does not explicitly consider scenario complexity, but it does include explicit consideration of training, experience, procedure guidance, indications of conditions, and familiarity with the pattern of stimuli, which may get at complexity related issues (and these are discussed under the appropriate criteria below). | N/A |
| **Indications of Conditions** | The operator's cue to initiate F&B per FRH1 when wide range levels on 2 SGs are less than 50%-- this analysis assumes this occurs within 1 to 2 minutes of the scenario. In this case, the analyst noted that this should be a straightforward diagnosis as operators should be looking at the steam generator level indication and correctly enter procedure | 0 |
| **Execution Complexity** | ASEP only considers two options for execution complexity: step-by-step or dynamic. This scenario involves step-by-step actions, which imply simple execution (actions could be performed without procedure guidance) and was chosen for this scenario. | N/P |
| **Training & Experience** | Analyst chose the lower bound for diagnosis HEP as operators indicated during their interviews that they are aware of the need for F&B under these circumstances. Note: ASEP assumes that a novice operator is quickly replaced with an experienced operator for the purposes of assigning the HEP. Also, the analyst selected step-by-step execution complexity which implies "a routine, procedurally guided set of steps performed one step at a time without a requirement to divide ones attention between the task in question and other tasks. With high levels of skill and practice a step-by-step task may be performed reliability without recourse to written procedures." This could mean that the operators are well-trained in such routine tasks. | N/P |
| **Procedural Guidance** | Assuming procedures are being followed. Analyst chose "step-by-step" for execution complexity, so this implies that there is procedural guidance. (Again, the definition of step-by-step is "a routine, procedurally guided set of steps," e.g. procedures exist for such tasks.) | N/P |

| | | |
|---|---|---|
| **Human-Machine Interface** | Covered by "adequacy of time" and "indications of conditions"-specific rules are assigned depending on control panel layout. For example, in Table 8-1, step 5b states "Assess 1 minute as the required travel and manipulation time combined for each control room (CR) control action taken on the primary operating panels which are normally in visual access of the CR operator." In this way, HMI considerations in ASEP are deferred to adequacy of time calculations. | N/A |
| **Work Processes** | ASEP does not explicitly consider work processes—in this table, its influences are considered under team dynamics. | N/A |
| **Communication** | ASEP does not explicitly consider communication—in this table, its influences are considered under team dynamics. | N/A |
| **Team Dynamics** | | N/P |
| **Other** | Consideration of multiple events; subsequent events given higher HEP. In other words, dependency and recovery taken into account. In other words, the default for ASEP is that the HEP covers the entire crew (see step 9g), but there is some provision for breaking out individuals if justified. The one exception can be found in Table 8-5, where recovery credit can be assigned for a "second person who checks the performance of the original performed"—the analyst gives credit for this recovery in this analysis. | N/A |

*G2.2.4 Comparison*

There is no data for this HFE; no crew had an automatic reactor trip. Thus, there is no basis for comparison.

## G2.3 HFE 1C

*G2.3.1 Summary of Qualitative Findings*

The HRA team recognizes the difficulty the operators would face in diagnosing an SGTR under these circumstances. The analyst identifies that this scenario involves F&B and feeding a steam generator with AFW flow, which is assumed to cause the SGTR (due to thermal shock). In this situation, it would be hard to determine there is a tube rupture based on radiation detection in the main steam lines, especially if the MSIVs are closed (radiation detection in the SG blow down lines would not be possible since these lines would be isolated). In addition, it would not be easy to detect an increase in the secondary level from the flow through the ruptured tube as there is a flow of 600 gpm into the SG secondary. Similarly, detection on the primary side would also be difficult when F&B is in operation as the crew could attribute the loss of inventory to SI flow. Thus, while performing the actions to isolate the ruptured SG is more straightforward, detecting the SGTR itself could be very difficult.

*G2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP= .03 (lognormal, EF= 10)

Diagnosis HEP:

- $T_m$= 40 minutes. 40 minutes is chosen for maximum time to diagnose and complete actions.
- Assumes 5-minute delay due to reading written procedures (ASEP step 5 a) and 5 to 10 minutes to actually perform the action (note: the estimation of 5 minutes to perform the actions is given by the operator, so, per ASEP step 5 e, this figure is doubled to get 10 minutes). This leads to $T_a$= travel time + action time = 5 + 10. $T_a$= 15 minutes.
- $T_d$= $T_m$ –$T_a$, thus $T_d$ (time for diagnosis)= 40 – 15= 25 minutes
  - Analyst's note: The analyst performed sensitivity studies when there was variability in values and selected HEPs based on the results of the studies. In this case, the analyst selected the minimum diagnosis time of 25 minutes as the basis for the evaluation. When reading Figure 8-1, it is difficult to ascertain any significant difference when using either 25 or 30 minutes as the diagnosis time.
- Using Figure 8-1 in ASEP manual and Td, the Diagnosis HEPs are as follows:
  - Nominal diagnosis HEP= 3E-3
  - Lower bound HEP= 3E-4
  - Upper bound HEP= 3E-2 (*This is selected based on difficulty of diagnosis.)

Analyst's note: Uncertainty on when a cue would be received was reflected in the analysis by using the upper bound of the diagnosis HEP when the cue was assumed to occur at 0 minute. The resulting HEP (3E-2) is equivalent to the nominal HEP (a nominal HEP should be used when the timing of the cue is reasonably known) when a cue is received within 10 minutes of the SGTR ($T_d$ = 15 minutes). Although, the ASEP rules do not specifically provide for this adjustment for the reason cited, some way was needed to account for variability in the possible cue timing since that information is highly variable for this scenario.

Post-Diagnosis Actions HEP:

- As post-diagnosis actions are determined to be "step-by-step" (actions could be performed without procedural guidance) with "moderately high stress" (since multiple systems failed and the actions ASEP assumes moderately high stress for a minimum of two hours), analyst uses ASEP table 8-5 to determine that the probability of failure to isolate SGTR is 0.02 per action (3) actions required to isolate SGTR, which are completely dependent) with recovery being 0.2.
- This leads to Post-Diagnosis Action HEP = (0.02)(1.0)(.2)= 4E-3

Total Failure Probability (approximation): $P_d + (1-P_d)P_a$ = 3E-2 + 4E-3= **3E-2**

*G2.3.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Time for diagnosis drives HEPs for diagnosis. The crew has less time than in HFE3A to perform the same actions, so the diagnosis HEP bounds are much wider than in HFE3A.). Also, ASEP specifies at least moderately high stress in the first two hours, and all of these actions need to be done within two hours (as stated in the analysis, the time window to perform the required actions was estimated to be approximately 40 minutes). | ND |
| **Time Pressure** | Reflected in the base HEP selected based on the adequacy of time and the actions subsequent to an initial human failure event | N/A |
| **Stress** | ASEP only considers two degrees of stress: moderately high and extremely high stress. Either option is then combined with execution complexity. The analyst chose moderately high stress for this case. | N/P |
| **Scenario Complexity** | ASEP does not explicitly consider scenario complexity, but it does include explicit consideration of training, experience, procedure guidance, indications of conditions, and familiarity with the pattern of stimuli, which may get at complexity related issues (and these are discussed under the appropriate criteria below), which would make diagnosis relatively more difficult than HFE 3A. | N/A |
| **Indications of Conditions** | The analyst noted that it would be difficult to determine there is a tube rupture based on radiation detection in the main steam lines especially if the MSIVs are closed (detection of radiation in the SG blow down lines would not be possible since these lines would be isolated). With a flow of 600 gpm into the SG secondary, it would not be easy to detect an increase in the secondary level from the flow through the ruptured tube. | MND |
| **Execution Complexity** | ASEP only considers two options for execution complexity: step-by-step or dynamic. This scenario involves step-by-step actions, which imply simple execution (actions could be performed without procedure guidance). | N/P |
| **Training & Experience** | The analyst selected step-by-step execution complexity which implies “a routine, procedurally guided set of steps performed one step at a time without a requirement to divide ones attention between the task in question and other tasks. With high levels of skill and practice a step-by-step task may be performed reliability without recourse to written procedures.” This could mean that the operators are well-trained in such routine tasks. Note: ASEP assumes that a novice operator is quickly replaced with an experienced operator for the purposes of assigning the HEP. | N/P |
| **Procedural Guidance** | Assuming procedures are being followed. Analyst chose “step-by-step” for execution complexity, so this implies that there is procedural guidance. (Again, the definition of step-by-step is “a routine, procedurally guided set of steps,” e.g. procedures exist for such tasks.) | N/P |

| | | |
|---|---|---|
| **Human-Machine Interface** | Covered by "adequacy of time" and "indications of conditions"-specific rules are assigned depending on control panel layout. For example, in Table 8-1, step 5b states "Assess 1 minute as the required travel and manipulation time combined for each control room (CR) control action taken on the primary operating panels which are normally in visual access of the CR operator." In this way, HMI considerations in ASEP are deferred to adequacy of time calculations. | N/A |
| **Work Processes** | ASEP does not explicitly consider work processes—in this table, its influences are considered under team dynamics. | N/A |
| **Communication** | ASEP does not explicitly consider communication—in this table, its influences are considered under team dynamics. | N/A |
| **Team Dynamics** | | N/P |
| **Other** | Consideration of multiple events; subsequent events given higher HEP. In other words, dependency and recovery taken into account. In other words, the default for ASEP is that the HEP covers the entire crew (see step 9g), but there is some provision for breaking out individuals if justified. The one exception can be found in Table 8-5, where recovery credit can be assigned for a "second person who checks the performance of the original performed"—the analyst gives credit for this recovery in this analysis. | N/A |

*G2.3.4 Basic comparison of qualitative predictions—operational stories*

The analyst accurately identified that it would be difficult to determine that there is a tube rupture based on radiation detection in the main steam line. Additionally, the analyst predicted that it would be difficult to detect an increase in the secondary level from the flow through the ruptured tube (with a flow of 600 gpm into the SG secondary). The crews did have difficulty immediately detecting the SGTR, as the analysis predicted, but the ASEP methodology is perhaps not extensive enough to anticipate all the difficulties the crews encountered. However, the HRA team did not consider the effects of the actions taken earlier in the scenario. For instance, the crews had to first follow FRH1 steps to stop safety injection and close the pressurizer PORVs as a function of RCS subcooling, which was a crucial step to maintain the SG pressure below the setpoint.

The crews were filling the ruptured SG to be able to stop F&B in procedure FR-H1, and as long as they were, the tube rupture was masked. Eventually all crews stopped AFW to the ruptured SG on the basis of having a high enough level to stop F&B and in order to cross-connect to start feeding the other SGs. At this point, the crews were unaware of the SGTR. After having stopped AFW, the crews quickly noticed that the level continued to rise and thus determined that they had a SGTR. Once the crews diagnosed the SGTR, they took different routes to isolate the ruptured SG, based on what procedures they were currently using. All four crews isolated the ruptured SG—three from knowledge and one from procedural guidance—but only one accomplished this within the 40 minute time criterion. ASEP is not necessarily equipped to predict the varying crew results (e.g. different procedural paths) and minor procedural missteps.

Quantitatively, adjustments to the diagnosis HEP were made based on the expertise of the analyst rather than exclusively based on the method, specifically the upper bound diagnosis HEP was chosen based on the difficulty of the diagnosis, which is not a liberty ASEP explicitly allows. While this probably led to a more accurate HEP, the same results may not have been procured by a more novice analyst.

*G2.3.5 Basic comparison of qualitative predictions—drivers*

Though not explicitly called out in the qualitative assessment, the analyst seemed to identify both negative and positive driving factors in this scenario. The analyst seemed to distinguish Execution Complexity, Training and Experience, Stress, Team Dynamics, and Procedural Guidance as either positive or nominal. The empirical data was more or less consistent with all of these conclusions except the last. Both the analyst and the data suggested a relatively straightforward Execution Complexity—the analyst implied this by selecting step-by-step complexity versus dynamic complexity. In addition, both the data and the analyst stated that the crews should be well-trained on the components of this scenario, thus a positive factor. Another nominal/positive factor that the analyst indicated is stress—as ASEP only considers two degrees of stress, the analyst chose the less conservative, moderately high stress level. This nominal effect seems consistent with the data's zero effect of stress, as is also the case with Team Dynamics. The analyst's last N/P factor, Procedural Guidance, was based on his decision again to choose step-by-step execution complexity (the definition of step-by-step is "a routine, procedurally guided set of steps," in other words procedures exist for such tasks.). This implies that there is procedural guidance for this scenario, and thus N/P effect would be appropriate. The data indicated, however, that the procedures were a negative driver. Three crews could not go to procedure E-30 because they were held up in the higher priority procedure FR-H1 and FR-P1. Three crews isolated the ruptured SG from knowledge (one crew's having support from ZA-18 (EOP users' guide). One crew was procedurally driven to go to E-30. They incorrectly exited the FR-H1 without stopping the last HHSI pump.

The analyst indirectly identified the adequacy of time, scenario complexity, and indications of conditions as negative drivers—all consistent with the empirical data. Since the SGTR was masked by AFW for some time following the rupture, the crews had less than 40 minutes to diagnose the SGTR, work through the procedures, and accomplish the required actions. Given that even the fastest crew barely made the time criterion, adequacy of time seemed empirically to be a negative driver. This rationale is consistent with the analyst's selection of the upper bound diagnosis HEP. Additionally, though ASEP does not explicitly consider scenario complexity, it does include explicit evaluation of training, experience, procedural guidance and indications of conditions. In this case, the analyst noted that diagnosis would be more difficult than in a "straightforward" SGTR (as in HFE 3A) due to complexity of circumstances. Thus, the factors that contribute to scenario complexity are deemed as negative drivers according to the analysis. This is consistent with the data's marking scenario complexity as the main negative driver as the rupture was initially masked by the AFW flow to the ruptured SG and the low steam flow prevented radiation alarms. These considerations are also related to the assessment's final negative driver—indications of conditions. The analyst noted that it would be difficult to determine that there is an SGTR based on radiation detection in the main steam lines.

The drivers were, therefore, relatively consistent between the data and the assessment. ASEP, however, does not explicitly consider HMI, work processes, or communication, which incidentally proved to be zero factors in the crew performance.

*G2.3.6 Quantitative Comparison*

The HEP is driven by the diagnosis HEP, for which the analyst selected the upper bound. In this regard, uncertainty on when a cue would be received was reflected in the analysis by using the upper bound of the diagnosis HEP when the cue was assumed to occur at 0 minutes. The resulting HEP (3E-2) is equivalent to the nominal HEP (a nominal HEP should be used when the timing of the cue is reasonably known) when a cue is received within 10 minutes of the SGTR ($T_d$ = 15 minutes). Although, the ASEP rules do not specifically provide for this adjustment for the reason cited, the analyst noted that some way was needed to account for variability in the possible cue timing since that information is highly variable for this scenario.

The analyst predicted an HEP of .03 for this HFE, and only one crew accomplished the success criteria within the 40 minute time frame. (Three crews accomplished the success criteria if the 40 minute time limit is not considered.) Both the analysis and the empirical data ranked this HFE as second in the difficulty scale.

### G2.4 HFE 2A

*G2.4.1 Summary of Qualitative Findings*
The HRA team notes that the loss of the 120 VAC panel 1201 should be easily recognizable by the operators based on past experience (an actual event occurred) and training. In addition, many annunciators should result in the crew's checking for the loss of vital AC bus. According to the operators interviewed, there is no abnormal procedure related to the loss of all CCW, and no procedure that directly ties loss of CCW to need for protecting charging pumps and impact on RCP seal cooling and injection. However, the interviews indicated that such effects are within the knowledge base of the operators. For this scenario, loss of RCP cooling and injection should result in a high RCP thermal barrier alarm and entry into an RCP off normal procedure. Nonetheless, identifying and completing the necessary actions in the time required would be very difficult.

*G2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
HEP= 0.6 (Lower Bound HEP = 0.4, Upper Bound HEP =1.0)

Diagnosis HEP:

- $T_m$= 7-9 minutes. 9 minutes is chosen as maximum time to diagnose and complete actions.
- Assumes 5-minute delay due to reading written procedures (ASEP step 5 a) and 2 minutes to actually perform the necessary actions in the control room. This leads to Ta = 5 + 2 = 7 minutes.
- $T_d = T_m – T_a$, Thus $T_d$ (time for diagnosis) = 9 – 7= 2 minutes
- Using Figure 8-1 in ASEP manual and $T_d$, the Diagnosis HEPs are as follows:
  - Nominal diagnosis HEP= 0.5 (*This is selected based on interviews.)
  - Lower bound HEP= 0.3
  - Upper bound HEP= 1.0

Post-Diagnosis Actions HEP:

- As post-diagnosis actions are determined to be "dynamic" with "extremely high stress" (because multiple systems have failed and the actions must be performed quickly), analyst uses ASEP table 8-5 to determine that the probability of failing to trip the RCPs and initiate the PDPs is 0.25 per action (2 actions, assumed to be independent) with a recovery of 0.5.
- This leads to Post-Diagnosis Action HEP
  - (exact)= (.25 +.25 – (.25)^2)(.5)= .21875

Total-Failure Probability: $P_d + (1-P_d)P_a$ =

- (exact) = .5 + (1-.5)(.21875) = **0.60937**

- Note: The ASEP guidance (Step 11) explicitly calls for adding the diagnosis and post-diagnosis action HEPs. Here, the analyst used the exact approximation method, which may not be unreasonable but is not strictly what the ASEP method directs.

*G2.4.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Time for diagnosis drives HEPs for diagnosis—the nominal HEP was chosen. The 2 minutes allowed for diagnosis resulted in a very high HEP (taken from ASEP table). Also, the fact that actions needed to be performed very quickly allowed extremely stress to be designated for post-diagnosis actions as "extremely high stress" which led to the high post-diagnosis action HEP. | MND |
| **Time Pressure** | Reflected in the base HEP selected based on the adequacy of time and the actions subsequent to an initial human failure event | N/A |
| **Stress** | ASEP only considers two degrees of stress: moderately high and extremely high stress. The analyst chose extremely high stress. | ND |
| **Scenario Complexity** | ASEP does not explicitly consider scenario complexity, but it does include explicit consideration of training, experience, procedure guidance, indications of conditions, and familiarity with the pattern of stimuli, which may get at complexity related issues (and these are discussed under the appropriate criteria below). | N/A |
| **Indications of Conditions** | The analyst noted that the occurrence of multiple annunciators typically results in checking for the loss of a vital bus. | 0 |
| **Execution Complexity** | ASEP only considers two options for execution complexity: step-by-step or dynamic. Either option is combined with an option for stress (4 combinations total). Analyst designated the post-diagnosis actions as dynamic. | ND |
| **Training & Experience** | Analyst chose the nominal diagnosis HEP based on operator statements indicating that the impacts of the loss of CCW are within the knowledge base of the operators. This implies that the event is covered in training, practiced, and that all the operators know the pattern of stimuli associated with the event. Further, this implies that the operators would have been through the event during simulator requalification. Note: ASEP assumes that a novice operator is quickly replaced with an experienced operator for the purposes of assigning the HEP. | N/P |
| **Procedural Guidance** | Assuming procedures are being followed, but ASEP assumes a 5 minutes delay due to procedure reading (this affects diagnosis HEP) The analyst noted that there is a deficiency in procedural guidance that would tie a loss of CCW to the need for protecting the charging pumps and negative impacts on the RCP seal cooling injection | ND |

| | | |
|---|---|---|
| **Human- Machine Interface** | Covered by “adequacy of time” and “indications of conditions”- specific rules are assigned depending on control panel layout. For example, in Table 8-1, step 5b states “Assess 1 minute as the required travel and manipulation time combined for each control room (CR) control action taken on the primary operating panels which are normally in visual access of the CR operator.” In this way, HMI considerations in ASEP are deferred to adequacy of time calculations. | N/A |
| **Work Processes** | ASEP does not explicitly consider work processes—in this table, its influences are considered under team dynamics. | N/A |
| **Communication** | ASEP does not explicitly consider communication—in this table, its influences are considered under team dynamics. | N/A |
| **Team Dynamics** | | N/P |
| **Other** | Consideration of multiple events; subsequent events given higher HEP. In other words, dependency and recovery taken into account. In other words, the default for ASEP is that the HEP covers the entire crew (see step 9g), but there is some provision for breaking out individuals if justified. The one exception can be found in Table 8-5, where recovery credit can be assigned for a “second person who checks the performance of the original performed”—the analyst gives credit for this recovery in this analysis. | N/A |

*G2.4.4 Basic comparison of qualitative predictions—operational stories*

In general, the analyst’s qualitative assessment was consistent with the crews’ operational stories. The analyst predicted that the operators would recognize the loss of the 120 VAC distribution panel based on past experience and training. All four crews did, in fact, quickly recognize the failing distribution panel and took actions to take equipment in manual. The analyst further identified through interviews that there is no abnormal procedure related to the loss of all CCW and thus no procedure that directly links a loss of CCW to the need for protecting the charging pumps and the negative impacts on the RCP seal cooling and injection. Based on operator interviews, however, the analyst stated that the operators would know the impacts on the charging pumps and the RCPs. Empirically, while one crew tripped the RCP based on procedural guidance, the remaining three crews ordered the trip based on knowledge when they detected the loss of CCW and sealwater. Finally, the analyst noted that, for this particular scenario, the loss of RCP and cooling and injection should result in a high RCP thermal barrier alarm and entry into an RCP off normal procedure. Three of the four crews did start procedure “Reactor Coolant Pump Off Normal,” and two of these crews suggested starting the PDP.

While the qualitative analysis did cover parts of the operational stories, it did not explicitly call out the challenges posed by the short length of time that the operators had to execute the required actions. Being late in detecting the loss of CCW and sealwater not only made the crews trip the RCPs late, it also gave them less time to start the PDP before the RCP sealwater temperatures reached 230 degrees.

*G2.4.5 Basic comparison of qualitative predictions—drivers*

While the main qualitative concepts in the analysis did overlap with that of the empirical data, the two seemed to classify the driving factors somewhat differently. As this was a very challenging scenario, neither the assessment nor the operational data identified many nominal/positive drivers, if at all. As is the nature of ASEP, the lack of adequacy of time appeared to be the main negative driver, while the data determined that the main MNDs were scenario complexity and training and experience. As ASEP does not explicitly consider scenario complexity (rather, includes this in training and experience, procedural guidance, and indications of conditions) therefore the analysis did not identify this as a factor affecting performance. The analyst noted that there is a deficiency in procedural guidance to tie the loss of CCW to the need for protecting the charging pumps and negative impacts on the RCP seal cooling injection; thus procedures seemed a negative driver. Additionally, the analyst identified training and experience to nominal/positive. This is based on the fact that the analyst chose the nominal diagnosis HEP based on operator statements indicating that the impacts of the loss of CCW are within the knowledge base of the operators. This implies that the event is covered in training, practiced, and that all the operators know the pattern of stimuli associated with the event. Further, this implies that the operators would have been through the event during simulator requalification. However, the data proves this to be inaccurate. The operators were not used to this kind of scenario and the required prioritization.

*G2.4.6 Quantitative Comparison*

The high predicted HEP of 0.6 was relatively accurate in that none of the crews succeeded in meeting the criteria for this HFE. All crews stopped the RCPs, but none did it within the 1-minute time criterion. No crews started the PDP. Both the empirical data and the ASEP analysis classified this scenario as the most challenging.

## G2.5 HFE 3A

*G2.5.1 Summary of Qualitative Findings*

The HRA team acknowledges that diagnosing an SGTR in this scenario would be significantly easier than in HFE 1C. It would be possible to identify a ruptured SG based on radiation detection in the main steam lines or SG blow down lines. In addition, detection on the primary side would also be possible since the loss of inventory would result in an increase in charging flow. (There would be no masking from F&B, as in HFE 1C.) As the maximum makeup to the VCT is 200 gpm, the VTC would rapidly deplete and be indicated in the control room. (Note: During an interview with an operator, the analyst learned that there may be a shorter time frame with regards to diagnosis since there is a limit for radiation release in the licensing bases analysis that would require isolation by 30 minutes. If this is the case, then the HEP in HFE1C may be more appropriate.)

*G2.5.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP= .004 (lognormal, EF= 10)

Diagnosis HEP:

- $T_m$= 2- 3 hours (given in scenario) or 120 – 180 minutes. It seems as though analyst chose a midway point of 150 minutes as the maximum time to diagnose and complete actions.
- Assumes 5-minute delay due to reading written procedures (ASEP step 5 a) and 5 to 10 minutes to actually perform the action (note: the estimation of 5 minutes to perform the actions is given by the operator, so, per ASEP step 5 e, this figure is double to get 10 minutes). This leads to $T_a$ = travel time + action time = 5 + 10. $T_a$ = 15 minutes.
- $T_d$= $T_m$- $T_a$, thus $T_d$ (time for diagnosis)= 150 – 15 =135 minutes
- Using Figure 8-1 in ASEP manual and Td, the Diagnosis HEPs are as follows:

- Nominal diagnosis HEP= 5E-5 (*This is selected based on operator statements indicating the potential for diagnosing a tube rupture based on radiation indications would be possible under these circumstances.)
- Lower bound HEP= 1E-6
- Upper bound HEP= 2E-3

Post-Diagnosis Actions HEP:

- As post-diagnosis actions are determined to be "step-by-step" (because actions could be performed without procedural guidance) with "moderately high stress", analyst uses ASEP table 8-5 to determine that the probability of failure to diagnosis and isolate SGTR is 0.02 per action (3) actions required to isolate SGTR, which are completely dependent) with recovery being 0.2
- This leads to Post-Diagnosis Action HEP= (0.2)(1.0)(.2)= 4E-3

Total Failure Probability (approximation): $P_d + (1-P_d)P_a$ =

5E-5 + 4E-3= **4E-3**

*G2.5.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Time for diagnosis drives HEPs for diagnosis—the nominal HEP was chosen. Also, ASEP specifies at least moderately high stress in the first two hours, and all of these actions need to be done in two hours. | N/P |
| **Time Pressure** | Reflected in the base HEP selected based on the adequacy of time and the actions subsequent to an initial human failure event | N/A |
| **Stress** | ASEP only considers two degrees of stress: moderately high and extremely high stress. Either option is then combined with execution complexity. The analyst chose moderately high stress for this case. | N/P |
| **Scenario Complexity** | ASEP does not explicitly consider scenario complexity, but it does include explicit consideration of training, experience, procedure guidance, and indications of conditions (and these are discussed under the appropriate criteria below). In this case, the analyst noted that this should be a relatively easy diagnosis (significantly more so than HFE 1C). | N/P |
| **Indications of Conditions** | The analyst noted that it would be possible to determine that there is an SGTR based on radiation detection in main steam lines or SG blow down lines. Also, detection on primary side would also be possible since the loss of inventory would result in increased charging flow. The VCT would rapidly deplete. | N/P |
| **Execution Complexity** | Step by step in ASEP implies simple execution (actions could be performed without procedure guidance) | N/P |

| | | |
|---|---|---|
| **Training & Experience** | Analyst chose nominal diagnosis HEP because of operator statements indicating that diagnosing a tube rupture based on radiation indications would be possible under these circumstances Note: ASEP assumes that a novice operator is quickly replaced with an experienced operator for the purposes of assigning the HEP. | N/P |
| **Procedural Guidance** | Assuming procedures are being followed, but ASEP assumes a 5 minutes delay due to procedure reading (this affects diagnosis HEP). | 0 |
| **Human- Machine Interface** | Covered by "adequacy of time" and "indications of conditions"- specific rules are assigned depending on control panel layout. For example, in Table 8-1, step 5b states "Assess 1 minute as the required travel and manipulation time combined for each control room (CR) control action taken on the primary operating panels which are normally in visual access of the CR operator." In this way, HMI considerations in ASEP are deferred to adequacy of time calculations. | N/A |
| **Work Processes** | ASEP does not explicitly consider work processes—in this table, its influences are considered under team dynamics. | N/A |
| **Communication** | ASEP does not explicitly consider communication—in this table, its influences are considered under team dynamics. | N/A |
| **Team Dynamics** | | N/A |
| **Other** | Consideration of multiple events; subsequent events given higher HEP. In other words, dependency and recovery taken into account. In other words, the default for ASEP is that the HEP covers the entire crew (see step 9g), but there is some provision for breaking out individuals if justified. The one exception can be found in Table 8-5, where recovery credit can be assigned for a "second person who checks the performance of the original performed"—the analyst gives credit for this recovery in this analysis. | N/A |

*G2.5.4 Basic comparison of qualitative predictions—operational stories*

The analyst accurately identified that it would be significantly easier for the operators to determine that there was a steam generator tube rupture than it would be in HFE1C. The analyst predicted that it would be possible to determine that there was a tube rupture based on radiation detection in the main steam lines or SG blow down lines. All the crews, in fact, quickly detected the radiation alarms and determined that there was a leak in the steam generator C. The analyst selected the nominal diagnosis HEP, rather than the upper bound, based on operator statements indicating that they would be able to detect such an SGTR under the given circumstances. All crews tripped the reactor between 1 and 4 minutes after the radiation alarms, and isolated the SGs no longer than 21 minutes after the initiating event.

*G2.5.5 Basic comparison of qualitative predictions—drivers*

Neither the ASEP analysis nor the raw data indicated any negative drivers. They both identified the following drivers to be nominal or positive: indications of conditions (based on the radiation alarms) and training and experience (as the crews frequently train on this scenario in the simulator and there were no added complications). While the data indicates that the procedural guidance (PG) was nominal/positive based on the fact that procedures exists to address this scenario, the assessment seemed to classify PG as a non-driver, which could potentially be considered nominal for this case. The data also identified the HMI to be a nominal positive factor, but ASEP does not explicitly address HMI and was thus not applicable for the assessment.

*G2.5.6 Quantitative Comparison*

The post-diagnosis actions drove the HEP of 4E-3. This lower HEP seemed consistent with the fact that no crews failed this scenario. In addition, the analyst's HEP for this scenario ranked the lowest (least difficult rating) of all of the scenarios—as did the empirical ranking.

## G2.6 Comparison Summary

*G2.6.1 Predictive Power*

In general, ASEP seemed to be an easy to use, straightforward method with moderate quantitative predictive power yet more limited qualitative predictive power. HEPs were in large part influenced by the operators' adequacy of time—both for diagnosis and post-diagnosis actions. (For instance, adequacy of time influences whether post-diagnosis actions are deemed to cause "extremely high stress" if actions need to be performed very quickly.) Additionally, ASEP addresses, both directly and indirectly, five main aspects of a scenario: adequacy of time, stress, execution complexity, procedural guidance, training and experience. For post-diagnosis activities, based on such characteristics as the operators' training, the procedural guidance, and how long the crew has to perform the actions, the analyst selects one of two options for execution complexity (step-by-step or dynamic) and one of two options for stress level (moderately high stress or extremely high stress). While the decision-making process is relatively simple, it limits the predictive power of the analyst. For instance, in many of the cases the analyst designated the procedural guidance in a scenario to be nominal or positive, when the empirical data indicated that the procedures were, in fact, a negative driver. This appears to be due in large part to the simplistic way ASEP provides guidance for the evaluation of procedures; it only accounts for whether or not procedures exist and whether they are expected to be followed. ASEP does not provide much guidance for analyzing whether or not the procedural guidance adequately directs the operators.

ASEP is a very structured, straightforward method to use, but it has two main weaknesses: it lacks depth of analysis (limited guidance for detailed analysis) and it has limited flexibility to account for nuances of a situation or to incorporate expert judgment. While the method mentions many drivers, it does not cover them in great detail. In addition, there are cases in which the analyst might anticipate a potential complication that operators will face, but the method does not provide guidance for how to incorporate such additional information. In this respect, it is extremely important for the analyst to perform sensitivity studies when completing an HRA and understand that the results should not be considered as very precise.

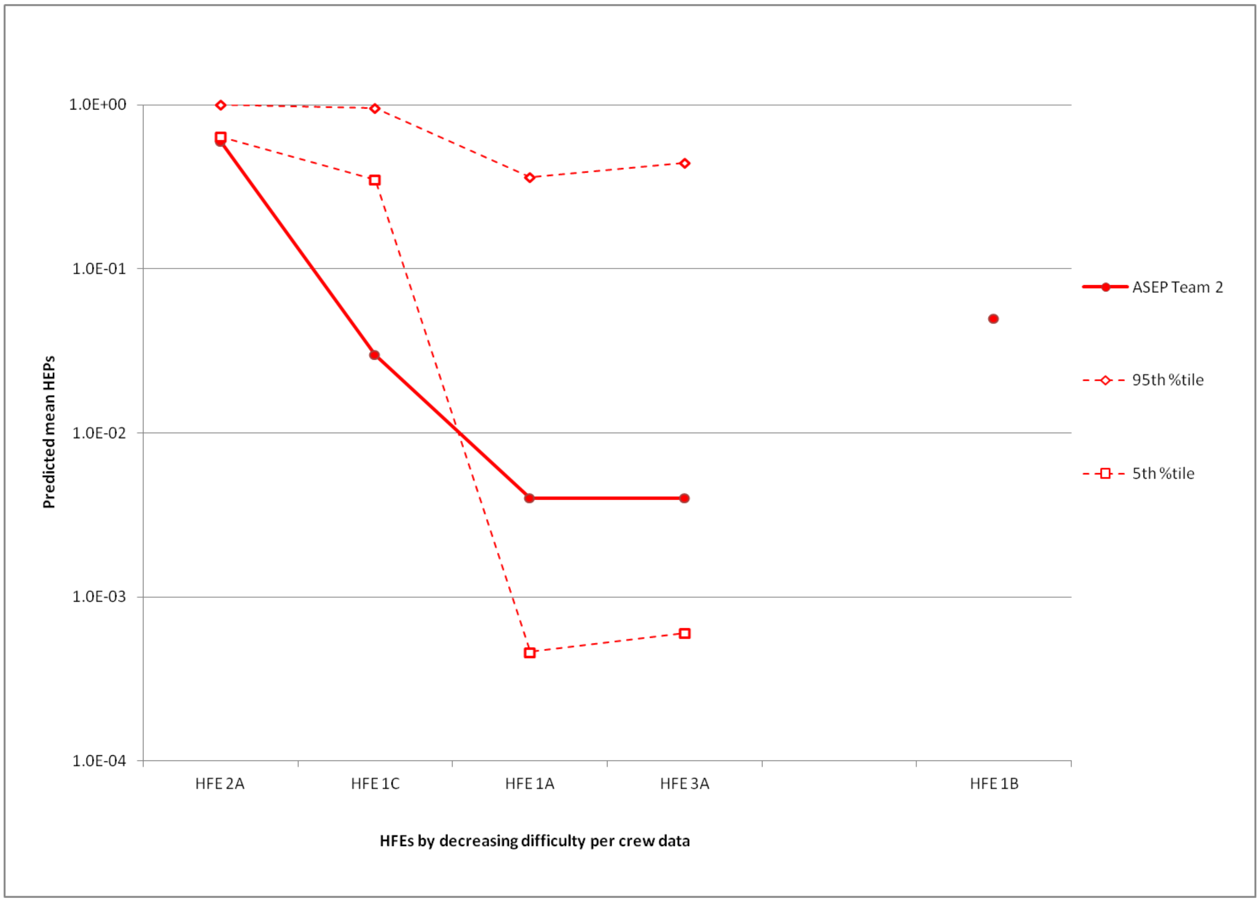



Figure 0.2 Predicted Mean HEPs with Uncertainty Bounds by Team 2 with ASEP

*G2.6.2 Traceability*

One of the major strengths of ASEP is that, due to its structured steps, it is very traceable. It is possible for one to reproduce another's quantitative ASEP results given they make the same assumptions. Some questions arise concerning the qualitative assumptions an analysts makes that drive the quantitative results. These questions require analysts to make judgments in selecting step-by-step versus dynamic or moderately high stress versus extremely high stress and therefore not all analysts will necessarily make the same judgments. Another potential source of uncertainty or variability in results comes from what indications of conditions an analyst credits and what cues he or she uses for timing calculations. Nonetheless, once these points are clarified, it is relatively straightforward to calculate the same HEP.

*G2.6.3 Adequacy of the guidance*

The procedures for developing nominal HRA of post-accident tasks, the ASEP guidance in Table 8-1, is a relatively straightforward 12 step process. Some of the steps get into greater depth than others and require more investigation on the part of the analyst, but in general, the process follows a fairly direct path: estimate the travel and diagnosis time, adjust diagnosis HEP, select post-diagnosis HEPs from table, adjust post-diagnosis HEP, calculate total failure probability. On the surface, it is a simple method to learn, and a novice analyst can produce results based on the procedural guidance. It does not, however, provide guidance for much additional interpretation or manipulation of the HEP based on information not directed to by the method (e.g., procedural problems given a unique scenario). This seems to make it more challenging for an experienced analyst to produce the in depth results of which he or she is capable.

In addition, the HRA is highly dependent on a graphic—Figure 8-1 for the Nominal Diagnosis Model or Figure 7-1 for the screening model. This graph can be difficult to read, and thus the results cannot be interpreted very precisely. In this respect, it is very important for the analyst to conduct sensitivity studies when calculating the HEP.

*G2.6.4 Usefulness of the HRA results for human error reduction*

ASEP, at its core, has a very quantitative focus; it is driven by adequacy of time and is highly dependent on a graph (time reliability correlation) that relates time and diagnosis probability. This in itself does not provide much guidance for error reduction. Additionally, for usability's sake, factors are lumped together and result in only two main decisions by the analyst. Training, procedures, and team dynamics are thus treated very simplistically. An assessment of training/experience and an assessment of procedures feed into one decision by the analyst: choosing step-by-step or dynamic execution complexity. (For instance, step-by-step execution complexity implies "a routine, procedurally guided set of steps performed one step at a time without a requirement to divide ones attention between the task in question and other tasks. With high levels of skill and practice a step-by-step task may be performed reliability without recourse to written procedures." An analyst may assume that the operators are well-trained in such routine tasks.) Again, while this binary option makes the job of the analyst less complicated, it limits flexibility and doesn't support focus on specific, potentially very significant, drivers. As such, qualitative findings can get lost in the calculation as they are manifested through other choices and thus have no quantitative effect on the HEP and may not be clear to support error reduction. An example of this is how training/experience was designated as a nominal/positive driver in HFE2A but had no effect on the HEP. While the HEP proved to be consistent with the empirical data, according to this analysis, one would not have recognized training as needing improvement, and the data suggested it was a main negative driver.

In summary, the following ratings seem appropriate to characterize ASEP's predictive power, traceability, and procedures (1-5 scale):

- Predictive Power
  - Qualitative predictive power: 2, moderately poor
  - Quantitative predictive power: 3, fair
- Traceability: 5, good
- Adequacy of guidance: 4, moderately good

## G.3 <u>CBDT/ASEP (NRI)</u>

Each HFE's performance is affected by three contributors: information and cognitive errors, delays in starting manipulations, and execution errors. Using the NRI approach, the total HEP for an HFE is the sum of the contribution of these contributors (information errors, diagnosis/delays, and execution errors).

Information processing failure probability is related to detection, diagnosis, and decision as to a plan of action; in the quantification, this is referred to as the *identification* part of the task. For *Identification*, the NRI team used the CBDT to get the HEP. The total *Identification* HEP is the sum of the HEPs from the 8 decision trees. This phase considers a large number of PSFs. *Diagnosis/delay* refers to the delay in starting manipulations based on previously performed information processing; this is referred to as the *diagnosis* part of the task. For *Diagnosis,* the NRI team used the time reliability curves from THERP (ASEP table 8-2). This table is based entirely on Available Time.

Execution error probability is related to executing a planned action; this is referred to as the execution phase of the task. For *Execution,* the NRI team used an adapted version of ASEP. They use two PSFs for this: Stress and Execution Complexity.

### G3.1 HFE1A

**Type of task:** Type C – Post-initiating event accidents.
**Task description:** It is supposed that the crew tripped reactor manually within approximately 30 - 45 seconds of the loss of feed water, therefore they have approximately 45 minutes for F&B establishing before CD.

According to FR-H1FR-H1, F&B shall be initiated when WR level on any two SGs are less than 50%. This criterion should be reached approximately 1 to 2 minutes after the LOFW.

**Procedures:** E-0: "Reactor Trip Or Safety Injection", in step 16 (Initiate Monitoring of Critical Safety Functions) transfer to F-03 "Heat Sink Safety Function Status Tree" - RED condition, transfer to FR-H1FR-H1 "Response to Loss of Secondary Heat Sink" (main steps 10-13).

**Goal of human action**: To establish F&B within 45 minutes of the reactor trip.

#### *G3.1.1 Summary of Qualitative Findings*

Difficulties identified by the NRI team are:

- Making the correct diagnosis of the situation
  - The stress level is increased due to loss of FW
- Timely actuation of F&B
  - Crews may spend too much time trying to restart the flow to the SGs.
  - Due to increased workload caused by MFW and AFW trip, crews may skip step 16 in E-0.

According to the NRI qualitative statements, the most important factors that negatively impact crew performance will be scenario complexity (middle-high workload), stress, and execution complexity (both the dynamics of the execution task and the internal negative dependency between required steps for execution). Slightly negative factors include time adequacy, scenario complexity (monitoring), and procedural guidance (logic, graphical distinction between steps).

Positive factors are indications of conditions (accuracy and availability of necessary information), training (yearly scenario-specific training, training on specific procedure steps), HMI (good location and ease of locating indicators), work processes (use of placekeeping aids), and procedural guidance (obvious instructions).

*G3.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

The total HEP for HFE1A is **1.20E-02** (Error Factor = 5).

Breakdown:

- Identification: 3.75E-3 (from CBDT decision trees)
  - Insufficient attention of operator (data not attended to): 4.5E-4
  - Procedural Step Skipping: 3.00E-3
  - Procedure Logic: 3.00E-4
- Diagnosis: 5.00E-4 (from ASEP)
- Execution: 7.73E-3 (from ASEP)

The main contributors to HEP in their quantitative analysis are *procedural step skipping (identification error)* and *execution errors*. The negative PSF that drives procedural step skipping is lack of graphical distinction between procedure steps (procedural guidance). The PSFs that drive execution errors are stress and execution complexity.

*Diagnosis errors (delays)*, *insufficient operator attention*, and *procedure logic* each make a smaller contribution to the HEP. The negative PSF that drives diagnosis errors is time. The negative PSF that drives insufficient operator attention is workload rated middle-high (scenario complexity). Procedural logic is driven by the use of “AND” and “OR” statements in the procedures (procedural guidance).

*G3.1.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Available time. The available time for this action is supposed to be 43 to 44 minutes (= 45 minutes to core damage minus 1 to 2 minutes, which is the time to reach 50% of SG WR level). The available time is a negative driver for diagnosis/delays. | ND |
| **Stress** | Stress was identified as a negative driver in the NRI qualitative analysis. It is also a dominant contributor to the execution phase HEP. | ND |
| **Scenario Complexity** | • Workload is a characteristic of complex scenarios. Workload is supposed to be middle-high (average between low and high), since the crews solve the trip of feedwater pumps (MFW and AFW) and may spend some time trying to restart the flow to the SGs. CBDT states that workload can result in reduced operator attention, which increases HEP.<br>• Monitoring[4]. Operators should monitor Critical Safety Functions and SG level during the scenario. Monitoring tasks are associated with a higher HEP since operators must perform monitoring actions frequently enough to catch the required parameter state. | ND |

[4] The NRI team determined that Monitoring was relevant to performance. The reviewer determined Monitoring to be an aspect of Scenario Complexity based on review of the CBDT documentation.

| | | |
|---|---|---|
| **Indications of Conditions** | Indication availability and accuracy. The important information concerning SG level is available in the CR and it is accurate. (It is assumed that the missing indication of recirculation valve's position (which is open by mistake) will not affect considerably crew's action.) Good indications of conditions reduce the HEP. | N/P |
| **Execution Complexity** | • Execution complexity was identified as a negative driver in the NRI qualitative analysis. It is also a dominant contributor to the execution phase HEP. It was determined that the response procedure execution steps were partially dynamic, which increases HEP..<br>• Dependency between the response procedure execution steps also increases the HEP. | ND |
| **Training and Experience** | • Training in obtaining and interpreting the required indication. The crews are quite familiar with this type of scenario - it is supposed to be trained for about once a year. This reduces the HEP.<br>• Training on steps. The crews have received training on the correct interpretation of the relevant steps, which reduces the likelihood of skipping procedure steps. This reduces the HEP.<br>• ASEP method assumes that novice operators (< 6 months experienced) are replaced by experienced operators, so experience is not a driver in the method. It is also not a factor in CBDT.<br>• The NRI team stated that operators have extended experience with CR layout and location of the relevant indicators. | N/P |
| **Procedural Guidance** | This is the main negative driver for identification errors based on the quantitative results provided by NRI.<br>Negative aspects:<br>• Graphical distinction. The relevant steps (16 in E-0 and 2, 9-13 in FR-H1FR-H1) are not distinct from other steps, which raises the probability of the step skipping.<br>• Procedure logic. AND or OR logic occurs in some parts of the relevant procedures, which raises the error probability of procedure interpretation.<br>Positive aspects:<br>• Obvious instructions. The relevant instructions are considered as "Obvious" since E-0, F-03 and FR-H1FR-H1 are procedures with separate, stand-alone numbered steps. The steps contain standard, unambiguous wording and presents all information required to identify actions directed and objects. According to CBDT, this reduces the likelihood of error.<br>• Single procedure. Operators are using one procedure at a time. This reduces the HEP. | ND |

| | | |
|---|---|---|
| **Human- Machine Interface** | • Good indicators are easy to locate. There were not observed any human factor deficiencies in the layout, demarcation or labeling of the control boards during simulator excursion. Similarly, no deficiencies connected to indicators were observed. This reduces the HEP.<br>• All the essential indications are supposed to be placed on the front panels and alarmed (the main indicators considered in this scenario are: 1. narrow range SG level, 2. wide range SG level, and 3. feedwater flow), which reduce possibility of operators' error due to insufficient attention and thus reduces HEP. | N/P |
| **Work Processes** | Use of placekeeping aids. The operators were observed to be using placekeeping aids at the simulator. This reduces the probability of skipping procedure steps and thus reduces the HEP. | N/P |
| **Communication** | All CR personnel adhered to a formal communication protocol during the simulator scenario observed by NRI. The NRI team inferred that crews would use formal communication during the empirical scenarios. Use of formal communication reduces the probability of miscommunicating data and thus reduces HEP. | N/P |
| **Team Dynamics** | | 0 |
| **Other** | In all scenarios, operators are assumed to believe that the instructions presented are appropriate to the situations. This reduces the HEP (via negligible contribution of deliberate violation DT). | N/P |

*G3.1.4 Comparison of Drivers to Empirical Data*

The NRI team correctly identified Scenario Complexity and Procedural Guidance as negative drivers. These were both identified as negative drivers in the empirical data, albeit for slightly different reasons. Scenario complexity was a negative driver in the HRA because of moderate workload and the need to monitor CSFs. The NRI team determined that the workload would be moderate (averaged between low and high workload) because the crew needed to address the feedwater pumps (MFW and AFW) and the lack of flow to the SGs. In the empirical data, scenario complexity was a negative driver because the inadequate indications of conditions masked the plant conditions, which caused the crew to delay starting F&B. There was evidence that increased workload did cause some teams to delay initiation of F&B due to task prioritization (crew members trying to start the AFW pumps to avoid need for F&B). However, there is no evidence that this increased workload adversely impacted their success in the identification/diagnosis of the condition.

Procedural guidance was a negative driver in the NRI analysis because of procedure readability issues (lack of graphical distinction among procedure steps and difficult logic in procedures E-0 and FR-H1FR-H1). The empirical data does not mention these aspects; rather the poor procedural guidance that affected performance was related to poor guidance (missing information) in the Critical Safety Function status trees.

The NRI team identified Indications of Conditions as a positive influence. This is inconsistent with the empirical data, which shows that poor indications of conditions were a negative driver. The NRI team assumed that the missing indicator for recirculation valve position would not have a significant effect on the crew's actions.

The NRI team identified three negative drivers that were not considered negative drivers in the empirical data: Adequacy of Time, Stress, and Execution Complexity. The NRI team felt that the 45 minutes time window was a slightly negative performance driver (in the quantitative analysis, this affects delay probability). This was not reflected in the empirical data:  all of the crews were able to diagnose the LOFW and start F&B in less than 20 minutes. They also thought that stress level would be increased, and this increased stress would have a negative effect on performance. This was not reflected in the empirical data, although crew S did have a self-imposed sense of urgency.

According to the NRI team, Execution Complexity was a negative driver because of the dynamic nature and internal dependency of the response procedure execution (steps 10-13 in FR-H1FR-H1); in the empirical data this was not a main negative driver. However, there is evidence that one crew (crew S) made an error in following the procedure (FR-H1FR-H1, step 2), but this was quickly recovered (due to good teamwork and work processes) and did not have a significant impact on the scenario.

The NRI team correctly identified Training and Experience and HMI as positive drivers. They indicated that the scenario would be trained on approximately once a year, which helps crews correctly obtain and interpret required indication. They also noted that the crew would be trained on use of the procedures. The NRI team also identified work processes, communication, and team dynamics as either positive influences or not a driver, which is consistent with the net effect of these PSFs, which were not drivers of performance in the empirical data.

*G3.1.5  Comparison of Qualitative Analysis to Empirical Data*

The NRI team thought that operators would have difficulty making the correct diagnosis of the situation, especially faced with increased stress level due to the loss of FW. The crews did have difficulty making the correct diagnosis (due to masked indications of conditions, not due to elevated stress), but this was not sufficient to fail the HFE.

The NRI team also thought the crews would have difficulty with timely actuation of the F&B procedure. They identified two reasons the crews would struggle with this: crews could spend too much time trying to restart the flow to the SGs, or crews could skip step 16 in E-0 (Monitor Critical Safety Functions) due to increased workload (caused by the MFW failure and AFW trip). Increased workload did cause some of the crews to slightly delay initiation of F&B by not responding the first time the lack of increasing level in SG B was announced. (Crew R: two crew members were busy trying to restart the AFW pumps. Crew T was not monitoring the Critical Safety Functions early in the scenario when they noticed a red path on heatsink [no reason provided in the documentation]. Crew T also delayed cross-connecting the running AFW pump to a different SG because they only had two reactor operators available.). Neither of these delays was significant enough to fail the HFE.

The NRI team expected that the short time window (which is correlated with increased stress) could result in errors of commission or omission during manipulations. While Crew S did make an error of omission in procedure FR-H1FR-H1, step 2 (due to self-imposed sense of urgency), this error was quickly recovered and was not significant enough to fail the HFE.

*G3.1.6  Impact on HEP*

All of the PSFs identified by the NRI team are elements of the CBDT & ASEP approach. All of the identified negative drivers increased the HEP in CBDT & ASEP quantification. None of the factors identified as positive drivers increased the HEP.

There was slight disagreement between NRI qualitative assessment of drivers (ranked "negative" versus "slightly negative) and quantitative impact of the drivers (proportional impact on HEP). For example, procedural guidance was listed as a slightly negative driver, but it was a dominant contributor to the HEP. Likewise, workload was listed as a negative factor, but made a relatively small contribution to the HEP. However, this disagreement only affected the magnitude of impact on HEP, not the direction of impact (positive vs. negative) of the driver.

Overall, the HEP for this event was quite low. The HEP for this HFE (1.2E-02) was approximately 6.2 times greater than the lowest possible HEP (1.94E-03).[5]

### G3.2 HFE1B

**Type of task:** Type C - Post-accidents
**Task description:** Failure to establish F&B within 13 minutes of the reactor trip, given that the crews do not manually trip the reactor before an automatic reactor trip occurs.
We suppose that the crew did not trip reactor manually before automatic reactor trip occurred (50-60 seconds after loss of FW), therefore they have approximately 13 minutes for F&B establishing before CD. According to FR-H1FR-H1, F&B shall be initiated when WR level on any two SGs are less than 50%. This criterion should be reached approximately 1 to 2 minutes after the LOFW.

**Procedures**: E-0: "Reactor Trip Or Safety Injection", in step 16 (Initiate Monitoring of Critical Safety Functions) transfer to F-03 "Heat Sink Safety Function Status Tree" - RED condition, transfer to FR-H1FR-H1 "Response to Loss of Secondary Heat Sink" (main steps 10-13).

**Goals of human action**: To establish F&B within 13 minutes of the reactor trip.

#### *G3.2.1 Summary of Qualitative Findings*
Difficulties identified by the NRI team are:

- Making the correct diagnosis
  - Time is very short
- Timely execution of F&B procedure.
  - Crews may spend some time trying to restart the flow to the SGs, which may be critical, especially in this case, when assumed available time for the whole action is 11-12 minutes.
  - Due to increased workload caused by MFW and AFW trip, crews could skip step 16 in E-0 (Monitor Critical Safety Functions), although this is unlikely.

According to the NRI qualitative statements, available time is the dominant negative driver for this event. Scenario complexity (high workload), increased stress, and execution complexity (both the dynamics of the execution task and the internal negative dependency between required steps for execution) are also negative drivers. Slightly negative drivers are scenario complexity (monitoring) and procedural guidance (logic, graphical distinction between steps).

---

[5] The lowest possible HEP that can be obtained using the method (1.94E-03) was calculated by the reviewer to help compare the HFEs. The HEP 1.94E-3 is obtained by summing the lowest possible HEP from the Identification, Diagnosis/Delay, and Execution phases. The lowest HEP for identification is 1.00E-03 (In CBDT, 1.00E-03 is the lowest HEP in tree E (Skip Procedure Step); all other trees are set to a path with negligible contribution to HEP). The lowest HEP for Diagnosis/Delay is 0 (negligible). The lowest HEP for Execution is 9.4E-04 (In ASEP, 9.4E-04 is the HEP for a standard task with normal stress. The reviewer ignored the possibility for reducing this value due to recovery.)

Positive factors are indications of conditions (accuracy and availability of necessary information), training (yearly scenario-specific training, training on specific procedure steps), HMI (good location and ease of locating indicators), work processes (use of placekeeping aids), and procedural guidance (obvious instructions).

*G3.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

The total HEP for HFE1b is **1.32E-01** (Error Factor = 3)

Breakdown:

- Identification: 4.05E-3
  - Insufficient attention of operator (data not attended to): 7.5E-4
  - Procedural Step Skipping: 3.0E-3
  - Procedure Logic: 3.00E-4
- Diagnosis: 1.00E-1
- Execution: 2.82E-2

Based on the quantitative results, the main driver for failure is *diagnosis/delay*. The negative PSF that drives these errors is time.

Execution errors make a smaller contribution to the HEP. The negative PSFs that drive execution errors are stress and execution complexity. *Procedural step skipping errors* make a smaller contribution to HEP. The negative PSF that drives procedural step skipping is lack of graphical distinction between procedure steps (procedural guidance). *Insufficient operator attention*, and *procedure logic* each also made a very small contribution to the HEP. The negative PSF that drives insufficient operator attention is workload rated middle-high (scenario complexity). Procedural logic is driven by the use of “AND” and “OR” statements in the procedures (procedural guidance).

*G3.2.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Available time was identified as the main negative driver in the NRI qualitative analysis. It is also a dominant contributor to the HEP (via delays).<br>For HFE1B, the available time is 11 to 12 minutes (= 13 minutes to core damage minus 1 to 2 minutes, which is the time to reach 50%SG WR level). | MND |
| **Stress** | Stress. Stress level was evaluated as increased. This increases the HEP in execution tasks. | ND |
| **Scenario Complexity** | • Monitoring. Operators should monitor Critical Safety Functions and SG level during the scenario. Monitoring tasks are associated with a higher HEP since operators must perform monitoring actions frequently enough to catch the required parameter state.<br>• Workload is a characteristic of complex scenarios. Workload is supposed to be high, since the crews, solving the trip of feedwater pumps (MFW and AFW), may spend some time trying to restart the flow to the SGs, which may be critical, because the available time window is very short in this scenario. CBDT states that workload can result in reduced operator attention, which increases HEP. | ND |

| | | |
|---|---|---|
| **Indications of Conditions** | Indication availability and accuracy. The important information concerning SG level is available in the CR and it is accurate. (It is assumed that the missing indication of recirculation valve's position (which is open by mistake) will not affect considerably crew's action). Good indications of conditions reduce the HEP. | N/P |
| **Execution Complexity** | • Execution complexity was identified as a negative driver in the NRI qualitative analysis. It is a contributor to the execution phase HEP. It was determined that the response procedure execution steps were dynamic and partially dynamic, which increases HEP.<br>• Dependency between the response procedure execution steps also increases the HEP. | ND |
| **Training and Experience** | • Training in obtaining and interpreting the required indication. The crews are quite familiar with this type of scenario - it is supposed to be trained for about once a year. This reduces the HEP.<br>• Training on steps. The crews have received training on the correct interpretation of the relevant steps, which reduces the likelihood of skipping procedure steps. This reduces the HEP.<br>• ASEP method assumes that novice operators (< 6 months experienced) are replaced by experienced operators, so experience is not a factor in the method. It is also not a factor in CBDT.<br>• The NRI team stated that operators have extended experience with CR layout and location of the relevant indicators | N/P |
| **Procedural Guidance** | This is the main negative driver for identification errors based on the quantitative results provided by NRI.<br>Negative aspects:<br>• Graphical distinction. The relevant steps (16 in E-0 and 2, 9-13 in FR-H1) are not distinct from other steps, which raises the probability of the step skipping.<br>• Procedure logic. AND or OR logic occurs in some parts of the relevant procedures, which slightly raises the error probability of procedure interpretation.<br>Positive aspects:<br>• Obvious instructions. The relevant instructions are considered as "Obvious" since E-0, F-03 and FR-H1 are procedures with separate, stand-alone numbered steps. The steps contain standard, unambiguous wording and presents all information required to identify actions directed and objects. According to CBDT, this reduces the likelihood of error.<br>• Single procedure. Operators are using one procedure at a time. This reduces the HEP. | ND |

| | | |
|---|---|---|
| **Human- Machine Interface** | • Good indicators are easy to locate. There were not observed any human factor deficiencies in the layout, demarcation or labelling of the control boards during simulator excursion. Similarly, no deficiencies connected to indicators were observed. This reduces the HEP.<br>• All the essential indications are supposed to be placed on the front panels and alarmed (the main indicators considered in this scenario are: 1. narrow range SG level, 2. wide range SG level, and 3. feedwater flow), which reduce possibility of operators' error due to insufficient attention and thus reduces HEP. | N/P |
| **Work Processes** | Use of placekeeping aids. The operators were observed to be using placekeeping aids at the simulator. This reduces the probability of skipping procedure steps and thus reduces the HEP. | N/P |
| **Communication** | All CR personnel adhered to a formal communication protocol during the simulator scenario observed by NRI. The NRI team inferred that crews would use formal communication during the empirical scenarios. Use of formal communication reduces the probability of miscommunicating data and thus reduces HEP. | N/P |
| **Team Dynamics** | | 0 |
| **Other** | In all scenarios, operators are assumed to believe that the instructions presented are appropriate to the situations. This reduces the HEP (via negligible contribution of deliberate violation DT). | N/P |

*G3.2.4 Comparison of Drivers to Empirical Data*

All crews manually tripped the reactor in response to LOFW scenario, so there is no empirical data on crew performance for scenario HFE1b.

*G3.2.5 Comparison of Qualitative Analysis to Empirical Data*

All crews manually tripped the reactor in response to LOFW scenario, so there is no empirical data on crew performance for scenario HFE1b.

*G3.2.6 Impact on HEP*

All of the PSFs identified by the NRI team are elements of the CBDT & ASEP approach. All of the identified negative drivers increased the HEP in CBDT & ASEP quantification. None of the factors identified as positive drivers increased the HEP.

There was slight disagreement between NRI qualitative assessment of drivers (ranked “negative” versus “slightly negative) and quantitative impact of the drivers (proportional impact on HEP). For example, workload was a negative driver, but had a very small contribution to the HEP. However, this disagreement only affected the magnitude of impact on HEP, not the direction of impact (positive vs. negative) of the driver.

Overall, the HEP for this HFE (1.32E-01) was approximately 68 times greater than the lowest possible HEP (1.94E-03).

### G3.3 HFE1C

**Type of task:** Type C – Post-accidents
**Task description:** Failure of crew to isolate the ruptured steam generator and control pressure

below the SG PORV setpoint to avoid SG PORV opening.
In this scenario, we assume that the crew successfully initiated F&B and they were able to establish AFW to one or several SGs. However, when they did so, an SGTR occurred in the first SG that was fed.

The crew will want to fill a SG to be able to exit FR-H1, and the tube rupture may be masked by AFW flow to the SG, as long as it is being fed. The leak size of the ruptured tube is about 500 GPM at 100% power, but the flow will depend on the differential pressure between the RCS and the ruptured SG.

There is initially no secondary radiation because there is a minimum steam flow. The blow down (BD) and sampling is secured because of the SI. The crew is working in FR-H1, and may have criteria for FR-P.1 as a consequence of the F&B. The time window to perform the required actions is estimated to be approximately 40 minutes**.**
**Procedures:** FR-H1 "Response to Loss of Secondary Heat Sink"**,** E-30 "Steam Generator Tube Rupture" (steps 2-4).
**Goals of human action**: To isolate the ruptured steam generator and control pressure below the SG PORV setpoint.

*G3.3.1 Summary of Qualitative Findings*
Difficulties identified by the NRI team:

- Making the correct diagnosis of the situation
  - Due to masking effect, when operators initially do not have complete information about SGTR
- Performing the correct actions on time
  - Due to short time window.

According to the NRI qualitative analysis, indications of conditions (in-accurate indications and the lack of alarms), scenario complexity (workload), stress, and execution complexity (dynamic steps in response procedure) are negative driver. Slightly negative drivers are adequacy of time, scenario complexity (monitoring), and procedural guidance (logic, graphical distinction between steps, mismatch between procedures and conditions, and lack of warnings regarding this mismatch).

Positive factors are HMI (good location and ease of locating indicators), procedural guidance (obvious instructions), indications of conditions (availability of indications), training (on specific procedure steps), and work processes (use of placekeeping aids).

*G3.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
The total HEP for HFE1C is **1.47E-01** (Error Factor = 3)

Breakdown:
- Identification: 8.13E-2
  - Availability of information (data not available): 5.0E-2
  - Insufficient attention of operator (data not attended to): 1.5E-2
  - Information misleading: 1.0E-2
  - Procedural Step Skipping: 6.0E-3
  - Procedure Logic: 3.00E-4
- Diagnosis: 1.00E-2
- Execution: 5.60E-2

Based on the quantitative results, the main contributors to HEP are *availability of information*, *insufficient operator attention, misleading information*, *diagnosis/delays*, and *execution errors*. The negative PSF that drives diagnosis/delay errors is time. The negative PSFs that drive execution errors are stress and execution complexity. *Availability of information, insufficient operator attention, and information misleading* are failure mechanisms that contribute to identification errors. The negative PSFs that drive *availability of information* are inaccurate indications for the secondary radiation (indications of conditions) and lack of procedural guidance about alternate information sources (procedural guidance). The negative PSFs that drive *insufficient operator attention* are workload rated high (scenario complexity), monitoring tasks (scenario complexity), and lack of alarms for secondary circuit radiation and SG level (indications of conditions). The negative PSFs that drive *misleading information* are the misleading cues about radiation in steamline (indications of conditions), lack of guidance in the procedure regarding the degraded cue (procedural guidance), and the mismatch between the cues and the procedures (scenario complexity).

*Procedural step skipping* and *procedure logic* make smaller contributions to the HEP. The negative PSF that drives procedural step skipping is lack of graphical distinction between procedure steps (procedural guidance) and use of multiple procedures (scenario complexity). Procedural logic is driven by the use of “AND” and “OR” statements in the procedures (procedural guidance).

*G3.3.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Available time. The available time for this action is supposed to be 40 minutes. Available time is a dominant contributor to the delay portion of the HEP. | ND |
| **Stress** | Stress. Stress level was evaluated as increased. This increases the HEP in execution tasks. | ND |
| **Scenario Complexity** | - Workload is a characteristic of complex scenarios. Workload is supposed to be high during this complicated version of SGTR. CBDT states that high workload can result in reduced operator attention, which increases the HEP.<br>- Multiple procedures. According to CBDT, use of multiple procedures is an indication that the workload is high, which increases the probability of slip-type errors.<br>- Monitoring - operators have to monitor RCS temperature and SG pressure during maintaining SG pressure below the setpoint. Monitoring tasks are associated with a higher HEP since operators must perform monitoring actions frequently enough to catch specific parameter values.<br>- Mismatch between cues and procedures. Some parameter values (radiation on the steamline) are not as stated in procedure (high steamline radiation is a criteria for SGTR). The mismatch between cues and procedures increases the complexity of the scenario. Mismatch between cues and procedures increases the HEP (information misleading). | ND |

| | | |
|---|---|---|
| **Indications of Conditions** | - Indications are initially misleading (not accurate) and there is no alternative source of information provided in the procedures. There is initially no secondary radiation because of a minimum steam flow. From that reason was selected that the indication is not accurate and no alternate information sources provided in the procedures. Poor indications are a main contributor to data not available and information misleading failure mechanisms.<br>- Alarm. The relevant parameters (secondary circuit radiation and SG level) are initially not alarmed. This increases the HEP through the failure mechanism insufficient operator attention (data not attended to). | ND |
| **Execution Complexity** | Execution complexity was identified as a negative driver in the NRI quantitative analysis. It is a main contributor to the execution phase HEP. It was determined that the response procedure execution steps were dynamic, which increases HEP. | ND |
| **Training and Experience** | - Training on steps. The crews have received training on the correct interpretation of the relevant steps. This reduces HEP.<br>- ASEP method assumes that novice operators (<6months experienced) are replaced by experienced operators, so experience is not a factor in the method. It is also not a factor in CBDT.<br>- The NRI team stated that operators have extended experience with CR layout and location of the relevant indicators. | N/P |

| | | |
|---|---|---|
| **Procedural Guidance** | This is a negative driver for identification errors based on the quantitative results provided by NRI.<br>Negative aspects:<br>- Graphical distinction. The relevant steps (13 in E-0 and 2-4 in E-30) are not distinct from other steps. This increases HEP.<br>- Some cue states and parameter values are not as stated in procedures (no radiation indicated in steamline) and procedures do not provide clear warnings or alternatives. This raises the HEP.<br>- Procedure logic. In some parts of the activity described with HEP quantified, several actions may be required within one step (AND logic expected), also OR logic can be found in some steps of E-30 procedures (e.g. step 2), but no redundant alternatives are expected explicitly (combination of OR and AND logic within one step).<br>- Multiple procedures. The conservative assumption was made that operators would be using multiple procedures, which increases HEP.<br>Positive aspects:<br>Obvious instructions. The relevant instructions are considered as "Obvious" since the relevant procedures include separate, stand-alone numbered steps. The steps contain standard, unambiguous wording and presents all information required to identify actions directed and objects. According to CBDT, this reduces the likelihood of error. | ND |
| **Human- Machine Interface** | Good indicators are easy to locate. There were not observed any human factor deficiencies in the layout, demarcation or labeling of the control boards during simulator excursion. Similarly, no deficiencies connected to indicators were observed. This reduces the HEP. | N/P |
| **Work Processes** | Using of placekeeping aids. The operators were observed to be using placekeeping aids at the simulator. This reduces the probability of skipping procedure steps and thus reduces the HEP. | N/P |
| **Communication** | All CR personnel adhered to a formal communication protocol during the simulator scenario observed by NRI. The NRI team inferred that crews would use formal communication during the empirical scenarios. Use of formal communication reduces the probability of miscommunicating data and thus reduces HEP. | N/P |
| **Team Dynamics** | | 0 |
| **Other** | In all scenarios, operators are assumed to believe that the instructions presented are appropriate to the situations. This reduces the HEP (via negligible contribution of deliberate violation DT). | N/P |

*G3.3.4 Comparison of Drivers to Empirical Data*

The NRI team correctly identified Adequacy of Time, Scenario Complexity, Indications of Conditions, and Procedural Guidance as negative drivers. These were all identified as negative drivers in the empirical data, but the reasons varied. NRI identified Adequacy of Time, Scenario Complexity, and Indications of Conditions as drivers for reasons consistent with the reasons provided in the empirical data. However, the reasons for identifying Procedural Guidance as a main driver differed from the empirical data.

The NRI team felt that the 40 minutes time window was a slightly negative driver; limited time increased the HEP via delay errors. They used a conservative estimate for the contribution of time because of the masked radiation indicators. The empirical stated that the 40 minute criterion was barely adequate to diagnose the SGTR, work through the procedures, and execute the actions in a complex scenario. The NRI team correctly identified time as a negative driver in the data, although it appears that they underestimated the magnitude of this effect.

Scenario complexity was identified as negative driver in the NRI analysis. Workload was high and operators were working in multiple procedures. The operators were required to monitor RCS temperature and SG pressure (monitoring tasks must be completed more often than non-monitoring tasks). There was a mismatch between the available cues (low steamline radiation) and the cues in the procedure (high steamline radiation is a criteria for SGTR). The mismatch between cues and procedures is a result of the off-normal conditions (where the steamline is isolated due to previous emergency). In the empirical data, the scenario complexity was a main negative driver because of the workload (responding to previous emergency situation and the current situation) and because there was missing information (lack of radiation alarms).

Poor indications of conditions were a negative driver in the NRI analysis. The information on SG level was initially masked by AFW flow. There were no secondary side radiation alarms received because of minimum steam flow. NRI deemed the indications to be misleading (inaccurate) and there were no alternative sources of this information. They also noted that the secondary circuit radiation and SG level were not alarmed (or not near enough to their setpoints to trigger an alarm). This is consistent with the empirical data, which stated that the missing radiation alarms and masked SG level were negative performance drivers.

Procedural Guidance was identified as a slightly negative driver because of differences between the procedures and the scenario, and procedure readability issues. In the scenario, the steamline radiation sensor did not alarm because of low flow in the steamline (which is a result of isolating the main steamline). The procedures did not provide warnings or alternatives for situations where the plant cues did not match procedural guidance (steamline radiation is one of the symptoms of SGTR). The procedure readability issues include lack of graphical distinction among procedure steps, and difficult logic in procedure E-30. Operators were also working in multiple procedures.

The NRI team identified two negative drivers that were not considered negative drivers in the empirical data: Stress and Execution Complexity. The NRI team rated stress as “increased”, which increases the probability of execution errors. In the empirical data, stress was rated as “not a driver.” However, the crews did express that they were feeling run down as the scenario progressed, because it was a long and complicated scenario.

The NRI team included execution complexity as a negative driver because several of the response procedure steps were dynamic; this increases execution error probability. In the empirical data, this was deemed to be a positive driver.

The NRI team correctly identified Training and Experience and HMI as positive drivers. They indicated that the crew has received training on the correct interpretation of procedure steps. There were no observed human factors deficiencies in the HMI and the operators have extended experience with the layout and location of the relevant indicators.

The NRI team identified team dynamics, work processes, and communication as either positive influences or not a driver, which is consistent with the net effect of these PSFs, which were not drivers of performance in the empirical data.

*G3.3.5 Comparison of Qualitative Analysis to Empirical Data*

The NRI team thought that operators would have difficulty making the correct diagnosis of the situation due to the masking of the SGTR. This is consistent with the empirical data, which demonstrates that the masking of the SGTR condition played a key part in delaying the crews' diagnosis of the SGTR.

The NRI team also thought that the 40 minutes time window was quite short; this short time window could impact the ability of the crew to perform the required actions on time. This is also consistent with the empirical data, because 3 of the 4 crews failed to perform the required actions within 40 minutes. The delay in diagnosing the SGTR (due to masking) further shortened the available time to execute response actions. Only one crew successfully diagnosed the SGTR and completed the required actions within the available time window.

The NRI quantitative analysis identified the most important execution steps as isolating the SG and maintaining SG pressure below the setpoint by cooling down the RCS. They did not consider the actions (procedural steps) needed to stop SI and close the PZR PORVs. In the empirical data, one crew did not stop SI and subsequently failed the scenario by opening the PZR PORV.

*G3.3.6 Impact on HEP*

All of the PSFs identified by the NRI team are elements of the CBDT & ASEP approach. All of the identified negative drivers increased the HEP in CBDT & ASEP quantification. None of the factors identified as positive drivers increased the HEP.

There was slight disagreement between NRI qualitative assessment of drivers (ranked "negative" versus "slightly negative) and quantitative impact of the drivers (proportional impact on HEP). For example, available time was identified as a slightly negative driver, but it made a significant contribution to the HEP. However, this disagreement only affected the magnitude of impact on HEP, not the direction of impact (positive vs. negative) of the driver.

Overall, the HEP for this HFE (1.47E-01) was approximately 76 times greater than the lowest possible HEP (1.94E-03).

### G3.4 HFE2A

**Type of task:** Type C - Post-accidents
**Task description:** Trip the RCPs after the loss of CCW and start the Positive Displacement Pump (PDP) to provide seal injection before sealwater inlet or lower sealwater bearing temperatures are greater than 230 °F (per ES01 Step 6 or 0POP04-RC-0002: Reactor coolant pump off-normal) to avoid potential (not necessarily immediate) RCP seal LOCA. Time to reach 230 °F is about 7-9 minutes from the loss of CCW.
**Procedures:** E-0: "Reactor Trip Or Safety Injection" – transfer to ES-01 from step 4_a3), ES-01: "Reactor Trip Response" step 6_c3): Start the PDP, RC-02: "Reactor Coolant Pump Off Normal", step 3.0: STOP affected RCP(s) within 1 minute.
**Goals of human action**: To trip the RCPs and start the PDP to prevent RCP seal LOCA.

*G3.4.1 Summary of Qualitative Findings*

Difficulties identified by the NRI team are:

- Making the correct diagnosis of the situation on time
- Performing required (procedural) steps

Both of these difficulties are attributed to the dynamic scenario with unusual combination of events, and the very short time window for performance.

According to the NRI qualitative statements, the most important factors that negatively impact crew performance will be time adequacy (insufficient), training (very infrequent scenario-specific training), scenario complexity (high workload and the use of multiple procedures), stress, procedures (graphical distinction, procedure logic), and execution complexity (dynamic response procedure steps).

Positive factors are indications of conditions (accuracy and availability of necessary information), HMI (good location and ease of locating indicators), procedural guidance (obvious instructions), training (on specific procedure steps), and work processes (use of placekeeping aids).

*G3.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

The total HEP for HFE2A is **1.27E-01** (Error Factor = 3)

Breakdown:
- Identification: 1.2E-2
  - Procedural Step Skipping: 6.0E-3
  - Procedure Logic: 6E-3
- Diagnosis: 1.00E-1
- Execution: 1.54E-2

The main contributors to HEP in the quantitative analysis are *diagnosis/delay errors.* The negative PSF that drives dianogsis/delay errors is time.

*Execution errors* and *Identification errors* make a smaller contribution to the HEP; Identification errors occur via *procedural step skipping* and *procedure logic*. The negative PSFs that drive execution errors are stress and execution complexity. The negative PSFs that drive procedural step skipping are lack of graphical distinction between procedure steps (procedural guidance) and use of multiple procedures (scenario complexity). Procedural logic errors are driven by the use of “NOT” and “AND” statements in the procedures (procedural guidance).

*G3.4.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Available time. The available time for this action is supposed to 7-9 minutes, which is the time from loss of CCW. This is a very short time. Available time is a dominant contributor to delays, which dominate the HEP for HFE2A. | ND |
| **Stress** | Stress. Stress level was evaluated as increased. This increases the HEP in execution tasks. | ND |
| **Scenario Complexity** | - Workload is a characteristic of complex scenarios. Workload is supposed to be high during this scenario, since operators may spend some time solving problems emerging from failures of the distribution panel, SG A feedwater regulation, CCWP A seized shaft and the bus E1C fault. Furthermore the available time (7-9 minutes) is also very short.<br>- Checking – operators check the status of CCW and PDP. They do not need to monitor any parameter. Checking tasks are associated with lower HEP.<br>- Multiple procedures. Operators will be using 2 procedures, which increases HEP.<br><br>*High workload increases the HEP, but checking tasks decrease the HEP. This PSF appears to have a negligible contribution to the HEP since the DTs that include workload are all negligible for HFE2a. | ND* |
| **Indications of Conditions** | Indication availability and accuracy. We assume that the required information concerning the loss of CCW is available in the CR and it is accurate. This reduces HEP. | N/P |
| **Execution Complexity** | Execution complexity is a main contributor to the execution phase HEP. It was determined that the response procedure execution steps were dynamic, which increases HEP. Additionally, the operators work with two procedures (E-0 followed by ES-01 and RC-02) in parallel, which also increases HEP. | ND |

| | | |
|---|---|---|
| **Training and Experience** | Negative aspects:<br>Training in obtaining and interpreting the required indication. In this scenario, operators have to face very specific combination of events, which is believed to be trained for about once in 5 years (as came out from the interviews with simulator instructors).<br><br>Positive aspects:<br>- Training on procedure steps. The crews are supposed to have some practice with executing the relevant steps in similar (not necessarily same) scenarios. This reduces the HEP.<br>- ASEP method assumes that novice operators (less than 6 months experienced) are replaced by experienced operators, so experience is not a factor in the method. It is also not a factor in CBDT.<br>- The NRI team stated that operators have extended experience with CR layout and location of the relevant indicators.<br><br>Insufficient training increases the HEP, but this negative state appears to have a negligible contribution to the HEP in multiple DTs. The positive effect of training on procedure steps slightly reduces the HEP. | ND<br>and<br>N/P |
| **Procedural Guidance** | This is a negative driver for identification errors based on the quantitative results provided by NRI.<br>Negative aspects:<br>- The relevant steps (4 in E-0, 6 in ES-01 and 3.0 in RC-02) are not distinct from other steps.<br>- Procedure logic. Not logic is included in the steps #4_a3 of E-0 and #6_c3 of ES-01. AND or OR logic occurs in some parts of the relevant procedures (e.g. step 9 in FR-H1: Check… SG wide range level… OR Pressurizer pressure…), which raises the error probability of procedure interpretation.<br>- Multiple procedures. Operators will be using 2 procedures, which increases HEP.<br>Positive aspects:<br>Obvious instructions. The relevant instructions are considered as “Obvious” since E-0 and ES-01 and RC-02 are procedures with separate, stand-alone numbered steps. The steps contain standard, unambiguous wording and presents all information required to identify actions directed and objects. According to CBDT, this reduces the likelihood of error. | ND |
| **Human- Machine Interface** | Good indicators are easy to locate. There were not observed any human factor deficiencies in the layout, demarcation or labeling of the control boards during simulator excursion. Similarly, no deficiencies connected to indicators were observed. All the essential indications are placed on the front panels and are supposed to be alarmed. This reduces the HEP. | N/P |

| | | |
|---|---|---|
| **Work Processes** | Using of placekeeping aids. The operators were observed to be using placekeeping aids at the simulator. This reduces the probability of skipping procedure steps and thus reduces the HEP. | N/P |
| **Communication** | All CR personnel adhered to a formal communication protocol during the simulator scenario observed by NRI. The NRI team inferred that crews would use formal communication during the empirical scenarios. Use of formal communication reduces the probability of miscommunicating data and thus reduces HEP. | N/P |
| **Team Dynamics** | | 0 |
| **Other** | In all scenarios, operators are assumed to believe that the instructions presented are appropriate to the situations. This reduces the HEP (via negligible contribution of deliberate violation DT). | N/P |

*G3.4.4 Comparison of Drivers to Empirical Data*

The NRI team correctly identified Adequacy of Time, Scenario Complexity, Procedural Guidance, and Training as negative drivers. NRI believed that the 7-9 minutes time window was very short, and the effect of this short time dominated the HEP (via delays). In the empirical data, none of the 4 crews started the PDPs within the available time window, which was attributed in part to the late identification of the situation (due to the complicated scenario). All of the crews were able to detect the loss of CCW and sealwater within the 7 minutes criteria, but none of the crews was able to execute response actions within the remaining time.

Scenario complexity was identified as a negative driver in the NRI analysis because of a high workload and the use of multiple procedures. The NRI tem predicted that operators would spend time solving problems related to failures of DP1201, SG A FW regulation, CCWP A seized shaft, and the bus E1C fault. This all needed to be accomplished with a 7-9 minutes time window. This negative driver is reflected in the empirical data, which states that having multiple failures at the same time made it difficult to prioritize the demands. The data also notes that crews indicated that they had difficulty because they knew they needed to use multiple procedures, but they didn't have a large enough crew to address all of the procedures at once.

The NRI team felt that Procedural Guidance would be a negative driver because of the use of multiple procedures and procedure readability issues (lack of graphical distinction among procedure steps and difficult logic in E0 and ES01). Procedural guidance was a negative driver in the empirical data, but it was a negative driver for different reasons than those indicated by NRI. In the empirical data, the procedures were insufficient to address the scenario because of two issues: the instruction to start PDP is suggested too late in the high-priority procedure (ES-0.1, a POP5 procedure) and the other procedure that contains the instruction to start PDP is a lower priority procedure (POP4). The preceding steps in the POP5 procedure take time to complete, and the time limit for this event was short, so none of the crews made it through the POP5 procedure.

The NRI team identified Training as both a negative and a positive driver. They indicated that the crew was facing a very specific combination of events, which they train on approximately once every 5 years. However, they also indicate that the crew has been trained on the relevant response procedure steps, but not necessarily in the same scenario. This is consistent with the empirical data, which suggests that training is a negative driver because loss of CCW and sealwater is trained on approximately once every 2 years, but this training is always coupled with a loss of offsite power (LOSP) scenario. The crew had no training on loss of CCW and sealwater in any scenario without a LOSP, so they did not expect loss of CCW and sealwater in a non-LOSP scenario.

The NRI team identified Indications of Conditions and Work Processes as positive drivers, but these were both negative drivers in the empirical data. NRI stated that all of the indications about the loss of CCW are available in the control room and are accurate. The empirical data state that "A lot of train A indications were gone because of the 1201 failure." The NRI team felt that work processes would be adequate because the sample crew was observed using placekeeping aids and adhering to formal communication protocol. In the empirical data, the crews were less than adequate at monitoring the control boards and acknowledging alarms, and they had difficulty in implementing the POP4 procedure. The empirical data mentions that some crews had difficult executing the procedure, which could be attributed to execution complexity.

The NRI team identified stress and execution complexity as negative drivers, but these was not considered negative drivers in the empirical data. The NRI team evaluated the stress level as increased. The empirical data notes that there were observations of stress among the crews, but it did not attribute any effect on the scenario to the observed stress. The NRI team also evaluated the level of execution complexity as increased, because the response procedure steps were dynamic, and operators were working in multiple procedures. In the empirical data, execution complexity is not a driver. The empirical data notes that the execution complexity varies depending on which procedure is in use and who read the procedure (an RO would find specific steps to be more complex than a US). The data also notes that crews indicated that they were having difficulty because they knew that they needed to use multiple procedures.

The NRI team correctly identified HMI as a positive driver. They also identified Communication, Team Dynamics as either positive influences or not a driver, which is consistent with the net effect of these PSFs, which were not drivers of performance in the study.

*G3.4.5 Comparison of Qualitative Analysis to Empirical Data*

The NRI team thought that the operators would have difficulty in making the correct diagnosis of the situation on time and performing required (procedural) steps. Both of these difficulties are attributed to the dynamic scenario with unusual combination of events, and the very short time window for performance. This is consistent with the empirical data, wherein a very complex scenario resulted in late diagnosis and failure to complete the required procedure steps within the short time window.

*G3.4.6 Impact on HEP*

All of the PIFs identified by the NRI team are elements of the CBDT & ASEP approach. However, there was some inconsistency between the qualitative analysis and the quantitative analysis, with regards to the Scenario Complexity and Training.

The NRI team also identified specific aspects of Scenario Complexity and Training as negative drivers, but these aspects had a negligible effect on the final HEP. One aspect of scenario complexity (the use of multiple procedures) is a negative driver in their qualitative analysis and also increases the HEP. However, the NRI team also indicates that another aspect of scenario complexity (high workload) is a negative driver for the scenario. In the quantitative analysis, the high workload does not increase the HEP; the decision trees that contain workload all have a negligible contribution to HEP. Similarly, the NRI team identified Training (obtaining and interpreting the required indication) as a negative driver, but this also has a negligible contribution to HEP. Another aspect of training (the crews have practiced executing the procedure steps in the simulator) was identified as a positive driver and appropriately decreases the HEP.

Adequacy of Time, Stress, Execution Complexity, and Procedural Guidance were each indicated as negative drivers in the qualitative analysis and also increased the HEP. None of the factors identified as positive drivers increased the HEP.

Overall, the HEP for this HFE (1.27E-01) was approximately 66 times greater than the lowest possible HEP (1.94E-03).

### G3.5 HFE3A

**Type of task:** Type C - Post-accidents
**Task description:** Failure of crew to isolate the ruptured steam generator and control pressure below the SG PORV setpoint before SG PORV opening.
The time window to perform the required actions is estimated to be 2 to 3 hours. The actions include:

- isolate the ruptured SG (feedwater and main stream isolation valves closed)
- maintain RCS pressure below the setpoint by cooling down the RCS (cooling the secondary by dumping steam and depressurizing the RCS).

**Procedures:** E-0: "Reactor Trip Or Safety Injection", transfer to E-30 in steps 13 (radiation in steamline, SG level) or 22 (SG level), E-30: "Steam Generator Tube Rupture"
**Goals of human action**: To isolate the ruptured SG and control pressure below the SG PORV setpoint before SG PORV opening.

#### *G3.5.1 Summary of Qualitative Findings*

According to NRI statements, the most probable failure mechanism (but still with very low probability) due to CBDT methodology seems to be procedural step skipping. This failure mechanism is followed by problems with procedure logic and not sufficient attention of operators. NRI does not expect, that especially these two failure mechanisms would be observed during simulator runs, since the crews are very well familiar with the base version of SGTR. Another failure mechanism can occur during manipulations (error of commission, error of omission), especially during the second step, where monitoring of SG pressure is required.

For events with long time windows (> 1 hour), the HRA team states that, since the symptom based procedures provide the crew with all necessary diagnosis support, this contribution [of delay] to the total human failure probability is expected to be negligible when crew actions are driven with this type of procedures.

According to the NRI qualitative statements, the most important factors that negatively impact crew performance will be scenario complexity (monitoring) and procedural guidance (logic, graphical distinction between steps).

Positive factors are adequacy of time, indications of conditions (accuracy and availability), training (frequent scenario-specific training), scenario complexity (workload), work processes (use of placekeeping aids), HMI (good location and ease of locating indicators), and procedural guidance (obvious instructions).

*G3.5.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

The total HEP for HFE3A is **9.49E-03** (Error Factor = 5)

- Breakdown:
  - Identification: 3.45E-3
    - Insufficient attention of operator (data not attended to): 1.5E-4
    - Procedural Step Skipping: 3.0E-3
    - Procedure Logic: 3.00E-4
  - Diagnosis: Negligible
  - Execution: 6.04E-3

Based on the quantitative results, the main drivers for failure are *procedural step skipping* (*identification errors*) and *execution errors*. The negative PSF that drives procedural step skipping is lack of graphical distinction between procedure steps (procedural guidance). The negative PSF that drive execution errors is execution complexity (partially dynamic response steps).

*Insufficient operator attention* and *procedure logic* each make a smaller contribution to the HEP. The negative PSF that drives insufficient operator attention is monitoring activities (scenario complexity). Procedure logic is driven by the use of “AND” and “OR” statements in the procedures (procedural guidance).

*G3.5.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Available time: The time window to perform the required actions is estimated to be 2 to 3 hours. This is a large time window. Available time is the only PSF that contributes to HEP for delays; for this amount of time, the HEP contribution of delays is negligible. | N/P |
| **Stress** | Stress is supposed to be nominal for this activity and has a negligible effect on HEP. | 0 |
| **Scenario Complexity** | - Workload is a characteristic of complex scenarios. Workload is supposed to be low during this base version of SGTR. This reduces the HEP.<br>- Monitoring. Operators should monitor SG level during the scenario. Monitoring tasks are associated with a higher HEP since operators must perform monitoring actions frequently enough to catch the required parameter state. | ND |
| **Indications of Conditions** | - Indication availability and accuracy. The important information concerning SG level is at disposal in the CR and it is accurate. Good indications reduce HEP.<br>- All the essential indications are alarmed. | N/P |
| **Execution Complexity** | Execution complexity was not identified as a driver based on qualitative statements. However, execution complexity is a main contributor to the execution phase HEP. It was determined that the response procedure execution steps were step-by-step and partially dynamic, which slightly increases HEP. | 0 |
| **Training and Experience** | - Training in obtaining and interpreting the required indication. The crews are very well familiar with this type of scenario. This kind of scenario is supposed to be trained for about 4 times a year (as came out from the interviews with simulator instructors).<br>- Training on steps. The crews have received training on the correct interpretation of the relevant steps. This reduces HEP.<br>Good training decreases the HEP (The effect of the DTs where training appears is negligible)<br>- ASEP method assumes that novice operators (<6months experienced) are replaced by experienced operators, so experience is not a factor in the method. It is also not a factor in CBDT.<br>- The NRI team stated that operators have extended experience with CR layout and location of the relevant indicators. | N/P |

| | | |
|---|---|---|
| **Procedural Guidance** | This is a negative driver for identification errors based on the quantitative results provided by NRI.<br>Negative aspects:<br>- Graphical distinction. The relevant steps (13 in E-0 and 2-4 in E-30) are not distinct from other steps.<br>- Procedure logic. In some parts of the activity described with HEP quantified, several actions may be required within one step (AND logic expected), also OR logic can be found in some steps of E-30 procedures (e.g. step 2), but no redundant alternatives are expected explicitly (combination of OR and AND logic within one step).<br>Positive aspects:<br>- Obvious instructions. The relevant instructions are considered as “Obvious” since E-0 and E-30 are procedures with separate, stand-alone numbered steps. The steps contain standard, unambiguous wording and presents all information required to identify actions directed and objects. This reduces HEP.<br>- Single procedure. Operators are using one procedure at a time. This reduces the HEP. | ND |
| **Human- Machine Interface** | - Good indicators are easy to locate. There were not observed any human factor deficiencies in the layout, demarcation or labeling of the control boards during simulator excursion. Similarly, no deficiencies connected to indicators were observed.<br>- All the essential indications are placed on the front panels and are alarmed. The main alarms considered in this scenario are: 1. secondary circuit radiation alarms, 2. damaged SG abnormal level, 3. significant drop of pressurizer pressure, 4. significant drop of pressurizer level. | N/P |
| **Work Processes** | Using of placekeeping aids. The operators were observed to be using placekeeping aids at the simulator. This reduces the probability of skipping procedure steps and thus reduces the HEP. | N/P |
| **Communication** | All CR personnel adhered to a formal communication protocol during the simulator scenario observed by NRI. The NRI team inferred that crews would use formal communication during the empirical scenarios. Use of formal communication reduces the probability of miscommunicating data and thus reduces HEP. | N/P |
| **Team Dynamics** | | 0 |
| **Other** | In all scenarios, operators are assumed to believe that the instructions presented are appropriate to the situations. This reduces the HEP (via negligible contribution of deliberate violation DT). | N/P |

*G3.5.4 Comparison of Drivers to Empirical Data*

NRI correctly identified Indications of Conditions, Training and Experience, and HMI as positive drivers. They note that the indications are available in the control room and they are accurate. The NRI team notes that there are alarms for the main indications (secondary circuit radiation, SG level, PZR pressure drop, and PZR level drop). In the empirical data, the radiation alarms helped the crews identify the SGTR quickly. The NRI team also noted that the crews are very familiar with this scenario, since they are trained on it approximately 4 times a year. This mirrors the empirical data, which states that this is a frequently trained scenario with no added complications.

The NRI team correctly identified Stress and Execution Complexity as PSFs that are not a driver in the empirical data. The NRI team identified Adequacy of Time, Work Processes, and Communications as nominal/positive drivers. In the empirical data, these were not performance drivers. The NRI team identified Team Dynamics as not a driver. In the empirical data, Team Dynamics was a positive driver. In CBDT, the effect of a PSF rated nominal/positive is the same as a PSF rated not a driver.

The NRI analysis identified Scenario Complexity and Procedural Guidance as potential negative drivers. Procedural guidance was a negative driver in the NRI analysis because of procedure readability issues (lack of graphical distinction among the procedure steps and difficult logic in E-30). NRI felt that the Workload aspect of scenario complexity would be low. However, they also noted that operators were required to monitor the SG level (as opposed to checking the level once). Monitoring tasks are associated with higher HEP than checking tasks, so this aspect of scenario complexity is a negative driver. In the empirical data, scenario complexity was not a driver and procedural guidance was a positive driver[6].

*G3.5.5 Comparison of Qualitative Analysis to Empirical Data*

The NRI team predicted a low probability of failure for this action, which is consistent with the empirical data (all of the crews succeeded). The operators are unlikely to make an error in this scenario, but if the operators do make an error, it is likely due to skipping a step in a procedure or making an error during action execution.

*G3.5.6 Impact on HEP*

All of the PSFs identified by the NRI team are elements of the CBDT & ASEP approach. All of the identified negative drivers increased the HEP in CBDT & ASEP quantification. None of the factors identified as positive drivers increased the HEP.

Overall, the HEP for this event was very low. The HEP for this HFE (9.49E-03) was approximately 5 times greater than the lowest possible HEP (1.94E-03).

### G3.6 Comparison Summary

*G3.6.1 Predictive Power*

The predictive power of the NRI CBDT & ASEP analysis method in this study was judged to be fair. For the most part, the final HEPs were consistent with the identified drivers.

There was good correspondence between the negative drivers relevant to the quantification and the qualitative descriptions provided by the analysis team, although the magnitude of each driver's effect was not always consistent between the qualitative and quantitative analyses.

[6] Part of this discrepancy is due to the way that PSFs are used in the CBDT. In the CBDT method, when factors are considered to be positive, the effect of the factors is considered negligible (they do not decrease HEP), but negative factors increase the HEP. CBDT uses 30 questions to predict the HEP; several of these questions map onto each of the PSFs used for the empirical data, and there is no explicit weighting of the PSFs. The presence of a single negative aspect of the PSF is sufficient to make the PSF a negative driver for the HEP; the positive aspects of the PSF cannot offset the negative aspect.

The correspondence between the drivers identified in the method and the drivers identified in the crew data was fair. The method usually identified some of the main drivers, but the reason for the PSF being identified as a driver in the NRI analysis was often different than the reason that it was a driver in the empirical data. In most cases, the NRI analysis identified additional negative drivers which were not observed in the data.

The HEPs for the empirical HFEs fall into two groups, with an order of magnitude difference between the HEPs for harder HFEs and easier HEPs. There was good correspondence between these groups and the rankings of the HFEs based on the empirical data. There was no significant difference between HFEs within the same group.  For the most difficult HFEs, the HEPs were optimistic.

#### G3.6.1.1  Qualitative Predictive Power – in Terms of Drivers

In this study, the CBDT & ASEP analysis identified some of the important drivers that influenced performance in the scenarios.  In many cases there was superficial agreement between the method and the crew data in terms of important drivers, but the reason for a factor being identified as a driver in the NRI analysis did not appear to be the same as that identified in the data (this is especially true for Procedural Guidance). In most cases the method identified additional negative drivers that were not observed in the empirical data.

The identification of drivers was guided by questions asked in the decisions trees and by factors addressed in ASEP. Negative conditions lead to higher HEPs. Via the 30 questions in CBDT, the method generally identified at least one aspect of the negative drivers identified in the empirical data, but it generally fell short of identifying the most critical aspect of the driver. For example, procedural guidance was frequently identified as a negative driver by both the NRI team and in the empirical results. However, using CBDT the NRI team often identified the procedure formatting as a negative driver (i.e., the lack of graphical distinction between steps, which could increase errors via procedural step skipping). In the empirical data, the procedure content (CSF trees do not address misaligned valves) or procedure priority (FR-H1 takes priority over E-30) was relevant rather than the formatting of the procedures.

It appears that the CBDT questions drove the analysts to identify a number of drivers that could produce human errors, but that these drivers were not sufficient to produce an HFE (the errors could have been easily recovered, as was the case with the quickly recovered error of omission by Crew S in HFE 1A).

All of the negative drivers identified by NRI had an impact on the HEP, although the magnitude of each driver’s effect was not always consistent between the qualitative and quantitative analyses. For example, NRI qualitatively stated that graphical distinction among procedure steps was a slightly negative driver and workload was a negative driver for HFE1A. In the quantification, graphical distinction was a dominant contributor to the HEP and workload had a relatively small contribution to the HEP.

#### G3.6.1.2  Qualitative Predictive Power – in Terms of Operational Expressions

The qualitative analysis in the CBDT & ASEP method largely consists of identifying the driving factors for performance.

The discussion of operational expressions provided by NRI mainly summarized which phase of human performance (identification, diagnosis/delay, or execution error) was most relevant to the particular HFE. In some cases the NRI team discussed a limited number of failure mechanisms that could influence errors in the three phases. These discussions were driven mostly by analyst insight (which appear to be inspired by aspects of the method, but were not directly driven by the method).

G3.6.1.3 Quantitative Predictive Power

The NRI team predicted higher HEPs for the HFEs where crews failed and also predicted lower HEPs for the HFEs where the crew succeeded. In the empirical data, the HFEs were ranked from most difficult to least difficult as (2A, 1C, 1A, 3A). The NRI team predicted high HEP for 1C and 2A and low HEP for 1A and 3A.

The HEP results from the NRI analysis roughly fall into two groups: high HEP (approx. 1E-01) and low HEP (approx. 1E-02). The HEPs for HFE 1C (highest) and 2A were very close; the difference between the high HEPs was a factor of 1.5 [HFE 1B also fell into this group, but was not ranked in the empirical data]. The HEPs for 1A and 3A (lowest) were a factor of 12.75 smaller than the HEP for 1C and 2A, but the difference between the low HEPs for 1A and 3A was a factor of 1.27. Given the uncertainty on each HEP (expressed as an error factor), the differentiation between the two groups is significant, but the differentiation within each group is not significant.

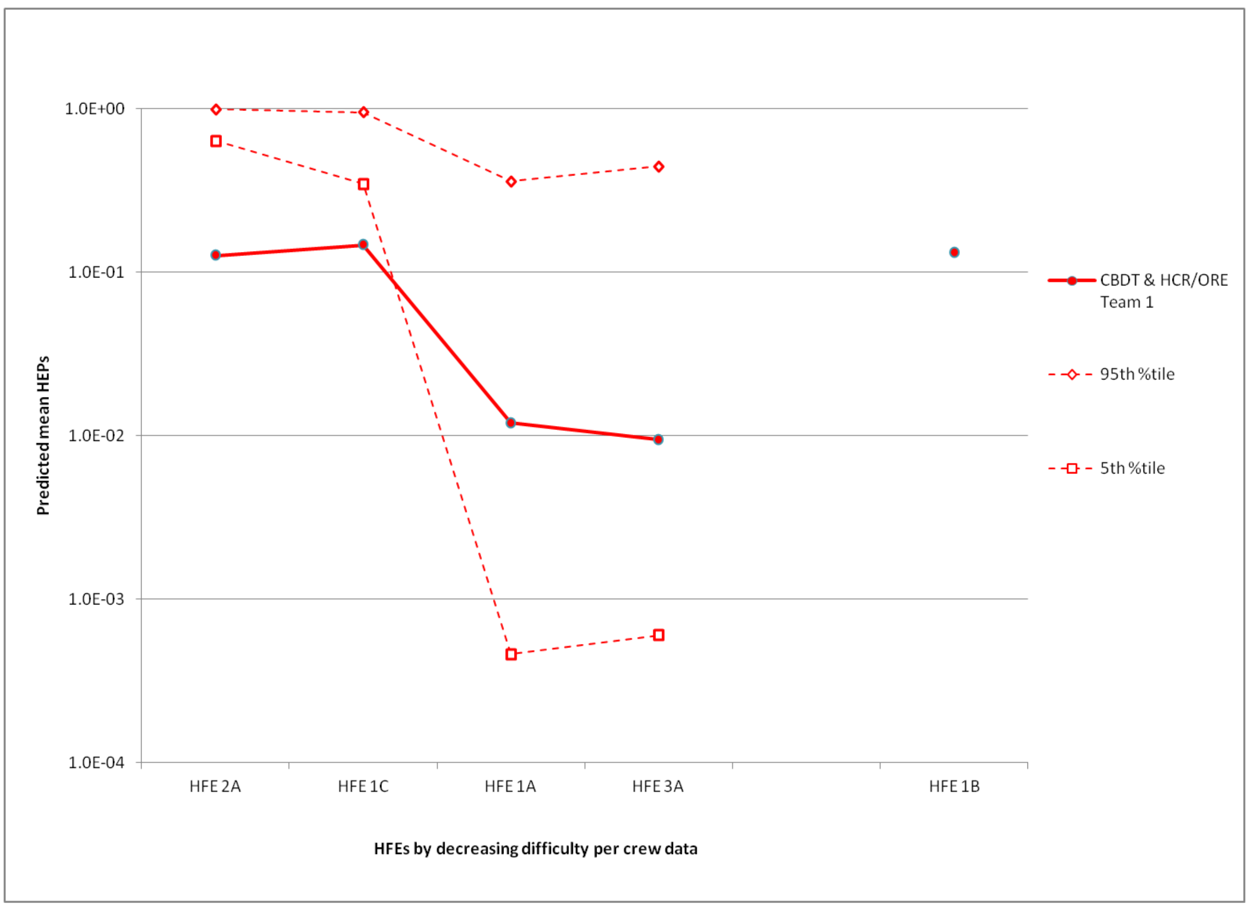


Figure 0.3 Predicted Mean HEPs with Uncertainty Bounds by NRI HRA Team with CBDT/ASEP

*G3.6.2  Assessment of Guidance and Traceability*
Method traceability is moderately good. CBDT uses the state of the PSFs (qualitative analysis is conducted by answering 30 questions) to quantify the failure mechanisms. The guidance for using the qualitative analysis to produce quantitative estimates is moderately good.  However, the HEPs are heavily rooted in THERP, so the quantitative aspects of the method are as traceable as THERP. The method does not always provide sufficient documentation to explain why the selected questions are relevant to human error or how individual PSFs are weighted in the quantification.

Adequacy of guidance is fair to moderately poor. The set of questions used to elicit the context is not comprehensive and there is not sufficient guidance to answer many of the questions. There is no guidance on how to conduct task analysis on the HFEs.

In CBDT, the end points (HEPs) for the decision tree branches were quantified based on THERP. In the documentation there is limited detail about the elicitation process. The decision trees document how various PSF combinations relate to specific failure mechanisms, but do not document the magnitude of the effect of individual PSFs. It is unclear how the individual PSFs were weighted or how their relationships were determined.

CBDT does not provide sufficient guidance to answer many of the questions. CBDT asks questions about whether the indications of conditions are available and accurate, but the guidance for how to answer this question appears to be inadequate, based on the NRI application. In HFE 1A, NRI determined that the incorrect indications of recirculation valve position would not considerably affect the crew's action. This decision influenced the HEP via the *data not available* DT, which was deemed to be negligible because the indications were assumed to be accurate. According to the empirical data, the masked indication was an important negative driver for crew performance. This suggests that CBDT needs better guidance on how to select different decision branches. The CBDT question "Are the indications accurate" does not provide sufficient guidance to help analysts identify which indications are relevant and how to characterize their accuracy.

The CBDT approach does not provide a comprehensive assessment of the factors that could affect human error. The set of questions does not include several facets of the factors identified as performance drivers in the empirical data.

There are also many facets of the empirical PSFs that are not considered in the CBDT method, but which are relevant to crew performance. For example, in HFE 2A, the crews could not work through the procedures fast enough. Procedures are a frequent negative driver in the CBDT trees, but the method mostly focused on the readability of the procedures, not the appropriateness of the procedures for the scenario.

It also appears that the CBDT method does not consider all of the sources of information that crews use to solve problems. One of the main drivers of HFE 1A was procedural guidance (the Critical Safety Function trees). The CBDT methodology only considers the paper procedures. CBDT also assumes that the paper procedures are sufficient to diagnose the situation without significant cognition from the operators.

*G3.6.3  Insights for Error Reduction*
CBDT provides a number of insights into reducing errors, but the method misses some of the important drivers.  It appears to be conservative for severe contexts.  The method could benefit

from including the factors that lead to limiting scenarios with a very high likelihood of failure. The current set of questions in CBDT does a moderately good job of identifying factors that could result in random human errors, but falls short of identifying the wide range of factors responsible for more significant human failures, especially in tasks that are not guided by procedures. CBDT provides moderately good insights for reducing random human errors, but does a moderately poor job of providing insights for severe scenarios involving substantial complexity and teamwork.

*G3.6.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses*
The method could benefit from additional guidance on how to conduct qualitative task analysis. The method could also benefit from an expanded set of questions, and more explicit guidance for answering the existing questions

The combination of ASEP and CBDT could be double-counting errors in the diagnosis phase, since diagnosis is an aspect of both the identification tasks (quantified in CBDT) and the diagnosis/delay tasks (quantified using ASEP/THERP). Further guidance is needed to ensure that analysts using CBDT and ASEP do not overestimate HEP because of the overlap between the methods.

It appears that this methodology helps analyst evaluate a number of conditions that could result in random human errors. However, many of these human errors can be quickly recovered, so they do not necessarily produce an HFE.

The method appears to under-predict HEP for the most severe contexts. The CBDT method appears to be most beneficial when used to identify the possibility for random errors.

The methodology could benefit from task analysis guidance. The CBDT questions identify causes for human errors that occur at a low level (e.g., skipping a step in a procedure), but the data demonstrated that many of these errors have a minor impact on the scenario, i.e. they do not create an HFE. One of the main problems with the NRI analysis was that they assumed that the masked indicators for HFE 1A would not have a significant impact on the crew. This type of analyst error could be avoided if CBDT provided better guidance to help analysts identify the key information (alarms, indicators, parameters, procedures, procedure steps, etc) required for the crew to succeed. Likewise, CBDT could benefit from guidance that helps analysts determine the key opportunities for human errors to create HFEs.

In the application of the method, there was inconsistency between the qualitative and quantitative assessment of the impact of the PSF. The direction of the effect (positive vs. negative driver) was always consistent between NRI qualitative statements and the quantitative results, but the magnitude of the effect (e.g., negative vs. slightly negative) was not consistent.

In CBDT, the PSFs are treated on a binary scale. In CBDT, workload is either high or low. For the NRI analysis, they often assessed workload as “moderate.” To quantify the HEP, they averaged between the HEP for “high” workload and the HEP for “low” workload.

Another opportunity for variability in CBDT results from the problem with binary PSFs, coupled with the fact that the impact (weight) of individual PSFs. Analysts must rate a PSF as either positive or negative; there is no opportunity for analysts to influence the HEP via a “slightly negative” PSF impact. This presents an opportunity for analyst-to-analyst variability; one analyst may determine that a PSF has a “slightly negative” impact on the scenario should be treated conservatively in the decision tree. Another analyst may determine that the effect of the PSF is so slight, that it should be treated less conservatively. CBDT does not provide guidance to resolve these situations.

CBDT quantification produces an HEP for different combinations of PSFs rather than quantifying the individual impact of each PSF. Since CBDT quantifies the HEP based on the state of multiple factors at once, it is difficult to quantify the effect of positive factors beyond stating that they do not increase the HEP more than they would in their negative state. CBDT could benefit from using a value-neutral set of PSFs, where both negative and positive PSFs alter the HEP.

CBDT traceability could be improved by quantifying the impact of individual questions (quantifying at the branch point instead of the end state). This would allow analysts to see the impact of each individual PSF. At a more basic level, traceability could be improved by providing additional sources of evidence for HEP estimates.

## G.4 HRA Calculator (NRC)

### G4.1 HFE 1A

*G4.1.1 Summary of Qualitative Findings*

The NRC Calculator team defined the operator action success criteria as recognizing total loss of feedwater and establishing F&B. Three cues were identified by the team:

- Failure of main feedwater pumps and plant trip
- Decreasing level in the SGs
- Failure of AFW pumps to start and run

The relevant procedures identified were E0 and FRH1. Most actions were considered to be control room actions. Three subtasks were identified:

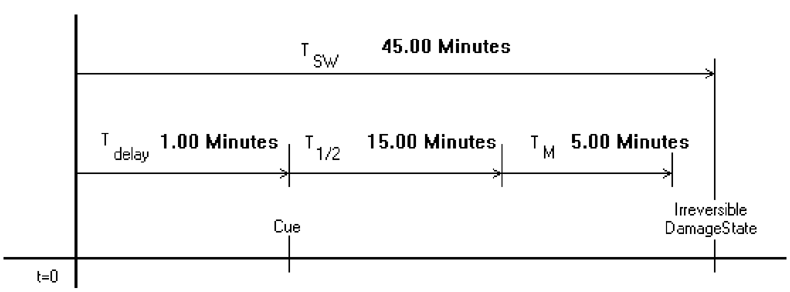
- Diagnose failure of AFW
- Actuate Safety Injection
- Open both of the PZR PORVs

Recovery was not considered for either diagnosis or execution.

The RO monitoring the parameters will indicate loss of feed and will then be directed to manually trip the reactor within 30 seconds. For this scenario, AFW pump in recirc will show flow from AFW but the SGs will not show increase in flow. All SGs will be decreasing in flow in the same fashion. An RO monitors parameters (critical safety functions) while the primary RO is performing verification steps in E0. Alarms will annunciate, notifying that the feedwater has tripped.

The NRC Calculator team noted that the driver for this scenario would be the cognitive portion of the HFE. However, they note that operators are trained to monitor feed flow (E0, ES-01, and monitoring critical safety functions), have indication which will lead them to F&B, and will not hesitate to start F&B. They also noted that the relevant PSFs are: moderate complexity and high stress.

The timing used for this HFE was as follows:




- t_delay is 1 minute; the cue, in this case is decreasing water level on the SG, which the team estimated would take ~1 minute to manifest and be seen.
- T_1/2 is 15 minutes; this is based on the team's estimate for diagnosis (recognizing misleading indication) and entering FR-H1.

- T_m is 5 minutes; this is based on the team's estimate for execution (starting SI and opening a pressurizer PORV)

*G4.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP = 0.01, EF= 5

- Cognitive contribution: Pcog = 6.2E-03; CBDT
- Execution contribution: Pexe = 4.3E-03
  - Actuate SI: 4.3E-4 x5 (stress factor)
  - Open PORV: 4.3E-4 x5 (stress factor)
- No credit for recovery of either cognitive or execution steps
- No dependencies identified

From the quantitative analysis process, two cognitive failure mechanisms were identified by the method as drivers Pcd and Pce:

| Pc Failure Mechanism | Branch | HEP |
|---|---|---|
| $Pc_a$: Availability of Information | a | neg. |
| $Pc_b$: Failure of Attention | d | 1.5E-04 |
| $Pc_c$: Misread/miscommunicate data | a | neg. |
| **$Pc_d$: Information misleading**<br>False AFW flow indication would delay the operators' diagnosis on knowing the complete loss of heat sink. However, the operators would perform F&B when criteria reached. The equal reduction of all SGs' levels would eventually lead operators to learn that all feed water were not available. | **b** | **3.0E-03** |
| **$Pc_e$: Skip a step in procedure**<br>The procedure step of checking SG water levels are not graphically distinguished from other step. | **c** | **3.0E-03** |
| $Pc_f$: Misinterpret instruction | a | neg. |
| $Pc_g$: Misinterpret decision logic | k | neg. |
| $Pc_h$: Deliberate violation | a | neg. |
| **Sum of $Pc_a$ through $Pc_h$ = Initial Pc =** | | 6.2E-03 |

While these two drivers have similar numerical impacts, the team indicated that recognition of the misleading direction (and thus entry into FR-H1) would be the critical point for the HFE.

While the cognitive error was identified by the analyst as the driving factor, the relative numerical contribution of the execution portion was comparable to that of the cognitive portion (6.2E-03 cognitive contribution vs. 4.3E-03 execution contribution). In the appendix of their form which documented the Calculator results/documentation, the stress was designated as "High" due to the fact the plant response was not as expected [impact x5].

*G4.1.3 Summary Table of Driving Factors*

The following scale was used by the assessors to rate the impact of the various factors:

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | More than adequate time. In this method the adequacy of time is primarily manifested by which cognition method is chosen – CBDT or HCR/ORE. In this case CBDT was chosen. Dependence levels for crediting recovery actions are based on time; in this case a medium dependency was evaluated for the cognitive portion, but the recovery was not actually credited. | N/P |
| **Time Pressure** | | N/A |
| **Stress** | (THERP) High because plant does not respond as expected. Contributed significantly to the execution portion. | ND |
| **Scenario Complexity** | Cognitive complexity due to misleading information is addressed by CBDT and THERP (handled as a multiplier in THERP). | MND |
| **Indications of Conditions** | Availability and clarity of indications with respect to the scenario are explicitly considered by CBDT in several places. Pcd (Information Misleading) was one of the major contributors in the quantitative analysis, and was identified as the most likely driver in the qualitative analysis. In this case this is the same as scenario complexity.<br>Location and physical quality of indications was also addressed, but here it is discussed under Human Machine Interface. | MND |
| **Execution Complexity** | Explicitly considered by THERP tables based on specific types of complexity (e.g., long checklist v. short checklist, etc.); execution was not complex in this scenario. | N/P |
| **Training** | Explicitly credited in relevant CBDT trees. Not a factor in this analysis. | 0 |
| **Experience** | Not addressed by this method. | N/A |
| **Procedural Guidance** | Evaluated explicitly as part of the guided qualitative analysis; both availability and quality of guidance appears as decision points in relevant CBDT trees. Pcd tree is a driver because all cues are not stated for this scenario and Pce is a driver because the step is not graphically distinct. | MND |
| **Human- Machine Interface** | Indicator location and physical quality addressed in CBDT. Select HMI issues relevant to execution are also covered by THERP. Not applicable to this HFE. | N/P |
| **Work Processes** | Availability of check off provisions within the procedure is also assessed in the execution portion. That was not found to be a driver in this scenario except in that it kept the execution error probability low because checkoff was used. | N/P |
| **Communication** | Use of formal communications explicitly covered by CBDT in Pcc. | N/P |
| **Team Dynamics** | Not addressed by this method. | N/A |

| | | |
|---|---|---|
| **Other** | None. | N/A |

*G4.1.4 Comparison of Drivers to Empirical Data*

At the high level, the drivers indicated by the team match the empirical data to a moderate level. They match in that both see this scenario as moderately complex as the cues do not match the scenario. The dominant failure mechanisms for this assessment were information misleading (due to cues not as stated) and skip a step in the procedure (both cognitive and execution). Misleading information in combination with the mismatched procedure was a driver for both the assessment and the empirical results. Skipping a step in the procedure was manifested in the empirical data: one crew skipped a step (failed to stop RCPs before starting F&B) but recovered from it.

The empirical data showed that one crew failed to stop the RCPs before starting F&B, but they recovered and were able to complete the action successfully. The HRA did not credit recovery despite the long time frame.

| Factor | Empirical | HRA | Comments |
|---|---|---|---|
| **Stress** | 0 | ND | This is a strong numerical driver. THERP automatically defines stress level based on a pre-determined set of criteria. In this case, the plant did not respond as expected (recirc valve open), so a stress factor was applied (~30% of the total HEP value). |
| **Scenario Complexity** | ND | MND | The main driver for HRA scenario was the combination of the fact that there was no clear indication of the loss of AFW and there was a mismatch between the plant and the procedures (yellow path vs. decreasing SG levels). These are the same drivers identified by the empirical team. The difference is the designation of “main” ND in the HRA analysis – these were considered “main” because they collectively provided the largest numerical contribution to the resulting HEP. |
| **Indications of Conditions** | ND | MND | |
| **Procedural Guidance** | ND | MND | |
| **Training/Experience** | N/P | 0 | Not a driver in HRA; strong positive factor in empirical data |
| **Adequacy of Time** | N/P | N/P | Both acknowledge there was plenty of time. |
| **Execution Complexity** | N/P | N/P | Nominal in both cases. |
| **Human- Machine Interface** | N/P | N/P | Nominal in both cases |
| **Work Processes** | N/A | N/P | Nominal in HRA; indeterminate in empirical analysis. |
| **Communication** | N/A | N/P | Nominal in HRA; indeterminate in empirical analysis. |
| **Time Pressure** | ---- | N/A | |
| **Team Dynamics** | N/A | N/A | |

*G4.1.5 Comparison of Qualitative Analysis to Empirical Data*

The operational story in the qualitative analysis was fairly consistent with the actual data – the crew figured out that AFW was lost (as expected) and that led them to FRH1 where F&B was started without difficulty. The timing of the operational story from the HRA was on the pessimistic side; timing data for the HRA was based on team judgment using the scenarios. The HRA estimated 16 minutes (after reactor trip) to reach FRH1, with execution complete at 21 minutes. It took the crews 8-13 minutes to diagnose and enter FRH1 and they established F&B in 10-17 minutes.

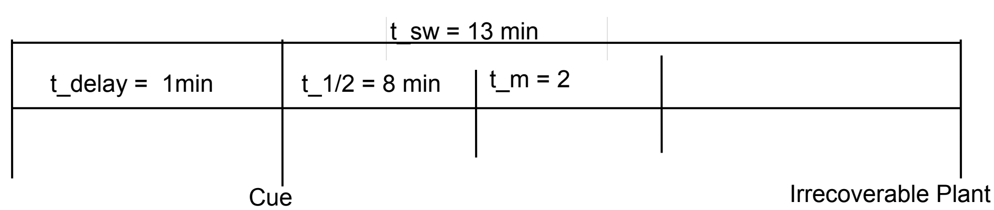
*G4.1.6 Impact on HEP*

Overall, the HEP for this failure was 1.0E-2; this is a mid-range HEP, indicating that the operators are generally well able to find a success path, but some difficulties exist. The HEP was fairly equally split between cognitive and execution errors (cognitive: 6.2E-3; execution: 4E-03). While the experimental data yielded no failures, there were some insights that support this mid-range HEP with an even split between cognitive and execution error: one crew did not know that the recirculation valve was open (cognitive difficulty) and another crew skipped a step in the procedure (execution) which they were able to recover from.

## G4.2 HFE 1B

*G4.2.1 Summary of Qualitative Findings*

The NRC Calculator team noted that this HFE was similar to 1A with the exception of revised timing. In addition to the two relevant PSFs identified in 1A (moderate complexity and high stress), the team identified that time available would also be a relevant PSF. The team indicated the primary driver for this scenario would be the limited time for diagnosis and execution. The quantitative findings reinforce the driver is due to limited time, specifically for cognition. The timing used for this HFE was as follows:




- 8 minutes for diagnosis involves recognizing misleading indication and entering FR-H1. This is reduced from 1A because, with the automatic trip happening later than the manual trip, it is expected that the critical function tree will have satisfied the red path criteria (although the path will not actually be red because out the open recirc valve) by the time the crew has completed the immediate actions in E0. This will take the team directly to FRH1, whereas in 1A, the team is expected to get to FRH1 from the procedures.
- 2 minutes for execution involves starting SI and opening a pressurizer PORV. This is reduced from 1A because the crew is expected to feel time pressure, knowing they have a short time frame, and that will lead them to cool down as close to the max cool down rate as possible, where in 1A they would cool down more slowly.

*G4.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP = 3.9E-1, EF= 1

- Cognitive contribution: Pcog = 3.9E-01; HCR/ORE
  - PWR CP3 sigma value of 0.77
- Execution contribution: Pexe = 4.3E-03
  - Actuate SI: 4.3E-4 x5 (stress factor)
  - Open PORV: 4.3E-4 x5 (stress factor)
- No credit for recovery

CP3 is defined as a response following an event that gives rise to a primary cue that has to be achieved *before* some plant parameter reaches a critical value. This critical value can be regarded as a soft prompt or secondary cue. In this case, the primary cue is the reactor trip and recognition of no AFW flow. The actions for F&B then need to be performed before core damage.

*G4.2.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | In this method the adequacy of time is primarily manifested by which cognition method is chosen – CBDT or HCR/ORE. In this case HCR/ORE was chosen. Dependence levels for crediting recovery actions are based on time; no credit was given for recovery | MND |
| **Time Pressure** | | N/A |
| **Stress** | (THERP) High because plant does not respond as expected. Contributed to the execution portion, but the execution was a negligible contributor to the total HFE. | 0 |
| **Scenario Complexity** | Scenario complexity also explicitly addressed in THERP for its impact on execution (separate from execution complexity). In this case, because the plant does not respond as expected, a stress factor (x5) is applied. This contributed to the execution portion, but the execution was a negligible contributor to the total HFE. Not considered in the HCR/ORE method. | 0 |
| **Indications of Conditions** | Addressed in CBDT, but not used in this case. | 0 |
| **Execution Complexity** | Explicitly considered by THERP tables based on specific types of complexity (e.g., long checklist v. short checklist, etc.); execution was not complex in this scenario. | N/P |
| **Training** | Addressed in CBDT, but not used in this case. | 0 |
| **Experience** | In HCR/ORE, this can be addressed by calculating a scenario specific sigma, but this analysis used the predetermined sigma. | N/P |
| **Procedural Guidance** | Addressed in CBDT, but not used in this case. | 0 |
| **Human- Machine Interface** | Select HMI issues relevant to execution are also covered by THERP. Not applicable to this HFE. | N/P |

| | | |
|---|---|---|
| **Work Processes** | Availability of check off provisions within the procedure is also assessed in the execution portion (THERP). That was not found to be a driver in this scenario, but there was a small positive impact due to available checkoff provisions. | N/P |
| **Communication** | Addressed in CBDT, but not used in this case. | 0 |
| **Team Dynamics** | Considered implicitly in building the timeline. In HCR/ORE, this can be addressed by calculating a scenario specific sigma, but this analysis used the predetermined sigma. | N/P |
| **Other** | Sigma value of HCR/ORE is based on the type of cue/action. | ND |

*G4.2.4 Comparison of Drivers to Empirical Data*
Empirical data not available for this scenario.

*G4.2.5 Comparison of Qualitative Analysis to Empirical Data*
Empirical data not available for this scenario.

*G4.2.6 Impact on HEP*
Empirical data not available for this scenario.

### G4.3 HFE 1C

*G4.3.1 Summary of Qualitative Findings*

In this scenario, the operators have successfully established F&B and re-established AFW flow to one or several of the SGs, though when the flow reaches the first SG, a SGTR occurs. The HFE is: operator fails to cooldown and depressurize following SGTR and successful F&B. The operator action success criteria is to recognize the increasing pressure and level in the ruptured steam generator and isolate the steam generator prior to SG PORV opening.

The cues for this action as identified by the team are:

- Increasing pressure and level in ruptured steam generator
- Radiation monitoring
- Decreasing RCS level or increasing SI flow

Two sub-tasks were identified:

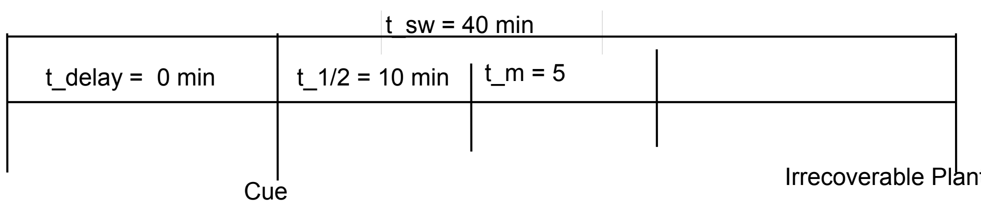
- Diagnose ruptured SG
- Isolate ruptured SG

The team did not assess or credit any recovery factors. During the qualitative analysis, two PSFs were identified as relevant: moderate complexity and high stress. The degree of clarity of the cue was assessed as “average.”

Once AFW is re-established that will kick the crew out of FRH.1 back to the previous procedure and step in effect, which is E0. At this same time the SGTR will occur because of the injection. E0 will take the operators to E10 and then to E-30. Note: this is a misinterpretation of the given scenario. The scenario did not specify that the SGTR would occur once they were out of FRH.1.

When STGR occurs it may be difficult to diagnose the SGTR because the AFW flow is just being established and might mask the rupture. RCS may give indication. There is initially no secondary radiation because there is a minimum steam flow. The blow down (BD) and sampling is secured because of the previous SI. When SG reaches 50% the control room would stop flow but due to the SGTR the SG level would still be increasing, indicating something is wrong. The site may also get radiation alarm indication at this time

The operators must recognize the ruptured SG before the SGs are filled and the SG PORVs lift releasing radiation to the atmosphere. The time available for the operators depends on the length of time until the SG PORVs are lifted. For this analysis, a value of 40 minutes is used. The timing analysis is as follows:




In this case, the cue is the increasing level in the ruptured SG, which is pretty instantaneous.

- It takes 10 minutes to recognize the cue, diagnose and get to the appropriate step in E-30 (via E10). The team notes that the operators are monitoring the SG levels closely and should diagnose quickly.
- Responding to the SGTR will occur after AFW flow is established and the operators have exited FRH1. Because of this misinterpretation of the scenario, t_delay and t_1/2 may not be consistent with what the team would have assessed if they had to figure in working through FRH.1.

*G4.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP = 1.0E-2, EF= 5

- Cognitive contribution: Pcog = 6.0E-03; CBDT
  - No recovery
- Execution contribution: Pexe = 4.4E-03
  - Identify Ruptured SG: 4.3E-4 x5 (stress factor)
  - Isolate SG: 4.3E-4 x5 (stress factor)
  - Note: because of the misinterpretation of the scenario, certain steps required to RCS depressurization and cooling are not quantified in this scenario.
- No credit for recovery

From the quantitative analysis process, two cognitive failure mechanisms were identified by the method as drivers Pcb and Pce:

| Pc Failure Mechanism | Branch | HEP |
|---|---|---|
| $Pc_a$: Availability of Information | a | neg. |
| **Pcb: Failure of Attention**<br>Monitored, Not Alarmed | **e** | **3.0E-03** |
| $Pc_c$: Misread/miscommunicate data | a | neg. |
| Pcd: Information misleading | a | neg. |
| **$Pc_e$: Skip a step in procedure**<br>The procedure step is not graphically distinct. | **c** | **3.0E-03** |
| $Pc_f$: Misinterpret instruction | a | neg. |
| $Pc_g$: Misinterpret decision logic | k | neg. |
| $Pc_h$: Deliberate violation | a | neg. |
| **Sum of $Pc_a$ through $Pc_h$ = Initial Pc =** | | 6.2E-03 |

The relative numerical contribution of the execution portion was comparable to that of the cognitive portion (6.0E-03 cognitive contribution vs. 4.4E-03 execution contribution).

*G4.3.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | More than adequate time. In this method the adequacy of time is primarily manifested by which cognition method is chosen – CBDT or HCR/ORE. In this case CBDT was chosen. Dependence levels for crediting recovery actions are based on time; in this case recovery was not credited. | N/P |
| **Time Pressure** | | N/A |
| **Stress** | (THERP) High because plant does not respond as expected. Contributed significantly to the execution portion. | ND |
| **Scenario Complexity** | Components of Cognitive Complexity explicitly addressed by CBDT. Scenario complexity also explicitly addressed in THERP for its impact on execution (separate from execution complexity). In this case, because the plant does not respond as expected, a stress factor (x5) is applied. | ND |
| **Indications of Conditions** | Availability and clarity of indications with respect to the scenario are explicitly considered by CBDT in several places. In this case, the saliency of the cues is a major driver (Pcb).<br>Location and physical quality of indications also addressed, but here it is discussed under Human Machine Interface. | ND |
| **Execution Complexity** | Explicitly considered by THERP tables based on specific types of complexity (e.g., long checklist v. short checklist, etc.); execution was not complex in this scenario. | N/P |
| **Training** | Explicitly credited in relevant CBDT trees. Not a factor in this analysis. | 0 |
| **Experience** | Not addressed by this method. | N/A |

| | | |
|---|---|---|
| **Procedural Guidance** | Evaluated explicitly as part of the guided qualitative analysis; both availability and quality of guidance appears as decision points in relevant CBDT trees. In Pce, the HEP was affected by poor layout of procedure. | ND |
| **Human- Machine Interface** | Indicator location and physical quality addressed in CBDT and found to be a negative driver (Pcb). Select HMI issues relevant to execution are also covered by THERP; these were not applicable to this HFE. | ND |
| **Work Processes** | Availability of check off provisions within the procedure is also assessed in the execution portion. That was not found to be a driver in this scenario except in that it kept the execution error probability low because checkoff was used. | N/P |
| **Communication** | Use of formal communications explicitly covered by CBDT in Pcc. | N/P |
| **Team Dynamics** | Not addressed by this method. | N/A |
| **Other** | None. | N/A |

*G4.3.4 Comparison of Drivers to Empirical Data*

There were several substantial drivers in the empirical data that were not predicted by the quantitative or qualitative analysis. Scenario complexity, indications of conditions and procedural guidance at the high level were all considered negative drivers for both the analysis and the empirical data. However, when the details are examined, the underlying failure mechanisms and the anticipated effect do not match. Details are given in the table below for these three factors. The qualitative analysis did note that the scenario was complex, the cues did not match the response and this combination could make the diagnosis difficult, which is indeed what happened.

The timing estimates from the HRA team also did not match the scenario well. Adequacy of time was identified as a negative driver for the empirical data, but it was considered nominal/positive in the HRA analysis (30 minute time margin). The team predicted it would take 10 minutes to diagnose the SGTR and 5 minutes to isolate the SG. Three out of four crews diagnosed the SGTR in 20-23 minutes, with one team taking 73 minutes to declare the SGTR (the delay was to perform activity sample to verify). Once the SGTR was diagnosed, it took one crew 6 minutes to isolate the SG, two crews took 50-60 minutes and the third crew only took 11 minutes (but they failed to secure the HHPs, resulting in a release). The discrepancy in timing likely stems from the fact that the HRA team misinterpreted the scenario and assumed the crew was out of FRH.1 when the SGTR occurred.

The qualitative analysis was not detailed enough to predict the impact of the drivers or predict specific failure mechanisms. The quantitative analysis (CBDT trees and THERP tables) predicted several drivers/failure mechanisms that were not supported by empirical data.

It is worth noting that the cognitive errors dominated the empirical data, but the HRA analysis showed 60/40 split between the cognitive and execution contributions.

| Factor | Empirical | HRA | Comments |
|---|---|---|---|
| **Stress** | 0 | ND | THERP automatically defines stress level based on a pre-determined set of criteria. In this case, the plant did not respond as expected (SGTR masking), so a stress factor was applied (x5 applied to execution). |
| **Scenario Complexity** | MND | ND | Plant response not as expected. For the HRA this was a negative driver for the execution portion. It did not show up as a driver for the cognitive portion in the CBDT trees. For the empirical results, the complexity arose from a combination of the masking and being in FRH1 without clear directions on how to address the SGTR (impacts cognition more than execution). |
| **Indications of Conditions** | ND | ND | The empirical data noted that masking was a negative driver. The qualitative assessment did note that masking might make diagnosis difficult; however, this did not show up in the quantitative analysis. In fact, the cue clarity was designated as "average" and in the CBDT trees they indicated the indications were available, accurate and all cues were as stated. The numerical driver in this case had to do with failure of attention due to the saliency of the cues (Pcb; monitor cue, not alarmed). There is no indication in the empirical data that failure of attention was a driver. |
| **Procedural Guidance** | ND | ND | The empirical data saw one team fail to stop the last HHSI pump due to a misunderstanding of what an "active loop" was. Two teams could not get to the actions procedurally through E-30 (isolated from knowledge) because they were held up in higher priority procedures. The qualitative analysis anticipated the crews to get to the actions through the procedures without issue once diagnosis was performed. However, procedural guidance was a negative driver for the cognitive portion of the qualitative analysis because the step in the procedure was not graphically distinct (Pce; skip a step in the procedures). |
| **Adequacy of Time** | ND | N/P | Three of the four groups did not make the 40 minute timeline (though two of those crews were ultimately successful in preventing release); the forth group barely made the time criterion. The HRA determined that time would be adequate, with a 30 minute time margin available. |
| **Human- Machine Interface** | N/P | ND | (see indications of conditions ->Pcb) |
| **Training/Experience** | N/P | 0 | Nominal/positive for the empirical data. It did not come up in the qualitative or quantitative analysis. |
| **Execution Complexity** | N/P | N/P | Nominal/positive in both. |

| | | | |
|---|---|---|---|
| **Work Processes** | 0 | N/P | Nominal in HRA; indeterminate in empirical analysis. |
| **Communication** | 0 | N/P | Nominal in HRA; indeterminate in empirical analysis. |
| **Time Pressure** | --- | n/a | |
| **Team Dynamics** | 0 | n/a | |

*G4.3.5 Comparison of Qualitative Analysis to Empirical Data*

The empirical data showed much variability in how the crews progressed through the scenario, however, the operational story in the qualitative analysis did not reflect this uncertainty, nor the potential conflict with FRH1/FRP1. This may be an interpretation of the scenario as when the scenario said that the SGTR would occur once AFW is re-established, the analysis team interpreted that to mean that AFW has been re-established, meeting the criteria to exit FRH1 back to the previous procedure in effect (E0 in this case). While it is not entirely clear from the quantitative analysis, it seems the timeline starts roughly the same time FRH1 is exited. This clearly does not match the scenario as the operators must fill a SG before they can exit FRH1. As we saw above, many of the drivers were also mismatched; overall the qualitative analysis did not match the empirical data well.

*G4.3.6 Impact on HEP*

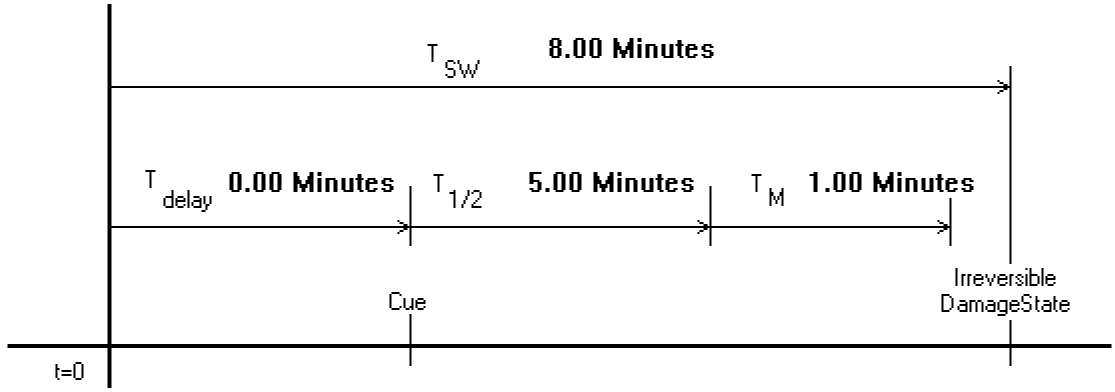
This team predicted an HEP of 0.01; given three out of four teams failed to maintain low SG pressure and isolate the SG in 40 minutes (and one of those teams failing to prevent release entirely), this HEP seems optimistic. This is expected given the mismatch between the empirical data and the qualitative analysis.

### G4.4 HFE 2A

*G4.4.1 Summary of Qualitative Findings*

This HFE represents the failure of the operators to trip the RCPs and establish charging flow to the RCP seals to prevent a seal LOCA. The initial cue for this action is the rising RCP seal temperatures. The operators are trained to monitor the RCPs and to trip the RCPs if necessary and establish seal cooling if the seals exhibit high temperatures or have lost cooling. At this time, the operators will start charging pump “B.” Two minutes later the pump will trip giving the operators alarms for seal temperature and flow. Increase seal temperatures or excessive leakoff flow will direct the operators to the RCP alarm response procedure (ARP) 0POP04-RC-0002 “Reactor Coolant Pump Off Normal.” The ARP directs the operators to the trip the affected RCP and re-establish cooling if necessary.

The timing used for this HFE was as follows:




This scenario is time limited (6 minute to diagnose and perform actions; 7-9 minutes available), so HCR/ORE was used to quantify the cognitive portion of this error. In this case, the cognitive error dominates the HEP. Insights from the operator confirm that the cognitive element dominates this scenario: it is described as a “complicated evolution” where the operators must recognize the RCP Seal LOCA is eminent based on the available indications and then respond to it while working in parallel with E0 or ES-01 (according to the operator insights, E0/ES-01 by itself will not get the operators to a success path to trip the RCP).

While not a driver, the execution portion was determined to have moderate stress due to high workload.

*G4.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP = 2.8E-01; EF = 1

| HFE Description | HCR/ORE | Execution Prob. | Total HEP |
|---|---|---|---|
| Operators fail to trip RCPs | 2.8E-01 | 2.6E-03 | 2.8E-01 |
| Operators fail to start the PDP to provide cooling to the RCP seals | n/a | 2.6E-03 | 2.6E-03 |
| **HFE 2-Total** | **2.8E-01** | **5.2E-03** | **2.8E-01** |

- **Fail to trip RCPs [2.8E-01]:**
  - Cognitive: 2.8E-01; HCR/ORE
    - PWR CP1 sigma value of 0.57 [when operators see RCP seal temperatures, they will respond].
  - Execution: 2.6E-3; THERP
    - Stress factor of 2 (high workload)
    - Error of omission
  - No recovery, no dependency
- **Fail to start the PDP to provide cooling to RCP seals [2.6E-3]:**
  - Execution: 2.6E-3; THERP
    - Stress factor of 2 (high workload)
    - Error of omission
  - No recovery, no dependency

*G4.4.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | In this method the adequacy of time is primarily manifested by which cognition method is chosen – CBDT or HCR/ORE. In this case HCR/ORE was chosen. Dependence levels for crediting recovery actions are based on time; no credit was given for cognitive recovery | MND |
| **Time Pressure** | | N/A |
| **Stress** | (THERP) Moderate because of high workload. Contributed to the execution portion, but the execution was a negligible contributor to the total HFE. | 0 |
| **Scenario Complexity** | Not explicitly addressed by HCR/ORE. It was a driver in the qualitative analysis, but it is not clear from the analysis if the timing was adjusted to account for the complexity. Effect cannot be determined. | 0 |
| **Indications of Conditions** | Addressed by the calculator, but not a contributor to this scenario. Not explicitly addressed by HCR/ORE. | 0 |
| **Execution Complexity** | Explicitly considered by THERP tables based on specific types of complexity (e.g., long checklist v. short checklist, etc.); execution was not complex in this scenario. | N/P |
| **Training** | Addressed by the calculator, but not a contributor to this scenario. Not explicitly addressed by HCR/ORE. However, was credited in the qualitative analysis by crediting a non-proceduralized action of tripping the RCP. | N/P |
| **Experience** | In HCR/ORE, this can be addressed by calculating a scenario specific sigma, but this analysis used the predetermined sigma. | N/P |
| **Procedural Guidance** | Addressed in CBDT, but not explicitly by HCR/ORE. | 0 |
| **Human- Machine Interface** | Select HMI issues relevant to execution are also covered by THERP; they were considered nominal in the THERP analysis. | N/P |
| **Work Processes** | Availability of check off provisions within the procedure is also assessed in the execution portion. Checkoff provisions were used. | N/P |
| **Communication** | Addressed in CBDT, but not explicitly by HCR/ORE. | 0 |
| **Team Dynamics** | How a team works through the procedures and how well they multitask is considered implicitly by building the timeline. In HCR/ORE, this can be addressed by calculating a scenario specific sigma, but this analysis used the predetermined sigma. | N/P |
| **Other** | Sigma value of HCR/ORE is based on the type of cue/action. | ND |

*G4.4.4 Comparison of Drivers to Empirical Data*

Of the failure contributors to the empirical results, only one contributor was predicted fully by the HRA: limited time. Scenario complexity was also predicted in the qualitative analysis, but it did not turn out to be an explicit contributor in the quantitative analysis. Scenario complexity and training/experience were the main drivers of the empirical data; training and experience was not seen as a negative factor in either the qualitative or quantitative HRA.  Stress was a predicted negative contributor (although only minor due to high workload) and it was observed in the simulation runs.  While stress was observed, the effect of that stress on the performance of the crew could not be determined.  However, in the empirical data, monitoring of the control boards and acknowledging alarms was less than adequate, this could be connected to the high workload that resulted from the loss of the distribution panels.  In the HRA, this excessive workload was covered by the stress category.  Overall, cognitive failure dominates the empirical data and also dominates this HEP.

| Factor | Empirical | HRA | Comments |
|---|---|---|---|
| **Scenario Complexity** | MND | 0 | The empirical analysis concluded that many things happening at the same time made it difficult to detect the priority items. The qualitative analysis of the HRA indicated this would be a "complicated evolution"; however, HCR/ORE dominated the HEP, so complexity was not an explicit numerical driver. |
| **Training/Experience** | MND | N/P | Crew is not trained on Loss of CCW and sealwater to the RCPs very often; when trained it is usually done in a LOOP scenario. The training and experience on the needed prioritization is what made this a negative driver. Training/experience was not mentioned as a factor in the qualitative HRA. |
| **Adequacy of Time** | ND | MND | While time was a driver in the empirical data, the conclusion was "if [the crew] had more training there would have been enough time." However, the empirical analysis determined procedural guidance as a negative driver based on the fact that "crews had problems reaching the critical actions in time". This was the main driver in the HRA because it drove the decision to use HCR/ORE and there was a very small time margin available. |
| **Work Processes** | ND | N/P | In the empirical data, monitoring of the control boards and acknowledging alarms was less than adequate. Also had some difficulties when POP4 read by the RO. |
| **Procedural Guidance** | ND | 0 | |
| **Indications of Conditions** | ND | 0 | |
| **Other** | n/a | ND | In the quantitative analysis the type of cue drives the CP value applied in the calculator. Cue type was not considered by the empirical analysis |
| **Human- Machine Interface** | N/P | N/P | |
| **Execution Complexity** | 0 | N/P | |
| **Team Dynamics** | 0 | N/P | |
| **Stress** | 0 | 0 | Stress was observed in the empirical data, but its impact was unclear. The HRA considered moderate stress due to high workload. This was only a negligible contributor to the HEP. |
| **Communication** | 0 | 0 | |
| **Time Pressure** | n/a | n/a | |

*G4.4.5 Comparison of Qualitative Analysis to Empirical Data*

In the HRA, the cue for action was considered to be an increase in the RCP seal temperature, which the operators are trained to recognize and understand the implication of the eminent seal failure. From the cue, it was predicted the crew would take 5 minutes to diagnose the loss of CCW and sealwater and then start the alarm response procedure 0POP4-RC-0002. From there it would take 1 minute to trip the RCPs and start the PDP, presumably per the procedure, although the operators are trained to trip the RCPs by knowledge as well. This would be done while performing E0/ES-01 in parallel. Workload in the scenario would be high, causing some stress and the evolution could be perceived as complicated by the crew.

The empirical data supports the process described by the qualitative HRA, but not the timeline. In the empirical data, three of the four crews did indeed use the 0POP4-RC-0002 procedure once the loss of CCW and sealwater was detected. However, these crews did not begin the procedure for 7-10 minutes after the loss of CCW and sealwater; 4-6 minutes of that time passed before the loss of CCW was detected. At that point, two of the three crews had already exceeded seal temperatures of 230 °F and one team had 30 seconds until that seal temperature was exceeded. No team was able to successfully use the alarm response procedure to trip the RCP and start the PDP in time. However, three crews did trip the RCP based on knowledge (less than 1-2 minutes from detection).

Overall the qualitative analysis matched the expected progression of the crew through the procedures, although the timeline was overly optimistic. Given such a small time margin in the HRA, that optimism was the difference between considering the action feasible vs. infeasible. While acknowledging high workload, tight timeline and complication of the scenario, the HRA underestimated the effect of these factors on elongating the timeline.

*G4.4.6 Impact on HEP*

This team predicted an HEP of 0.28. The empirical data shows that all four crews failed. This low predicted HEP reflects the expected difficulty of successful action, but is perhaps still optimistic.

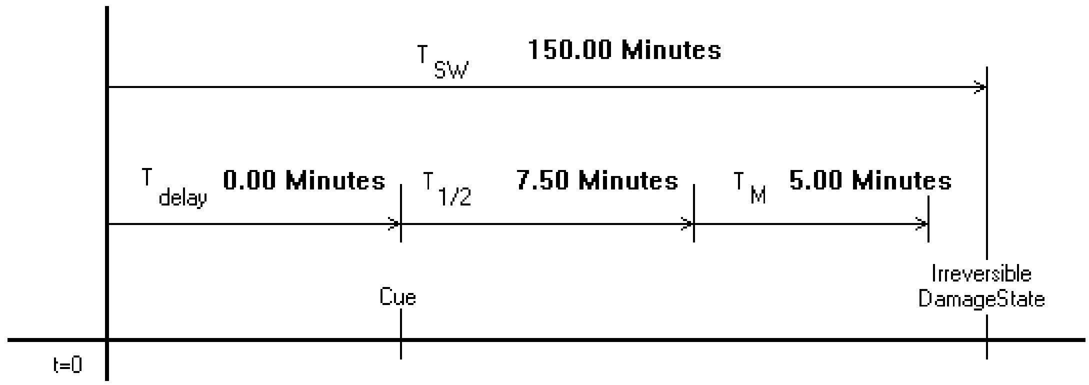
### G4.5 HFE 3A

*G4.6.1 Summary of Qualitative Findings*

This scenario is a design basis scenario which is well understood by the operators and trained on frequently (at least 6 times a year). The initial cue for this action is the rupture; SG pressure and level is increasing higher than the other steam generators. The secondary cue is the actuation of the radiation monitors. This is consistent with how the operators train; there are no difficulties expected with this scenario. Both the cognitive and execution steps were determined to be low complexity and all other PSFs optimal.

The total HEP is fairly low (5.7E-03) and it is equally split between cognitive and execution contributions. The dominant failure mechanism for the cognitive error is skipping a step in the procedure, because the procedure is not graphically distinct. Likewise, the dominant error mechanism for the execution steps were errors of omission associated with checklist provisions.

Timing for this scenario was as follows:




*G4.6.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP = 5.8E-03; EF = 5

| HFE Name | HFE Description | HCR/ORE | Execution Prob. | Total HEP |
|---|---|---|---|---|
| HFE 3-1 | Diagnose SGTR | 3.2E-03 | n/a | 3.2E-03 |
| HFE 3-2 | Identify Ruptured SG | n/a | 1.3E-03 | 1.3E-03 |
| HFE 3-3 | Isolate SG | n/a | 1.3E-03 | 1.3E-03 |
| **HFE 3-Total** | | **5.7E-03** | **2.6E-03** | **5.8E-03** |

- **3-1: Fail to Diagnose SGTR [3.2E-03]:**
  - Cognitive: 3.2E-03; CBDT
    - Skip a step in procedure (3.0E-03); step not graphically distinct
    - Failure of attention (1.5E-04); monitor parameter
  - Despite the long time frame, no credit was given for cognitive recovery
- **3-2: Fail to Identify Ruptured SG [1.3E-3]:**
  - Execution: 1.3E-3; THERP
    - Error of omission
    - PSFs optimal
  - No recovery, no dependency
- **3-3: Fail to Isolate Ruptured SG [1.3E-3]:**
  - Execution: 1.3E-3; THERP
    - Error of omission
    - PSFs optimal
  - No recovery, no dependency

*G4.6.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | In this method the adequacy of time is primarily manifested by which cognition method is chosen – CBDT or HCR/ORE. In this case there was plenty of time and CBDT was chosen. | N/P |
| **Time Pressure** | | N/A |

| | | |
|---|---|---|
| **Stress** | Low stress (THERP). | N/P |
| **Scenario Complexity** | Low complexity. However, this is not explicitly addressed by the method. | N/A |
| **Indications of Conditions** | While there were clear and expected indications of the conditions, in CBDT the HEP was reduced because the parameter was a "monitor" parameter instead of a "check" parameter. | ND |
| **Execution Complexity** | Explicitly considered by THERP tables based on specific types of complexity (e.g., long checklist v. short checklist, etc.); execution was not complex in this scenario. | N/P |
| **Training** | Addressed by the calculator, but not a contributor to this scenario. | N/P |
| **Experience** | Not considered in CBDT/THERP. | N/A |
| **Procedural Guidance** | Both quality and availability addressed in CBDT and THERP. The main mechanism both in CBDT and THERP for this HFE was skipping a step in the procedure or checklist. For CBDT, not having a graphically distinct step was the main numerical driver. | MND |
| **Human- Machine Interface** | Select HMI issues relevant to execution are also covered by THERP. Not applicable to this HFE. | N/A |
| **Work Processes** | Addressed by THERP, but not a contributor to this scenario. | N/P |
| **Communication** | Use of formal communications explicitly covered by CBDT in Pcc. | N/P |
| **Team Dynamics** | Not addressed by this method. | N/A |
| **Other** | None. | N/A |

*G4.6.4 Comparison of Drivers to Empirical Data*

The empirical data did not indicate any negative drivers for this scenario, but did note that positive drivers included indications of conditions, procedures, training and teamwork. This is fairly consistent with the qualitative HRA, which indicated, because this is a design basis scenario which is frequently trained on, there would be no substantial negative drivers. The HRA predicted that the team would be well equipped to successfully perform these actions. In the quantitative analysis the negative factors that dominated were associated with minor slips (error of omission in execution, skipping step in procedures).

*G4.6.5 Comparison of Qualitative Analysis to Empirical Data*

The qualitative analysis matched well the empirical data in that the ease of the scenario due to good training was adequately captured in the HRA.

*G4.6.6 Impact on HEP*

The HEP for this scenario was 5.7E-03; this is a fairly low HEP and corresponds reasonably to the empirical data which saw no failure out of three participating crews.

## G4.7 Comparison Summary

### *G4.7.1 Predictive Power*

The overall predictive power of the EPRI HRA Calculator method as employed by the NRC team in this study was judged to be fair. The final HEPs seem to be fairly consistent in trend, ranking and level of differentiation between HEPs. Overall, the assessment for the moderate-easy HFEs (1A, 3A) was much closer to the empirical data in terms of drivers, operational story and timing than the more difficult HFEs (1C, 2A). For the moderate/easy HFEs, the drivers and operational story ranked moderately good to good. However, the analysis of the difficult HFEs did only a fair job in describing the story, did a moderately poor job in matching the drivers and did a poor job matching the scenario timing. The table below provides a summary of how this analysis performed against the empirical data across scenarios.

| HFE | HEP Ranking | | HEP Values | | Consistency with Empirical Data* | | |
|---|---|---|---|---|---|---|---|
| | Empirical | HRA | Empirical | HRA | Drivers | Op Story | Timing |
| **2A** | 1 | 1 | 4/4<br>Very difficult | 0.28 | 2.5 | 3 | 1 |
| **1C** | 2 | 2 | 3/4<br>Difficult | 0.01 | 2 | 2 | 1 |
| **1A** | 3 | 2 | 0/4<br>Fairly Difficult | 0.01 | 4 | 5 | 4 |
| **3A** | 4 | 4 | 0/4<br>Easy | 0.0058 | 5 | 5 | 4 |
| * 1= poor; 2= moderately poor; 3=fair; 4=moderately good; 5=good | | | | | | | |

#### G4.7.1.1 Qualitative Predictive Power in Terms of Drivers

The EPRI HRA Calculator uses three distinct methods: HCR/ORE, CBDT and THERP. Cognitive steps are quantified using the higher value between HCR/ORE and CBDT, while THERP quantifies errors during execution. The HCR/ORE method uses a time reliability correlation, so under normal applications the only quantitative driver that can be distinguished is timing. For the CBDT method, however, identification of important drivers is guided by evaluation of the factors addressed in a set of decision trees with recovery factors applied at the end. THERP is based on a set of specific criteria for different types of actions. In CBDT and THERP, negative conditions lead to higher HEPs, with the factors that have the greatest negative effect on the resulting HEP being the main drivers. The main contributors to the numerical value do not always correspond with the factors identified in the operation story; some drivers (primarily scenario complexity) were identified in the operational story, but did not come up as a numerical driver when the method was applied. For HCR/ORE, the distinction between perceived drivers in the operational story and numerical drivers is not as clear because it is possible to extend the timing estimates to account for factors such as scenario complexity. Derivation of timing estimates used in this analysis was not documented in sufficient detail to determine what factors were considered in the timing. It did not appear that the questions addressed in the method always guided the analysts to address the critical aspects of the scenario that ended up affecting actual crew performance. For example, in HFE 2A, training was a main negative driver because there was a mismatch between their training and the scenario (the scenario required them to prioritize in a way their training had not prepared them to); training is addressed in CBDT, but only when the indications do not match the scenario, not, in this case, where the training does not match the scenario in critical ways.

For the challenging scenarios which saw failures in the empirical data (1C and 2A) there was much similarity in how well the HRA matched the empirical data (i.e., generally there were some mismatches). In both cases some drivers were correctly identified; of these, some major drivers were identified in the operational story, but did not translate to a numerical driver. Also in both cases the operational story only matched the empirical data fair at best, being very optimistic in the timing. It seems in both cases the operational story identified that the scenario was complex, but this complexity did not translate into the quantitative assessment either through timing adjustment or the CBDT trees (the latter only applicable in 1C). It is worth noting that for scenario 1C, there was a misinterpretation of the scenario, so the scenario analyzed did not match entirely the scenario run in the simulator; however, the conclusions regarding the method's ability to translate scenario complexity into a quantitative driver is still valid.

HFE 1A matched at a high level the right categories for drivers, but did not capture all the major mechanisms. HFE 3A had no strong negative drivers in either the HRA or the empirical data; the positive drivers were captured by the HRA. Overall, though, this method did a fair to moderately good job in predicting the overall drivers.

#### G4.7.1.2 Qualitative Predictive Power in Terms of Operational Expressions

While the overall qualitative predicative power of the high level risk drivers was moderately good to fair, the analysts' ability to predict specific failure mechanisms was only moderately poor to fair. For the cognitively difficult HFEs (HFE 1C, 2A) the failure mechanisms associated with scenario complexity (e.g., not working procedures appropriately in parallel or not completing necessary actions based on knowledge) were not predicted at all. For HFE 1C the cognitive errors dominated, but the resulting HEP, while appropriately high, showed 60/40 split between the cognitive and execution contributions. HFE 2A was quantified using HCR/ORE; HCR/ORE does not provide insights into failure mechanism, but rather provides a correlation of HEP given task timing information. In this case timing was an important driver of the empirical data. The operational story in this case did not provide sufficient detail to predict specific failure mechanisms.

HFE 1A was the best in terms of predicting specific failure mechanisms. Three failure mechanisms were identified and two of them were manifest in the empirical data, although did not lead to actual failures. "Information misleading" was identified by CBDT and the empirical data noted that diagnosis was delayed due to the misleading information; this was not a failure because plenty of time was available. "Skip a Step in the Procedure" was identified by CBDT as a cognitive failure mechanism and "Omission of item on long list" was identified by THERP as an execution failure mechanism. In reality, one crew missed stopping the RCP in the procedure. This error, however, was recovered; recovery was not credited in the analysis. For

HFE 3A no failure mechanism were observed in the empirical data, this matches the prediction given the small sample size and low predicted HEP.

G4.7.1.3 Quantitative Predictive Power

Overall, with respect to rank and level of differentiation, the quantitative results did a moderately good job in matching the empirical data in the extreme cases (very difficult and easy), but only a fair job in matching cases with moderate difficulty. The most difficult HFE (2A) saw 4/4 crews fail; the predicted HEP was 0.28. The easiest HFE (3A) saw 0 crew failures; the predicted HEP was 5.8E-3. The two middle HFEs (1C and 1A, characterized as "Difficult" and "Fairly Difficult") saw 3/4 crew failures and 0 crew failures, respectively; both received a HEP of 0.01. It is worth noting that this analysis did not credit recovery for any actions even when this credit was allowed by the method. If recover was credited, this would have reduced HFE 1A to below HFE 1C as expected by the results.

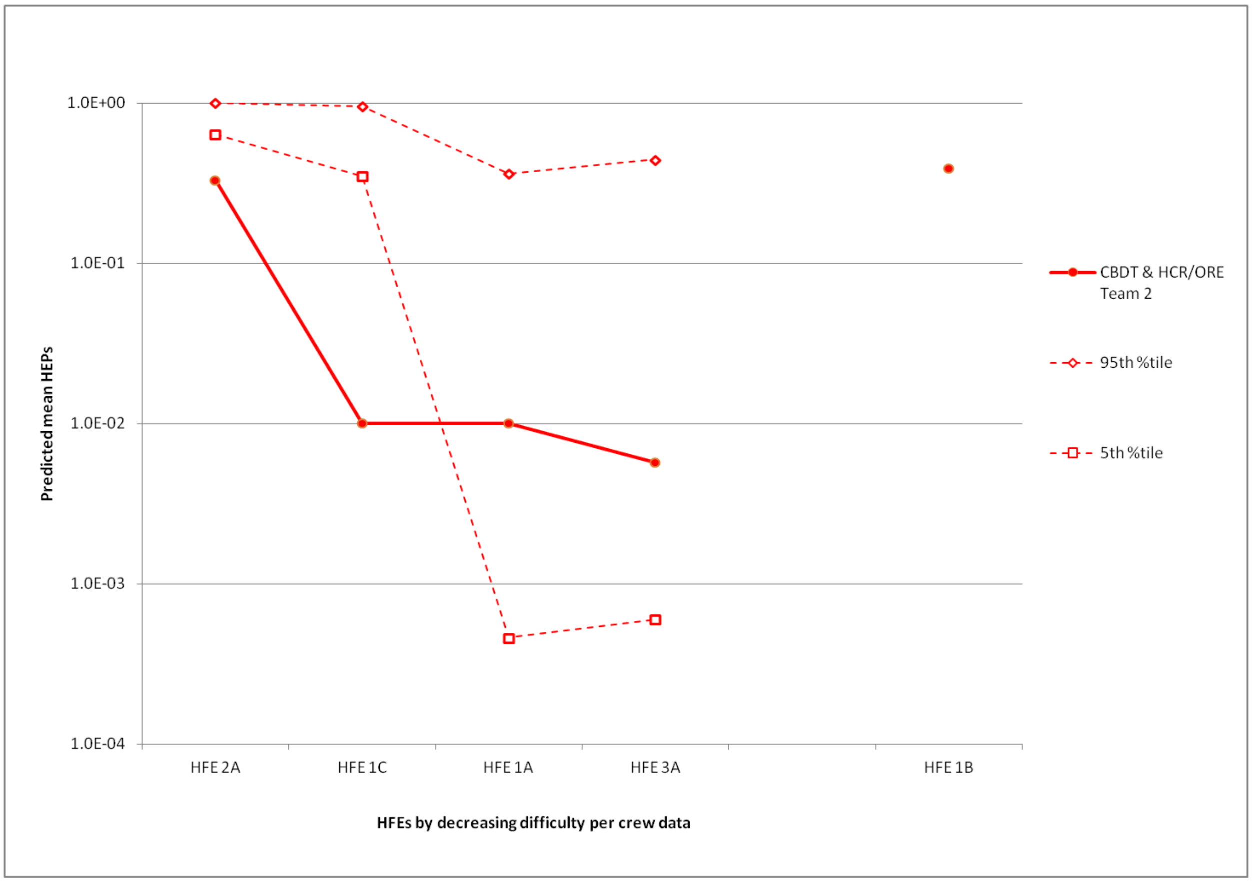



Figure 0.4 Predicted Mean HEPs with Uncertainty Bounds by the NRC HRA Team with HRA Calculator

*G4.7.2 Assessment of Guidance and Traceability*

The derivation of the HEPs within the method and what is important to performance is very traceable for CBDT and THERP. How the various factors are weighted in determining the final HEP can be determined by examining the contribution of various factors from the decision trees or THERP tables. The ability to trace the basis for the judgments regarding the branch points in the trees, choice of THERP table or choice of sigma value for HCR/ORE relies on the analysts' documentation.

The area of least traceability, which produced much variability in the end HEPs, is in developing the operational story and the timing analysis. There is no substantial guidance on how to perform operator interviews. While the timing elements are well defined, how those times are obtained and how they are interpreted (e.g., what is a "cue") is not clear in the guidance. In this analysis, the operational stories were not detailed enough to understand how the timing was derived. For application of HCR/ORE, this lack of traceability for the timing analysis translates directly to a lack of traceability in the end HEP.

Every method in Calculator is based on applying a generic set of rules to the scenario for quantification. While this makes the method quite traceable, as we saw in the data, sometimes factors that are identified in the qualitative analysis are not reflected in the method. Furthermore, there is little guidance provided for how to break up an HFE into subcomponents that need to be quantified separately.

*G4.7.3 Insights for Error Reduction*

For certain types of errors, this method can be powerful in providing insights for error reduction. There are generally two categories of error reduction: 1) improving human factors to reduce errors (e.g., improve location and clarity of indications, improve format of procedures, etc.) and 2) reducing the difficulty of cognitively challenging errors (e.g., improved training, clearer procedures, etc.). This method clearly provides insight into the former category of error reduction when CBDT and THERP are used. This method however provides a more limited set of insights for reducing error due to complex scenarios. The primary source of insights for error reduction of these scenarios comes from the qualitative analysis, which is supposed to include operator interviews and/or simulator observations. When HCR/ORE is used, the only insight to error reduction that can be gleaned is to find a way (either through training, revised procedures, etc.) to improve the timeline. Insights for error reduction were judged to be fair in this application of the EPRI HRA Calculator.

One factor seen in the empirical data that reduced the failure rate was the ability to recover from errors. This method credits different types of recovery and will point to the most effective (least dependent) recovery options, but the analysis performed by this team did not look at recovery in any of the scenarios.

*G4.7.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses*

The biggest strength of this method is that the judgments made by the analysts in applying the Calculator and obtaining the HEPs are clearly traceable in the sense that each method in the Calculator is based on a set of clear rules to be applied based on well-defined criteria. Overall the application of this method produced fair results. The quantitative predictions were moderately good, but the qualitative predictions were only fair. The method performed best overall (identifying risk drivers, identifying failure mechanisms and quantification) for easy HFEs; HFEs that presented high cognitive complexity, particularly when operators had to juggle multiple procedures and perform knowledge based actions (i.e., HFE 1C and 2A), did not compare as well with this method.

Consistently across the scenarios, the biggest discrepancy between the analysis and the empirical data was the timing information. While there is clear guidance on the definition of each timing point (i.e., definitions exist for t_delay, t_m, t_1/2 and t_sw), there is not sufficient guidance on how to assess those values, particularly on how to use the operational interviews to obtain those values (e.g., t1/2). It seems in this analysis, the time used was based on the time it would take to get the cues and get through the steps, without accounting for the elongation of the time frame due to distractions, parallel actions, etc.

The method generally seemed to provide a reasonable set of influencing factors to address in order to predict crew performance and provide insights for error reduction. However, there were aspects of the scenario that affected crew performance that were not detected based on the use of the influencing factors described in the method and the associated guidance. This seems to be particularly true for cases where there is moderate to high complexity paired with moderate to low time availability. High complexity is covered well by the CDBT method in the various tree branches, where time-limited scenarios are covered by HCR/ORE. Scenarios where both are a driving factor are not well accounted for because the analyst uses the highest HEP produced between the two methods.

## G.5 <u>HRA Calculator (EPRI)</u>

### G5.1 HFE 1A

*G5.1.1 Summary of Qualitative Findings*

The Scientech Calculator team broke down HFE 1A into 2 failure modes: A1) Fail to enter FRH1; and A2) Fail to establish F&B within 45 minutes.

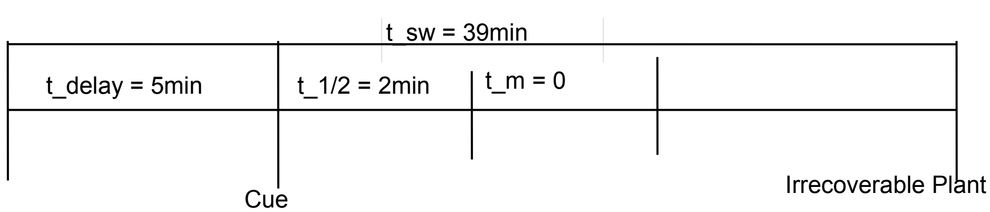
The Scientech Calculator team noted that the driver for this scenario would be failure to enter FR-H1 (A1), a cognitive failure, due to the poor cues and indications. To this point, two underlying assumptions are key: 1) operators trust their cues and indications; and 2) operators will follow their procedures.

Two cues will lead to the team to FRH1 (0POP05-EO-F003): 1) NR Level in at Least One SG Greater Than 14% <u>AND</u> 2) Total AFW Flow to SGs Less Than 576 GPM. The STA will notice the first cue when they enter CR and begin monitoring CSFSTs. For the second cue, the crew will notice the AFWST level is not decreasing [confirmed by operator interview; also operators have training 2-3 times a year on failure of a recirc valve and are familiar with the scenario]. This will lead to diagnosis of failed AFW indicators which through operator training and redundant steps in the procedure, will lead to realization of no AFW flow. Relevant procedures are ES-01 and F003. Once that is diagnosed they will go directly to FRH1 and start F&B.

For this scenario, the team decided the following:

- The degree of clarity for this second cue is determined to be poor.
- Time pressure will be low (long time frame; no training on time requirement)
- Stress will be high (plant response not as expected)
- Workload will be high (multiple procedures to restore AFW/MFW/condensate) such that "operators could be distracted and neglect the cues to implement F&B". Note: once F&B is considered necessary, they will drop all other procedures and focus on F&B, so the workload for execution portion becomes low.
- Cognitive complexity is high. (Execution complexity N/A)

Timing is as follows for A1: Note that the time window is 39 minutes instead of 45 minutes because 6 minutes are required for execution, so the time for diagnosis (which is all A1 is concerned with) is limited to 39 minutes.




*G5.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP = P(A1)+P(A2) = 3.2E-2, EF= 5

- **A1: Fail to enter FRH1 [3.1E-02]:**
    - Cognitive: 3.1E-2; CBDTM used (HRA team used the maximum HEP obtained from applying CBDTM and HCR/ORE. The large time window resulted in a negligible HEP from HCR/ORE.)
        - Recovery available for each failure mechanism (via either self review or extra crew) reducing the HEP from 7.8E-02 to 3.1E-02
        - Dependency low to no dependency due to extra crew and long time frame for self-review
    - Execution: N/A

- **A2: Fail to establish F&B within 45 min [1.3E-03]:**
    - Given successful entry in FRH1
    - Cognition: 1.6E-4; CBDTM used (HRA team used the maximum HEP obtained from applying CBDTM and HCR/ORE. The large time window resulted in a negligible HEP from HCR/ORE.)
        - Recovery available via self-review for failure mechanism (skip a step in procedure) reducing the HEP from 3.08E-03 to 1.6E-04
        - Dependency for recovery is low because if step 2 is skipped, step 9 will direct them again to check the level of the SG.
    - Execution: 1.1E-3; THERP
        - 2 execution steps each with a verification step. Recovery via verification reduces execution HEP from 2.0E-02 to 1.1E-03
        - “High” stress (x5) due to plant not behaving as expected.
        - Dependency for recovery is low because of long recovery time

Four failure mechanisms were identified by the method as drivers in A1 (table below). A2 did not contribute significantly to the total HEP, and so these drivers are not included in this discussion.

| Pc Failure Mechanism | Branch | HEP |
|---|---|---|
| **$Pc_a$: Availability of Information**<br>The indication for AFW flow will be incorrect and there is no warning or annuciator response to indicate the misaligned recirc valve. However, the operators are trained 2-3 times per year to deal with mis-positioned recirc valves. | **e** | **5.0E-02** |
| **$Pc_b$: Failure of Attention**<br>High workload since the operators could be working in multiple attachments of ES-01. A red path on the CSFSTs is not alarmed from the control room. | m | **1.5E-02** |
| $Pc_c$: Misread/miscommunicate data | a | neg. |
| **$Pc_d$: Information misleading**<br>Cues for AFW flow are not as stated. A flow rate will be showing in the control room however, there will be no flow going to SGs; AFWST not decreasing. No procedure warning for misalighed recirc valve, but operators are trained. | **c** | **1.0E-02** |

| | | |
|---|---|---|
| **$Pc_e$: Skip a step in procedure**<br>Step not graphically distinct. Operators will only be working in ES-01 | **c** | **3.0E-03** |
| $Pc_f$: Misinterpret instruction | a | neg. |
| $Pc_g$: Misinterpret decision logic | k | neg. |
| $Pc_h$: Deliberate violation | a | neg. |

Recovery mechanisms with low to no dependency were found for each of the four failure mechanisms identified; with 32 minutes available for recovery, these mechanisms were credited:

- Pca and Pcd: Extra crew credited because SG level is a parameter which is monitored by all crew members. (No Dependency)
- Pcb and Pce: Self review credited since the STA is constantly looping through the CSFSTs; over 35 minutes the trees would be looped through twice. (Low Dependency)

*G5.1.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | More than adequate time. In this method the adequacy of time is primarily manifested by which cognition method is chosen – CBDT or HCR/ORE. In this case CBDT was chosen because it is not time limited. Dependence levels for crediting recovery actions are based on time. | N/P |
| **Time Pressure** | Determined in qualitative analysis that there is no time pressure because of the long time frame and the fact that crews do not have a training requirement to complete FRH1 in a set amount of time. | N/A |
| **Stress** | (THERP) High because plant does not respond as expected. While this was identified as important, this is an execution PSF; the driver for the scenario is the diagnostic failure. | ND |
| **Scenario Complexity** | Components of Cognitive Complexity explicitly addressed by CBDT. Scenario complexity also explicitly addressed in THERP for execution; in this case in the components that make us apply a stress factor (x5). | MND |
| **Indications of Conditions** | Availability and clarity of indications with respect to the scenario are explicitly considered by CBDT in Pca (Availability of Information) and Pcd (Information Misleading). In this case these two failure mechanisms were the top drivers. In this case this is the same as scenario complexity.<br>Location and physical quality of indications also addressed, but here it is discussed under Human Machine Interface. | MND |
| **Execution Complexity** | Explicitly considered by THERP tables based on specific types of complexity (e.g., long checklist v. short checklist, etc.); execution was not complex in this scenario. | N/P |

| | | |
|---|---|---|
| **Training** | Evaluated explicitly as part of the guided qualitative analysis and also appears as decision point in relevant trees. Both specific and general training is addressed. In this case training was credited in several trees. In Pcd HEP was strongly moderated by this compared to the branches without training. | N/P |
| **Experience** | Not addressed by this method. | N/A |
| **Procedural Guidance** | Evaluated explicitly as part of the guided qualitative analysis; both availability and quality of guidance appears as decision points in relevant CBDT trees. In Pca HEP was strongly affected by lack of procedural guidance about alternative sources of information when an indicator is believed to be unreliable. | MND |
| **Human- Machine Interface** | Indicator location and physical quality addressed in CBDT. Select HMI issues relevant to execution are also covered by THERP. Not applicable to this HFE. | N/P |
| **Work Processes** | Availability of check off provisions within the procedure is also assessed in the execution portion. Check-off provisions were used. | N/P |
| **Communication** | Use of formal communications explicitly covered by CBDT in Pcc. | N/P |
| **Team Dynamics** | Considered only implicitly in qualitative analysis in building the timeline. | N/A |
| **Other** | Workload was considered in Pcb and assessed to be High in this scenario, and thus a contributor to the cognitive error. No other driving PSFs were identified<br>For execution, THERP considers environment and workload as well; neither was a contributor in this case; workload was not considered to be high during the execution step. | ND |

*G5.1.4 Comparison of Drivers to Empirical Data*

At the high level, the drivers indicated by the team match the empirical data quite well in predicting the specific performance issues. The main drivers were correctly identified; however, two additional drivers (only one of which had a substantial impact on the HEP) were identified by the HRA that did not appear to be drivers in the empirical analysis. In this case, the driver that was identified by the HRA which had an impact on the HEP (workload) was identified by the operator as a negative factor during the interview. See table below for more detail.

The empirical data yielded no failures, so specific failure mechanisms cannot be compared here. However, one failure mechanism was predicted by the quantitative analysis that manifested itself in the empirical data: one crew skipped a step (failed to stop RCPs before starting F&B) but recovered from it. In this HRA, skipping a step in the procedure was a failure mechanism identified (~1% contributor); recovery was also determined to be likely for this mechanism, as we saw manifest in the empirical data. The failure mechanism expected to dominate was failure to enter FRH1 due to the poor cues. Nothing in the empirical data supports this specific mechanism.

| Factor | Empirical | HRA | Comments |
|---|---|---|---|
| **Stress** | 0 | ND | This may be a definition difference. THERP automatically defines stress level based on a pre-determined set of criteria. In this case, the plant did not respond as expected (recirc valve open), so a stress factor was applied, however, this had a fairly negligible contribution to the total HEP (~3% of the total HEP value). It was not identified as important in the qualitative analysis either. |
| **Other (Workload)** | N/A | ND | Workload was not identified as a driver of the empirical results. It was identified in the qualitative analysis as a negative factor (though not a driver) based on operator insights. It contributes ~38% to the total HEP. |
| **Scenario Complexity** | ND | MND | The main driver for HRA scenario was the combination of the fact that there was no clear indication of the loss of AFW and there was a mismatch between the plant and the procedures (yellow path vs. decreasing SG levels). These are the same drivers identified by the empirical team. The difference is the designation of "main" ND in the HRA analysis – these were considered "main" because they collectively provided the largest numerical contribution to the resulting HEP and because it was determined to be the primary negative factors in the qualitative analysis. |
| **Indications of Conditions** | ND | MND | |
| **Procedural Guidance** | ND | MND | |
| **Training/Experience** | N/P | N/P | For the qualitative analysis, the operator insights from this team picked up that this was a strong positive driver: because the recirc issue has come up in operation, it is now trained on. Crediting training had a substantial effect on the total HEP. The empirical team also singled this out as a strong positive driver. |
| **Adequacy of Time** | N/P | N/P | Both acknowledge there was plenty of time. This was a particularly positive driver in the HRA because it allowed recovery to be credited. |
| **Execution Complexity** | N/P | N/P | Nominal in both cases. |
| **Human- Machine Interface** | N/P | N/P | Nominal in both cases |
| **Work Processes** | N/A | N/P | Nominal in HRA; indeterminate in empirical analysis. |
| **Communication** | N/A | N/P | Nominal in HRA; indeterminate in empirical analysis. |
| **Time Pressure** | ---- | N/A | |
| **Team Dynamics** | N/A | N/A | |

*G5.1.5 Comparison of Qualitative Analysis to Empirical Data*

A majority of the comparison of the qualitative analysis to the data was presented in the previous section regarding the scenario drivers. The operational story in the qualitative analysis was consistent with the actual data – the crew figured out that AFW was lost (as expected) and that led them to FRH1 where F&B was started without difficulty. The timing of the operational story from the HRA was on the optimistic side; timing data for the HRA was based on the operator interview. The HRA estimated 7 minutes (after reactor trip) to reach FRH1, 1 minute to diagnose need for F&B and begin actuating SI, with execution complete at 13 minutes. It took the crews 8-13 minutes to diagnose and enter FRH1, 1.5-3.5 minutes to begin SI actuation, establishing F&B in 10-17 minutes. These timing differences have no impact to the quality of the analysis as the time available was large. The qualitative analysis also pointed out there were several points to recover if the initial cues were missed.

The qualitative analysis did mention that the workload would be high prior to entry to FRH1 because the crew will be busy trying to check and restore MFW/AFW. The empirical data confirms that the crews wanted to verify they didn't have AFW to SG B, and at least 3 crews called a PO to check the valve. However, in the empirical data, these actions do not seem to provide sufficient workload to be a hinderance.

*G5.1.6 Impact on HEP*

Impact on the HEP is discussed in detail in the comparison with drivers section. Overall, the HEP for this failure was 3.2E-2; this is a mid-range HEP, indicating that the operators are generally well able to find a success path, but some difficulties exist. This seems consistent with the empirical data, where all the crews were able to be successful, but one crew did not know that the recirculation valve was open and another crew skipped a step in the procedure which they were able to recover from.

## G5.2 HFE 1B

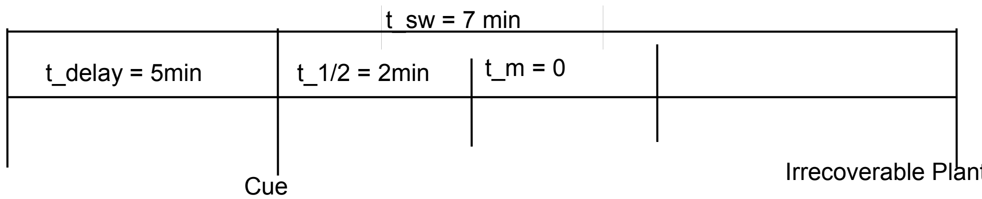
*G5.2.1 Summary of Qualitative Findings*

Differences between this HFE and 1A are all reflected in differences in timing, so the qualitative analysis is the same, except for the timing. The Scientech Calculator team broke down HFE 1B into the same 2 failure modes as 1A: B1) Fail to enter FRH1 and B2) Fail to establish F&B within 13 min. Unlike 1A, because of the timing constraints, B1 and B2 are equal contributors.

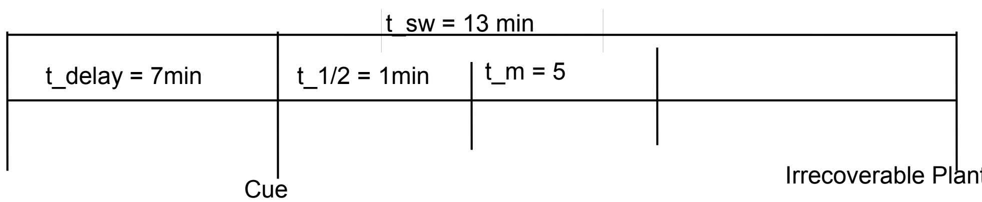
Two dominating factors were identified for this HFE: 1) timing and 2) impact on cues and indications. For this scenario, the team decided the following:

- The degree of clarity for this second cue is determined to be poor.
- Stress will be high (plant response not as expected)
- Workload will be high (multiple procedures to restore AFW/MFW/condensate) such that "operators could be distracted and neglect the cues to implement F&B". Note: once F&B is considered necessary, they will drop all other procedures and focus on F&B, so the workload for execution portion becomes low.
- Cognitive complexity is high. (Execution complexity N/A)
- Time pressure will be low because the crews are not aware of this time critical action. They do not train to reach F&B criteria with in a specific time. Therefore, the operators will not be rushing through the procedures to meet the PRA defined success criteria. The procedures as written, do lead the operators to a success path if the there is more time available.

Timing is as follows for B1: Note that the time window is 7 minutes instead of 13 minutes because 6 minutes are required for execution, so the time for diagnosis (which is all B1 is concerned with) is limited to 7 minutes.




Timing for B2: t_delay is based on 5 minutes required for the STA to enter the room and recognize the storage tank levels are not dropping and an additional 2 minutes to get to FRH1 (determine the AFW pump is on recirc and have a team meeting to transition).




The cues for this HFE are not immediately apparent and can be considered poor because although the AFW flow reading will indicate flow, due to a misaligned recirc valve, no water will be reaching the SGs. Based on operator interviews, the operators stated that they would look at CST tank level as part of normal monitoring of critical safety functions. The NR level is providing a correcting reading.

### *G5.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP = 7.5E-1, EF= 1

For quantification, the **Max** of CBDTM and HCR/ORE were used. Because of the short time window the HRC/ORE dominated in both cases.

| HFE Name | HFE Description | CBDTM | HCR/ORE | Execution Prob. | Total HEP | Error Factor |
|---|---|---|---|---|---|---|
| HFE 1B-1 | Operators fail to enter FRH1 | 7.8E-2 | 5.0E-1 | N/A this HFE models cognition only | 5.0E-1 | 1 |
| HFE 1B-2 | Operators fail to establish F&B within 13 mins. | 3.0E-3 | 5.0E-1 | 1.1E-3 | 5.0E-1 | 1 |
| **HFE 1B-Total** | | **8.1E-2** | **7.5E-1** | **1.1E-3** | **7.5E-1** | **1** |

- **B1: Fail to enter FRH1  [5.0E-1]:**
    - Cognitive:  5.0E-01; HCR/ORE
        - PWR CP1 sigma value of 0.57
    - Execution: N/A
    - No recovery

- **B2: Fail to establish F&B within 13 min  [5.0E-1]:**
    - Given successful entry in FRH1
    - Cognition: 5.0E-01; HCR/ORE
        - PWR CP1 sigma value of 0.57
        - No recovery; insufficient time.
    - Execution: 1.1E-3; THERP
        - 2 execution steps each with a verification step, and each with a stress factor of 5 applied.  Recovery via verification reduces execution HEP from 2.0E-02 to 1.1E-03
        - Dependency for recovery is low because of long recovery time

CP1 is the category for the "DO immediately" actions.  If Wide range SG level is less than 50% the operators will perform F&B. They will not need to monitor the cues for F&B while in FRH1 because these conditions will already exist before the operators enter FRH1.

*G5.2.3  Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | In this method the adequacy of time is primarily manifested by which cognition method is chosen – CBDT or HCR/ORE.  In this case HCR/ORE was chosen. Dependence levels for crediting recovery actions are based on time; no credit was given for cognitive recovery | MND |
| **Time Pressure** | | N/A |
| **Stress** | (THERP) High because plant does not respond as expected. Contributed to the execution portion, but the execution was a negligible contributor to the total HFE. | 0 |
| **Scenario Complexity** | Scenario (cognitive) complexity was determined by the qualitative analysis to be high, however, because HCR/ORE was the dominant value, scenario complexity does not have a numerical effect. | 0 |
| **Indications of Conditions** | Addressed in CBDT, but not used in HCR/ORE. | 0 |
| **Execution Complexity** | Explicitly considered by THERP tables based on specific types of complexity (e.g., long checklist v. short checklist, etc.); execution was not complex in this scenario. | N/P |
| **Training** | Addressed in CBDT, but not used in HCR/ORE. | 0 |
| **Experience** | In HCR/ORE, this can be addressed by calculating a scenario specific sigma, but this analysis used the predetermined sigma. | N/P |
| **Procedural Guidance** | Addressed in CBDT, but not used in HCR/ORE. | 0 |

| | | |
|---|---|---|
| **Human- Machine Interface** | Select HMI issues relevant to execution are also covered by THERP. Not applicable to this HFE. | N/P |
| **Work Processes** | Availability of check off provisions within the procedure is also assessed in the execution portion. Check-off provisions were used. | N/P |
| **Communication** | Addressed in CBDT, but not used in HCR/ORE. | 0 |
| **Team Dynamics** | Considered implicitly in building the timeline. HCR/ORE considers all PIF (except time and cue type) to be nominal. | N/P |
| **Other** | Sigma value of HCR/ORE is based on the type of cue/action. | MND |

*G5.2.4 Comparison of Drivers to Empirical Data*
Empirical data not available for this scenario.

*G5.2.5 Comparison of Qualitative Analysis to Empirical Data*
Empirical data not available for this scenario.

*G5.2.6 Impact on HEP*
Empirical data not available for this scenario.

## G5.3 HFE 1C

*G5.3.1 Summary of Qualitative Findings*

In this scenario, the operators have successfully established F&B and re-established AFW flow to several of the SGs, though when the flow reaches the first SG, a SGTR occurs. The HFE is: operator fails to cooldown and depressurize following SGTR and successful F&B. A successful outcome for this HFE is contingent on the operators diagnosing the SGTR, isolating the ruptured SG and performing the necessary cooldown and depressurization within 40 minutes.

This scenario starts at step 14 in procedure FRH1. FRH1 is entered on a red path on the critical safety function status trees. Anytime a red path is obtained the procedure entered (in this case, FRH1) cannot be exited until the proceduralized exit conditions exist.[7]

With the AFW flow feeding the SG, this will disguise the SGTR for a period of time. It isn't until the operators reach the RCS subcooling steps (28 & 29) of FRH1 that they get the appropriate cues to diagnose a SGTR. During these two steps, the operators are attempting to reduce the number of charging pumps and HHSI pumps running to exit FRH1, but with the SGTR, the operators will find themselves caught in a loop since the AFW flow will not be an adequate source to maintain a low enough temperature and exit FRH1.

At this point the operators may realize they have a SGTR, however, their training will not allow them to leave FRH1 without approval from a higher position (potentially cause a 50.54 violation). Once they have this permission, they will exit FRH1 and enter E-30. However, obtaining management permission to exit FRH1 was estimated to take 30

[7] While they cannot exit FRH1 until they meet the exit criteria, the interviewee states that (per ZA18) it is possible to enter another procedure in parallel as long as it does not interfere with FRH1. This analysis does not consider this possible parallel action, however, the interview suggests that there are steps in dealing with the tube rupture that might conflict with FRH1, precluding parallel action.

minutes. The execution for cooldown and depressurization is expected to take on the order of 20 minutes based on design basis SGTR training requirements. Therefore the operators would not have enough time to complete this action within the scenario defined time window of 40 minutes.

In this case it is the crew's training and the belief in the adequacy of their procedures that will cause them to fail. Perhaps a crew with less discipline would abandon FRH1 and skip to E-30. But based on the operator interviews, it was clear the operators would repeat RCS subcooling until they were given clearance through the proper channels to exit FRH1 and address the SGTR.

If it wasn't for the procedures, the Scientech team believes that operators would have a high success rate. The cues, although disguised initially, would appear before it was too late to perform the critical steps to isolate, cooldown and depressurize the SG.

The remaining PSFs are considered optimal including environmental conditions and the complexity of what is required. The operators undergo regular training to diagnose the address the ruptured SG.

*G5.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP = 1.0

The underlying assumption behind this method, confirmed by operator interviews, is that operators will trust and follow their procedures. Based on this assumption, the timing analysis indicates that this action is not feasible. Therefore, this HEP was assessed as a 1.0 due to the fact that the operators will not leave the procedure they are in until they have permission and this will not occur within the 40 minute time window.

*G5.3.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | This HEP was set to 1.0 because it was considered infeasible due to inadequate time given how they use their procedures. | ND |
| **Time Pressure** | | N/A |
| **Stress** | PSFs are not applicable to this scenario. | 0 |
| **Scenario Complexity** | PSFs are not applicable to this scenario. | 0 |
| **Indications of Conditions** | PSFs are not applicable to this scenario. | 0 |
| **Execution Complexity** | PSFs are not applicable to this scenario. | 0 |
| **Training** | PSFs are not applicable to this scenario. | 0 |
| **Experience** | PSFs are not applicable to this scenario. | 0 |
| **Procedural Guidance** | If it wasn't for the procedures, the Scientech team believes that operators would have a high success rate. However, the | MND |

| | | |
|---|---|---|
| | FRH1 scenario takes priority and they cannot get out of that in the 40 minute time frame. | |
| **Human- Machine Interface** | PSFs are not applicable to this scenario. | 0 |
| **Work Processes** | PSFs are not applicable to this scenario. | 0 |
| **Communication** | PSFs are not applicable to this scenario. | 0 |
| **Team Dynamics** | PSFs are not applicable to this scenario. | 0 |
| **Other** | | N/A |

*G5.3.4 Comparison of Drivers to Empirical Data*

While the categorization of drivers did not necessarily match (i.e., "scenario complexity" was identified as the MND in the empirical data, but "procedural guidance" was the MND in this assessment), the HRA actually matched the empirical data quite well. It correctly identified that 40 minutes was not sufficient to get to and perform the appropriate procedural steps for cool down and depressurization. The root cause for this inadequate time was also properly identified – addressing the SGTR procedurally would be delayed because addressing FRH1 takes priority. However, the qualitative analysis said that the crew would not isolate the ruptured SG outside of the procedures in the allocated time; to do this would require permission that would take ~30 minutes to obtain. In reality, three of the crews isolated the ruptured SG from knowledge, outside the higher priority procedures. Overall, the crews found the SGTR challenging because they felt "stuck" in the FR procedures.

The qualitative analysis also noted that the cues would be disguised and delay the diagnosis, but, given the training they have on SGTRs, the diagnosis should not be difficult to do once the cues were no longer masked. This is consistent with the empirical data.

One interesting note is that the operational story described that the crew would get caught in a loop with step 28 of FRH1, where they could not secure the last HHSI pump because the AFW flow will not be adequate to sufficiently lower the temperature. This description is, in essence, the same thing that happened to the crew which misinterpreted the phrase "active loop". In this case the HRA assessment team also neglected to understand the implication of the phrase "active loop", and that there was a possible success path.

*G5.3.5 Comparison of Qualitative Analysis to Empirical Data*

As indicated in the previous section, the qualitative analysis predicted several of the empirical results very well, not just in identifying the drivers at the high level, but also correct in predicting that the degraded or failed performance would manifest itself in a confusion in how to address the SGTR. The operational story did not, however, account for the observed crew variability or the fact that the SGTR would be isolated outside the procedures.

*G5.3.6 Impact on HEP*

The HEP for this HFE was 1.0. This closely reflects the difficulty the crews had in diagnosing the SGTR, cooling down and depressurizing the SGs within 40 minutes of the SGTR. The empirical data showed that three crews failed this criteria and one barely made it (37 minutes).

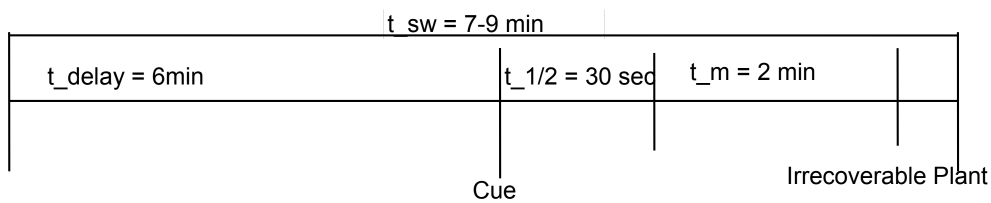
It is worth noting that the method gives no credit for knowledge-driven activities and that the empirical crew that did succeed did so via a knowledge-based path.

## G5.4 HFE 2A

### *G5.4.1 Summary of Qualitative Findings*

The Scientech Calculator team broke down HFE 2 into two HFEs because there are two individual cognitions: 1) operators fail to trip the RCPs and 2) operators fail to start the PDP pump. Because this scenario is time constrained to 7-9 minutes, and time was determined to be a driving factor, the team decided to quantify the HFE as two separate scenarios – one using a 7 minutes timeline and one using a 9 minutes timeline.

Based on the operator interview, in this scenario, once the reactor trips, the crew will perform the first four steps of E0 and then take a step back and look at the system. There are clear cues that the CCW has been lost, so the operators will immediately trip the RCP. They will then continue to ES-01 in the procedures and when they get to 6.c.3 they will start the PDP. The following is the timing for failure to start the PDP (note: the t_delay includes time to trip the RCP as well as the time due to distractions from dealing with the lost distribution panel):




Based on the timeline above, the Scientech Calculator team concluded that the time required to get to a perform the appropriate steps for success is greater than 7 minutes, therefore, the HEP for the 7 minutes scenario was set to 1.0 solely based on timing.

The operator action was considered feasible for this HFE given a 9 minutes time window. The dominant contribution to this HFE was the failure to start the PDP, which yielded a 1.2E-1 HEP (compared to a 4.2E-3 HEP for failure to trip the RCP). Failure to trip the RCP is not a dominant failure mechanism because once the operators realize they have lost CCW, they will immediately trip the RCP: "Operator training develops an instinct that will cause an operator to trip the RCPs without referring to a procedure in the interest of time." Driving factors, then, for failure to start the PDP were:

- Timing: When the distribution panel is lost, they lose several key controlling channels, most notably for SGs A and B and the pressurizer level control. On the loss of pressurizer level control operators will lose critical time when they have to complete the RNO portion of step 6.a.
- Workload: Between the distribution panel failure, the reactor trip the crew will have to work through multiple procedures in tandem to properly diagnose the cues and resolve the situation.
- Moderate Stress based on the combination of time and workload.

### *G5.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

7 minute time window: HEP = 1.0
9 minute time window: HEP = 1.2E-01; EF = 1

Because the 7minutes scenario was considered infeasible due to time, only the quantitative findings for the 9 minutes scenario will be discussed here.

For quantification, the maximun of CBDTM and HCR/ORE were used. Because of the short time window the HRC/ORE dominated in both cases.

| HFE Name | HFE Description | CBDTM | HCR/ORE | Execution Prob. | Total HEP | Error Factor |
|---|---|---|---|---|---|---|
| HFE 2-1 | Operators fail to trip RCPs | 3.0E-03 | 4.2E-03 | 0 | 4.2E-03 | 10 |
| HFE 2-2 | Operators fail to start the PDP within 9 minutes | 6.0E-03 | 1.1E-01 | 7.8E-03 | 1.2E-01 | 1 |
| **HFE 2-Total** | | **9.0E-03** | **1.1E-01** | **7.8E-03** | **1.2E-01** | **1** |

- **B1: Fail to trip RCPs [4.2E-3]:**
  - Cognitive: 4.2E-03; HCR/ORE
    - PWR CP1 sigma value of 0.57 [when operators acknowledge loss of CCW alarms they will respond to it].
  - Execution: N/A
  - No recovery

- **B2: Fail to start the PDP within 9 minutes [1.2E-1]:**
  - Cognition: 1.1E-01; HCR/ORE
    - PWR CP1 sigma value of 0.57
    - No recovery; insufficient time.
  - Execution: 7.8E-3; THERP
    - 2 execution steps with a step to monitor temperature; stress factor of 2 applied due to high workload. Recovery via temperature monitoring reduces execution HEP from 1.0E-02 to 7.8E-03
    - Dependency for recovery is high.

CP1 is the category for the “DO immediately” actions. When operators acknowledge loss of CCW alarms they will respond to it without waiting for a procedure step.

*G5.4.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | In this method the adequacy of time is primarily manifested by which cognition method is chosen – CBDT or HCR/ORE. In this case HCR/ORE was chosen. Dependence levels for crediting recovery actions are based on time; no credit was given for cognitive recovery | MND |
| **Time Pressure** | | N/A |
| **Stress** | (THERP) Moderate because of high workload. Contributed to the execution portion, but the execution was a negligible contributor to the total HFE. | ND |

| | | |
|---|---|---|
| **Scenario Complexity** | Addressed by the calculator, but not a contributor to this scenario. Not explicitly addressed by HCR/ORE. | 0 |
| **Indications of Conditions** | Addressed by the calculator, but not a contributor to this scenario. Not explicitly addressed by HCR/ORE. | 0 |
| **Execution Complexity** | Explicitly considered by THERP tables based on specific types of complexity (e.g., long checklist v. short checklist, etc.); execution was not complex in this scenario. | N/P |
| **Training** | Addressed by the calculator, but not a contributor to this scenario. Considered nominal by HCR/ORE. Training was credited in the qualitative analysis by crediting a non-proceduralized action of tripping the RCP. | N/P |
| **Experience** | In HCR/ORE, this can be addressed by calculating a scenario specific sigma, but this analysis used the predetermined sigma. | N/P |
| **Procedural Guidance** | Addressed in CBDT, but not used in this case. | 0 |
| **Human- Machine Interface** | Select HMI issues relevant to execution are also covered by THERP. Not applicable to this HFE. | N/P |
| **Work Processes** | Availability of check off provisions within the procedure is also assessed in the execution portion. Check-off provisions were used. | N/P |
| **Communication** | Addressed in CBDT, but not used in this case. | 0 |
| **Team Dynamics** | Considered implicitly in building the timeline. In HCR/ORE, this can be addressed by calculating a scenario specific sigma, but this analysis used the predetermined sigma. | N/P |
| **Other** | Sigma value of HCR/ORE is based on the type of cue/action. | ND |

*G5.4.4 Comparison of Drivers to Empirical Data*

While the main drivers in did not necessarily match well in how they were categorized in the table below, a closer look shows that the HRA was reasonably successful in predicting some of the drivers for this scenario. Amount of time available compared to the time needed to get to the critical step in the procedure was found to be a negative driver for both analyses. Similarly, the workload due to the loss of indications and associated stress was a negative driver.

The main drivers, however, were not well characterized by the analysis. Scenario complexity was a major driver in the empirical data; this was not reflected in the HRA. The HRA qualitative analysis showed a straight forward path that the operators would follow; it indicated that the operators would have no problem tripping the RCPs and working their way through ES01 from a complexity standpoint. Training and experience was another factor that the empirical team found to be important that the HRA did not. In this case the analysts of the empirical data suggested that the ability of the team to prioritize was a major contributor, and that was due to inadequate training/experience. Furthermore, they assert that time would have been adequate if the operators were better trained. This differs significantly from the HRA. The qualitative portion of the HRA asserts (and the impact of the driver is confirmed by its numerical impact on the HEP) that there was simply insufficient time to get to the appropriate step in ES01 to start the PDP prior to the temperature exceeding 230 degrees regardless of training. There was no indication in the qualitative analysis that the operators would start the PDP outside the procedures.

| Factor | Empirical | HRA | Comments |
|---|---|---|---|
| **Scenario Complexity** | MND | 0 | The empirical analysis concluded that many things happening at the same time made it difficult to detect the priority items. The HRA did not foresee any difficulty in diagnosis or prioritization. |
| **Training/Experience** | MND | N/P | Crew is not trained on Loss of CCW and sealwater to the RCPs very often; when trained it is usually done in a LOOP scenario. The training and experience on the needed prioritization is what made this a negative driver. The HRA assessed training as a positive factor in identifying the Loss of CCW. |
| **Adequacy of Time** | ND | MND | The analysis presented 2 cases: one with a 7 minute time window (HEP of 1.0) and one with a 9 minute window (HEP of 1.2E-1). In both cases the main driver of the quantitative analysis was the time it would take to get to the appropriate step in ES-01 to start the PDP. This is also the driver identified in the qualitative analysis. While time was a driver in the empirical data, the conclusion was "if [the crew] had more training there would have been enough time." However, the empirical analysis determined procedural guidance as a negative driver based on the fact that "crews had problems reaching the critical actions in time". |
| **Procedural Guidance** | ND | 0 | |
| **Indications of Conditions** | ND | 0 | |
| **Work Processes** | ND | N/P | In the empirical data, monitoring of the control boards and acknowledging alarms was less than adequate. Also had some difficulties when POP4 read by the RO. Stress was observed, but its impact was unclear. In the HRA, the work process was considered nominal because the operational story assumed the crew would get to the PDP via ES01, so the ease or difficulty of POP4 was never addressed by the analysis. However, they did note that workload would be high and this would cause delay/distraction and moderate stress. None of these were the main driver, but were contributing factors. |
| **Stress** | 0 | ND | |
| **Other** | n/a | ND | In the quantitative analysis they type of cue drives the CP value. Cue type was not considered by the empirical analysis |
| **Human- Machine Interface** | N/P | N/P | |
| **Execution Complexity** | 0 | N/P | |
| **Team Dynamics** | 0 | N/P | |

| | | | |
|---|---|---|---|
| **Communication** | 0 | 0 | |
| **Time Pressure** | n/a | n/a | |

*G5.4.5 Comparison of Qualitative Analysis to Empirical Data*

The qualitative analysis only moderately matched the actual progression. The HRA determined that the crew would diagnose the loss of CCW fairly quickly (within 3 minutes) and then, based on knowledge, they would immediately trip the RCPs (30 seconds). In actuality, it took roughly 4-6 minutes to detect the loss of CCW and sealwater (only 3 of the teams stopped the RCP before the seal temperatures reached 230 °F). Three crews tripped the RCPs based on knowledge in ~1 minute; the fourth crew waited to trip the RCP per a procedural step in 0POP4-RC-0002, taking ~2 minutes.

To start the PDP, the HRA operational story indicates that the crew would work through ES01 and get to step 6.c.3 in that procedure where they would then start the PDP. It was estimated to take 6 minutes to get to the necessary procedural step and 2.5 minutes to actually read the step and begin the PDP. Thus, the crew is expected to fail if there is only a 7 minute time window; they can barely make it (30 seconds time margin) if there is a 9 minute time window. In reality, no crew attempted to start the PDP through ES01; three of the crews attempted to get there through 0POP4-RC-0002, and one crew did not see seal cooling as an issue. Given the different procedural paths described in the two analyses, and also given that no crew was successful in starting the PDP, there is no way to compare the predicted time for this portion of the analysis.

*G5.4.6 Impact on HEP*

Despite some of the mismatch in the qualitative analysis, this HFE did a very good job in predicting an appropriate HEP. The predicted HEP was 1.0 based on a 7 minutes time window and 1.2E-1 for a 9 minutes time window. In actuality 4/4 crews failed. For those 4 crews, three crews exceeded a seal temperature of 230 °F in ~7.5 minutes after loss of CCCW and sealwater, and the fourth crew exceeded the seal temperature in ~8.5 minutes.

### G5.5 HFE 3A

*G5.5.1 Summary of Qualitative Findings*

The Scientech Calculator team broke down HFE 3 into 3 separate parts: 1) operators fail to diagnose SGTR [fail to enter E-30], 2) operators fail to identify and isolate ruptured steam generator, and 3) operators fail to maintain the RCS steam pressure below the setpoint by cooling down the RCS. The team indicated that this HFE has the lowest cognitive complexity of all the scenarios in this study. Given the combination of low cognitive complexity (good training, clear cues, excess time, and no additional errors) and the numerous execution steps required for success, execution failure was considered the dominant failure mode for this HFE.

According to the qualitative analysis, while the execution is expected to dominate the HEP, the total HEP is expected to be low. The driver is the sheer number of execution steps required to successfully isolate, cooldown and depressurize the steam generator. The total HEP is expected to be low because, despite the number of steps, there is ample time for recovery and there are recovery steps built into the procedure that requires the operators to monitor criteria that will clue them if they have made a mistake. Operators are very well trained on diagnosis and execution associated with this scenario as it is a design basis scenario for the plant.

*G5.5.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
HEP = 1.9E-03; EF = 5

| HFE Name | HFE Description | Cognition Probability CBDTM | Cognition Probability HCR/ORE | Execution Prob. | Total HEP* | Error Factor |
|---|---|---|---|---|---|---|
| HFE 3-1 | Operators fail to diagnose SGTR | 9.06E-06 | Negligible | n/a | 1.0E-04 | 10 |
| HFE 3-2 | Operators fail to isolate ruptured steam generator. | 9.06E-06 | Negligible | 1.7E-04 | 1.8E-04 | 10 |
| HFE 3-3 | Operators fail to maintain the RCS pressure below the setpoint by cooling down the RCS. | 9.06E-06 | Negligible | 1.6E-03 | 1.6E-03 | 5 |
| **HFE 3-Total** | | **2.7E-05** | **Negligible** | **1.8E-03** | **1.9E-03** | **5** |

*Note: a lower bound is applied here; probabilities lower than 1E-04 this were set at 1E-04.

- **1: Fail to diagnose SGTR [1E-04]:**
    - Cognitive: 9.6E-6; CBDTM used because large time window dominated
        - Only failure mechanism identified was “skip a step in procedure”, which was mitigated by the fact that the EOPs include placekeeping aids and checkoff provisions.
        - Recovery available (self-recovery) reducing the HEP from 3.0E-03to 9.6E-06
        - No dependency: recovery time available is long (114 minutes equates to no dependency in the calculator for CBDT) and they have a procedural cue [if fail the diagnosis in step 13 of EOP-0 then have another chance at step 23].
    - Execution: N/A

- **2: Fail to identify and isolate ruptured steam generator [1.8E-04]:**
    - Given successful entry to E-3
    - Cognition: 9.6E-6; CBDTM used because large time window dominated
        - Only failure mechanism identified was “skip a step in procedure”, which was mitigated by the fact that the EOPs include placekeeping aids and checkoff provisions.
        - Recovery available (self-recovery) reducing the HEP from 3.0E-03to 9.6E-06
        - No dependency: self-review is credited because the cue for this HFE is also proceduralized in the foldout pages which the operator will also be reviewing (in addition to E-30).
    - Execution: 1.7E-4; THERP
        - 4 execution steps necessary for success, including a self-review. The operators will also be monitoring the SG Narrow Range level, which will provide them the cue for recovery if they skip a step (see table below).
        - Zero dependence for recovery actions; 110 minutes available for recovery.

| HFE 3-2 Execution Probability Breakdown | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Critical Step** | **Recovery Step** | **Action** | **HEP (exe)** | **HEP (rec)** | **Dep .** | **Cond HEP (Rec)** | **Total for Step** |
| 3.a | | Adjust ruptured SGs PORV controller setpoint to between 1260 PSIG and 1265 PSIG | 1.7E-3 | | | | 7.3E-7 |
| | 4.a | NR level – GREATER THAN 14% in ruptured SG | | 4.3E-4 | ZD | 4.3E-4 | |
| 3.e | | Check SG 1D - RUPTURED | 4.3E-4 | | | | 1.8E-7 |
| | 4.a | NR level – GREATER THAN 14% in ruptured SG | | 4.3E-4 | ZD | 4.3E-4 | |
| 3.h | | Close ruptured SGs MSIVs and MSIBs | 1.7E-3 | | | | 7.3E-7 |
| | 4.a | NR level – GREATER THAN 14% in ruptured SG | | 4.3E-4 | ZD | 4.3E-4 | |
| 4.b | | Stop AFW flow | 1.7E-3 | | | | 1.7E-4 |
| | Self Review | | | 1.0E-1 | ZD | 1.0E-1 | |
| | | **Total Unrecovered** | **5.6E-3** | **Total Recovered** | | | **1.7E-4** |

- **3: Operators fail to maintain the RCS pressure below the setpoint by cooling down the RCS. [1.3E-03]:**
  - Given successful entry to E-30 and completion through step 7; SGTR diagnosed and ruptured SG isolated.
  - Cognition: 9.6E-6; CBDTM used because large time window dominated
    - Only failure mechanism identified was "skip a step in procedure." Procedures for cooldown and depressurization don't include any characteristics that make them stand out, however, there is a line to check as the operators progress through E-30.
    - Recovery available (self-recovery) reducing the HEP from 3.0E-03 to 9.6E-06
    - No dependency: self-review is credited because the operators will have to check the core exit TCs in step 7.i to ensure they are less than the established temperatures in step7.a during cooldown. While depressurizing, Addendum D will be used as a check to correct any errors.
  - Execution: 1.3E-3; THERP
    - 10 execution steps necessary for success. There are several steps which provide cues for recovery if they skip a step (see table below).
    - Zero dependence for recovery actions; 90 minutes available for recovery.

| HFE 3-3 Execution Probability Breakdown | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Critical Step** | **Recovery Step** | **Action** | **HEP (exe)** | **HEP (rec)** | **Dep.** | **Cond HEP (Rec)** | **Total for Step** |
| 7.a | | Determine required core exit temperature | 7.9E-3 | | | | 7.9E-4 |
| | 7.i | Core exit T/Cs LESS THAN required core exit temperature | | 1.0E-1 | ZD | 1.0E-1 | |
| 7.c | | Block low steamline pressure SI | 1.7E-3 | | | | 1.7E-4 |
| | 7.i | Core exit T/Cs LESS THAN required core exit temperature | | 1.0E-1 | ZD | 1.0E-1 | |
| 7.g | | Place steam dump 'INTLK SEL' switches to BYPASS INTERLOCK | 1.7E-3 | | | | 1.7E-4 |
| | 7.i | Core exit T/Cs LESS THAN required core exit temperature | | 1.0E-1 | ZD | 1.0E-1 | |
| 7.h | | Dump steam to condenser from intact SG(s) at maximum rate | 1.7E-3 | | | | 1.7E-4 |
| | 7.i | Core exit T/Cs LESS THAN required core exit temperature | | 1.0E-1 | ZD | 1.0E-1 | |
| 15.b | | Stop RCS cooldown | 1.7E-3 | | | | 2.9E-6 |
| | 15.c | Maintain core exit T/Cs LESS THAN required temperature | | 1.7E-3 | ZD | 1.7E-3 | |
| 18.b | | Place group 'C' pressurizer heater control switch to PULL TO LOCK | 1.7E-3 | | | | 2.9E-6 |
| | D.1 | Continue to depressurize until cooldown conditions are met | | 1.7E-3 | ZD | 1.7E-3 | |
| 18.c | | Place all other pressurizer heater group control switches to OFF | 1.7E-3 | | | | 2.9E-6 |
| | D.1 | Continue to depressurize until cooldown conditions are met | | 1.7E-3 | ZD | 1.7E-3 | |
| 18.d | | Initiate maximum pressuizer spray | 1.7E-3 | | | | 2.9E-6 |
| | D.1 | Continue to depressurize until cooldown conditions are met | | 1.7E-3 | ZD | 1.7E-3 | |
| 18.f.1 | | Normal spray valves CLOSED | 1.7E-3 | | | | 2.9E-6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | D.1 | Continue to depressurize until cooldown conditions are met | | 1.7E-3 | ZD | 1.7E-3 | |
| 18.f.3 | | Auxiliary spray valves CLOSED | 1.7E-3 | | | | 2.9E-6 |
| | D.1 | Continue to depressurize until cooldown conditions are met | | 1.7E-3 | ZD | 1.7E-3 | |
| | | **Total Unrecovered** | **2.4E-2** | **Total Recovered** | | | **1.3E-3** |

*G5.5.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | More than adequate time. In this method the adequacy of time is primarily manifested by which cognition method is chosen – CBDT or HCR/ORE. In this case CBDT was chosen because it is not time limited. Dependence levels for crediting recovery actions are based on time. | N/P |
| **Time Pressure** | Determined in qualitative analysis that there is no time pressure. | N/A |
| **Stress** | (THERP) Low because plant responds as expected, nominal workload and other PSFs are optimal. | N/P |
| **Scenario Complexity** | Components of Cognitive Complexity explicitly addressed by CBDT. Scenario complexity also explicitly addressed in THERP for execution. This is a low complexity scenario | N/P |
| **Indications of Conditions** | Availability and clarity of indications with respect to the scenario are explicitly considered by CBDT. In the CBDT, the indications were considered to be poor, but the CBDT value was not a numerical contributor. | 0 |
| **Execution Complexity** | Explicitly considered by THERP tables based on specific types of complexity (e.g., long checklist v. short checklist, etc.); execution was not complex in this scenario, however there are many execution steps and the cumulative probability of the various steps (each individual one having a low HEP) is the main contributor to this HFE. | MND |
| **Training** | Not addressed explicitly, but addressed in the qualitative analysis. This is a design basis scenario, so it is well trained on. Also, because there is a lot of data on this scenario, the timing analysis is less conservative. | N/P |
| **Experience** | Not addressed by this method. | N/A |
| **Procedural Guidance** | Both quality and availability addressed in CBDT and THERP. The main mechanism both in CBDT and THERP for this HFE was skipping a step in the procedure or checklist. | ND |
| **Human- Machine Interface** | Indicator location and physical quality addressed in CBDT. Select HMI issues relevant to execution are also covered by THERP. In the CBDT, the indications were considered to be poor, but the CBDT value was not a numerical contributor. | 0 |

| | | |
|---|---|---|
| **Work Processes** | Availability of check off provisions within the procedure is also assessed in the execution portion. Check-off provisions were used. | N/P |
| **Communication** | Use of formal communications explicitly covered by CBDT in Pcc. | N/P |
| **Team Dynamics** | Considered only implicitly in qualitative analysis in building the timeline. | N/A |
| **Other** | | N/A |

#### *G5.5.4 Comparison of Drivers to Empirical Data*

The empirical data did not indicate any negative drivers for this scenario, but did note that positive drivers included indications of conditions, procedures, training and teamwork. This is consistent with the qualitative HRA, which identified clear cues, excess time and good training as the positive factor (low cognitive complexity). The HRA predicted that the team would be well equipped to successfully perform these actions. In the quantitative analysis the negative factors that dominated were associated with minor slips (error of omission in execution, skipping step in procedures).

#### *G5.5.5 Comparison of Qualitative Analysis to Empirical Data*

The qualitative analysis matched well the empirical data in the overall assessment of the HFE. However, the HRA used a 10 minute time line to identify and isolate the SG, based on EOPT-03.13 Rev.21 – Design Basis Timeline for a SGTR, from initiating event to isolation of feed flow. This 10 minutes was broken down into 6 minutes to get to the cue (including diagnosis of SGTR, team meeting and transfer to correct procedures), 1 minute to identify the ruptured SG and 3 minutes for execution of the appropriate action (isolate the SG). It took an additional 20 minutes to cooldown and depressurize. In reality it took the operators 8-12 minutes after reactor trip to transfer to the correct procedure and 14-21 minutes after the initiating event to isolate the SG. It took the operators ~20 minutes to cooldown and depressurize. Thus, the HRA timeline was a bit optimistic.

#### *G5.5.6 Impact on HEP*

The HEP for this scenario was 5.8E-03; this is a fairly low HEP and corresponds reasonably to the empirical data which saw no failure out of three participating crews.

### G5.6 Comparison Summary

#### *G5.6.1 Predictive Power*

The overall predictive power of the EPRI HRA Calculator method as employed by the NRC team in this study was judged to be moderately good. The final HEPs seem to be fairly consistent in trend, ranking and level of differentiation between HEPs. Overall, the assessment was moderately good to good at predicting risk drivers and in the quantitative predictive power, and was fair to moderately good at predicting the operational expressions of the drivers. The table below provides a summary of how this analysis performed against the empirical data across scenarios.

| HFE | HEP Ranking | | HEP Values | | Consistency with Empirical Data* | | |
|---|---|---|---|---|---|---|---|
| | Empirical | HRA | Empirical | HRA | Drivers | Op Story | Timing |
| **2A** | 1 | 1 | 4/4<br>Very difficult | 1.0, 0.12 | 3.5 | 3.5 | 4 |
| **1C** | 2 | 1 | 3/4<br>Difficult | 1.0 | 3 | 3 | 4 |
| **1A** | 3 | 3 | 0/4<br>Fairly Difficult | 0.032 | 5 | 5 | 4 |
| **3A** | 4 | 4 | 0/4<br>Easy | 0.0019 | 5 | 5 | 3 |
| * 1= poor; 2= moderately poor; 3=fair; 4=moderately good; 5=good | | | | | | | |

G5.6.1.1 Qualitative Predictive Power in Terms of Drivers

The EPRI HRA Calculator uses three distinct methods: HCR/ORE, CBDT and THERP. Cognitive steps are quantified using the higher value between HCR/ORE and CBDT, while THERP quantifies errors during execution. For HFEs that include both cognitive and execution portions, the HEP is determined by adding the cognitive HEP (HCR.ORE or CBDT) to the execution HEP (THERP). The HCR/ORE method uses a time reliability correlation, so under normal applications the only quantitative driver that can be distinguished is timing. For the CBDT method, however, identification of important drivers is guided by evaluation of the factors addressed in a set of decision trees with recovery factors applied at the end. THERP is based on a set of specific criteria for different types of actions. In CBDT and THERP, negative conditions lead to higher HEPs, with the factors that have the greatest negative effect on the resulting HEP being the main drivers.

Overall, this application of the EPRI HRA Calculator did a moderately good to good job in predicting the drivers of the empirical data. The place where the drivers were only fairly well predicted (HFE 1C, 2A) was the HFE were the operational story did not match entirely (further described in Section 0). Both of these HFEs received a HEP of 1.0 based on the timing analysis from the operational story (action not feasible), so none of the methods within the Calculator were utilized to determine drivers.

G5.6.1.2 Qualitative Predictive Power in Terms of Operational Expressions

In addition to doing a good job in predicting drivers, this analysis did a fair to moderately good job in predicting the operational expressions of those drivers.

The analysis did a good job in determining the specific failure mechanisms for some of the crews for HFE 1C. However, the operational story did not reflect the crew variability seen in the data, so the analysis was unable to predict all the drivers. HFE 2A was quantified using HCR/ORE; HCR/ORE does not provide insights into failure mechanism, but rather provides a correlation of HEP given task timing information. In this case timing was an important driver of the empirical data. The operational story in this case identified correctly that, while the RCP pumps would be stopped in the timeframe, but starting the PDP would be very difficult in the 9 minute time frame and not possible in the 7 minutes time frame.

For HFE 1A several possible failure mechanisms were identified and two of them were manifest in the empirical data, although did not lead to actual failures. "Information misleading" was identified by CBDT and the empirical data noted that diagnosis was delayed due to the misleading information; this was not a failure because plenty of time was available. "Skip a Step in the Procedure" was identified by CBDT as a cognitive failure mechanism and, in reality, one crew missed stopping the RCP in the procedure (but recovered). In addition, training/experience was seen as a strong mitigator in both the analysis and the empirical data. HFE 3A saw no failure mechanism in the empirical data, this matches the prediction given the small sample size and low predicted HEP.

G5.6.1.3 Quantitative Predictive Power

Overall, the quantitative results did a moderately good job in matching the empirical data both in rank and appropriate level of differentiation. The most difficult HFE (2A) saw 4/4 crews fail. The HEP was 1.0 for a 7 minutes time frame and 1.2E-1 for a 9 minutes time frame (8.5 minutes determined to be the minimum time needed for success). The actual time frame in the simulation was 7.5-8.5 minutes, with the seal temperature exceeding 230 °F at 7.5 minutes in 3 of the 4 crews. The second most difficult HFE (1C) saw 3/4 crews fail. The HEP for this scenario was 1.0, so it was ranked the same as HFE 2A.

The easiest HFE (3A) saw 0 crew failures; the predicted HEP was 1.9E-3. The second easiest HFE (1A) was determined to be "fairly difficult" but saw 0 crew failures; this received an HEP of 3.2E-2.

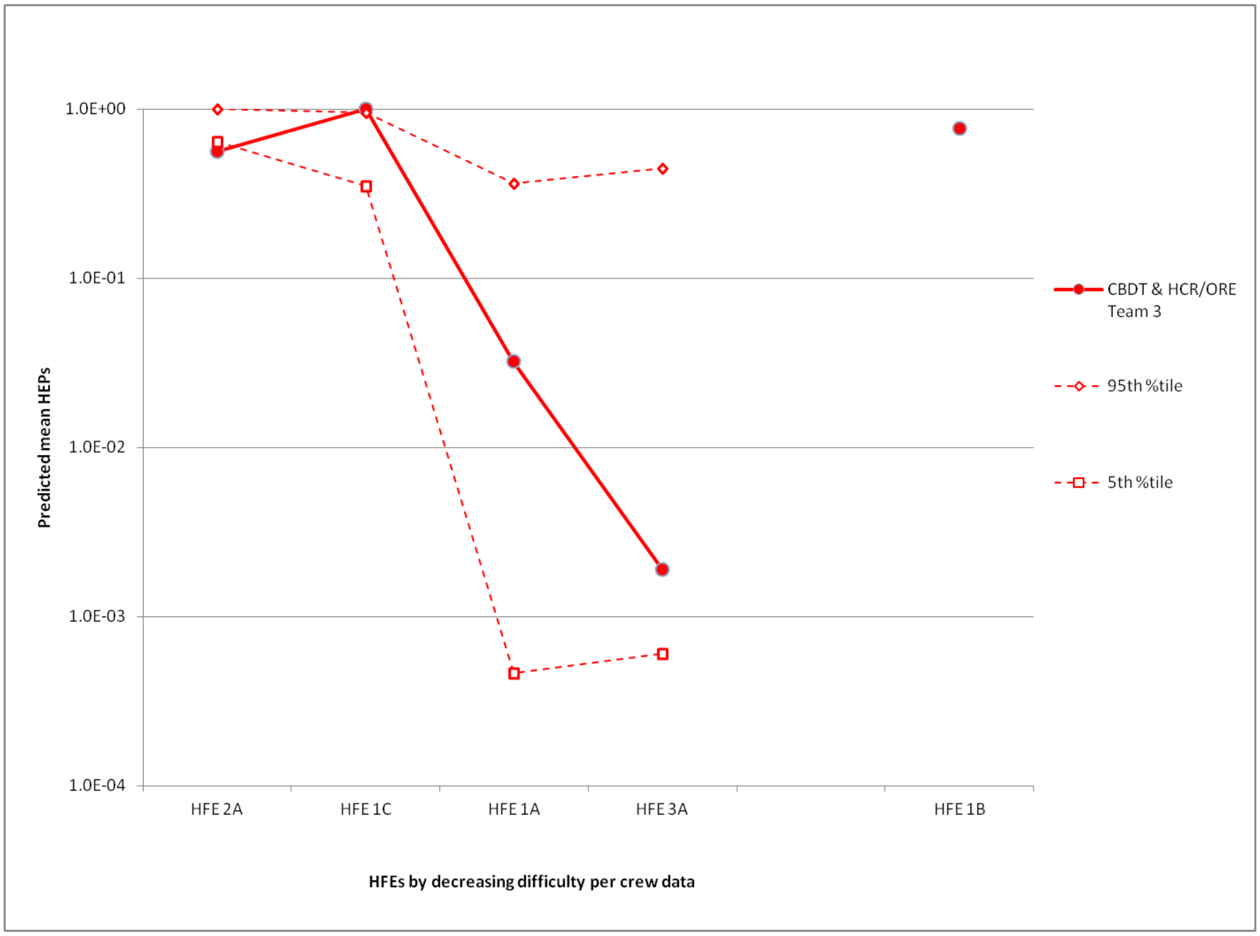



Figure 0.5 Predicted Mean HEPs with Uncertainty Bounds by Scientech HRA Team with EPRI Calculator

*G5.6.2 Assessment of Guidance and Traceability*

The derivation of the HEPs within the method and what is important to performance is very traceable for CBDT and THERP. How the various factors are weighted in determining the final HEP can be determined by examining the contribution of various factors from the decision trees or THERP tables. The ability to trace the basis for the judgments regarding the branch points in the trees, choice of THERP table or choice of sigma value for HCR/ORE relies on the analysts' documentation. HCR/ORE provides only as much traceability as the analyst provides in the development of the timing.

The area of least traceability is in development of the operational story and the timing analysis. There is also little guidance provided for how to break up an HFE into subcomponents that need to be quantified separately.

*G5.6.3 Insights for Error Reduction*

For certain types of errors, this method can be powerful in providing insights for error reduction. There are generally two categories of error reduction: 1) improving human factors to reduce errors (e.g., improve location and clarity of indications, improve format of procedures, etc.); and 2) reducing the difficulty of cognitively challenging errors (e.g., improved training, clearer procedures, etc.). This method clearly provides insight into the former category of error reduction when CBDT and THERP are used. For example, in HFE 1A, CBDT tree $P_{ca}$

(Information Misleading), specific training was credited to offset the impact of misleading information. If no specific training was available, a much higher HEP would have been assigned, so the analyst could tell from the tree that one major way to reduce error is to implement training.

This method however provides a more limited set of insights for reducing error due to complex scenarios. The primary source of insights for error reduction of these scenarios comes from the qualitative analysis, which is supposed to include operator interviews and/or simulator observations. In this analysis, the operational stories were defined from one operational interview, and reflected only one progression of the scenario. In the two most difficult HFEs (1C and 2A), the data produced a spectrum of scenario progressions. Thus, when the operational story matched the crews' progression, the drivers were very well predicted, but other crews who progressed through the same scenario differently did not present the same drivers. Without predicting the variability of the crew progression, the root cause of the various drivers – and thus a way to reduce the error – is difficult to glean. Again, both of these HFEs received a HEP of 1.0 based on the timing analysis from the operational story (action not feasible), so none of the methods within the Calculator were utilized to determine drivers.

When HCR/ORE is used, the only insight to error reduction that can be gleaned is to find a way (either through training, revised procedures, etc.) to improve the timeline. Insights for error reduction were judged to be fair in this application of the EPRI HRA Calculator.

One factor seen in the empirical data that reduced the failure rate was the ability to recover from errors. This method credits different types of recovery and will point to the most effective (least dependent) recovery options; recovery credit was a large part of the overall analysis.

*G5.6.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses*
The biggest strength of this method is that the judgments made by the analysts in applying the Calculator and obtaining the HEPs are clearly traceable in the sense that each method in the Calculator is based on a set of clear rules to be applied based on well-defined criteria.

Overall the application of this method produced moderately good results. The method performed best overall (identifying risk drivers, identifying failure mechanisms and quantification) for easy HFEs; HFEs that presented high cognitive complexity, particularly scenarios that saw much crew variability (HFEs 2A and 1C), did not compare as well with this method. The basic assumptions behind this method are that operators will follow procedures and generally trust their cues. This basic assumption means little to no credit is given for actions involving non-proceduralized knowledge-based actions.

The method seemed to provide a reasonable set of influencing factors to address in order to predict crew performance and provide insights for error reduction. However, much rides on the operational story and the initial determination of feasibility.

## G.6 SPAR-H (INL 1)

Overall, the analysis is easy to understand and reads pretty well. Although the documentation of the analysis is a little bit difficult to overview, maybe caused by the fact that we asked for a different kind of documentation (Form A) than what the method uses as a standard. Thus the Form A consists of copied paragraphs from the main document for the analysis. The main document contains some general PSF info in chapter 6 and scenario specific PSF info in chapter 7. Both of these are included per HFE in the Form A sheet. In Form A, everything regarding one HFE is written in one place, so general judgments on general PSFs are repeated. This is ok so all the info on one HFE is one place.

The chapters 8 and 9 in the main analysis document are more background for the analysis, and they are not directly commented upon here. They seem to be a good background for the qualitative analysis that underlies the PSF judgments though.

As seen in other SPAR-H submittals in the former study, the qualitative analyses of the scenarios are not diving into deep details of the scenarios and procedure use. Instead the qualitative description present is described as part of the PSF analysis.

Both diagnosis and action are analyzed for all the scenarios.

### G6.1 HFE 1A

*G6.1.1 Summary of Qualitative Findings*

Both diagnosis and action multipliers are calculated for all the HFEs.

There is no additional qualitative analysis in addition to the one that is done in relation to the PSF analysis. This could also be seen as the paragraphs in item 3) in form A was taken from the PSF analysis part in the main document.

From item 3):

*"This scenario was felt by the analysis team to be moderately complex; all main feed pumps are tripped, the reactor trips or the crew trips the reactor manually, and then the crew enters the appropriate procedure and begins the well trained to task of establishing F&B - - activating SI and opening the PZR PORVs."*

The qualitative analysis is on a high level, probably sufficient for a simple HFE.

*"The SPAR analysis could be slightly more interesting if the analyst were to consider the HFE1a to include the probability associated with the operators getting to the proper diagnosis within 30 seconds, and thus establishing the pace and pattern of the overall HFE. Operators stated that as part of simulator training that the crew trains to this portion of scenario - - to get to LOFW diagnosis and manual trip within 30 seconds. However, the HFE1a,b as presented requires calculation for crew realizing the need to establish F&B after a reactor trip by either manual action or by automatics has occurred."*

This describes actually another HFE than the one defined. It is maybe the case that this would be interesting to look at, but it doesn't really affect the current analysis since the HFEs were already defined.

In the description of the complexity PSF the team describes the situation with misleading indication of AFW pump 12. They do not describe any procedure transitions.

*G6.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
HEP = 0.015 [mean]; (Lower Bound HEP = 6.0E-05, Upper Bound HEP = 5.7E-2) [uncertainty]

Note: All HFEs calculated included aspects of diagnosis and action. In SPAR-H modelling space this is considered a joint HFE.

**Quantification of relevant factors:**

| *PSF Type* | *HFE1a (diag/act)* |
|---|---|
| Time | .1/1 |
| Stress | 2/2 |
| Complexity | 2/2 |
| Experience | .5/.5 |
| Procedures | 1/1 |
| Ergonomics | 10/10 |
| Fitness | 1/1 |
| Work Processes | .8/.5 |
| **Calc HEP** | 0.016/0.01 |
| **Adjusted for multiple negative PSFs (=or>3)/ Final HEP** | 0.014/.009* |
| **Total Joint HFE (Diagnosis + Action)** | 0.015 |

*G6.1.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | "For HFE 1a, the diagnosis portion of the HFE is calculated as limited to the crews understanding of the need for F&B and identification and use of the appropriate procedures while monitoring plant conditions. The time factor assigned from SPAR - H is extra time (.1); taking an upper decision limit of 31 minutes to correctly diagnose the situation and then acting in 10-15 minutes ( ratio of the action nominal time available to time required, (1) is possible within the 45 minutes allotted." | N/P |
| **Time Pressure** | Part of the stress PSF of SPAR-H. The SPAR-H method does not consider "Time Pressure" separately from "Stress." | N/A |
| **Stress** | "In interviews conducted by the HRA team, the operators stated that the stress level for HFE1a is high, "deciding to breach the RCS and take away a fission barrier is not an everyday occurrence" and a SPAR-H PSF multiplier of 2 was assigned." | ND |

| | | |
|---|---|---|
| **Scenario Complexity** | Scenario complexity is associated with the diagnosis part of the PSF "Complexity" in SPAR/H.<br><br>"The complexity is viewed as high. There is some misleading indication present, there is flow indicated from AFW pump 12, but the water never reaches the SG, and thus, the plant computer doesn't show a red path on the heat sink status tree. Also, indication is present in different parts of the control room and adds to complexity. SPAR-H suggests that analysis should not double count the effects of PSFs, thus, indicator layout is not counted against performance under the ergonomics PSFs category, and was thought to be intertwined with complexity rather than to function as an independent effect upon crew performance. Specifically, the operator has to refer to additional computer screens to obtain some key information." | ND |
| **Indications of Conditions** | Part of "Ergonomics/HMI" PSF in SPAR-H.<br><br>"Ergonomics was given a poor rating (10) because potentially important valve position status associated with a field-located recirculation valve that could an important cue for operators, was not available in the control room. This could be construed as missing indication, so that a multiplier of 50 could be applied, however, there are other sources of information that allow for the crew to come to the proper conclusions. Also, flow status indication (on pump12) was misleading and was felt to contribute to the cognitive complexity of events, however, this influence has been assigned under complexity." | MND |
| **Execution Complexity** | Execution complexity is associated with the action part of "Complexity".<br><br>The analysis team assigns both the action and the diagnosis part of the SPAR-H PSF Complexity the same weight: multiplier 2 or "moderately complex". However, as in HFE1C, they do not explain why the action part is also complex, but they have only one explanation, which seems to be more directed to the difficulties in the diagnosis. | 0 |
| **Training** | The factor in SPAR-H is "Experience/Training".<br><br>"Training insights. Experience of the crew and their training was observed in unrelated simulator runs and found to be high. We were also told that incoming operators were trained to LOFW followed by to sharpen their diagnosis skills. Given the time horizon of 45 minutes, this training by the licensee should contribute directly to the ability of the crew to respond to the event. Accordingly a SPAR-H training PSF modifier of .5 was assigned." | N/P |
| **Experience** | The factor in SPAR-H is "Experience/Training". | N/P |

| | | |
|---|---|---|
| **Procedural Guidance** | "A SPAR-H PSF procedures multiplier (.5) for availability and applicability of symptom based was applied, however, after discussion it was decided that there were enough additional procedures that could be accessed during the event that a nominal procedures value (1) was most appropriate. Also, the team noted that symptom-based procedures for these types of events are an industry standard in the US and that "extra credit" given by SPAR-H for symptom based procedures perhaps should be redefined such that symptom based is nominal by industry practice and that deviation from this case constitutes a negative situation (need for a positive multiplier)." | N/P |
| **Human- Machine Interface** | Part of "Ergonomics/HMI" PSF in SPAR-H. No additional comments further than what mentioned under "Indications of conditions" above. | 0 |
| **Work Processes** | "The overall crew simulator scenario observed indicated good communication and work processes and a SPAR-H PSF value of .8 for work processes was applied." | N/P |
| **Communication** | In SPAR-H, included in "Work Processes".<br><br>"It was assumed that work practices, training, and experience met industry standards. However, the operators we interviewed had a high level of experience and communicated well. This information was extrapolated to the Scenarios under evaluation. Work practices and communications – In the non-related scenario runs that we observed during our visit to the simulator there was a well followed sequence of call backs initiated by the unit supervisor to and from the operators he directed. We also observed this for multiple procedures in use." | N/P |
| **Team Dynamics** | In SPAR-H, included in "Work Processes" | N/P |
| **Other** | Fitness for Duty. "From the scenario description, fitness for duty factors had no bearing on this analysis and the nominal value (1) was assigned." | 0 |

In many of the PSF evaluations, the team does not explicitly state the differences between the diagnosis part and the action part of the analysis. They do this for the time available, but e.g., for the complexity they don't explain what is complex in the diagnosis and what is complex for the action part of the HFE. An explanation is maybe that in chapter 10 in the main analysis they discuss complexity impact on both diagnosis and action and seem to have concluded that it applies to both.

Regarding the separation of the PSFs complexity and the ergonomics in SPAR-H, the team clearly states what is considered in which PSF and states the way they have chosen to analyse it so not to double-count any effects. This is good in this analysis. It maybe shows the overlap of the definitions of these two PSFs though. However, many PSFs do overlap in real life and as long as they explain the way in which they have done their analysis, this is ok. For analysts not aware of the overlaps between these PSFs, the SPAR-H documentation needs better guidance as to what to include in which PSF.

*G6.1.4 Comparison of Drivers to Empirical Data*

The INL 1 team (Gertman) SPAR-H analysis identified three negative drivers on the performance of HFE 1A: Stress, Complexity (scenario complexity as I interpret it) and Indication of conditions. The Indications of Conditions is assessed as being the main negative driver, as explained when they set the SPAR-H PSF Ergonomics to a weight of 10.

In the empirical data, the three PSFs Scenario complexity, Indications of conditions and Procedural guidance were noted to be negative drivers. The reasons behind these are mostly captured by the SPAR-H analysis: The masking of loss of AFW including the misleading indication on the open recirculation valve, and the CSF heat sink status tree that does not address the misaligned valve. The latter is noted under procedures in the empirical data, but this SPAR-H analysis accounts for it under complexity. Thus the nominal/positive procedure prediction deviates from the negative PSF in the empirical data. The team notes that had there been time for a second visit by the team the adequacy of status tree procedures to address the valve in question may have been reviewed in greater detail.

A question mark is put on the action part of the complexity analysis, which in this comparison is categorized as execution complexity. It has the same weight as the diagnosis part, but with the same explanation. This explanation is maybe not relevant for the action part, and the SPAR-H analysis thus missed the observed N/P execution complexity. The SPAR-H team notes: "Complexity for us was determined from operator interviews, for example and for the most they stated whether things were relatively complex for action and decision".

Stress is not observed in the data, but predicted negative by this analysis. The team replies: "We think we are correct, the operators were discussing the level of stress they would feel in the plant and not in the simulator.  I would stand by their impressions including the high profile of the event and misleading indication". Thus the whole issue of stress should maybe be de-emphasized in this comparison, since this is the issue that may be really different from the real life to the simulator, and we have observations from the simulator and compare to predictions of real life.

Regarding positive drivers, the INL 1 team SPAR-H analysis correctly identifies adequacy of time and training/experience. They also predict work processes to be positive, but this is not observed in the data.

*G6.1.5  Comparison of Qualitative Analysis to Empirical Data*

The SPAR-H analysis is factor (PSF) oriented, as expected. It is also oriented towards a high level, which is applied across scenarios, at least on a high level in the scenario. Thus they do not analyze (at least they have not documented, other than in the chapters 8 and 9, documenting the operator interviews) scenario details like specific steps in the procedures where difficulties may occur, for most of the PSFs. For the "Indication of condition" and the "Complexity" PSFs they have noted specific valve and flow indications that are important in this scenario though. The judgments of these PSFs are built on scenario insight gained through review of material and table top walk through and interview.

*G6.1.6  Impact on HEP*

The main contributor to increasing the HEP is the ergonomics PSF taking into account the indication of the recirculation valve that was not available in the control room, only locally in the plant. This gives a multiplier of 10. The misleading indication of the open valve is included in the complexity PSF and gives a multiplier of 2. Whether these weights indicate a realistic relation of importance in this scenario is maybe an open discussion. The amount of time for the diagnosis part gives a 0.1 multiplier for this and makes the diagnosis and action contributions similar in the total HEP.

The calculated HEP was 0.015. In the data, all crews managed the HFE, and the rating was fairly difficult to difficult. It was ranked more difficult than HFE 3A but easier than 1C and 2A.

### G6.2 HFE 1B

*G6.2.1 Summary of Qualitative Findings*

The same qualitative analysis as for HFE 1A was the basis for the analysis, except the shorter time available due to the automatic reactor trip, and how this influences certain PSFs like Adequacy of time and stress.

*G6.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP = 0.304 [mean]; (Lower Bound HEP = 2.1E-03, Upper Bound HEP = 8.6E-01) [uncertainty]

**Quantification of relevant factors:**

| *PSF Type* | *HFE1b* |
|---|---|
| Time | 1/1 |
| Stress | 5/2 |
| Complexity | 2/2 |
| Experience | .5/.5 |
| Procedures | 1/1 |
| Ergonomics | 10/10 |
| Fitness | 1/1 |
| Work Processes | .8/.5 |
| **Calc HEP** | 0.4/0.01 |
| **Adjusted for multiple negative PSFs (=or>3)/ Final HEP** | 0.295/.009* |
| **Total Joint HFE (Diagnosis + Action)** | 0.304 |

*G6.2.3 Summary Table of Driving Factors*

The values assigned to the SPAR-H PSFs were the same as for HFE 1A except time available and stress. See the table below.

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | "With the time available for initiating F&B reduced to 13 minutes, operator and crew had only nominal time available for the diagnosis portion and action portions of this HFE." | 0 |
| **Time Pressure** | Part of the stress PSF of SPAR-H. | N/A |
| **Stress** | "Stress was thought to be extreme for HFE1b (x5); the crew is aware that the potential for release is far greater. The crew may also feel that the event has begun to get away from them and this is considered in the high stress PSF assignment. They also knew that they did not meet the performance criteria goals presented to them in reactor training (the unit has a goal of tripping the reactor in 15-30 seconds). Also, if the crew understands that multiple events are under way, this knowledge may contribute to their feelings of stress. For example, it is assumed that after automatic injection is initiated, that the crew knows that only 13 minutes remain before core damage. Some consideration was given to the fact that event HFE1c would be more stressful." | MND |
| **Scenario Complexity** | Unchanged from HFE 1A. | ND |

| | | |
|---|---|---|
| **Indications of Conditions** | "it seemed that information was arranged such that operators had to turn around and switch from panels to PCs to gather information and, hence, a poor ergonomics rating was assigned to diagnosis (10) and actions (10)."<br><br>Otherwise, unchanged from HFE 1A. | MND |
| **Execution Complexity** | Unchanged from HFE 1A. | ND |
| **Training** | Unchanged from HFE 1A. | N/P |
| **Experience** | Unchanged from HFE 1A. | N/P |
| **Procedural Guidance** | Unchanged from HFE 1A. | N/P |
| **Human- Machine Interface** | Unchanged from HFE 1A. | 0 |
| **Work Processes** | Unchanged from HFE 1A. | N/P |
| **Communication** | Unchanged from HFE 1A. | N/P |
| **Team Dynamics** | Unchanged from HFE 1A. | N/P |
| **Other** | Unchanged from HFE 1A. | 0 |

*G6.2.4 Comparison of Drivers to Empirical Data*

Since none of the crews had an automatic reactor trip, all crews manually tripped the reactor, there was no data on HFE 1B.

*G6.2.5 Comparison of Qualitative Analysis to Empirical Data*

As stated, there were no observed cases for this HFE for the operating crews.

*G6.2.6 Impact on HEP*

The team states:

*"The HFE values listed above represent sensitivity to differences between the time available (HF1a versus HF1b). The difference was in the direction expected. With less than one third of the time available to the crew, SPAR-H shows a factor of 20x higher failure rate."*

### G6.3 HFE 1C

*G6.3.1 Summary of Qualitative Findings*

From item 3):

*"The highest expected failure rate for the HFE1 sequence was HFE1C where competing events, poor ergonomics and procedural guidance dictating satisfying conditions for one event over another."*

*"HFE1c refers to failure to isolate the SG after a tube rupture occurs during the above scenario. The event is complicated by complexity from two sources, interpretation of existing indication and preoccupation with F&B with the potential result that operator attention is diverted from the SGTR situation. As a result, we felt that the crew may quite easily end up not isolating the SGTR.*

*As one operator put it, "core cooling trumps tube rupture," so that not taking care of isolating the SG right away may represent a strategy that is not in error. Thus, the HFE is expected to be higher and is represented as such in the SPAR-H analysis. Alternately, lack of SGTR response may be due to lowered awareness, the analysis team identified that focus upon task completion* *for F&B could result in diminished situation awareness regarding additional operation concerns such as SGTR isolation.*

*Determining whether "neglecting" SG isolation is the result of a conscious cognitive decision or a whether it might be the result of tunnel vision issues could be resolved through debrief of crew performance after simulator trials. These data were not available to the authors."*

*G6.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
HEP = 0.916 [mean]; (Lower Bound HEP = 2.7E-01, Upper Bounf HEP = 1.0) [uncertainty]

**Quantification of relevant factors:**

| *PSF Type* | *HFE1c* |
|---|---|
| Time | 1/1 |
| Stress | 5/5 |
| Complexity | 5/5 |
| Experience | 1/1 |
| Procedures | 1/1 |
| Ergonomics | 10/10 |
| Fitness | 1/1 |
| Work Processes | 1/1 |
| **Calc HEP** | 2.5/.25 |
| **Adjusted for multiple negative PSFs (=or>3)/ Final HEP** | .716/.200* |
| **Total Joint HFE (Diagnosis + Action)** | 0.916 |

*G6.3.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Other PSFs than stress, complexity and ergonomics are rated nominal. | N/P |
| **Time Pressure** | Part of the stress PSF of SPAR-H. | N/A |
| **Stress** | "Stress should not be underestimated, the plant is in F&B and will incur economic consequences, the operators stress is extreme for both diagnosis and action, and the stress multiplier = x5". | MND |
| **Scenario Complexity** | Scenario complexity is associated with the diagnosis part of the PSF "Complexity" in SPAR/H.<br><br>"HFE1c refers to failure to isolate the SG after a tube rupture occurs during the above scenario. The event is complicated by complexity from two sources, interpretation of existing indication and preoccupation with F&B with the potential result that operator attention is diverted from the SGTR situation. As a result, we felt that the crew may quite easily end up not isolating the SGTR. Complexity is assigned a rating of x5."<br><br>They also note on the method:<br><br>"The highest influence for complexity allowed by SPAR-H is an upper multiplier of "5". We felt the range associated with complexity is not broad enough to account for the well documented effects of complexity upon human performance. It would be more appropriate for HFEs such as "HFE1c" to allow a multiplier of at least "10."" | MND |

| | | |
|---|---|---|
| **Indications of Conditions** | Part of "Ergonomics/HMI" PSF in SPAR-H.<br><br>"The operator indication present in HFE1c was misleading, however, assignment of x50 in the presence of alternative forms of indication seemed excessive and the differences were mapped to complexity. The HEP was already at .90 and adding a large multiplier was not necessary to call attention to the vulnerability associated with the context of the situation." | MND |
| **Execution Complexity** | Execution complexity is associated with the action part of "Complexity".<br><br>The analysis team assigns both the action and the diagnosis part of the SPAR-H PSF Complexity the same weight: multiplier 5 or "highly complex". However, as in HFE1A, they do not explain why the action part is also complex, but they have only one explanation, which seems to be more directed to the difficulties in the diagnosis. | 0 |
| **Training** | The factor in SPAR-H is "Experience/Training".<br><br>The same overall judgement for training and experience was used for HFE 1C as for the other HFEs. | N/P |
| **Experience** | The factor in SPAR-H is "Experience/Training". | N/P |
| **Procedural Guidance** | As also noted for the other HFEs:<br><br>"Procedures were assumed to be industry standard, a quick review did not reveal a technical basis for assuming that they were better than industry standard.<br><br>Procedures usage: It was observed during observation of a simulator exercise that operators wrote on procedures, checked off for place keeping, and during an event the procedure could become a record. Operators may use the procedure to markup and use as a timeline."<br><br>"..depending on other contextual factors, procedure following may have the crew ignore isolating the SG for quite a while.." | N/P |
| **Human- Machine Interface** | Part of "Ergonomics/HMI" PSF in SPAR-H. No additional comments further than what mentioned under "Indications of conditions" above. | 0 |
| **Work Processes** | The same overall judgment for work processes and communication was used for HFE 1C as for the other HFEs. However, the multiplier in the analysis for HFE1C was set to 1, compared to 0.8/0.5 for HFE 1A and 1B. Apparently they have made a decision on changing the multiplier without documenting it in the analysis write-up. | N/P |
| **Communication** | In SPAR-H, included in "Work Processes". | N/P |
| **Team Dynamics** | In SPAR-H, included in "Work Processes" | N/P |
| **Other** | Fitness for Duty. "From the scenario description, fitness for duty factors had no bearing on this analysis and the nominal value (1) was assigned." | 0 |

After having evaluated the complexity and stress, it seems that they stop detailed evaluation of other PSFs, since they state *"Other PSFs have been evaluated as nominal. There is enough negative PSF influence that the HFE is very high (HFE = 0.9)".* Thus time etc is evaluated to be nominal in the analysis.

*G6.3.4 Comparison of Drivers to Empirical Data*

The INL 1 team (Gertman) SPAR-H analysis identified three negative drivers on the performance of HFE 1C: Stress, Complexity (scenario) and Indication of conditions. All of these are assessed as being main negative drivers, since the stress and complexity are maxed out (including an explanation on that the complexity weight ought to be higher even), and the SPAR-H PSF Ergonomics is set to a weight of 10 (including an explanation on that it could in isolation have been judged to be 50, but a consideration of the value of the HEP made them decide to leave it at 10).

In the empirical data, the main negative driver was considered to be the scenario complexity, since the tube rupture was initially masked by the AFW flow to the ruptured SG. It was also noted that when the tube rupture occurred, the crews already had an emergency situation and were working in FR-H1. These issues are well described by the INL 1 SPAR-H team when explaining the complexity PSF ("preoccupation with the F&B", etc), even though they don't go into details on which procedures that were in use and exactly how the flow was masking the tube rupture. Indication of conditions was thus also noted in the empirical data as a negative driver, as also identified by the SPAR-H analysis. Procedural guidance was noted as a negative driver in the data, since it was not easy to manage the SG isolation while they were "stuck" in the procedure for F&B (FR-H1 (and FR-P1)). The INL 1 SPAR-H team notes this goal conflict several places, even in a comment on procedures: *"procedure following may have the crew ignore isolating the SG for quite a while".* However, they did not change the weight for the procedures PSF, but kept it nominal. The last negative driver in the data was the adequacy of time due to the 40 minutes period set up as a goal in the HFE. Time was considered nominal in the INL 1 team SPAR-H analysis.

Stress is not observed in the data, but predicted negative by this analysis. (Whether this then has to do with the difference between training and real life remains unknown and should not be emphasized in this comparison). Stress might not be self-reported because it is expected.

*G6.3.5 Comparison of Qualitative Analysis to Empirical Data*

Also in this HFE, the SPAR-H analysis includes discussions on scenario developments when describing the complexity and indication of conditions PSFs. They correctly note the existence of two goals, F&B and SG isolation, and some of the difficulties around this. These considerations are still on a quite high level, not going into details of difficulties in procedures. They don't mention in which procedures difficulties may occur. Also, the procedure PSF seems to be treated on a high level. Note that this is the impression from the PSF analysis. In the interview write-ups they go more into details on procedures.

*G6.3.6 Impact on HEP*

The main contributors to a high HEP are the PSFs ergonomics, complexity, and stress. A quite simple analysis, but with quite good justification, was the basis for these weights. This may be an indication that for such difficult HFEs it is possible to get to a reasonable HEP with the SPAR-H method, without going into detail on the way in which the procedures are followed in detail, for example.

In the empirical data, 3 out of 4 failed given the 40 minutes time criterion, while 1 out of 4 if the time criterion was not included. It was rated difficult and ranked as the second most difficult, after HFE 2A.

## G6.4 HFE 2A

*G6.4.1 Summary of Qualitative Findings*
From item 3) in the analysis:

*"HFE2a is moderately complex, there is a loss of a distribution panel (DP), the RCPS have to be tripped, the operators must recognize low flow or no flow from CCW and have to start a positive displacement pump. The crew loses several pieces of equipment and must provide manual control and stay on top of providing feed flow to the SGs (A & B). The SG will trip on high level. The feed REG valve is failed open and cannot be closed. There are no CCW pumps available to the crew. If sealwater is not provided to the RCPs via the PDP, the lube oil temperature will exceed 230 degrees, and extensive damage to the pumps will occur."*

*G6.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
HEP = 0.033 [mean]; (Low Bound HEP = 1.4E-04, Upper Bound HEP = 1.2E-01) [uncertainty]

**Quantification of relevant factors:**

| *PSF Type* | *HFE2a* |
|---|---|
| Time | 1/1 |
| Stress | 2/2 |
| Complexity | 2/1 |
| Experience | 1/1 |
| Procedures | 1/1 |
| Ergonomics | 1/1 |
| Fitness | 1/1 |
| Work Processes | .8/.5 |
| **Calc HEP** | 0.032/.001 |
| **Adjusted for multiple negative PSFs (=or>3)/ Final HEP** | 0.033 |
| **Total Joint HFE (Diagnosis + Action)** | 0.033 |

*"Although crew performance related to this HFE seemed difficult at first, after applying SPAR-H criteria, success was deemed more likely for HFE2 than for HFE1c."*

*G6.4.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | "Time available and time required. The scenario specifies that only 7-9 minutes are available before seal injection is needed. Operators estimate that from directly after the RX trip, operators would go to immediate actions that are done from memory (2-3min). They would then pause assess the situation and would trip the coolant pumps. The time for this whole series is estimated to take 5 minutes. Pump controls are located together on the boards. For purposes of this analysis we assumed that the event would go very fast. In responding, operators will note that the 1A charging pump is out of service, should they demand the 1B charging pump, it will trip shortly thereafter. In terms of time to respond, the analysis team used the lower bound of 7 minutes and decided that there no room for additional margin, i.e., extra operator time as defined by SPAR-H. Therefore the SPAR-H time PSF was assessed as nominal time available, PSF multiplier = 1." | N/P |
| **Time Pressure** | Part of the stress PSF of SPAR-H. | N/A |

| | | |
|---|---|---|
| **Stress** | "Multiple faults and multiple equipment are present and the relatively fast pace of the event was cause to assign a multiplier of x2 for elevated operator stress during diagnosis. The same stress multiplier is assigned for actions since the two are so highly integrated." | ND |
| **Scenario Complexity** | Scenario complexity is associated with the diagnosis part of the SPAR-H PSF "Complexity."<br><br>"According to in-depth operator interviews, the scenario is moderately complex for diagnosis, (SPAR PSF assignment = 2)" | ND |
| **Indications of Conditions** | Part of "Ergonomics/HMI" PSF in SPAR-H.<br><br>"Visual indication (cues) is good for observing reduced flow to RCPs post-trip." | N/P |
| **Execution Complexity** | Execution complexity is associated with the action part of the SPAR-H PSF "Complexity."<br><br>"our review concluded that once the cognitive portion of the task was concluded that the SPAR PSF for action execution was of nominal complexity and guided by procedures (multiplier =1)." | N/P |
| **Training** | The factor in SPAR-H is "Experience/Training"<br>Considered nominal. | N/P |
| **Experience** | The factor in SPAR-H is "Experience/Training" | N/P |
| **Procedural Guidance** | Same overall analysis as for HFE 1A, 1B and 1C.<br><br>"Procedures were observed to be adequate but not above nominal in quality." | N/P |
| **Human- Machine Interface** | Part of "Ergonomics/HMI" PSF in SPAR-H. | 0 |
| **Work Processes** | Same overall analysis as for HFE 1A, 1B and 1C.<br><br>"Work processes were observed to be very good during simulator trials for another scenario and SPAR-H work process PSF assignment was assigned .8 for diagnosis and .5 for action according to SPAR-H guidelines." | N/P |
| **Communication** | In SPAR-H, included in "Work Processes" | N/P |
| **Team Dynamics** | In SPAR-H, included in "Work Processes" | N/P |
| **Other** | Fitness for Duty – assumed not an influence | 0 |

For this HFE, the analysis team separates between the action and the diagnosis part for several PSFs, as I missed in the HFE 1A, 1B and 1C analyses.

*G6.4.4 Comparison of Drivers to Empirical Data*

The INL 1 team (Gertman) SPAR-H analysis identified only two negative drivers on the performance of HFE 2A: High stress and moderate scenario complexity. Both of these were thus considered elevated, but not having a very high impact (both with a multiplier of 2). In the empirical data, both of these were observed, although stress was observed but not reckoned to be a driver for the performance of the crews. More importantly, scenario complexity was noted to be a main negative driver, since many things were happening at the same time that made it difficult to detect the priority items. The current SPAR-H analysis seems to have underestimated the impact of the complexity, even though they identified it.

In the empirical data, the main negative driver in addition to the scenario complexity was the training and experience. As stated in the chapter 4, empirical data document: "They were not used to this kind of scenario and the needed prioritizations." "They were used to train on loss of CCW and sealwater only in loss of offsite power scenarios." The INL 1 team SPAR-H analysis considered the training to be nominal.

Adequacy of time was also singled out as a negative driver in the empirical data. It is noted in relation to training though: "if they had had more training there would have been enough time". The allowed time span for all the actions in this scenario is a few minutes. In the SPAR-H analysis, they state that operators should be able to perform immediate actions from memory within the first 2-3 minutes. It seems that these assumptions coming from interviews at the plant do not hold in the present situation in the scenario, given the complexity of the situation and the lack of training on this specific scenario. Whether more detailed interviews, or a more detailed qualitative analysis would lead to another conclusion is hard to say.

Work processes was also deemed a negative driver in the empirical data, identifying problems in monitoring of control boards, acknowledging alarms and reading the POP4 procedure (Off Normal Procedures).

Indication of conditions was considered a negative driver in the empirical data, since a lot of indications were gone because of the distribution panel 1201 failure. The SPAR-H team stated that "the visual indications were good for observing reduced flow to RCPs post-trip". The team was informed through the interview that although other indication was lost there was at least one source where reduced flow could easily be observed. However, it appears through the empirical data that the crew overestimated their ability to find the relative data quickly and easily.

Procedural guidance is also noted to be a negative driver in the data. When the crews were late in detecting the loss of CCW and sealwater, they simply could not move through procedures fast enough to start the PDP in time. The SPAR-H analysis notes the procedures to be nominal. It seems that they have not analysed in detail the use of procedures in case the crews had a time problem from the start. The combined effect of lack of detailed training, the complexity of the situation and the time span of a few minutes, seemed to have been underestimated by the SPAR-H analysis.

*G6.4.5 Comparison of Qualitative Analysis to Empirical Data*

In the empirical data, it was clear that all the crews were late in detecting the loss of CCW and sealwater, causing them to trip the RCPs late and also gave them too little time to start the PDP before the RCP sealwater temperature reached 230 °F, then they simply could not move through the procedures fast enough. The main reason for them being late was that they lacked training on the specific type of scenario, the situation was very complex and they had a problem recognizing the urgency. The SPAR-H team doesn't question the combination of these factors and they don't analyze in details how some of these issues impact performance. The team stated that they relied on interviews when assessing the HFE. It might be two reasons for the mismatch based on this: the interviews paid to little attention to focus on details, or the self-assessment of the interviewees was not sufficient.

*G6.4.6 Impact on HEP*

The SPAR-H team states: *"Although crew performance related to this HFE seemed difficult at first, after applying SPAR-H criteria, success was deemed more likely for HFE2 than for HFE1c."* Given that this was considered to be the most difficult HFE, four out of four crews failed, this analysis was not correct. The impact of the factors and the combined impact of the factors (training, complexity, lack of time, etc) were not recognized by this analysis.

An interesting point to discuss is why the analysis of HFE 2A seems to have missed the difficulties that occurred for the crews, while the analysis of HFE 1C correctly identifies many of the difficulties in that HFE. The HRA team's response on this: "*Complexity and stress were both rated as non nominal, but perhaps not highly enough by the team. In terms of time as a factor…7-9 minutes were available and our estimate was 5 min to complete, we used 7 min the lower bound as the time available and still came up with 40% margin which is nominal in SPAR. Method modifications may consider defaults for particularly short time intervals, we were bit encumbered. SPAR-H PSF assessment looks at time required/time available. Our team error (and we couldn't verify in simulator) was whether the 5 min estimate we used based on operator interviews was low (it was). This would have shifted our estimate to 3E-1".* In the team's defence, for most PRAs, an HEP of E-2 is considered conservative or realistic and gives cause for more detailed analysis.

### G6.5 HFE 3A

*G6.5.1 Summary of Qualitative Findings*
From item 3) in the analysis:

*"SGTR is a well practiced and "trained to" scenario, thus, performance during events described in Scenario 3 does not present with any aspects that challenge the SAR review of this event. The type of SGTR break encountered in Scenario 3 is a circumferential as opposed to axial break.*

*The SPAR-H complexity factor was used to represent operator consensus that this was an obvious diagnosis, not easily confused with anything else. [SPAR-H for low power calls out rates for obvious diagnosis, but currently SPAR-H for at power events does not do so.] However, action execution was deemed nominal, and neither a negative or positive influencing factor for the current scenario."*

*G6.5.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
HEP = 0.0002 [mean]; (Lower Bound HEP = 1.0E-05, Upper Bound HEP = 7.7E-04) [uncertainty]
**Quantification of relevant factors:**

| *PSF Type* | *HFE3a* |
|---|---|
| Time | .1/.1 |
| Stress | 2/2 |
| Complexity | .1/1 |
| Experience | 1/1 |
| Procedures | 1/1 |
| Ergonomics | .5/1 |
| Fitness | 1/1 |
| Work Processes | .8/.5 |
| **Calc HEP** | 0.00008/.0001 |
| **Adjusted for multiple negative PSFs (=or>3)/ Final HEP** | 0.00018 |
| **Total Joint HFE (Diagnosis + Action)** | 0.0002 |

*G6.5.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | "The time required for HFE3A is relatively favorable for both diagnosis and action. Operators indicated that the time required to diagnose was on the order of 2 minutes and that the time required to act including isolation of the SG was on the order of 15-20 minutes with 2-3 hours being available. Examining the time available ratios for combinations of these ranges (3 hours and 15 minutes (12:1) and 2 hours and 20 minutes( 6:1)) suggests that there is extra time available for diagnosis. Deducting the diagnosis time, the time remaining for action is still equal to >5x the time available (lower bound ratio of 1.5 hours to 12 minutes)." | N/P |
| **Time Pressure** | Part of the stress PSF of SPAR-H. | N/A |
| **Stress** | "Two operators said that this would be a high stress event because of the potential of radiation release and what this would mean for the plant. SPAR-H multiplier was applied to both diagnosis and action components of this HFE." | ND |
| **Scenario Complexity** | "The SPAR-H complexity factor was used to represent operator consensus that this was an obvious diagnosis, not easily confused with anything else." | N/P |
| **Indications of Conditions** | "Ergonomics was rated as .5 as the operators stated that it was easy to see and access everything needed for diagnosis, however, the ergonomics for taking actions were observed to be nominal and a multiplier of 1 was used." | N/P |
| **Execution Complexity** | "action execution was deemed nominal, and neither a negative or positive influencing factor for the current scenario." | 0 |
| **Training** | "SGTR is a well practiced and "trained to" scenario".<br><br>Despite this statement, the team used the "nominal" SPAR-H weight is used, not the "high" with multiplier 0.5. | 0 |
| **Experience** | The factor in SPAR-H is "Experience/Training". | 0 |
| **Procedural Guidance** | "Procedures were rated 1 to coincide with the fact that although they were symptom-based they were also only industry standard and what was to be expected." | N/P |
| **Human- Machine Interface** | Nominal. | N/P |
| **Work Processes** | "During non-related simulator trials the crews were observed to use a good system of call backs (3-peat) and to work well as a team, taking instruction from the unit supervisor. As a result of those observations a work process shaping factor value of .8 and .5 were used in calculating the HFE." | N/P |
| **Communication** | In SPAR-H, included in "Work Processes". | 0 |
| **Team Dynamics** | In SPAR-H, included in "Work Processes". | 0 |
| **Other** | Unchanged from HFE 1B. | 0 |

*G6.5.4 Comparison of Drivers to Empirical Data*

The INL 1 team SPAR-H analysis identified stress as the only negative driver of performance in this scenario. This was not observed at the training simulator when the scenario was run.

The SPAR-H analysis identified adequacy of time, complexity, indications of conditions, procedures and work processes as being positive factors contributing to good performance in this scenario.

In the empirical data, the indication of conditions were considered positive, since the radiation alarms and indications helped the crews identify the SGTR quickly. Also, the HMI was rated positive. The procedures were also considered positive, since they supported this base-case scenario well. The crews also showed examples of good teamwork. Thus all the factors noted positive by the SPAR-H analysis was confirmed by the empirical data. In the empirical data, also the training and experience was noted positive, since this was a frequently trained scenario with no added complications.  This was also noted by the SPAR-H team, but not reflected in the PSF weight.

*G6.5.5  Comparison of Qualitative Analysis to Empirical Data*
This scenario was designed to be a standard ("vanilla") PSA scenario, and it turned out to be as well.  It is often trained and everything ran as expected following well prepared procedures and good indications.  The SPAR-H analysis predicted this, and it seems that for standard scenarios the need to go into details in scenario developments in order to find out the difficulties for the crews is less prominent than with more difficult scenarios.  So SPAR-H seems to be a sufficient tool for this kind of scenario.

*G6.5.6  Impact on HEP*
In the empirical data, all crews succeeded this HFE, and the scenario was simple for the crews. The HEP from the SPAR-H analysis is 2E-4, reflecting the low probability for failure.  This shows that SPAR-H has tools to decrease the failure probability from the nominal values.

## G6.6  Comparison Summary

*G6.6.1  Predictive Power*

G6.6.1.1  Qualitative Predictive Power – in Terms of Drivers

In HFE 1A, the analysis correctly identified the Complexity and the Indication of conditions as the main negative drivers. Procedural guidance was also noted to be negative in the empirical data, and in this SPAR-H analysis procedures were noted to be nominal. However, they did note the reason for why procedures were rated negative in the empirical data, which was due to the CSF heat sink status tree that does not address the misaligned valve. In the SPAR-H analysis they accounted for this fact under the Complexity PSF. So qualitatively, they captured all the main drivers for HFE1A.

In HFE 1C, they similarly correctly identified the reasons for negative Complexity and Indications of conditions PSFs, including the goal conflict between the F&B and the tube rupture.  They also noted the procedural difficulties on this point, but did however rate the procedures nominal.  At this point they had a very high HEP and did not look for further negative PSFs.

HFE 2A was the most difficult one according to the empirical data.  The SPAR-H analysis did not manage to identify the combined negative effect of lack of detailed training, the complexity of the situation including many alarms and loss of indications, and the time span of a few minutes.  Also, they did not predict that when the complex situation was detected late, the crews did not have time to move through the procedures.  It seems that their analysis was too superfluous to detect the drivers for this difficult HFE.

For HFE 3A, this SPAR-H analysis noted several positive PSFs as were also seen in the empirical data, so the analysis was good for this HFE.

In several HFEs, stress was not observed in the data, but predicted to be negative by this analysis.

In many of the PSF evaluations, the team does not explicitly state the differences between the diagnosis part and the action part of the analysis. They do this for the time available, but e.g., for the complexity they don't explain what is complex in the diagnosis and what is complex for the action part of the HFE. An explanation is maybe that in chapter 10 in the main analysis they discuss complexity impact on both diagnosis and action and seem to have concluded that it impacts both. It is however questionable whether it impacts both in the same way.

Regarding the separation of the PSFs complexity and the ergonomics in SPAR-H, the team clearly states what is considered in which PSF and states the way they have chosen to analyze it so not to double-count any effects. This is good in this analysis. It maybe shows the overlap of the definitions of these two PSFs though. However, many PSFs do overlap in real life and as long as they explain the way in which they have done their analysis, this is ok.

The overall judgment of the qualitative prediction of the drivers is good to fair. For all the HFEs but 2A they have described the drivers in a very good way, but for 2A they miss out. The reason for this might be that they have not analyzed this scenario as thoroughly as the others.

#### G6.6.1.2 Qualitative Predictive Power – in Terms of Operational Expressions

There is no additional qualitative analysis in addition to the one that is done in relation to the PSF analysis, at least not documented. This is typical for SPAR-H, since it is a purely factor based method.

The analysis correctly pointed out some of the difficult operational details in the scenarios when discussing the complexity and indication of conditions. Especially, they noted the misleading indication of the open recirculation valve and that this would mislead the operators. This was also found in the empirical data. However, they didn't note the exact place in the procedures where this would have an impact. They did not go to this level of detail (which the other SPAR-H team did), at least they did not document this. It seems that the judgments of these PSFs are built on pretty good scenario insight though.

To this assessor it seems that a more detailed scenario analysis including procedure branching and timing would have improved the analysis for the difficult HFE 2A. It is the impression of the assessor (in the assessment group) that SPAR-H doesn't manage to capture the nature of the difficulty of a difficult HFE if they don't do this. For HFE 2A they didn't manage to predict the difficulties that occurred. They might have managed this if they had analyzed it in more detail. This hypothesis is built on the fact that the rather difficult HFEs 1A and 1C were analyzed more in detail, and the analyses were good fits to the empirical data. It is the assessor's impression that this was accomplished due to a more detailed scenario understanding. The team stated that they relied on interviews when assessing the HFE. It might be two reasons for the mismatch based on this: That the interviews were too little focused on details, or that the self-assessment of the interviewees was not sufficient.

For the easy ("vanilla") PSA scenario and HFE though, an overall superfluous analysis seems to be enough and SPAR-H makes an ok assessment. This is the case if the crews don't come into trouble though. For example, HFE 3A is often trained and everything ran as expected following well prepared procedures and good indications. The SPAR-H analysis predicted this. It seems that for standard scenarios the need to go into details in scenario developments in order to find out the difficulties for the crews, is less prominent than with more difficult scenarios. One could conclude that SPAR-H is a sufficient tool for vanilla scenarios that are very often trained, or for scenarios that the analyst is highly familiar with.

The judgment of the qualitative predictive power in terms of the operational expressions is fair for the HFEs 1A, 1B and 1C, while for 2A and 3A it is judged to be moderately poor.

G6.6.1.3 Quantitative Predictive Power

In Figure 0.6, the HEPs from the analysis are plotted. The sequence of the plots are according to the ranking of the empirical data, in which HFE 2A was judged to be very difficult, 1C difficult, 1A fairly difficult to difficult, and 3A easy.

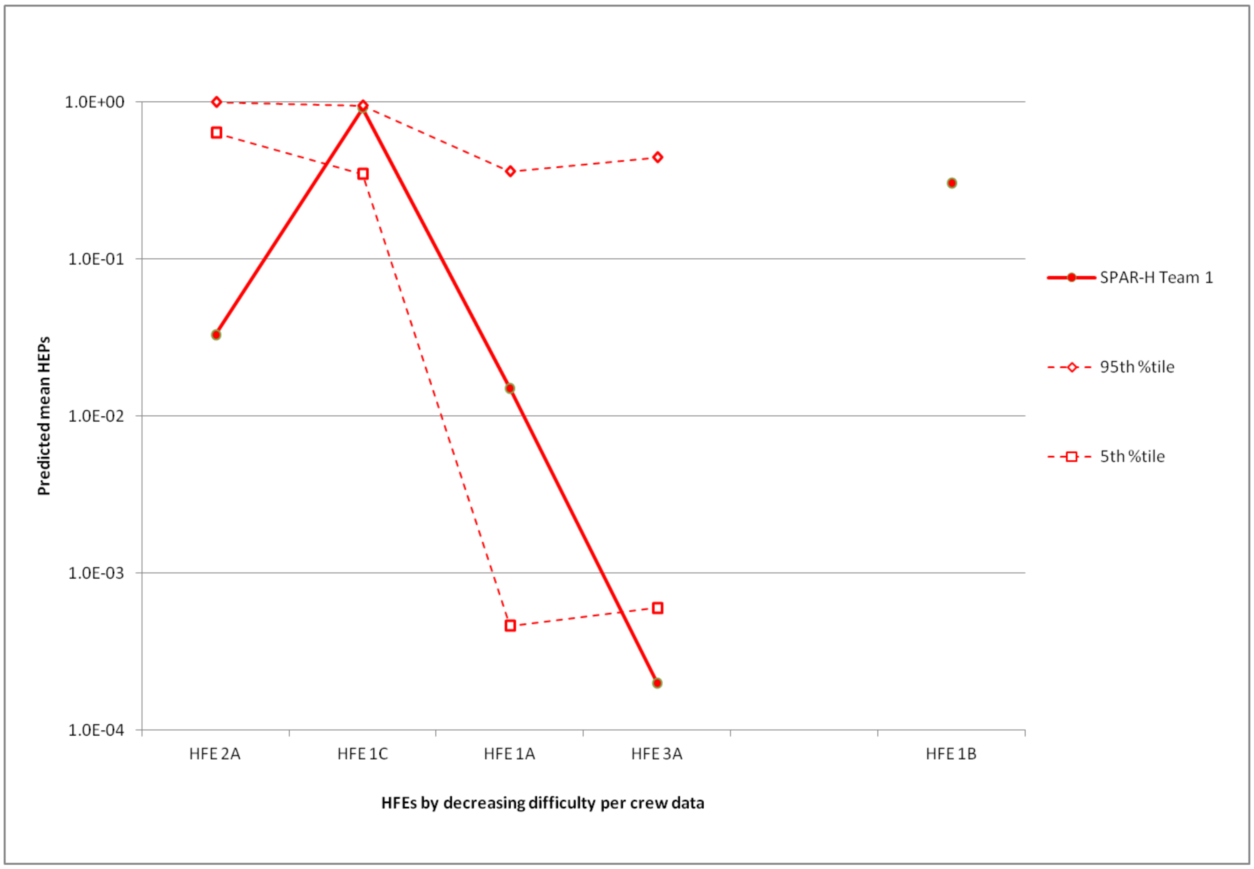



Figure 0.6 Predicted Mean HEPs with Uncertainty Bounds by INL1 HRA Team with SPAR-H

Four out of four crews failed HFE 2A, so it seems that the analysis really missed the severity of this most difficult HFE. The reasons are described above. Another thing with 2A compared to 1A, 1B and 1C, is that it is a self-standing HFE in a separate scenario. The way the analysis was documented, where they set up all the PSF weights of all the HFEs side by side in a table, it should be easy to compare the three HFEs in the similar scenario 1A, 1B and 1C for a realism check. HFE 2A is quite different from the other HFEs and would be more difficult to compare in this way even though it is included in the same table as the other HFEs. Another point regarding HFE 2A is the judgment of time available. SPAR-H considers time required/time available.  If there's a 40% margin this is considered nominal. In the case of 2A they misjudged this based on the operator interviews, so they estimated nominal while in reality it was not. SPAR-H should consider treating short time intervals as a special case, since the sensitivity for wrong analysis is higher for time estimates of human actions to be done within a few minutes.

HFE 1C was also considered difficult, and the value of the HEP is reasonable for this one, without us being able to judge the value as such due to a quite weak quantitative basis of the empirical data, with only four crews.

Regarding the ranking of the HFEs, the analysis fails to see that HFE 2A was more difficult than 1C, otherwise the ranking is fine.

It is a bit difficult to judge whether the predicted numbers are within the confidence bounds (Bayesian), since we have so sparse quantitative data. However, based on the qualitative judgments of the empirical data, the HEPs for 1C, 1A and 3A seem reasonable.  1C was rated difficult, and they got a high HEP. HFE 3A was considered easy, none of the crews had problems of any sort, and the predicted HEP is very low which seems reasonable. We cannot really judge whether 1A and 3A are within one order of magnitude, but they seem reasonable at least.

The analysis differentiates very much between the HFEs 1C, 1A and 3A.  This seems reasonable given the empirical data, and in this respect the SPAR-H (INL 1) analysis is good.

The quantitative predictive power for this analysis is overall judged to be fair.  Without the miss of 2A, it would have been good.

*G6.6.2  Traceability of the qualitative analysis and quantification process*

The quantification process is very transparent in SPAR-H, when first the weights of the PSFs are decided, it is easy to know how the numbers arise.  In this analysis they did not attach the SPAR-H worksheets, which contain the description of the PSFs with the exact multipliers, and the formulas for how these are used on the nominal value and adjusted in case of many PSFs and also how dependencies are taken into account.  Thus the transparency of this specific analysis was not as good as it could have been.  The assessor used an empty SPAR-H worksheet in order to get this clear.  So in every SPAR-H analysis these worksheets should be attached in order to simplify the traceability of the analysis.  The traceability of the quantification itself is deemed to be good for SPAR-H.

The traceability of how the qualitative analysis leads to the PSF weights is not that obvious in SPAR-H though. In this analysis they did a good description of which predicted crew difficulty went into which PSF. There is often an overlap between PSFs. In this analysis a typical issue could be covered by either the Complexity or the Ergonomics (Indication of conditions in our analysis) PSFs, it is even an overlap with Procedures.  They explained well what phenomenon was accounted for in which PSF, and also used this to avoid double-counting of effects.

Another issue also seen before in SPAR-H, is the justification of the weight of the multiplier for each PSF.  Traceability of this relies on to which extent they explain the scenario development in operational terms and what exactly produces difficulties for the teams (especially for the complexity issues).  If they do a detailed scenario analysis and documents this, it might be good enough.  However, if they only judge the overall PSF on an overall level, it is difficult to trace this back and to repeat it.

The traceability of the basis for the quantification, the PSF ratings and the qualitative analysis, is thus moderately poor.

*G6.6.3 Insights on guidance*

The guidance for the weight of the multiplier for each PSF could be improved. In this analysis, they have quite good justifications for their choices, but the guidance could be improved regarding this.

*G6.6.4 Insights for Error Reduction*

The analysis describes the misleading indicators and some other aspects contributing to degraded indications of conditions and complexity. This information can be used to improve the situation in the plant, either by including these in the training or checking the indicators. Otherwise, there is not too much information to help error reduction at the plant.

G.6.1.1 *Remarks on strengths and weaknesses*

The conclusions for SPAR-H will be built on all the analyses by all the SPAR-H teams for this project. Here are just some remarks so far.

SPAR-H is very sensitive to the choice of which PSF to use and also which weight is chosen for the PSF for the phenomenon that is analyzed. In this analysis this can be seen where they have the choice to use the complexity PSF or the Ergonomics PSF to describe a complex situation caused by misleading or even missing indicators in the displays. The missing indicators are very often linked to a general complex situation in the scenario, so the choice is not straightforward. The ergonomics PSF has much higher weights than the complexity PSF in SPAR-H, so this choice has impact on the HEP (see also below). When the PSF is chosen, one has to choose the weight, and this is not straightforward either. This is an expert judgment, and it should be based on detailed knowledge of the operational situation in the scenario. If not, SPAR-H can be used in a too simplistic way. The guidance for this is not very well described either, e.g., there are no anchor points described in the guidance.

The SPAR-H team discusses the point regarding the multipliers of complexity: *"The highest influence for complexity allowed by SPAR-H is an upper multiplier of "5". We felt the range associated with complexity is not broad enough to account for the well documented effects of complexity upon human performance. It would be more appropriate for HFEs such as "HFE1c" to allow a multiplier of at least "10.*

## G.7 SPAR-H (INL 2)

Overall, the rationale behind the NRC SPAR-H analysis (INL 2 team) is fairly easy to follow; however the primary PSF information affecting each HFE is found in the analysis forms for the basic events rather than in Form A. This made it a bit difficult to determine how many basic events are affected by a given PSF. To assist with that, for each scenario, the SPAR-H PSF table is presented, with the rating for a given PSF indicated by the basic event/ HFE where it was rated. For all scenarios, the HFEs are broken down into basic events that consist of diagnosis and/or action. The qualitative analysis broke the high level events into basic events, as recommended in the SPAR-H method. However, rather than using the 10 step ATHEANA process outlined in the method, the analysts developed a Crew Response Tree (CRT).

SPAR-H provides minimal guidance on the qualitative analysis. It recommends that the analysts follow the 10-step ATHEANA search process to determine the tasks and contexts to be rated. This process is not the only technique that can be used. The analyst can use any structured method to determine the tasks and contexts, or the analyst, if appropriate, can analyze a scenario at a high level. Therefore, the SPAR-H team conducted its own qualitative analysis for each HFE, in some cases performing a procedural walk-through, to develop a high-level Crew Response Tree (CRT) for the scenarios, identifying the main failure paths.

All scenarios were at 100% power; therefore, the at-power worksheets were used. Each scenario was analyzed as a combination of Diagnosis, Action, and mixed diagnosis/action tasks. The basic events defined by the analysis were assessed for dependence. HEPs were determined using the appropriate formula for dependence.

Table of relative rank of each scenario as rated in the empirical data and through the NRC SPAR-H analysis

| HFE | Task | Unit Sup. rank | Failure rate | Difficulty | NRC SPAR-H estimated HEP | NRC SPAR-H RANKING |
|---|---|---|---|---|---|---|
| 2A | Stop RCPs and start PDP in scenario 2 | 1 | 4/4 | Very difficult | 1.14E-01 | 2 |
| 1C | Identify and isolate ruptured steam generator in scenario 1 | 2 | ¼ | Difficult | 1.1 E-01 | 3 |
| 1A | Start bleed and feed in scenario 1 | 3 | 0/4 | Fairly difficult to difficult* | 1.24E-01 | 1 |
| 3A | Identify and isolate ruptured steam generator in scenario 3 | 4 | 0/3 | Easy | 7.0 E-2 | 4 |
| 1B | Failure to establish F&B within 13 minutes of the reactor trip, given that the crews do not manually trip the reactor before an automatic reactor trip occurs. | No data | n/a | No data | 9.83 E-01 | Highest HEP |

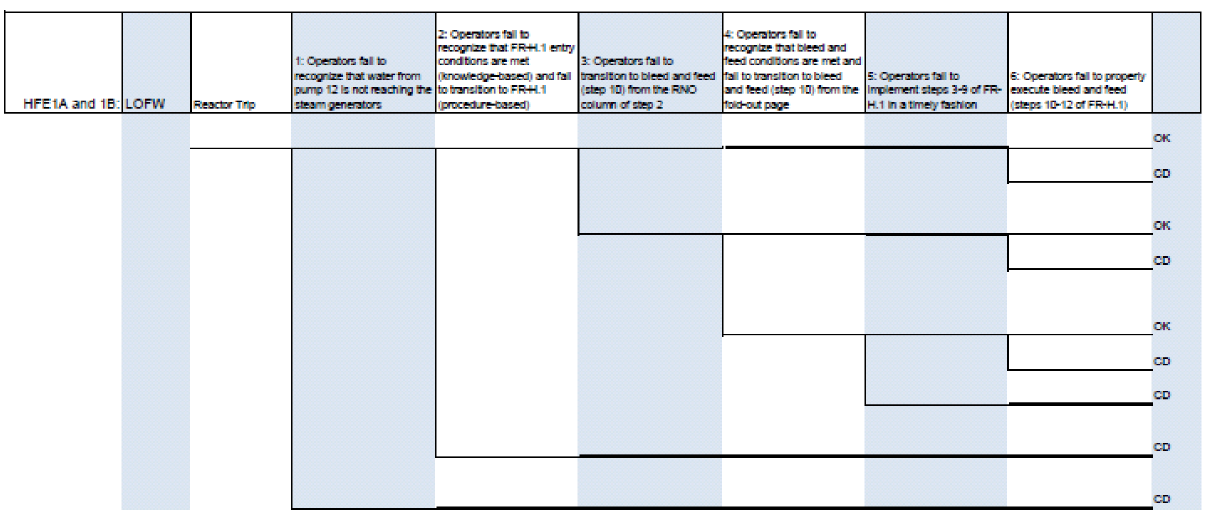
## G7.1 HFE 1A

*G7.1.1 Summary of Qualitative Findings*

In the case of HFE 1A, the event was broken down into 6 sub parts, for which an error rate was created, and the rates summed. Basic events HFE1A(1), (2), and (6) were considered by the analysts to be independent; HFE1A(3), (4), and (5) were dependent upon each other. Subparts (1), (2), (3), (4), and (5) were analyzed as pure diagnoses; subpart (6) (executing bleed and feed) was the action component of the HFE and was analyzed as a combination Diagnosis and Action following SPAR-H guidance. The basic events identified were:

HFE1A(1), Operators fail to recognize AFW pump 12 is not flowing to the SG;
HFE1A(2), Operators fail to transition to FR-H1;
HFE1A(3), Operators fail to transition to bleed and feed (FR-H1 steps 10-12) from the FR-H1 step 2 RNO;
HFE1A(4), Operators fail to transition to bleed and feed (FR-H1 steps 10-12) from the fold-out page of FR-H1;
HFE1A(5), Operators fail to implement steps 3-9 of FR-H1 in a timely manner;
HFE1A(6), Operators fail to properly execute bleed and feed.

The CRT for HFE 1A identified the event paths:




The breakdown of the events into basic events was driven not by guidance from SPAR-H, which recommends using the ATEHEANA process; rather the analysts developed a CRT. However, this may be an accurate reflection of how an event would be broken down by analysts performing HRA for a plant.

*G7.1.2 Quantification of relevant SPAR-H PSFs*

Summary table of SPAR-H PSFs Across All Basic Events for HFE1A:

| PSFs | PSF Levels | Action | Diagnosis |
|---|---|---|---|
| Available Time | Inadequate Time | P(failure) = 1.0 | |
| | Barely adequate time (~ 2/3 x nominal) | 10 | |
| | Nominal time | 1 | |
| | Extra time (between 1 and 2 x nominal and > 30 min) | 0.1 | HFE1A(1) DIAGNOSIS |
| | Expansive time (> 2 x nominal and > 30 min) | 0.01 | HFE1A(4) ACTION |
| | Insufficient Information | 1 | |
| Stress | Extreme | 5 | |
| | High | 2 | |
| | Nominal | 1 | |
| | Insufficient Information | 1 | |
| Complexity | Highly complex | 5 | |
| | Moderately complex | 2 | HFE1A(1) DIAGNOSIS; HFE1A(2) Diagnosis |
| | Nominal | 1 | |
| | Obvious diagnosis | 0.1 | |
| | Insufficient Information | 1 | |
| Experience /Training | Low | 10 | |
| | Nominal | 1 | |
| | High | 0.5 | HFE1A(6) ACTION |
| | Insufficient Information | 1 | |
| Procedures | Not available | 50 | |
| | Incomplete | 20 | |
| | Available, but poor | 5 | |
| | Nominal | 1 | |
| | Diagnostic/symptom oriented | 0.5 | HFE1A(3) DIAGNOSIS; HFE1A(5) DIAGNOSIS |
| | Insufficient Information | 1 | |
| Ergonomics /HMI | Missing/Misleading | 50 | HFE1A(1) DIAGNOSIS |
| | Poor | 10 | |
| | Nominal | 1 | |
| | Good | 0.5 | |
| | Insufficient Information | 1 | |
| Fitness for Duty | Unfit | P(failure) = 1.0 | |
| | Degraded Fitness | 5 | |
| | Nominal | 1 | |
| | Insufficient Information | 1 | |

| | | | |
|---|---|---|---|
| Work Processes | Poor | 2 | |
| | Nominal | 1 | |
| | Good | 0.8 | |
| | Insufficient Information | 1 | |

*G7.1.3 Summary table of Driving Factors for HFE1A*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Although this is a complex task which must be done somewhat timely, there is more than sufficient time to perform it [sub HEP (1)]. | N/P |
| **Time Pressure** | Although this is a complex task which must be done somewhat timely, there is more than sufficient time to perform it. | N/P |
| **Stress** | The operators are not under any physical duress. The control room environment is not inhospitable. Any perceived sense of urgency is unlikely to contribute to errors. Any "stress" in this scenario would be due to confusion and this is being taken into account under "ergonomics" and "complexity". | |
| **Scenario Complexity** | Operator must recognize that steam generator level has fallen over time even though adequate flow is indicated to the steam generator. Since the operator has multiple tasks (i.e., the operator is doing more than just monitoring steam generator levels) there is more than nominal complexity to this recognition.<br><br>There is a perceived competing success path to FR-H1: shutting the recirculation valve for pump 12. Although this action can be performed as part of FR-H1, pursuit of this perceived success path may delay entry into FR-H1. | ND |
| **Indications of Conditions** | The key driving factors identified by the qualitative analysis (not SPAR-H) for this HFE are the misleading indications on the flow from AFW pump 12 and the fact that recognizing that they have lost all feedwater is a knowledge-based decision | MND |
| **Execution Complexity** | Execution complexity would be reflected in the SPAR-H PSF of complexity for Actions. Only basic event 1A(6) has an action portion. The analysts state that while the task is uncommon, the equipment is commonly manipulated. Therefore, there is no effect of complexity on the action task. | 0 |
| **Training** | The steps for F&B are assumed to be well practiced. | N/P |
| **Experience** | Scenarios involving misleading indications are sometimes seen so the operators have been nominally trained to handle this situation even though they might not be highly practiced at it. This specific diagnosis is not one of the highly practiced actions (e.g. ATWS, swapping to sump recirc). | N/P |
| **Procedural Guidance** | Diagnostic/symptomatic procedures credited for basic events 1A(3), (5) | N/P |

| | | |
|---|---|---|
| **Human- Machine Interface** | The normal indication used by the operator is indicating that sufficient flow is reaching the steam generator when, in actuality, flow is insufficient. | MND |
| **Work Processes** | Presumed normal<br><br>A mitigating factor not specifically addressed by SPAR-H is that the cues for AFW flow are being looked at by multiple individuals. The STA typically performs the CSF status trees and the reactor operators monitor the main control board panels. Since there are two separate individuals with different backgrounds looking at the same data, it is more likely that the indication will be questioned. This “back-up” or “oversight” is not specifically credited by SPAR-H but does appear in other models. | N/P |
| **Communication** | Communication would be rolled into the SPAR-H PSF of Work processes | 0 |
| **Team Dynamics** | There may be some variability in crew performance due to team composition or response style. Crew characteristics are not explicitly addressed by SPAR-H, though they can be included in the Work Processes PSF if appropriate. Time is an important factor in this HFE, though it is not as tight as in HFE1B, but nevertheless a crew with a slow, methodical style may spend too much time on each procedure step and end up running out of time. If a crew has a knowledgeable STA or SRO, on the other hand, one who quickly is able to recognize the faulty indications and come to a correct situation assessment, such a crew is likely to perform quite well. Similarly, crews with more experience may perform better than crews with less operating experience | 0 |
| **Other** | Fitness for Duty: Presumed to be normal because there is no indication to assume otherwise. | 0 |

*G7.1.4 Comparison of Qualitative Analysis to Empirical Data*

In the NRC SPAR-H analysis the main negative driver as the PSFs indication of conditions and a secondary negative driver was scenario complexity. The empirical data also identified these as negative drivers in this scenario. However, the SPAR-H analysis indicated that the use of diagnostic/symptomatic procedures would be nominal or positive, while the empirical data indicated that Procedural guidance was a negative driver because the Critical Safety Function Status Trees do not address the misaligned valves. However, the SPAR-H analysts indicate that the recognition of the loss of feedwater is a knowledge based decision (implying that there is no procedural indication); the effect of the lack of procedural guidance is therefore included in the complexity PSF. In addition, the SPAR-H analysis includes the appearance of a competing success pathway to FR H-1, shutting the recirc valve for pump 12 in the complexity PSF. Had this been called out in the quantitative analysis as an insufficient procedure, the impact on the PSF would have been much greater than its impact when included in complexity.

In the empirical data, the critical negative PSFs were Scenario complexity, Indication of conditions, and Procedural guidance. The SPAR-H analysis captured the Scenario complexity and Indication of conditions quantitatively, giving the rationale in the analysis. However, while procedural guidance was called out in the qualitative analysis, it was not captured as a driver in the quantitative analysis. It is not clear why the Adequacy of time was called out in the empirical data as a positive factor. With respect to time, the analysts indicate that the decision to trip the reactor has provided the crew with extra time than they would have otherwise. The analysts stated that greatest mitigating factor for this HFE is the extra time allotted from the manual trip. Training, which was positive factor in the empirical data, and was a positive factor in the quantification.

The analysts describe factors including flow rate and procedural steps in their analysis, which led to the selection of PSF values. The analysis does go into detail regarding procedures and tasks performed by the crew in order to assess the PSFs. Whether the crew identifies that the average coolant temperature is not behaving as predicted is the critical indicator to prime the cue to question the flow indication of pump 12. The analysts state that recognition of the loss of all feedwater is a knowledge based task, the most difficult of the subtasks. They note that the procedure does not require crew to track the level of AFW; however, credit in the quantification is given for diagnostic symptom based procedures. The analysts state in their analysis that because the cues for AFW are reviewed by multiple team members, it is probable that the indication will be questioned. They note that this type of team process is not typically included in a SPAR-H analysis, and therefore it is not reflected in the quantitative analysis or PSFs. The analysts indicate that crew variability, which is not explicitly addressed by SPAR-H but which can be included in the Work Processes PSF, may affect performance. They note that team factors can significantly affect performance. These are congruent with the empirical findings.

The analysis states that the second most critical sub task of the scenario is the transition to FR-H1 (in HFE1A(2)), noting that this is a knowledge based task. They also note that the transition may be delayed if the crew tries to align pump 12 flowrate before transitioning. The analysts use this as part of their basis for the PSF complexity.

Finally, the analysts assume that the crew has good work processes as the scenario calls for a manual trip of the reactor, which they take as an indication that the crew is alert and proactive. One note, the analysts indicate that the PSF of available time affects the action portion of HFE1A(4); however the table indicates that the basic event is purely diagnosis. I believe that this is a typo as this analysis team was hurrying to complete the analysis and may have neglected to clear the boxes from previously used tables.

*G7.1.5 Impact on HEP*

The breakdown of the scenario into subtasks reduces the impact of PSFs on the entire HEP, while at the same time reduces the need for the analyst to control for more than 3 PSFs in a single basic event. The main contributors to the increase of the HEP were missing condition indicators and complexity. The missing indicator is assigned a multiplier of 50, but that factor effects only one basic event (HFE1A(1) diagnosis, while the complexity is assigned a factor of 2 and effects 2 basic events (HFE1A(1) and HFE1A(2)). The determination of base events reduces the effects of individual PSFs on the high level event.

One note should be made regarding how the analysts treated dependence between the basic events. In keeping with the SPAR-H method, the HEP was calculated from the basic events using the formula:

HEP 1A = HFE1A(1) + HFE1A(2) + [HFE1A(3) * HFE 1A(4) * HFE 1A(5)] + HFE1A(6).

Basic events 1A(3), 1A(4), and 1A(5) were assumed to be dependent; therefore their product was taken. HFE 1A(4) was determined to be moderately dependent on HFE 1A(3).

The analysts determined that HFE1A(4) is a recovery action for HFE1A(3), and therefore dependent, as follows. The analysts call out specific difficulties in transitioning to the F&B per step 2 of FR-H1. The analysts identify that if the crew fails to transition to F&B per step 2 of FR-H1, they will continue on in the procedure and eventually reach step 9 which provides an additional opportunity to transition to F&B. Prior to reaching step 9, it is possible they will transition to F&B via the Foldout Page (this HFE). The analysts chose not to link these two steps as complete dependence because they believe the foldout page would provide additional cues. Based on this, the analyst set dependency to moderate even though the strict SPARH definition of “different crew” was not met.

A similar decision was made to determine the degree of dependency of HFE1A(5) on the preceding step. The analysts determined that HFE1A(5) was a recovery action for HFE1A(3) and HFE1A(4) ; therefore there is dependency. The analysts did not assign complete dependence arguing that the performance of steps 3-8 would provide the crew a chance to “regain their bearing” and possibly perform step 9 correctly. Therefore analysts therefore chose “Not Close in Time” in the dependency because the of their assumption that performance of steps 3-8 would provide a break in the thought process and step 9 is worded and portrayed somewhat different from step 2, even though the strict SPAR-H definition of “not close in time” is not met.

By choosing ‘different crew’ rather than same crew, the HEP changed from 1.0 to the analysts’ estimate of 1.5E-1 for HFE1A(4). The selection of close in time rather than the analysts’ selection of ‘not close in time’ changes the HEP for HFE1A(5) from 1.0 to 5E-1. The total HEP for HFE 1A would change from the analysts’ estimate of

HEP 1A = {HFE1A(1) + HFE1A(2) + [HFE1A(3) * HFE 1A(4) * HFE 1A(5)] + HFE1A(6)}
= 0.101 + 0.021 + 0.00038 + 0.0005 = 0.12388 → 1.24E-01

To

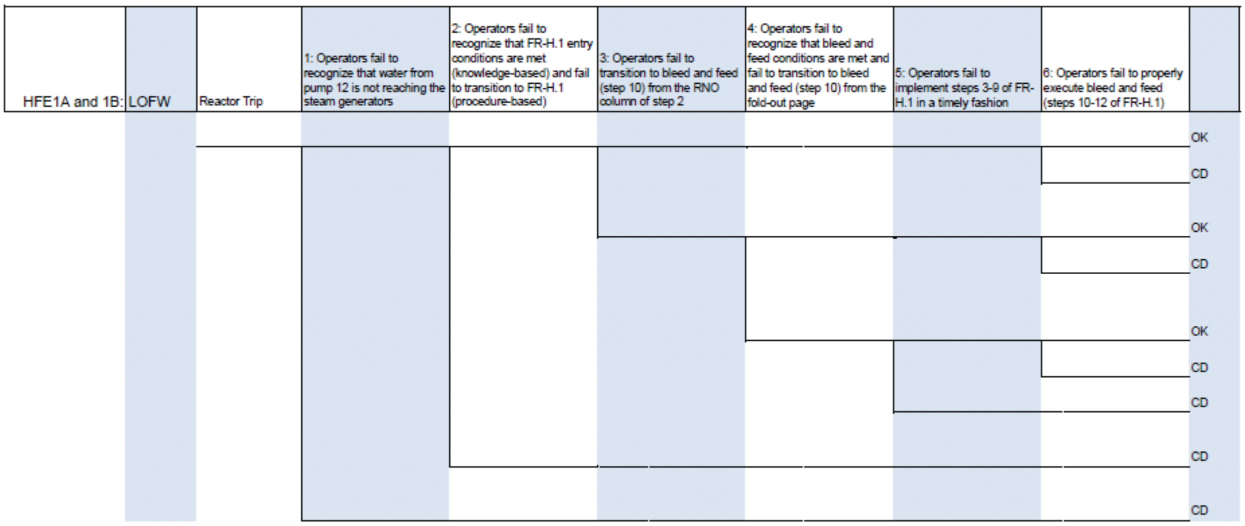
HEP 1A = {HFE1A(1) + HFE1A(2) + [HFE1A(3) * HFE 1A(4) * HFE 1A(5)] + HFE1A(6)}
= 0.101 + 0.021 + .0005+ 0.0005 = 0.123 → 1.24 E-01

Because the analysts determined dependency for the basic events, the liberal interpretation of the dependency rules has little effect on the total HEP.

### G7.2 HFE1B

*G7.2.1 Type of task:*
This scenario consisted of Diagnosis, Action and mixed Diagnosis and Action event. Again a CRT was developed. The HFE was segmented into 6 subparts, for each of which an HEP was calculated. The basic events defined were:

HFE1B(1): Operators fail to recognize AFW pump 12 is not flowing to the SG;
HFE1B(2): Operators fail to transition to FR-H1;
HFE1B(3): Operators fail to transition to bleed and feed (FR-H1 steps 10-12) from the FR-H1 step 2 RNO;
HFE1B(4): Operators fail to transition to bleed and feed (FR-H1 steps 10-12) from the fold-out page of FR-H1;
HFE1B(5): Operators fail to implement steps 3-9 of FR-H1 in a timely manner;
HFE1B(6): Operators fail to properly execute F&B.




### *G7.2.2 Summary of Qualitative Findings*

The key driving factors identified by the qualitative analysis was the tight thermal dynamic time window caused by waiting for the reactor to trip automatically on low SG level, and the misleading indications on the flow from AFW pump. The analysts also stated that recognition of complete loss of all feedwater is a knowledge-based decision, further driving the error factor. These factors are reflected in the PSF complexity. Analysts indicated that outside factors could cause the crew to focus on restoring AFW instead of initiating bleed and feed.

Recognition of loss of feedwater in the presence of misleading indications is a knowledge based decision requiring the operator to mentally recognize the following:

- The indication of pump flow does not equate to flow into the steam generator
- The steam generator level has not risen over a noted time period.

The operator is relying on their experience and training to both know to periodically check steam generator level in between other tasks and to recognize that the expected increase in level – which is not being actively recorded and trended - has not occurred.

Misleading indications and barely adequate time are the main error traps for subtask HFE1B(1). Even without the misleading indications the task of recognizing the failure of steam generator levels to rise is moderately complex so complexity is another source of error. Since the operators have waited for the reactor to trip automatically on low SG level (per the assumptions on this HFE) they have barely adequate time to perform this task. This increases the difficulty of this subtask.

HFE1B(1) was identified as the most critical subtask for success or failure in the scenario. After HFE1B(1), the next most difficult subtask is the diagnosis for transitioning to FR-H1 (HFE1A(2)). The analysts stated that this basic event was moderately complex because the operator must follow procedure cues prompted by control board instrumentation and also recognize that the normal cue (indicated AFW flow rates) cannot be followed to determine that AFW flow rates are less than the entry 576 gpm. The analysis noted that operators may delay entry into FR-H1 while they attempt to correctly align the pump12 flowrate. This temptation to delay entry into FR-H1 was part of the basis for in the moderate complexity rating.

Analysts felt that there were only two success paths to Feed & Bleed: step 2 of FR-H1 and the Foldout Page for FR-H1. In the analysis, Step 9 of FR-H1 was not believed to be a valid recovery path for step 2 because the thermal hydraulic analysis prediction of the automatic reactor trip scenario is that operators have only 13 minutes to initiate bleed and feed. Given the faulty indications on AFW pump 12 and the time it would take operators to navigate through E-0, ES-0.1, and transition to FR-H1, the analysis assessed that there was not enough time to reach the steps for bleed and feed if operators step through steps 3-9 one at a time. Recovery path HFE1B(5) was included in the quantitative analysis, and showed that the amount of time required to perform it is not adequate.

*HFE 1B(5) is a recovery action for HFE1B(3) and HFE1B(4). However, if the crews fail both HFE1B(3) and HFE1B(4), there is insufficient time to recover, and failure on this task is guaranteed, as indicated in the Action HEP above. Any additional dependence calculation is moot. However, were there adequate time, the dependence would be HIGH, for the same rationale as it was in HFE1A(5).*

*G7.2.3 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
HEP = <u>9.38E-01</u>
HEP 1B = {HFE1B(1) + HFE1B(2) + [HFE1B(3) * HFE 1B(4) * HFE 1B(5)] + HFE1B(6)}
= 0.911+ 0.021 + 0.0009 + 0.005 = 0.9379 → 9.38E-01

G7.2.3.1 Confidence Interval

| | HEP | Alpha | beta | 5th percentile | 95th percentile |
|---|---|---|---|---|---|
| HFE1B | 9.38E-01 | 5 | 0.3 | 7.31E-01 | 1.00E+00 |
| HFE1B(1) | 9.10E-01 | 4.5 | 0.4 | 6.57E-01 | 1.00E+00 |
| HFE1B(2) | 2.10E-02 | 0.495 | 23.1 | 8.08E-05 | 8.02E-02 |
| HFE1B(3) | 6.00E-03 | 0.5 | 82.8 | 2.38E-05 | 2.30E-02 |
| HFE1B(4) | 1.50E-01 | 0.45 | 2.6 | 4.29E-04 | 5.41E-01 |
| HFE1B(5) | 1.00E+00 | Na | na | na | na |
| HFE1B(6) | 5.00E-03 | 0.5 | 99.5 | 1.98E-05 | 1.92E-02 |

G7.2.3.2 Summary SPAR-H PSFs for HFE 1B

| PSFs | PSF Levels | Multiplier for Diagnosis | Comment |
|---|---|---|---|
| Available Time | Inadequate Time | P(failure) = 1.0 | HFE1B(5) ACTION |
| | Barely adequate time (~ 2/3 x nominal) | 10 | HFE1B(1) diagnosis; HFE1B(6) ACTION |
| | Nominal time | 1 | |
| | Extra time (between 1 and 2 x nominal and > 30 min) | 0.1 | |
| | Expansive time (> 2 x nominal and > 30 min) | 0.01 | |

| | | | |
|---|---|---|---|
| | Insufficient Information | 1 | |
| Stress | Extreme | 5 | |
| | High | 2 | |
| | Nominal | 1 | |
| | Insufficient Information | 1 | |
| Complexity | Highly complex | 5 | |
| | Moderately complex | 2 | HFE 1B(1) DIAGNOSIS; 1HFE1B(2) DIAGNOSIS |
| | Nominal | 1 | |
| | Obvious diagnosis | 0.1 | |
| | Insufficient Information | 1 | |
| Experience /Training | Low | 10 | |
| | Nominal | 1 | |
| | High | 0.5 | HFE1B(6) ACTION |
| | Insufficient Information | 1 | |
| Procedures | Not available | 50 | |
| | Incomplete | 20 | |
| | Available, but poor | 5 | |
| | Nominal | 1 | |
| | Diagnostic/symptom oriented | 0.5 | HFE1B(3) DIAGNOSIS |
| | Insufficient Information | 1 | |
| Ergonomics /HMI | Missing/Misleading | 50 | HFE1B(1) DIAGNOSIS |
| | Poor | 10 | |
| | Nominal | 1 | |
| | Good | 0.5 | |
| | Insufficient Information | 1 | |
| Fitness for Duty | Unfit | P(failure) = 1.0 | |
| | Degraded Fitness | 5 | |
| | Nominal | 1 | |
| | Insufficient Information | 1 | |
| Work Processes | Poor | 2 | |
| | Nominal | 1 | . |
| | Good | 0.8 | |
| | Insufficient Information | 1 | |

*G7.2.4 Summary table of Driving Factors for HFE1B*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Not enough time to transit through steps 3-9 if operators waited for automatic reactor trip ( HFE1B(5)) – AUTOMATIC FAILURE P =1.0 IF OPERATORS DO THIS. In addition, factor of 10 effect for HFE1B(6). | MND |
| **Time Pressure** | Derived from the SPAR-H PSFs of Available time and complexity. Barely adequate time to complete the diagnosis and still do all the remaining steps required for Feed & Bleed initiation. ( HFE 1B(1)) | MND |
| **Stress** | The operators are not under any physical duress. The control room environment is not inhospitable. Any perceived sense of urgency is unlikely to contribute to errors. Stressed caused by confusing indications is already factored in with “complexity” and “misleading | N/P |
| **Scenario Complexity** | Operator must recognize that steam generator level has fallen over time even though adequate flow is indicated to the steam generator. Since the operator has multiple tasks (i.e. the operator is doing more than just monitoring steam generator levels) there is more than nominal complexity to this recognition. (HFE 1B(1)) There is a perceived competing success path to FR-H1: shutting the recirculation valve for pump 12. Although this action can be performed as part of FR-H1, pursuit of this perceived success path may delay entry into FR-H1. (HFE1B(2) | ND |
| **Indications of Conditions** | The normal indication used by the operator is indicating that sufficient flow is reaching the steam generator when, in actuality, flow is insufficient. (Categorized under the SPAR-H PSF of ergonomics/hmi, with a factor of 50) (HFE 1B(1)) | MND |
| **Execution Complexity** | This would be the SPAR-H PSF complexity as applied to the Action tasks. This is not noted as a driver. | 0 |
| **Training** | The steps to perform F&B have likely been well practiced. (HFE 1B(6) – action) | N/P |
| **Experience** | Scenarios involving misleading indications are sometimes seen so the operators have been nominally trained to handle this situation even though they might not be highly practiced at it. This specific diagnosis is not one of the highly practiced actions (e.g. ATWS, swapping to sump recirc). (HFE 1B(1)) | N/P |
| **Procedural Guidance** | Step 2 in FR-H1 is a well written diagnostic step (HFE1B(3)) positive 0.5 | N/P |
| **Human-Machine Interface** | The normal indication used by the operator is indicating that sufficient flow is reaching the steam generator when, in actuality, flow is insufficient. (Categorized under the SPAR-H PSF of ergonomics/hmi, with a factor of 50) (HFE 1B(1)) | ND |
| **Work Processes** | Waiting for the reactor to trip automatically is an indication of a crew that is not taking a very proactive approach. However, this may not necessarily warrant a Work Processes rating of Poor, and the other factors of missing indications and little available time are driving the HEP up substantially. Adding a Poor Work Processes rating could be too conservative. (HFE 1B(1)) | N/P |

| | | |
|---|---|---|
| **Communication** | Can be included in SPAR-H PSF work processes | 0 |
| **Team Dynamics** | Can be included in SPAR-H PSF work processes | 0 |
| **Other** | Fitness for duty: Analysts indicated there was no reason to assume FFD had a role. | |

*G7.2.5 Comparison of Drivers to Empirical Data*

No empirical data. There were no observed cases for comparison.

*G7.2.6 Comparison of Qualitative Analysis to Empirical Data*

There were no observed cases on this HFE.

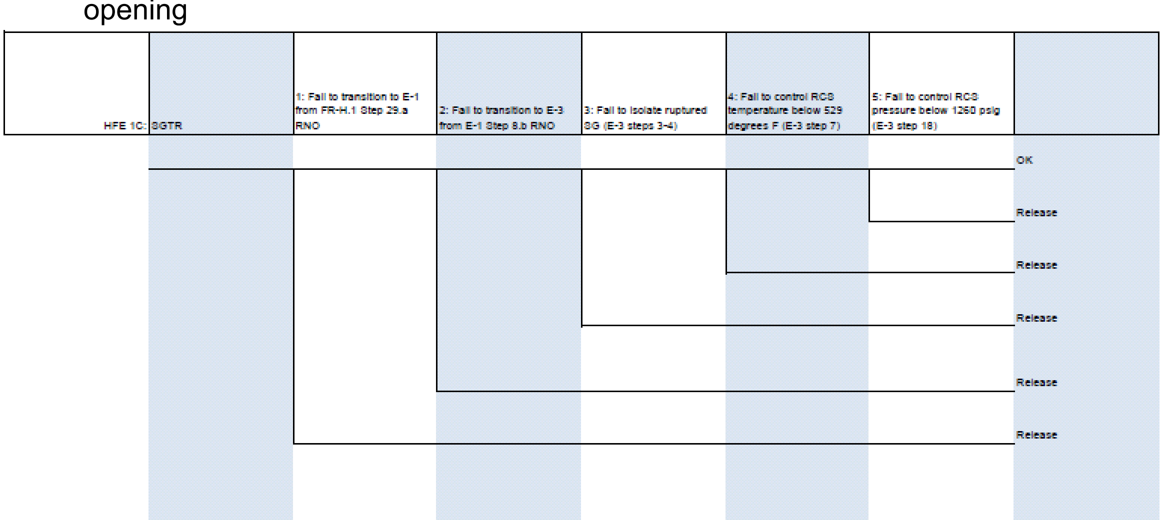
*G7.2.7 Impact on HEP*

The primary impact was derived from the lack of available time. Failure is assumed on one basic event, which drives the total HEP to close to 1.0.

### G7.3 HFE1C

This scenario was analyzed as a combination of diagnosis, action and mixed diagnosis-action tasks. The analysis was performed based on the expertise of the analysts developing a CRT. The basic events identified by the analysts were:

HFE1C(1): Fail to transition to E-10 from FR-H1 Step 29a RNO
HFE1C(2): Failure to transition to E-30 from E-10 step 8.b RNO
HFE1C(3): Failure to isolate the ruptured SG
HFE1C(4): Fail to control RCS temperature below 529 degrees F (E-30 step 7)
HFE1C(5): Failure to control RCS pressure below 1260 psig to prevent SG PORV opening




*G7.3.1 Summary of Qualitative Findings*

(Paraphrased from Item 3): When the SGTR occurs, the crew is still in FR-H1, in the process of recovering from the LOFW. The analysts anticipated that the crew could get stuck in a “do-loop” in FR-H1 steps 28 and 29, due to the need to obtain an adequate subcooling margin to move on. For the purpose of the assessment of available time, the analysts assumed that the crew would cycle through steps 28 and 29 three times before meeting the RNO criteria of step 29(a), which directs them to transfer to E-10. The analysts predicted that at some point during this "do-loop" cycle, the crew would become aware that they have a SGTR due to rising SG level. Success and failure hinged on whether the crew was able to move to E-30 within the procedures; analysts noted that it is not possible to transfer to E-30 from the Critical Safety Function Status Trees or from FR-H1 directly. The analysts used the time window defined by the scenarios, 40 minutes; however, the analysts noted that the actual time window would depend on the thermal-hydraulics of the plant. Analysts expected that once the crew recognized the SGTR, success would be dependent on completing E-30 before a SG PORV lifted, the difficulty of which depending on how much time the plant parameters allow them.

Analysts again noted that crew dynamics and work processes may play a role in this HFE, as well. (From item 3): *Crews who struggled with the LOFW may continue to have difficulty with the SGTR. If a crew is thrown by the unexpected plant response in recovery of the LOFW, they may have difficulty diagnosing the SGTR. However, crews with a knowledgeable STA or SRO who are quick to recognize the SGTR would perform better and have less difficulty. Like HFE1B, it may be that a crew that is willing to break from procedures and make the knowledge-based decision to go immediately to E-30 upon the recognition of the SGTR who perform the best.*

The main negative driver for this scenario was time pressure (Item 2): *the crew must establish aux feed, close PORVs via a "do loop", stop SI pumps via a concurrent "do loop", transition to E-10 and work through the first 7 steps, transition to E-03 and work through the first seven steps, cool down the RCS and finally depressurize the RCS. Complexity is also a negative driver for this HFE: some of these tasks are complex to think through (the "do loops") and some are tricky to perform (the depressurization). It will be tricky for the crews to diagnose the SGTR, given the lack of radiation alarms and the fact that they are expecting SG levels to rise. They will have to recognize that one steam generator is rising level faster than the others. Furthermore, the procedures may not be helpful for subtask HFE1C(1): the crew will spend a lot of time in do-loops, and it is likely that they will end up diagnosing the SGTR based on their knowledge and the rising SG level, rather than being aided to that diagnosis by the procedure. It is possible that the unfamiliarity of the situation (a SGTR during recovery from a LOFW) may confuse them for a little while. But once they've recognized the SGTR, the remaining separate substeps of this HFE are not overly complex or challenging in and of themselves; it just might be challenging to get through it all in a reasonable amount of time, before the SG PORVs lift.*

*G7.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

The total HEP for HFE1C is
HEP = <u>1.1E-01</u>
HEP 1C = HFE1C(1) + HFE1C(2) + HFE1C(3) + HFE 1C(4) + HFE 1C(5)
= 0.021 + 0.006 + 0.006 + 0.006 + .07 = 0.109 → 1.1E-01

G7.3.2.1 Confidence Interval

| | HEP | alpha | beta | 5th percentile | 95th percentile |
|---|---|---|---|---|---|
| HFE1C | 1.10E-01 | 0.465 | 3.8 | 3.50E-04 | 4.06E-01 |
| HFE1C(1) | 2.10E-02 | 0.495 | 23.1 | 8.08E-05 | 8.02E-02 |
| HFE1C(2) | 6.00E-03 | 0.5 | 82.8 | 2.38E-05 | 2.30E-02 |
| HFE1C(3) | 6.00E-03 | 0.5 | 82.8 | 2.38E-05 | 2.30E-02 |
| HFE1C(4) | 6.00E-03 | 0.5 | 82.8 | 2.38E-05 | 2.30E-02 |

| | | | | | |
|---|---|---|---|---|---|
| HFE1C(5) | 7.00E-02 | 0.475 | 6.3 | 2.34E-04 | 2.64E-01 |

G7.3.2.2 Summary of SPAR-H PSF for HFE1C

| PSFs | PSF Levels | Multiplier for Diagnosis | Comment |
|---|---|---|---|
| Available Time | Inadequate Time | P(failure) = 1.0 | |
| | Barely adequate time (~ 2/3 x nominal) | 10 | HFE1C(5) ACTION |
| | Nominal time | 1 | |
| | Extra time (between 1 and 2 x nominal and > 30 min) | 0.1 | |
| | Expansive time (> 2 x nominal and > 30 min) | 0.01 | |
| | Insufficient Information | 1 | |
| Stress | Extreme | 5 | |
| | High | 2 | HFE1C(1) DIAGNOSIS |
| | Nominal | 1 | |
| | Insufficient Information | 1 | |
| Complexity | Highly complex | 5 | HFE1C(5) DIAGNOSIS |
| | Moderately complex | 2 | HFE1C(5) ACTION |
| | Nominal | 1 | |
| | Obvious diagnosis | 0.1 | |
| | Insufficient Information | 1 | |
| Experience/Training | Low | 10 | |
| | Nominal | 1 | |
| | High | 0.5 | |
| | Insufficient Information | 1 | |
| Procedures | Not available | 50 | |
| | Incomplete | 20 | |
| | Available, but poor | 5 | |
| | Nominal | 1 | |
| | Diagnostic/symptom oriented | 0.5 | HFE1C(2) DIAGNOSIS; HFE1C(3) DIAGNOSIS; HFE1C(4) DIAGNOSIS |
| | Insufficient Information | 1 | |
| Ergonomics/HMI | Missing/Misleading | 50 | |
| | Poor | 10 | |
| | Nominal | 1 | |
| | Good | 0.5 | |
| | Insufficient Information | 1 | |

| | | | |
|---|---|---|---|
| Fitness for Duty | Unfit | P(failure) = 1.0 | |
| | Degraded Fitness | 5 | |
| | Nominal | 1 | |
| | Insufficient Information | 1 | |
| Work Processes | Poor | 2 | |
| | Nominal | 1 | |
| | Good | 0.8 | |
| | Insufficient Information | 1 | |

*G7.3.3 Summary Table of Driving Factors for HFE 1C*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | The SPAR-H PSF Time Available is rated as Time available is time required, indicating that the effect of time is felt as time pressure. | MND |
| **Time Pressure** | *By this point in the scenario the crew will likely be under time pressure. This is an involved task which takes more than a nominal amount of time to accomplish. ( HFE 1C(5) ACTION)* | MND |
| **Stress** | *While the operators are in a control room environment, they have just dealt with a LOFW event and are now faced with an inability to obtain an adequate subcooling margin (they were stuck in a loop at FR-H1 step 28-29 for a while). This may not be physical duress, but it would cause higher than normal stress. ( HFE1C(1))* | ND |
| **Scenario Complexity** | *This is a highly complex task, cognitively as well as physically, though there is a smaller penalty on the action portion of the task. Operator must depressurize the RCS to equilibrate RCS pressure with ruptured steam generator pressure while neither overfilling nor draining the pressurizer. This requires much cognitive activity to monitor and track RCS and pressurizer pressure. At this point, plant pressure is regulated by two steam bubbles: the one in the pressurizer and the one in the ruptured steam generator. However, the operator has no direct control over the steam bubble in the ruptured steam generator. ( HFE 1C(5) counted in both diagnosis and action)* | MND |
| **Indications of Conditions** | SPAR-H PSF HMI/Ergonomics: *Equipment is all easily accessible on the main control board but there is nothing about it that makes it stand out in a manner that warrants a credit* | NP |

| | | |
|---|---|---|
| **Execution Complexity** | *This is a highly complex task, which is accounted for in the Diagnosis section of this analysis. Physically, it is moderately complex to implement. Operator must depressurize the RCS to equilibrate RCS pressure with ruptured steam generator pressure while neither overfilling nor draining the pressurizer. At this point, plant pressure is regulated by two steam bubbles: the one in the pressurizer and the one in the ruptured steam generator. However, the operator has no direct control over the steam bubble in the ruptured steam generator (HFE1C(5) Action, complexity PSF)* | ND |
| **Training** | *E-30 is a commonly seen scenario, but it is not typically seen combined with a LOFW* | NP |
| **Experience** | E-30 is a commonly seen scenario, but it is not typically seen combined with a LOFW | NP |
| **Procedural Guidance** | *E-10 is a diagnostic/symptom oriented procedure.* ( HFE 1C(2)) | N/P |
| **Human- Machine Interface** | From Item 2: *It will be tricky for the crews to diagnose the SGTR, given the lack of radiation alarms and the fact that they are expecting SG levels to rise. They will have to recognize that one SG [level] is rising … faster than the others.* | Not quantified, but mentioned in analysis |
| **Work Processes** | From Item 3: *Crew dynamics and work processes may play a role in this HFE, as well. Crews who struggled with the LOFW may continue to have difficulty with the SGTR. If a crew is thrown by the unexpected plant response in recovery of the LOFW, they may have difficulty diagnosing the SGTR. However, crews with a knowledgeable STA or SRO who are quick to recognize the SGTR would perform better and have less difficulty. Like HFE1B, it may be that a crew who are willing to break from procedures and make the knowledge-based decision to go immediately to E-30 upon the recognition of the SGTR who perform the best* | Not included in quantification |
| **Communication** | | 0 |
| **Team Dynamics** | From Item 3: *Crew dynamics and work processes may play a role in this HFE, as well. Crews who struggled with the LOFW may continue to have difficulty with the SGTR. If a crew is thrown by the unexpected plant response in recovery of the LOFW, they may have difficulty diagnosing the SGTR. However, crews with a knowledgeable STA or SRO who are quick to recognize the SGTR would perform better and have less difficulty. Like HFE1B, it may be that a crew who are willing to break from procedures and make the knowledge-based decision to go immediately to E-30 upon the recognition of the SGTR who perform the best.* | Not included in quantification |
| **Other** | Fitness for Duty: presumed nominal, no reason to assume | 0 |

*G7.3.4 Comparison of Empirical Drivers to Data*

The drivers on performance that were included in the quantification were Time Pressure/Available time, Stress, Scenario Complexity and Execution Complexity and Procedural guidance. The qualitative analysis also indicated that unfamiliarity (which would be factored in terms of experience or training), Team factors, and Work Processes could affect crew performance. However, these factors were not included in the quantification, perhaps because the analysts, while they were aware they could affect performance, could not be certain that they would be present without further crew information. These analysts did participate in crew interviews; however, they may not have had the requirements of the SPAR-H analysis in mind as they had not yet been assigned the method.

Lack of available time was the initial driver in the empirical data; however, later it was realized that once the crew starts AFW, the available time before the PORVs lift changes. In their analysis, the analysts indicate the available time will be a function of the thermal hydraulics of the plant. They further state that the actions that the crew must take are not complex; the challenge is to get through all of the required actions in the time available. Part of that challenge lies in the possibility of getting stuck in a 'do-loop' at steps 28 & 29 of FR-H1. This challenge is not reflected in the quantification at the Procedural Guidance PSF; it is not clear why.

*G7.3.5 Comparison of Qualitative Analysis to Empirical Data*

Three crews isolated the ruptured SG, but only one did so within the 40 minute mark. As noted, the time available to the crew was a function of the plant thermal hydraulics. The analysts were accurate that performance would be affected by the misleading condition indicator. Crews did not perceive the rising SG until AFW was stopped. The difficulty in transitioning from FR-H1 to E-30 was also identified in the qualitative analysis, as was the utility of the Conditional Indications Page (FoldOut page). Crews also indicated that they felt run down by the long complicated scenario, the prediction of this was the basis for the Stress PSF.

The analysts appeared to develop a good understanding of the event. However, this understanding was driven by the expertise in operations of the team members rather than the SPAR-H method, as it does not provide guidance for the qualitative analysis. Even without explicit qualitative guidance from SPAR-H, the analysts were able to identify critical aspects of the qualitative analysis: the need to complete FR-H1 and control pressure and temperature. This understanding was also driven by the development of the CRTs, which may have assisted the analysts significantly. Similarly, they emphasized in their write up that the crew may only be able to identify the need to switch to E-30 based on prior knowledge, which is reflected in the SPAR-H PSF of complexity, and there was a possibility of the crew not completing the FR-H1 procedure in time because of being stuck in a do-loop in steps 28 and 29. However, it is not clear why the potential to get stuck is not reflected in their performance drivers such as the SPAR-H PSF of procedures. Had the analysts rated this negatively, this scenario would conceivably have been rated the most difficult.

Similarly, in the qualitative write up of the event, the analysts discuss that the unfamiliarity of the crew with the situation (a SGTR during recovery from a LOFW) may cause confusion and impact performance and that the diagnosis may be a result of the knowledge and experience of the crew. However, while this comment is included in the PSF table, the SPAR-H PSF Training and Experience is rated 'nominal' on all basic events. In addition, the qualitative write up indicates that there will be a lack of radiation alarms and that based on the existing HMI, the difference in rise rates for the SG may be difficult to identify. Once again, this is not commented in the PSF table nor is the SPAR-H PSF of HMI greater than nominal; however, the SPAR-H PSF Complexity would account for part of this. The analysts may have wished to avoid double counting for this effect. The analysts indicated that crew dynamics and work processes could affect the probability of error in the event, but this was not factored as a PSF in the HEP estimate.

Qualitatively, the analysts were correct to predict that the crew would not be able to transfer to the procedure that they knew they needed to perform because they had to complete other procedures first. They were also correct in their attribution of stress from having just been in a long complicated scenario. Most drivers were identified in the qualitative analysis but not all were factored into the quantitative analysis. In addition, the analysts indicated that while the HFE description indicated a time window of 40 minutes, the actual available time would be a function of the plant thermal-hydraulics.

*G7.3.6 Impact on HEP*

The qualitative analysis was very accurate in terms of the PSFs seen. However, the method did not provide a way for the analysts to include drivers that they thought might affect performance but for which they did not have evidence (e.g., work processes). This scenario was ranked as the most difficult by the crews, but was third most difficult in the analysis; however, the values of the HFEs do not differ much. Inclusion of factors such as work processes could have affected this. The analysts did note that the presence of an experienced STA or SRO could improve performance, but this was not quantified in the drivers. The analysts may have included this as a driver had they factored in that the STA was not present in the control room at the start of the scenario. In addition, breaking the scenario down from the high level to the detailed basic event may have also distilled the effects of critical PSFs. For example, the driver Stress is included only in HFE1C(1) Diagnosis. The analysts state that the basis for this:

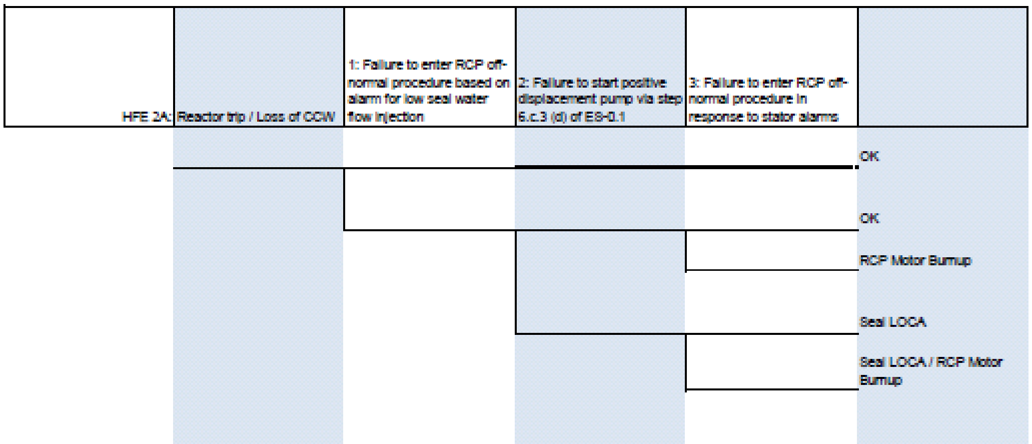
> *While the operators are in a control room environment, they have just dealt with a LOFW event and are now faced with an inability to obtain an adequate subcooling margin (they were stuck in a loop at FR-H1 step 28-29 for a while). This may not be physical duress, but it would cause higher than normal stress.*

The basis for this stress would be true of all of the basic events, but it is only factored in HFE1C(1). This may indicate a need to prompt the analyst to consider whether some PSFs are overarching of the set of basic events.

### G7.4 HFE2A

This scenario was analyzed as a combination of Diagnosis, Action and mixed action and diagnosis tasks. The analysis was performed based on the expertise of the analysts developing a CRT. The basic events identified were:

> HFE2A(1): Failure to enter RCP off-normal procedure based on alarm for low sealwater flow injection
> HFE2A(2): Failure to start positive displacement pump via step 6.c.3 (d) of ES-0.1
> HFE2A(3): Failure to enter RCP off-normal procedure in response to stator alarms




*G7.4.1 Summary of Qualitative Findings*

The analysts predicted that crews would have moderate difficulty with this task. The main negative driver was the complexity of identifying the loss of CCW and sealwater in the midst of all the other alarms and indications due to the reactor trip and loss of vital instrument bus. The limited time window they have to work in increased the difficulty.

From Item 3:

*The qualitative analysis identified one success path with one critical subtask: (1) Enter RCP Off Normal procedure based on the low sealwater alarm. The SPAR-H analysis identified one recovery path which held two critical subtasks: (2) Starting the Positive Displacement Pump per ES-0.1 step 6.c.3(d); and (3) Enter RCP Off Normal procedure based on the stator alarms. Each of these three subtasks was analyzed with the SPAR-H module of the HRA Calculator. Two minimal cut sets were identified: Failure of (1) and failure of (2); and failure of (1) and (3). The total HFE HEP was therefore calculated by the equation HEP 2A = (1){(2) + (3) – (2)(3)}.*

It is not clear that the development of the cut sets is part of the typical SPAR-H method; it may be a feature in the HRA calculator, with which this reviewer is unfamiliar. If it is a feature of the SPAR-H calculator, it is not clear that the NUREG has been updated. SPAR-H does discuss the use of cut sets

The qualitative analysis identified several additional paths that were not analyzed including crew failure of both HFE2A(1) and HFE2A(2), in which case a seal LOCA is assured and the only recovery remaining is whether the crew can save their RCP motors from burning up. Because there is no chance for success it was not analyzed. A second scenario was identified in which HFE2A(3) occurs prior to HFE2A(2) (not shown on the CRT). The analysis of the thermal-hydraulic timeline indicated that the seal LOCA criteria would be reached prior to receiving the stator alarms; therefore this scenario was not analyzed.

The two primary drivers for HFE2A were the confusion from the loss of the instrument bus and the CCPs and the time window of the scenario. In Item 2 the analysts state: *It is EASY to respond to a control board alarm when it is the only thing coming at you, but it is highly complex to pick that same alarm out as important when they are hit with it at the same time as all the normal alarms for a reactor trip PLUS all the invalid alarms for the loss of the instrument bus.*

Regarding the time window, the analysts note that there is a finite amount of time that the RCPs can run without CCW or Sealwater, and if the crews do not diagnose and mitigate the situation within that timeframe, they are faced with a seal LOCA and/or burning up their RCP motors. Given the complexity of the scenario, the analysts predicted that there was a small likelihood that the crew would be able to respond appropriately in the allowable time window.

*G7.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

HEP = 1.14E-01

HEP 2A = HFE2A(1) * {HFE2A(2) + HFE2A(3) – [HFE2A(2) * HFE 2A(3)]}
  = (0.51) * {0.15 + 0.098 – (0.15 * 0.098)} = 0.1144 → 1.14E-01

G7.4.2.1 Confidence Interval:

| | HEP | alpha | beta | 5th percentile | 95th percentile |
|---|---|---|---|---|---|
| HFE2A | 1.14E-01 | 0.465 | 3.6 | 3.66E-04 | 4.20E-01 |
| HFE2A(1) | 5.10E-01 | 0.5 | 0.5 | 6.51E-03 | 9.95E-01 |
| HFE2A(2) | 1.50E-01 | 0.45 | 2.6 | 4.29E-04 | 5.41E-01 |
| HFE2A(3) | 9.80E-02 | 0.465 | 4.3 | 3.05E-04 | 3.65E-01 |

G7.4.2.2 Summary Table of SPAR-H PSFs for HFE2A

| PSFs | PSF Levels | Multiplier for Diagnosis | Comment |
|---|---|---|---|
| Available Time (recommended choice based on timing information in bold) | Inadequate Time | P(failure) = 1.0 | |
| | Barely adequate time (~ 2/3 x nominal) | 10 | HFE2A(1) Diagnosis; HFE2A(1) ACTION; HFE2A(2) DIAGNOSIS; HFE2A(2) ACTION |
| | Nominal time | 1 | |
| | Extra time (between 1 and 2 x nominal and > 30 min) | 0.1 | |
| | Expansive time (> 2 x nominal and > 30 min) | 0.01 | |
| | Insufficient Information | 1 | |
| Stress | Extreme | 5 | |
| | High | 2 | |
| | Nominal | 1 | |
| | Insufficient Information | 1 | |
| Complexity | Highly complex | 5 | HFE2A(1) DIAGNOSIS; HFE 2A(3) DIAGNOSIS |
| | Moderately complex | 2 | |
| | Nominal | 1 | |
| | Obvious diagnosis | 0.1 | |
| | Insufficient Information | 1 | |
| Experience/ Training | Low | 10 | |
| | Nominal | 1 | |
| | High | 0.5 | |
| | Insufficient Information | 1 | |
| Procedures | Not available | 50 | |
| | Incomplete | 20 | |
| | Available, but poor | 5 | |
| | Nominal | 1 | |
| | Diagnostic/symptom oriented | 0.5 | |
| | Insufficient Information | 1 | |
| Ergonomics/ HMI | Missing/Misleading | 50 | |
| | Poor | 10 | |
| | Nominal | 1 | |
| | Good | 0.5 | |
| | Insufficient Information | 1 | |

| | | | |
|---|---|---|---|
| Fitness for Duty | Unfit | P(failure) = 1.0 | |
| | Degraded Fitness | 5 | |
| | Nominal | 1 | |
| | Insufficient Information | 1 | |
| Work Processes | Poor | 2 | |
| | Nominal | 1 | |
| | Good | 0.8 | |
| | Insufficient Information | 1 | |

*G7.4.3 Summary Table of Driving Factors HFE 2A*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Because of the subsequent and concurrent actions taken by the crew (e.g. E-0, responding to loss of CCW flow, responding to loss of an instrument bus, responding to loss of a CCP) it will take more than a nominal effort to reach step 6.c.3 (d) of ES-0.1 prior to exceeding allowable seal temperatures. (HFE 2A(2)) | MND |
| **Time Pressure** | Because of the other tasks occurring (e.g. performing the immediate actions of E-0) the crew should be under some time pressure to get this task completed in time. (HFE 2a(1)) | MND |
| **Stress** | The operators are not under any physical duress. The control room environment is not inhospitable. Any perceived sense of urgency is unlikely to contribute to errors. The "time stress" has already been taken into account. | 0 |
| **Scenario Complexity** | The operators have multiple priorities and must recognize that their top priority at this point is to respond to the RCP high stator alarm. Prioritizing tasks is a knowledge based decision. Based on the importance of the other tasks being performed and the importance of the other alarms which have not yet been responded to, it is not a trivial decision to postpone those other endeavours and choose this task to perform. (HFE 2A(2))<br><br>The operators have multiple priorities and must recognize that their top priority at this point is to respond to the RCP high stator alarm. Prioritizing tasks is a knowledge based decision. Based on the importance of the other tasks being performed and the importance of the other alarms which have not yet been responded to, it is not a trivial decision to postpone those other tasks and choose this task to perform. (HFE 2A(3)) | MND |
| **Indications of Conditions** | There is nothing which makes this annunciator stand out against the other annunciators which have alarmed. | 0 |
| **Execution Complexity** | This is a straight forward task. The only difficulty is in getting it done in the allotted time. The complexity due to limit time has already been taken into account. (HFE 2A(2)) | ND |

| | | |
|---|---|---|
| **Training** | Responding to alarms is a commonly trained on task, but this is not one of the highly practiced alarm responses (e.g., ECCS suction swap over from the RWST to the containment sump) | 0 |
| **Experience** | Younger or less experienced crews may have more difficulty recognizing the loss of CCW and sealwater in the midst of everything else they are dealing with. On the other hand, if a crew has more experience, or has a knowledgeable STA or SRO, and are quickly able to recognize the patterns of the alarms, realize that the sealwater alarm is activated, and come to a correct situation assessment of the loss of CCW and sealwater, such a crew is likely to perform quite well. However, there was no reason to assume other than nominal. | N/P |
| **Procedural Guidance** | The alarm procedure is a well written procedure which will get the required action done. The procedures are not particularly helpful in directing the crew to the proper alarm response procedure, however. ( HFE2A(1))<br><br>Although ES-0.1 is a diagnostic procedure, step 6.c.3(d) is deep into step 6 and does not stand out in any way. The procedure is adequate, but diagnostic credit is not being given. Another approach would be to give diagnostic credit here due to ES-0.1 being a diagnostic procedure, but to give a moderately complex penalty above due to step 6.c.3(d) not being reached until several failed paths. This approach was not chosen because "complexity" was analyzed based on the task being performed and not the procedure guidance for performing it. ( HFE 2A(2)) | N/P |
| **Human- Machine Interface** | There is nothing which makes this annunciator stand out against the other annunciators which have alarmed. | 0 |
| **Work Processes** | No reason to assume otherwise | 0 |
| **Communication** | In SPAR-H would be included in Work processes | 0 |
| **Team Dynamics** | In SPAR-H would be included in Work processes Analysts indicated that variability in crew performance could arise due to team composition or response styles, which are not explicitly addressed by SPAR-H (but which can be included in the SPAR-H PSF "Work Processes"). The analysts indicated that because time is an important factor in this HFE, a crew with a slow, methodical style could run out of time. However, there was no reason to assume this was present based on the information given. | 0 |
| **Other** | Fitness for duty: No reason to assume other than nominal | 0 |

*G7.4.4 Comparison of Drivers to Empirical Data*

The only SPAR-H PSFs identified in the analysis were Complexity and Available time. These translate to the drivers of Time Pressure, Time Adequacy and Scenario complexity. In the empirical data both of these were observed; however, complexity was rated 'moderate' in the empirical data. The characterization of scenario complexity in the empirical data ("many things happening at once") matches the characterization of the scenario complexity in the SPAR-H analysis. Execution complexity was included in the empirical drivers. This would correspond to the complexity PSF on an Action task in SPAR-H. None of the action tasks in the SPAR-H analysis included complexity as a factor. However, the execution complexity driver is described in terms of which procedure was used. In the analysis, procedures were rated nominal, as once the alarm procedure is entered it is identified as well written.

In the empirical data, time adequacy was rated in terms of training; the crew would have had sufficient time if they had had more training. Training and experience were not quantified as drivers in the SPAR-H analysis, although the potential impact of experience was recognized. Without specific information on the crew, the SPAR-H analysts were not able to select other than nominal. If these analysts had been able to interview the crews, they may have made a different selection. This underscores the importance of crew interviews for PSF selection. In the empirical data, work processes and communication were noted as potential drivers but not quantified. The SPAR-H analysis notes that work processes can include crew characteristics, but the analysts indicate that they have no reason to believe the crew is other than nominal.

The analysts did not indicate the missing condition indicator identified in the empirical data due to the 1201 failure. Procedural guidance was identified as a negative driver in the empirical data, in terms of the difficulty crews had in reaching critical actions in time. The SPAR-H analysis appears to combine this under Time Adequacy (*Because of the subsequent and concurrent actions taken by the crew (e.g. E-0, responding to loss of CCW flow, responding to loss of an instrument bus, responding to loss of a CCP) it will take more than a nominal effort to reach step 6.c.3 (d) of ES-0.1 prior to exceeding allowable seal temperatures.)*

*G7.4.5 Comparison of Qualitative analysis to Empirical Data*

The SPAR-H analysis accurately indicates that identifying loss of CCW in the midst of all the other alarms is very difficult for the crew and uses up significant time. The analysts indicate that there is a chance for recovery from failing to detect loss of CCW if the crew works rapidly once the RCP stator alarm occurs.

*G7.4.6 Impact on HEP*

The analysts predicted that the crews would have moderate difficulty on this scenario based on the belief that even if the crews missed the loss of CCW they could work quickly enough to succeed; however, this was found to be the most difficult scenario and all crews failed to get through the procedure.

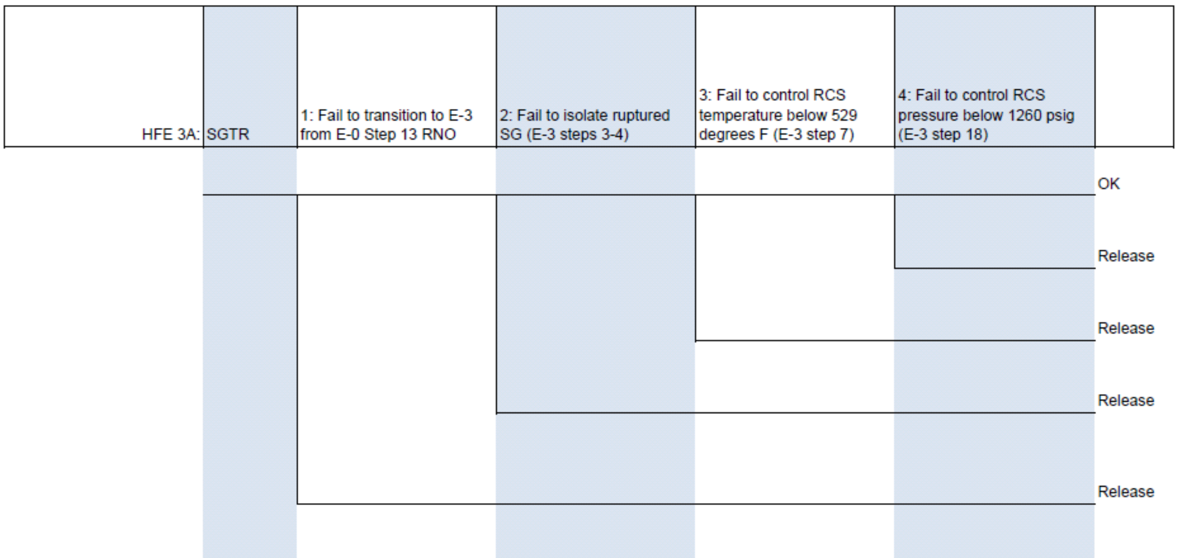
The qualitative analysis performed indicated several failure paths that were not analyzed because they did not have the possibility of success. One such path is seen in the lower path of the high-level CRT, and consists of a scenario in which the crew fails to enter RCP off-normal procedure AND failure to start the positive displacement pump, in which case a seal LOCA is assured and the only recovery remaining is whether the crew can save their RCP motors from burning up. This path was not analyzed further or included in our quantification because there was zero opportunity for success. It is intriguing that the qualitative analysis identified this path as all crews in the empirical data were late stopping the RCPs; one stopped the RCPs from procedure while the other three stopped RCPs from knowledge. This combined with the identification of training as a driver of performance would indicate that the pathway identified is a probable failure path. However, identification of this path was not driven by guidance on qualitative analysis from SPAR-H; rather it was dependent on the knowledge of the analysts. It is very clear that the strength of the SPAR-H analysis is a function of the good-ness of the qualitative analysis, and that the use of CRTs in this analysis lead to good application of the PSFs.

## G7.5 HFE3A

This scenario was broken down into 4 basic tasks using the CRT tool. The basic tasks were a combination of Diagnosis, Action and mixed Diagnosis-Action tasks. Four subtasks were identified:

HFE3A(1): Failure to transition to E-30 from E-0 Step 13 RNO

Mixed Diagnosis and action
HFE3A(2): Failure to isolate the ruptured SG (E-30 steps 3-4)
HFE3A(3): Failure to control RCS temperature below 529 °F (E-30 step 7)
HFE3A(4): Failure to control RCS pressure below 1260 psig (E-30 step 18)




*G7.5.1 Summary of Qualitative Findings*
This was predicted to be the easiest scenario, and all crews were predicted to succeed because it is frequently trained, has no complications such as inaccurate instrumentation, and good procedures. Any failures or problems were predicted to occur from physically controlling the depressurization. Even though this is well-trained, controlling a depressurization with a SGTR is tricky, due to having two steam bubbles to control, one in the ruptured SG and the other in the pressurizer. The analysts predicted that failure could also occur if a crew completely misdiagnoses the situation, if they misread or misunderstand the indications, but this was assumed to be unlikely due to training. The success criteria as prescribed by the HFE definition require that operators prevent the SG PORV from opening. The subtasks identified in the CRT are all necessary for success, and no recovery actions were modelled.

*G7.5.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
HEP 3A = 7.0E-2
HEP 3A = HFE3A(1) + HFE3A(2) + HFE 3A(3) + HFE 3A(4)
HEP 3A = 0.006 + 0.006 + 0.006 + 0.052 = 0.07 → 7.0E-2

G7.5.2.1 Confidence Interval

| | HEP | alpha | beta | 5th percentile | 95th percentile |
|---|---|---|---|---|---|
| HFE3A | 7.00E-02 | 0.475 | 6.3 | 2.34E-04 | 2.64E-01 |
| HFE3A(1) | 6.00E-03 | 0.5 | 82.8 | 2.38E-05 | 2.30E-02 |
| HFE3A(2) | 6.00E-03 | 0.5 | 82.8 | 2.38E-05 | 2.30E-02 |
| HFE3A(3) | 6.00E-03 | 0.5 | 82.8 | 2.38E-05 | 2.30E-02 |
| HFE3A(4) | 5.20E-02 | 0.485 | 8.8 | 1.89E-04 | 1.97E-01 |

G7.5.2.2 Summary Table of SPAR-H PSFs

| PSFs | PSF Levels | Multiplier for Diagnosis | Comment |
|---|---|---|---|
| Available Time (recommended choice based on timing information in bold) | Inadequate Time | P(failure) = 1.0 | |
| | Barely adequate time (~ 2/3 x nominal) | 10 | |
| | Nominal time | 1 | |
| | Extra time (between 1 and 2 x nominal and > 30 min) | 0.1 | |
| | Expansive time (> 2 x nominal and > 30 min) | 0.01 | |
| | Insufficient Information | 1 | |
| Stress | Extreme | 5 | |
| | High | 2 | |
| | Nominal | 1 | |
| | Insufficient Information | 1 | |
| Complexity | Highly complex | 5 | HFE3A(4) DIAGNOSIS |
| | Moderately complex | 2 | HFE3A(4) ACTION |
| | Nominal | 1 | |
| | Obvious diagnosis | 0.1 | |
| | Insufficient Information | 1 | |
| Experience/ Training | Low | 10 | |
| | Nominal | 1 | |
| | High | 0.5 | |
| | Insufficient Information | 1 | |
| Procedures | Not available | 50 | |
| | Incomplete | 20 | |
| | Available, but poor | 5 | |
| | Nominal | 1 | |
| | Diagnostic/symptom oriented | 0.5 | HFE3A(2) DIAGNOSIS; HFE3A(3) DIAGNOSIS; |
| | Insufficient Information | 1 | |
| Ergonomics/ HMI | Missing/Misleading | 50 | |
| | Poor | 10 | |
| | Nominal | 1 | |
| | Good | 0.5 | |
| | Insufficient Information | 1 | |
| Fitness for Duty | Unfit | P(failure) = 1.0 | |
| | Degraded Fitness | 5 | |
| | Nominal | 1 | |
| | Insufficient Information | 1 | |

| | | | |
|---|---|---|---|
| Work Processes | Poor | 2 | |
| | Nominal | 1 | |
| | Good | 0.8 | |
| | Insufficient Information | 1 | |

*G7.5.3 Summary Table of Driving Factors HFE 3A*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Nominal time & adequate procedures (called Available time in SPAR-H) | 0 |
| **Time Pressure** | Given that this event has no complications and good procedures, the crew should have sufficient time to respond. | 0 |
| **Stress** | | 0 |
| **Scenario Complexity** | This is a highly complex task, cognitively as well as physically, though there is a smaller penalty on the action portion of the task. Operator must depressurize the RCS to equilibrate RCS pressure with ruptured steam generator pressure while neither overfilling nor draining the pressurizer. This requires much cognitive activity to monitor and track RCS and pressurizer pressure. At this point, plant pressure is regulated by two steam bubbles: the one in the pressurizer and the one in the ruptured steam generator. However, the operator has no direct control over the steam bubble in the ruptured steam generator. ( HFE 3A(4) Diagnosis) | MND |
| **Indications of Conditions** | | 0 |
| **Execution Complexity** | This is a highly complex task, cognitively as well as physically, though there is a smaller penalty on the action portion of the task. Operator must depressurize the RCS to equilibrate RCS pressure with ruptured steam generator pressure while neither overfilling nor draining the pressurizer. This requires much cognitive activity to monitor and track RCS and pressurizer pressure. At this point, plant pressure is regulated by two steam bubbles: the one in the pressurizer and the one in the ruptured steam generator. However, the operator has no direct control over the steam bubble in the ruptured steam generator. ( HFE 3A(4) Action) | MND |
| **Training** | This is a well-trained event with no complications. | N/P |
| **Experience** | This is a well-trained event with no complications. | N/P |
| **Procedural Guidance** | Procedures are good and should direct operators to the correct course of action. ( HFE 3A(1)) | N/P |
| **Human- Machine Interface** | Interface provides all necessary information and operators are well-trained on the interface. | N/P |
| **Work Processes** | Crew work processes presumed to be nominal. | N/P |
| **Communication** | Can be included in work processes | 0 |
| **Team Dynamics** | Can be included in work processes | 0 |

| | | |
|---|---|---|
| **Other** | Fitness for Duty: No reason to presume anything other than nominal fitness for duty. | N/P |

*G7.5.4 Comparison of Drivers to Empirical Data*

All drivers in the empirical data were nominal or positive. There were no indicators of complexity on performance. As reflected in the SPAR-H analysis, the main effects were of positive effects of training and experience on performance. Crews ranked this the easiest of the scenarios, as the analysts predicted.

*G7.5.5 Comparison of Qualitative Analysis to Empirical Data*

SPAR-H predicted that this would be a very simple scenario, easy to perform with no complications. This was reflected in the empirical data.

*G7.5.6 Impact on HEP*

The HEP predicted by the analysis is relatively high when compared to the HEPs predicted for the more complicated scenarios. While the analysts did breakdown the scenario into basic events using CRTs, it is not immediately clear that this would be necessary. The qualitative analysis did not manage to discriminate this HFE sufficiently from the other HFEs. In the empirical data, this was show to be the easiest HFE, and a well trained one with no complications or failures. PSFs applied more generally across the scenario may yield a smaller HEP, or a number of nominal numbers could be summed to yield a very large failure probability. However, breakdown of the HFE into subevents did allow for a very detailed qualitative analysis; however if one breaks an event down into 10 basic independent events, the resulting summed HEP could be quite large. The breakdown is again based on the analysts' ability to perform the qualitative analysis and the technique that they use to define the basic events. In this case, the use of CRTs to define basic events lead to very detailed analysis of the scenarios, and examination of the role of the procedures in the events. Finally, the analysts did not give credit for time, training & procedures, or experience which would further reduce the HEP.

## G7.6 Comparison Summary

*G7.6.1 Predictive Power*

G7.6.1.1 Qualitative Predictive Power – in Terms of Drivers

Overall, the drivers matched well with the empirical data. This may be due to the analysts breaking down the high level events in to sub events to allow the analysts to gain significant knowledge of the tasks and contexts to be rated, as suggested in Section 4.2.2. In HFE 1A, which was estimated to be the most difficult by the analysts, the SPAR-H analysis identified Complexity, and Ergonomics/HMI as negative drivers. The main negative driver was Ergonomics/HMI. The empirical data also identified Complexity and Ergonomics as negative drivers in this scenario. Adequacy of time was called out in the empirical data as a positive factor, and the analysts stated that Available time was a positive factor in 2 of the basic events identified. Training was positive factor in the empirical data, and was a positive factor in the quantification. In HFE 1C, the analysis identified Available time, Stress and Complexity as the main negative drivers.

In several of the analyses SPAR-H analysis indicated that the use of diagnostic/symptomatic procedures would be nominal or positive, while the empirical data indicated that Procedural guidance was a negative driver. In HFE 1A, lack of procedures (knowledge based task) may have been counted in complexity, but it is not clear from the analysis. It may be a weakness of the method that when an effect is combined with another factor to avoid double counting, the overlap is not recorded.

At various points in the analyses, the analysts indicate that crew variability can be included in the Work Processes PSF and may affect performance. They note that team factors can significantly affect performance, but the analysts assume that both fitness for duty and work processes to be nominal, except in 1B, where the automatic trip is seen as an indication that the crew is not working well; however it is not included in the quantification. In HFE 1C, the qualitative description indicates that experience could be a factor affecting crew performance but it was not included in the quantification; again this may have reflected that the analysts did not have information about crew experience and thus had to assume it was nominal as with training and work processes. A weakness of the method may be that in a predictive analysis, the analyst may never have sufficient reason to rate these PSFs as negative because specific information about the crew to perform the task in the future is not easily known.

The analysts correctly identify that do-loops in procedures can slow crews or even cause errors; however, this is not quantified. The effect is instead indirectly recorded in other factors such as complexity. The rationale for not including a factor in the quantification or subsuming it into another factor is not always clear. It may be that this should be made more explicit. It should be noted that the HEP estimates for HFE 1A, 1C and 2A are very close in range to each other. HFE 3A was correctly predicted to be the easiest of the scenarios.

G7.6.1.2 Qualitative Predictive Power – in Terms of Operational Expressions

The analysts used the SPAR-H module in the EPRI HRA calculator to perform their analyses. The analysts created CRTs based on their understanding of the procedures, the critical paths, and diagrammed out the paths in Excel based on CRT guidance that is being developed as part of the SRM project. The guidance in SPAR-H on qualitative analysis is limited; the method suggests that scenarios be decomposed into basic sub events using the ATHEANA search process. It does not specify defining CRTs. One weakness of the SPAR-H method is that the analysis is not required to define sub events from a scenario; therefore, the analyst is free to apply the method only to the high level scenario definition. This would mean that the level of the PSF would be applied once to the diagnostic and action calculations, and that most likely the analyst would choose a lower value for a PSF that has a strong effect only on one aspect of the scenario.

The analysis was definitely aided by the definition of basic sub events for which the impact of PSFs were determined. This also led exploration and understanding of the procedures. SPAR-H does indicate that the analyst must have considerable understanding of the procedures, tasks and contexts in order to perform the analysis. In this case, the analysis of the scenarios was added by the operational experience of the analysis team. The analysis correctly points outs difficult aspects of the procedures (such as do-loops) that were problematic in the empirical results. While it was clear why the analysts chose the level of PSF included in the quantification, it was not always clear when a PSF was combined with another PSF.

It appears that the predictive power of the method may be driven in part by the time the analyst takes to develop understanding of the procedures, tasks and context; in that case the method's predictive power is rated good. However, it may be that the understanding developed is a function of the amount of operational experience that the analysis team has a priori. Further, the analysts' understanding of what occurred in the scenarios was primarily driven by the breakdown of the scenarios and the examination of the procedures. While this is recommended by SPAR-H, it is not required. It is possible that analysts may assess scenarios without a thorough examination of the operations.

G7.6.1.3 Quantitative Predictive Power

Table of relative rank of each scenario as rated in the empirical data and through the NRC SPAR-H analysis.

| **HFE** | **Task** | **Unit Sup. rank** | **Failure rate** | **Difficulty** | **NRC SPAR-H estimated HEP** | **NRC SPAR-H RANKING** |
|---|---|---|---|---|---|---|
| 2A | Stop RCPs and start PDP in scenario 2 | 1 | 4/4 | Very difficult | 1.14E-01 | 2 |
| 1C | Identify and isolate ruptured steam generator in scenario 1 | 2 | 1/4 | Difficult | 1.1 E-01 | 3 |
| 1A | Start bleed and feed in scenario 1 | 3 | 0/4 | Fairly difficult to difficult* | 1.24E-01 | 1 |
| 3A | Identify and isolate ruptured steam generator in scenario 3 | 4 | 0/3 | Easy | 7.0 E-2 | 4 |
| 1B | Failure to establish F&B within 13 minutes of the reactor trip, given that the crews do not manually trip the reactor before an automatic reactor trip occurs. | No data | n/a | No data | 9.83 E-01 | Highest HEP |

It is difficult to state whether the HEPs predicted by the analysis were predictive for several reasons. First, the ratings of 1A, 1C, and 2A are very close. The analysts did not *a priori* state which event they felt would be most difficult; however, they did state that 3A would be the easiest, which was supported by their analysis. The analysts estimated that scenario 1A would be the most difficult. However, 1A was ranked easier than either 1C or 2A by the crews. However, the analysts were correct estimating that 2A would be more difficult than 1C and that 3A would be the easiest scenario, and correct to indicate that 2A would be an extremely difficult scenario. While there are no data to support qualitative analysis of 1B, the analysts stated that a basis for their high estimate was predicated on the automated trip of the plant, which would indicate that the crew itself was not functioning effectively and thus would most likely not have sufficient coordination to respond in the short time available after the trip.

The quantitative predictive power for this analysis is overall judged to be fair. Had 1A been correctly identified as easier than 2A or 1C, it would have been good. Had the analysis decremented performance for poor procedures rather than crediting for symptom based procedures, the analysis would most likely have obtained the correct order; procedures were correctly identified in the qualitative analysis as critical, but were not included in the quantification as discussed.

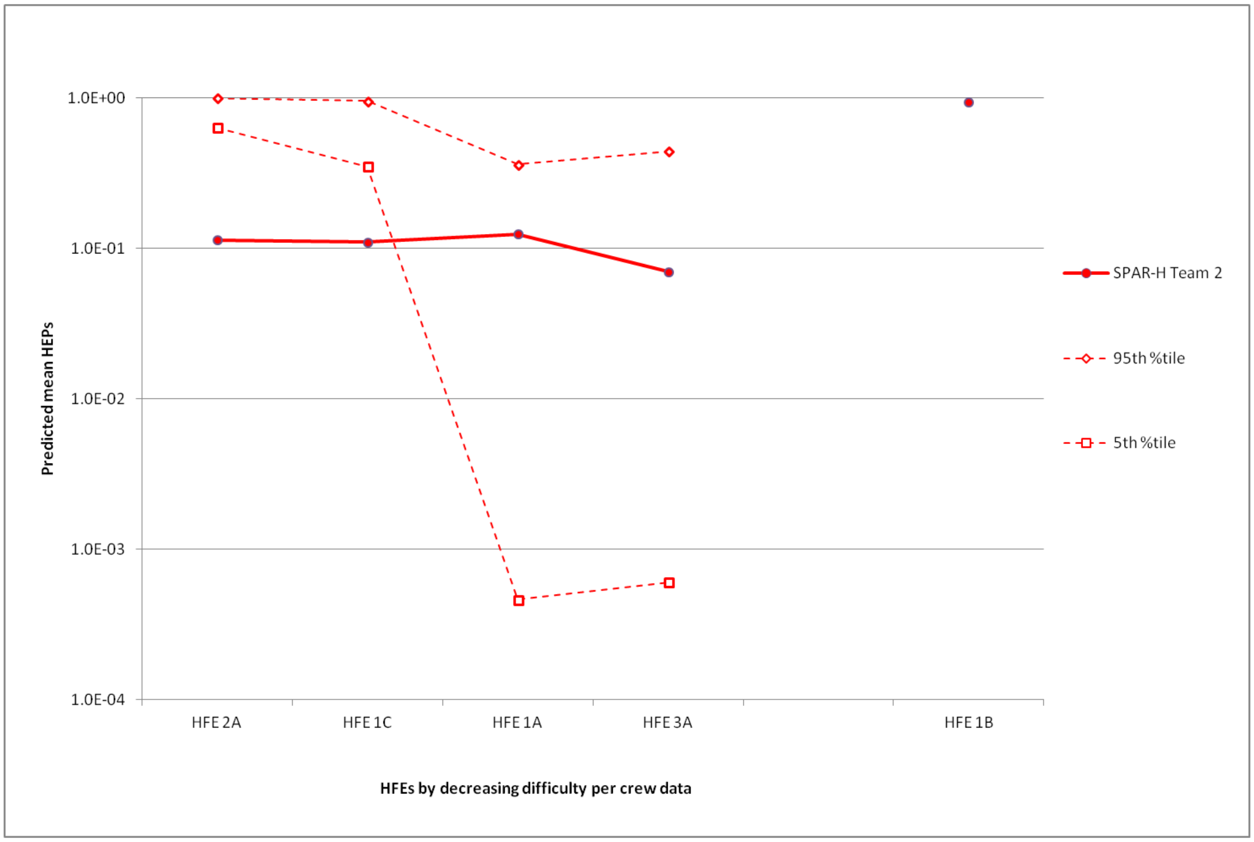


Figure 0.7 Predicted Mean HEPs with Uncertainty Bounds by INL2 HRA Team with SPAR-H

*G7.6.2 Traceability of the qualitative analysis and quantification process*

In this case, the analysts used the SPAR-H module of the EPRI HRA calculator to derive the HEPs. However, the SPAR-H analysis worksheets were also provided; the worksheets for the basic events contained comments that indicated why a given value of the PSF was selected. Therefore, the analysis was very transparent and could be easily reconstructed. Traceability of the quantification is rated 'good' for SPAR-H.

It was not clear in some cases why procedures were rated nominal or positive when the qualitative analysis had indicated that there was or could be a difficulty in the procedure. In some cases, this was assumed to be due to the overlap in the PSFs. Overlap between PSFs is expected; SPAR-H devotes significant discussion to the orthogonality and overlap between PSFs in section 2.7.5 and a table in Appendix G, indicating that dependence and overlap between PSFs; this section points out that dependence and overlap between PSFs could make the SPAR-H HEPs either too conservative or too optimistic. This is given as a caution to the analyst; at this time SPAR-H does not propose a technique to avoid impact of PSF dependence.

The traceability of the PSFs is rated good.

*G7.6.3 Insights on guidance*
Guidance on the necessity to derive basic events from scenarios should be improved. SPAR-H was not designed to perform as a qualitative tool; it assumes that the analyst has sufficient operations experience to perform the breakdown using something like the ATHEANA process. However, clarification of whether breakdown is necessary and should occur would be useful. Instruction on how to select the magnitude of PSFs or to avoid double counting when there is known, identified overlap between PSFs would improve the method.

*G7.6.4 Insights for Error Reduction*
The qualitative analysis identified a number of areas that could be explored to reduce error at the plant. These included identification of do-loops in procedures, potential effects of crew experience and work processes (even though these were not decremented in quantification).

*G7.6.5 Remarks on strengths and weaknesses*
SPAR-H was not developed to qualitatively assess a scenario; however, the need to breakdown a scenario into basic events has significant impact on the assessment. It would be much stronger if this aspect was developed; however, as SPAR-H was meant to provide a quick assessment, this may be at odds with its developed purpose. For analysts who are familiar with the scenarios in question and the operations, distillation of scenarios to basic events would be a simple and fast process. SPAR-H assumes that this type of analyst would need more assistance with the determination of the appropriate PSFs and their values. The qualitative analysis is highly traceable and replicable as long as the analyst comments as indicated in the worksheets. The HEP is very sensitive to the PSFs selected by the analyst. The method indicates that there can be significant overlap between PSFs but does not provide a technique to adjust the magnitude of the PSF when this is recognized nor does it provide guidance on selection of PSFs when a scenario involves related PSFs to avoid double counting.

## G.8 ATHEANA (Science Applications International Corporation (SAIC))

### G8.1 HFE 1A

Human failure event (HFE) 1A addresses the failure to establish F&B cooling within 45 minutes of the reactor trip, given that the crews initiate a manual reactor trip before an automatic reactor trip occurs.

*G8.1.1 Summary of Qualitative Findings*

The HRA analysts considered various ways in which the HFE could be realized:

- Failure of the operators to identify that flow from auxiliary feedwater (AFW) pump 12 was not being delivered to the steam generator (SG).
- Failure to close and verify closed the recirculation valve in a timely manner, accounting for comparing flow to the SG to the behaviour of SG level.
- Failure to implement F&B cooling in a timely manner.

The team noted that the mis-positioning of the recirculation valve and lack of indication mean that the failure of AFW flow to the SG might not be apparent unless operators notice SG levels decreasing. It is possible that having wide-range level less than 50% on two SGs would be the overriding entry condition for F&B cooling, and that condition is expected to be reached relatively quickly. Related to these considerations are the following:

- The crew has plenty of time to clear other alarms that might be competing for their attention.
- They have an indication that flow is continuing (although it is not) but SG level should still be going down, providing a cue to the operators.
- By the time the scenario is ongoing, the shift technical advisor (STA) should be present in the control room, providing another set of eyes to recognize the cue of SG level.
- The key is whether the operators get to the right procedure for initiating F&B cooling without being directed there from a yellow path in the Heat Sink Critical Safety Function Status Tree. Once they enter the proper procedure, the training is such that they will implement the process.
- One of the experts questioned whether or not there would be hesitancy on the part of the operators to implement F&B cooling, but the other experts indicated that this is trained and recognized as an important procedure and the insights from the operator interview also indicated they would follow the procedure.
- Subcooling margin should be a priority and should provide input to the diagnosis.
- The operators interviewed felt confident that given 45 minutes, the crew would be able to diagnose the actual conditions and implement F&B cooling.

Taking this information into account, a single failure scenario was quantified. This scenario was the failure to initiate F&B cooling, conditional on failing to recognize and correct the diversion of AFW flow through the recirculation path.

*G8.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

Per the ATHEANA process, each of three experts provided a three-point distribution of the probability for this HFE. The results were then discussed to arrive at the following consensus distribution:

Mean HEP = 5.5E-02

1%-tile = 1E-02
99%-tile = 1E-01

The lower bound was based on the consensus opinion that the operators would be successful 98% of the time. The upper bound was based on the consensus opinion that the operator could be stuck in another procedure, causing late time pressure.

*G8.1.3 Summary Table of Driving Factors*

The driving factors identified in the ATHEANA team's analysis are summarized in the table below.

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Analysts felt that 45 minutes was substantial time, subject to the potential for time pressure late in the scenario as noted for the next item. Adequacy of time and time pressure were ranked together as "good" with respect to the performance shaping factors (PSFs) considered. | N/P |
| **Time Pressure** | Time pressure considered to be low to moderate given the 45 minute timeframe cited in the HFE definition, although initial time spent in diagnosis could lead to time pressure late in the scenario. Adequacy of time and time pressure were ranked together as "good" with respect to the PSFs considered. | N/P |
| **Stress** | The need to follow a course other than that specifically associated with procedure FR-H1 could lead to increased stress. Overall, however, workload, time pressure, and stress were assessed to be "good" with respect to their impacts as PSFs. | 0 |
| **Scenario Complexity** | The analysts noted that operators interviewed considered the diagnosis of the need for F&B cooling to be slightly tricky due to the misleading AFW flow indication, but felt it would not be too great a difficulty.<br>The analysts felt that there was little complexity when the cue to enter the proper procedure is reached. The cue of wide range SG levels less than 50% overrides everything else. | 0 |
| **Indications of Conditions** | The quality of the indications was judged to be good in general, but within the context of the scenario, the indications would be misleading. Proper diagnosis would depend on the relying on indications of lack of a heat sink instead of other indications. There would be adequate and redundant indications of the need to restore SG level. | ND |
| **Execution Complexity** | Relatively simple execution; no issues identified. | N/P |
| **Training** | The crew is trained in the use of procedure FR-H1 about once per year, and at least every 2 years. Training did not specifically address the impact of the mis-positioned recirculation valve, but the analysts appeared to consider the training on response to the lack of heat sink as overriding. | N/P |
| **Experience** | Experience was not addressed separate from training. | 0 |

| Factor | Comments | Influence |
|---|---|---|
| **Procedural Guidance** | The cues related to the availability of AFW flow would not satisfy the clearest path from the Heat Sink Safety Function Status Tree to procedure FR-H1 for F&B cooling.<br>Interviews indicated that the operators would want to re-establish a heat sink so that they could exit procedure FR-H1, but that procedure would remain a priority until completed.<br>The analysts judged that, for scenarios with longer time frames, the procedures are adequate, but that that was less the case for the shorter-term scenario. | ND |
| **Human-Machine Interface** | The analysts judged that the control room employs control boards and displays standard to nuclear power plants, but not necessarily mimic displays found in many control rooms. The Quantitative Parameter Display System (QPDS) provides means to obtain a quick understanding of important plant parameters. | 0 |
| **Work Processes** | Not explicitly addressed outside team dynamics and communication. | 0 |
| **Communication** | During the simulator exercise observed by the analysts, crew communications were not always audible. The strategy on the part of the crew was for the supervisor to take the lead, read procedure steps aloud, direct activities, and request feedback from ROs on actions taken or parameters being monitored. | ND |
| **Team Dynamics** | The analysts observed that, in simulator exercises, the crew worked well together and there was a clear hierarchy of command. | N/P |
| **Other** | With respect to the availability of operating staff, it was noted that Units 1 & 2 have separate but identical control rooms. This means that if one unit is in an abnormal condition, substantially more than the minimum complement of operators can be available to assist. | N/P |

*G8.1.4 Comparison of Drivers to Empirical Data*

The previous section summarizes the various PSFs identified by the analysis team with respect to the extent to which they were drivers in the crews' performance relative to this scenario. This section compares the drivers identified by the analysts for this HFE in the table above to those derived from crew performance.

- **Adequacy of time**. In both the SAIC analysis made using ATHEANA and the actual crew response, time was judged to be a positive driver. In the simulator exercises, all four crews initiated F&B cooling within about 17 minutes from the total loss of feedwater. This was well within the 45 minutes available, and indicates that there could have been adequate time to address difficulties due to other factors if it had been needed.
- **Time pressure**. Time pressure was not addressed as a separate factor with respect to crew response. The SAIC analysts judged time pressure to be a nominal/positive driver, again in light of the fact that the operators had substantial time to respond to this situation.
- **Stress**. Although the analysts judged that the conditions posed by this event could heighten stress, they judged that it was a non-factor. This was true with respect to the assessment of crew response as well.

- **Scenario complexity**. With respect to crew response, the complexity of the scenario was assessed to be a negative driver. This conclusion reflected the judgment that response (i.e., the initiation of F&B cooling) would have been quicker if there had not been a misleading indication of AFW status. The SAIC analysts judged scenario complexity to be a non-factor. Although they acknowledged the misleading flow indication, they noted that the criteria for establishing F&B cooling were tied to SG levels, and that there was limited complexity associated with this set of parameters.
- **Indications of conditions**. Both the assessment of the crews' responses and the SAIC analysts' evaluation for this HFE found that indications constituted a negative driver for this scenario. This was due to the misleading indications of AFW flow.
- **Execution complexity**. The SAIC analysts noted that the actions required to initiate F&B cooling were relatively simple, and cited execution complexity as a nominal/positive driver. The assessment of crew response also characterized this as a nominal/positive driver without further comment.
- **Training and experience**. Training and experience were assessed to be nominal/positive factors in both assessments. The operators are trained in response to the loss of feedwater periodically, and were prepared to deal with the loss of heat sink, even if the indications were not as direct as they might have been.
- **Procedural guidance**. In slightly different ways, both the SAIC analysts and the assessment of crew response characterized the available procedural guidance as a negative driver. The SAIC analysts noted that the most direct path to procedure FR-H1, in which the direction to initiate F&B cooling is found, would not be satisfied because there was apparent flow from the AFW system. The assessment of crew response identified the lack of mention of misaligned valves in the Critical Safety Function Status Trees as an issue with the procedures. While this is not the type of consideration that would typically be addressed in the status trees, the issue is essentially the same.
- **Human-machine interface**. The SAIC analysts judged the human-machine interface to be a non-factor, weighing the lack of a mimic display against the digital display system used to convey important plant parameters. In the assessment of crew response, this aspect was characterized as nominal/positive, without further elaboration.
- **Team dynamics**. The SAIC analysts observed that the crew worked well together, and that there was a clear hierarchy of command. This was therefore a nominal/positive driver. It was assessed to be a non-driver based on observations of actual crew response.
- **Other factors**. The SAIC team explicitly addressed several factors that were not evaluated for the actual crew responses. These included communication (assessed to be a potential negative driver owing to a tendency for some crew interactions to be inaudible); and the availability of crew members from the other unit to provide support (a nominal/positive driver).

#### *G8.1.5 Comparison of Qualitative Analysis to Empirical Data*

Overall, it is judged that the SAIC analysts characterized the scenario in a manner that was generally consistent with the responses observed in the simulator exercises. The time available was more than adequate to overcome challenges posed by the available indications and the procedural guidance. In neither the SAIC analysis nor the assessment of crew response was a main negative driver cited. This is consistent with a scenario for which all crews completed the required response well within the time available.

#### *G8.1.6 Impact on HEP*

The SAIC analysis yielded a mean probability for the HFE of 0.055. This is a moderately high probability of failure. Although no main negative drivers were identified, this probability reflects consideration of the issues raised by the misleading indications for AFW flow and the resulting impact on the path that would lead the operators to procedure FR-H1 and, ultimately, to initiate F&B cooling.

### G8.2 HFE 1B

Human failure event (HFE) 1B addresses the failure to establish F&B cooling within 13 minutes of the reactor trip, given that the crews do not initiate a manual reactor trip before an automatic reactor trip occurs.

#### *G8.2.1 Summary of Qualitative Findings*

The HRA analysts considered the same ways in which the HFE could be realized as for HFE 1A:

- Failure of the operators to identify that flow from auxiliary feedwater (AFW) pump 12 was not being delivered to the steam generator (SG).
- Failure to close and verify closed the recirculation valve in a timely manner, accounting for comparing flow to the SG to the behaviour of SG level.
- Failure to implement F&B cooling in a timely manner.

The team noted that the mis-positioning of the recirculation valve and lack of indication mean that the failure of AFW flow to the SG might not be apparent unless operators notice SG levels decreasing. It is possible that having wide-range level less than 50% on two SGs would be the overriding entry condition for F&B cooling, and that condition is expected to be reached relatively quickly. Related to these considerations are the following:

- The crew has one indicator telling them that the need for AFW is satisfied and another leading to the yellow path for the Heat Sink Critical Safety Function Status Tree, but not the red path.
- While the crew is trained on the procedure, the context of this particular scenario makes it very difficult to get to the procedure since they have passed the 14% Narrow Range point for the SGs and are getting to 50% on Wide Range.
- The shift technical advisor (STA) will not be present in the control room for the first 5 minutes, which means that another set of eyes is not available to identify the cue of SG level.
- The key is whether the operators reach the procedure for establishing F&B cooling despite not being there from a yellow Heat Sink path. Once they enter the proper procedure, the training is such that they will implement the process.
- One of the experts questioned whether or not there would be hesitancy on the part of the operators to implement F&B cooling, but the other experts indicated that this is trained

and recognized as an important procedure and the insights from the operator interview also indicated they would follow the procedure.

- Subcooling margin should be a priority and should provide input to the diagnosis.
- The operators interviewed indicated that, given only 13 minutes, the crew could diagnose the actual conditions and implement F&B cooling. The operators did not engender significant confidence that this could be done reliably.

Taking this information into account, a single failure scenario was quantified. This scenario was the failure to initiate F&B cooling, conditional on failing to recognize and correct the diversion of AFW flow through the recirculation path.

*G8.2.1 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*

Per the ATHEANA process, each of three experts provided a three-point distribution of the probability for this HFE. The results were then discussed to arrive at the following consensus distribution:

Mean HEP = 3E-01
1%-tile = 5E-02
99%-tile = 6E-01

The lower bound was based on the consensus opinion that even at best the operator is unlikely to be able to assess the situation and take the necessary action within the short timeframe and the lack of clear procedure entry conditions. The upper bound was based on the consensus opinion that there are actions and scenarios that are worse than this, both within and outside the control room, but that the operator has essentially no time to take the action given the circumstances.

*G8.2.2 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | One input from operators was that there was that F&B cooling would certainly be implemented within 45 minutes, and could be within 10 - 15 minutes. The analysts judged that 13 minutes was not sufficient time considering the time that might be required to understand what is actually taking place and that 50% on the wide-range SG level indicators would be reached relatively quickly. | MND |
| **Time Pressure** | Workload could be considered to be high considering the time available and the challenging diagnosis conditions. The operators recognize implementing the F&B procedure as time-critical. They will therefore be reading, implementing, and repeating back procedure steps quickly. | ND |
| **Stress** | The need to follow a course other than that specifically associated with procedure FR-H1 could lead to increased stress. When included with time pressure and workload, this was assessed to be a poor PSF. | ND |

| Factor | Comments | Influence |
|---|---|---|
| **Scenario Complexity** | The analysts noted that operators interviewed considered the diagnosis of the need for F&B cooling to be slightly tricky due to the misleading AFW flow indication, but felt it would not be too great a difficulty.<br>The analysts felt that there was little complexity when the cue to enter the proper procedure is reached. The cue of wide range SG levels less than 50% overrides everything else. | 0 |
| **Indications of Conditions** | The quality of the indications was judged to be good in general, but within the context of the scenario, the indications would be misleading. Proper diagnosis would depend on the relying on indications of lack of a heat sink instead of other indications. There would be adequate and redundant indications of the need to restore SG level. | ND |
| **Execution Complexity** | Relatively simple execution; no issues identified. | N/P |
| **Training** | The crew is trained in the use of procedure FR-H1 about once per year, and at least every 2 years. Training did not specifically address the impact of the mis-positioned recirculation valve, but the analysts appeared to consider the training on response to the lack of heat sink as overriding. | N/P |
| **Experience** | Experience was not addressed separate from training. | 0 |
| **Procedural Guidance** | The cues related to the availability of AFW flow would not satisfy the clearest path from the Heat Sink Safety Function Status Tree to procedure FR-H1 for F&B cooling.<br>Interviews indicated that the operators would want to re-establish a heat sink so that they could exit procedure FR-H1, but that procedure would remain a priority until completed.<br>The analysts judged that, for scenarios with longer time frames, the procedures are adequate, but that that was less the case for the shorter-term scenario. | ND |
| **Human- Machine Interface** | The analysts judged that the control room employs control boards and displays standard to nuclear power plants, but not necessarily mimic displays found in many control rooms. The Quantitative Parameter Display System (QPDS) provides means to obtain a quick understanding of important plant parameters. | 0 |
| **Work Processes** | Not explicitly addressed outside team dynamics and communication. | 0 |
| **Communication** | During the simulator exercise observed by the analysts, crew communications were not always audible. The strategy on the part of the crew was for the supervisor to take the lead, read procedure steps aloud, direct activities, and request feedback from ROs on actions taken or parameters being monitored. | ND |
| **Team Dynamics** | The analysts observed that, in simulator exercises, the crew worked well together and there was a clear hierarchy of command. | N/P |

| Factor | Comments | Influence |
|---|---|---|
| **Other** | With respect to the availability of operating staff, it was noted that Units 1 & 2 have separate but identical control rooms. This means that if one unit is in an abnormal condition, substantially more than the minimum complement of operators can be available to assist. How effective this additional staff capacity might be given the relatively short time for this scenario was not addressed. | N/P |

*G8.2.3 Comparison of Drivers to Empirical Data*

In the simulator exercises, all crews successfully tripped the reactor before an automatic trip resulted. Therefore, there are no empirical results to which to compare the SAIC analysis.

### G8.3 HFE 1C

Human failure event (HFE) 1C considers failure of the crew to isolate the ruptured steam generator and control pressure below the setpoint for the SG pilot-operated relief valve (PORV). Controlling pressure in this manner prevents the need for the SG PORV to open, which could give rise to the potential for failure to re-close, creating a pathway from the RCS to the environment.

*G8.3.1 Summary of Qualitative Findings*

The HRA analysts considered the following way in which the HFE could be realized:

- Failure to respond to the SGTR, given lack of radiation alarms. This accounted as well for the “rule” to stay in procedure FR-H1, and the fact that the break flow might be masked by the slow AFW feed rate. They also considered crew distraction as a contributing factor.

Aspects that relate to the consideration of this failure include the following:

- It was not entirely clear from the scenario description, but given the time frame it appears that this action was considered for the scenario with shorter time for establishing F&B cooling (HFE 1B), in which the operators failed to trip the reactor before the automatic reactor trip occurred.
- Interviews indicated that the operators would want to exit procedure FR-H1, but it would remain a priority until exist conditions were satisfied. They would be able to take other actions as long as they don't conflict with FR-H1. Overall, however, the focus on procedure FR-H1 would tend to distract them from noticing the SG tube leak.
- The analysts observed that it would not be until the operators reached step 33 of FR-H1 that they would to directed to procedure ES-11, in which step 9 related to SI termination which directs them to monitor the possible need to reinitiate SI. At this point, the operators are still not directed to check RCS pressure, but rather only to look at pressurizer level. The RCS would be losing subcooling margin by that point, but there is no clear indication of the SG tube leak.
- The operators would not be directed to terminate SI until pressurizer level has recovered as part of restoring secondary cooling to allow termination of F&B cooling. Once the loss of RCS inventory to the secondary is terminated, the pressurizer will refill.

- When the operators terminate SI and are restoring normal alignments, they may eventually receive radiation alarms.
- Information collected during the operator interviews indicated that correct diagnosis would depend on the feed rate and how the crew is deployed. The operators could be distracted; the primary-side operator will be responsible for most of the actions relating to initiating and terminating SI. The diagnosis of the tube rupture will rely on the ability of the secondary-side operator to integrate the available information. With regard to the timing for this scenario, the shift technical advisor (STA) should be present in the control room, providing another set of eyes to recognize the cue of SG level.

This information was considered in addressing the potential failure to isolate the SG and to limit challenges to the SG PORV.

*G8.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
Per the ATHEANA process, each of three experts provided a three-point distribution of the probability for this HFE. The results were then discussed to arrive at the following consensus distribution:

Mean HEP = 2E-01
1%-tile = 4E-02
99%-tile = 6E-01

The lower bound was based on the consensus opinion that even at best the operator is unlikely to be able to assess the situation and take the necessary action within the short time frame (assuming that given the timing specified, the SGTR was entered from the scenario in which the reactor tripped automatically) and the lack of clear procedure entry conditions. The upper bound was based on the consensus opinion that there are actions and scenarios that are worse than this, both within and outside the MCR, but that the operator has essentially no clear procedural guidance in this case.

*G8.3.3 Summary Table of Driving Factors*
The relevance of the factors that affected the ATHEANA team's analysis is summarized in the table below.

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | The timing was considered to be reasonable. Adequacy of time and time pressure were ranked together as "good" with respect to the PSFs considered. | N/P |
| **Time Pressure** | Time pressure was not seen to be a major factor. Adequacy of time and time pressure were ranked together as "good" with respect to the PSFs considered. | N/P |
| **Stress** | The workload may be such that the operators may miss less direct indications of SGTR due to the preoccupation with F&B cooling. Stress could be increased due to the potential conflict associated with taking actions other than those called for in procedure FR-H1. The analysts considered the conditions to be less than optimal. | ND |

| Factor | Comments | Influence |
|---|---|---|
| **Scenario Complexity** | The operators interviewed considered the SGTR challenging to diagnose given the unfamiliar situation of a SGTR at the same time as a need for F&B cooling. The analysts concurred, based upon a review of the procedures and the available cues. | ND |
| **Indications of Conditions** | The quality of the indications was judged to be good in general, but within the context of the scenario, the indications would be misleading. Proper diagnosis would depend on the addressing a balance of responding to the lack of a heat sink and of other indications. There would be adequate and redundant indications of the need to restore SG level. | ND |
| **Execution Complexity** | Relatively straightforward execution; no issues identified. | N/P |
| **Training** | The crew is trained in the use of procedure FR-H1 about once per year, and at least every 2 years. Training did not specifically address the impact of the mis-positioned recirculation valve, but the analysts appeared to consider the training on response to the lack of heat sink as overriding. Training on response to a SGTR takes place about every six months. Training on both loss of heat sink and SGTR has apparently not been done; therefore the applicability of the training for this particular scenario was considered to be poor. | ND |
| **Experience** | Experience was not addressed separate from training. | 0 |
| **Procedural Guidance** | The procedural guidance will ensure that the operators remain with procedure FR-H1 until exit criteria are satisfied. The operators can take other actions as long as they don't conflict with FR-H1. Once the operators are able to enter the SGTR procedure (E-30), it would be very effective. There is, however, the question of how and whether they will reach that procedure. | ND |
| **Human- Machine Interface** | The analysts judged that the control room employs control boards and displays standard to nuclear power plants, but not necessarily mimic displays found in many control rooms. The Quantitative Parameter Display System (QPDS) provides means to obtain a quick understanding of important plant parameters. | 0 |
| **Work Processes** | Not explicitly addressed outside team dynamics and communication. | 0 |
| **Communication** | During the simulator exercise observed by the analysts, crew communications were not always audible. The strategy on the part of the crew was for the supervisor to take the lead, read procedure steps aloud, direct activities, and request feedback from ROs on actions taken or parameters being monitored. | ND |
| **Team Dynamics** | The analysts observed that, in simulator exercises, the crew worked well together and there was a clear hierarchy of command. | N/P |
| **Other** | With respect to the availability of operating staff, it was noted that Units 1 & 2 have separate but identical control rooms. This means that if one unit is in an abnormal condition, substantially more than the minimum complement of operators can be available to assist. | N/P |

*G8.3.4 Comparison of Drivers to Empirical Data*

This section compares the drivers identified by the analysts for HFE 1C in the table above to those derived from crew performance.

- ***Adequacy of time***. The SAIC analysis made using ATHEANA judged the adequacy of time to be a positive driver. With respect to the response of the crews, however, time seemed to be a greater issue and was assessed to be a negative driver. Three of the four crews failed to take measures within the pre-defined time period (40 minutes) to prevent lifting of the SG PORV. For two of these three crews, however, actions taken to control the situation limited the pressure increase in the steam generators, allowing them to avoid lifting the SG PORV, even though they did not do so within the 40-minute time frame. Given the extent to which the SGTR was masked by AFW flow and the number of procedural steps and actions the operators had to accomplish, it was judged that the time available for this scenario presented a significant challenge to the operators.
- ***Time pressure***. Time pressure was not addressed as a separate factor with respect to crew response. The SAIC analysts judged time pressure to be a nominal/positive driver, again in light of the fact that the operators had substantial time to respond to this situation.
- ***Stress***. The SAIC analysts concluded that the workload presented by this scenario could contribute to the potential for the operators to miss the partially masked indications of the SGTR. There could also be increased stress due to the need to satisfy the requirements from FR-H1 while also responding to the SGTR. Therefore, stress was cited as a negative driver. Some of the considerations noted relative to stress could alos be manifested through consideration of time adequacy and time pressure. With respect to the assessment of crew response, stress was assessed not to be a driver.
- ***Scenario complexity***. With respect to both the SAIC analysis and the assessment of crew response, the complexity of the scenario was assessed to be a negative driver (actually a main negative driver for the crew response). This resulted from the fact that the SGTR was a consequence of other events, rather than an initiating fault. Moreover, the AFW flow could mask the existence of the SGTR for some period of time.
- ***Indications of conditions***. Both the assessment of the crews' responses and the SAIC analysts' evaluation for this HFE found that indications of conditions constituted a negative driver for this scenario. The SAIC analysts focused on maintaining heat removal and proper SG levels at the same time. In the crew response, the indications that were most problematic related to the AFW flow that could mask the SGTR, combined with the lack of alarms on high radiation in the main steam lines due to low steam flow at this point in the scenario.
- ***Execution complexity***. The SAIC analysts noted that the actions were relatively straightforward, and cited execution complexity as a nominal/positive driver. The assessment of crew response also characterized this as a nominal/positive driver without further comment.
- ***Training and experience***. Training was assessed by SAIC to be a negative driver, because it did not necessarily cover the transition from a scenario involving a lack of heat removal to one involving a SGTR. Once the proper procedure was reached, good

guidance was available for responding to the SGTR. With regard to the crew response, training and experience were judged to be nominal/ positive drivers. Training covers both initial SGTR and SGTR occurring following a reactor trip, when there might be delayed or reduced indication of radiation in the steam lines.

- ***Procedural guidance***. Both the SAIC analysis and the assessment of crew response characterized the available procedural guidance as a negative driver. In both cases, it was noted that the path to enter the SGTR procedure (E-30) was not straightforward, leading to delay in implementing actions to limit pressure rise in the ruptured steam generator.
- ***Human-machine interface***. The SAIC analysts judged the human-machine interface to be a non-factor, weighing the lack of a mimic display against the digital display system used to convey important plant parameters. In the assessment of crew response, this aspect was characterized as nominal/positive, without further elaboration.
- ***Team dynamics***. The SAIC analysts observed that the crew worked well together, and that there was a clear hierarchy of command. This was therefore a nominal/positive driver. It was assessed to be a non-driver based on observations of actual crew response. In particular, for the one crew that did not prevent the PORV from opening, the shift technical advisor (STA) could have provided useful information that could have facilitated transferring to the proper procedure, but did not do so.
- ***Other factors***. The SAIC team explicitly addressed several factors that were not evaluated for the actual crew responses. These included communication (assessed to be a potential negative driver owing to a tendency for some crew interactions to be inaudible); and the availability of crew members from the other unit to provide support (a nominal/positive driver).

*G8.3.5 Comparison of Qualitative Analysis to Empirical Data*

Although the qualitative analysis performed by SAIC differed in some particulars from the responses observed in the simulator exercises, it was substantially on target. The biggest difference related to the characterization of time (judged to be a positive driver by SAIC and a negative driver in the actual crew responses). But SAIC accounted for some of the same impacts through stress (workload). The SAIC analysis noted the negative aspects associated with the complexity of the scenario, the indications of conditions, and the procedural guidance available to the operators.

*G8.3.6 Impact on HEP*

The SAIC analysis yielded a mean probability for the HFE of 0.2. This value was driven primarily by the complexity of the scenario, the lack of direct indications to guide response, and the lack of a clear path to the SGTR procedure. All of these contributed to challenges to the operating crews, including to the one crew that failed to prevent opening of the SG PORV.

### G8.4 HFE 2A

Human failure event (HFE) 2A addresses failure of the operators to trip the RCPs and start the positive displacement pump (PDP) to prevent a RCP seal LOCA following the total loss of RCP seal cooling.

*G8.4.1 Summary of Qualitative Findings*

The HRA analysts considered the following potential contributions to the occurrence of the HFE:

- Failure to identify the loss of 120V vital distribution panel 1201.
- Failure to control main feedwater to maintain appropriate level in the SGs.
- Failure to trip the reactor manually, including giving appropriate direction to do so; de-energizing the motor-generator sets for the control rod drive mechanism; or locally opening the trip breakers.
- Failure to recognize the loss of component cooling water (CCW), including failure to note alarms.
- Failure to start the PDP before the temperature of the RCP sealwater exceeds 230 ºF (within about 9 min after the loss of CCW), recognizing that charging is not working and accounting for step 6 of procedure ES01.

Aspects that relate to the consideration of these failures include the following:

- The operators will receive a signal on lampbox 3M02 for distribution panel DP1201, but they have to read it to realize which panel is affected; the alarms for the other distribution panels are close together, and the quality of the human-machine interface is not very good.
- If the operators miss the manual reactor trip step in procedure POP04-VA-0001 they might continue on and could get stuck in Addendum 4, which is a series of tables that distract the operator from what needs to be done in this particular scenario.
- When the reactor trips, step 4 of procedure E0 (Reactor Trip and Safety Injection) would direct the operators to procedure ES-01 (Reactor Trip Response) since SI would not be required. This happens before step 27 of E0, which calls for securing the RCPs. Based on the interviews conducted, the operators would know (from their training and from being sensitized to the importance of RCP seal cooling) that they should trip the RCPs, but the procedure path will not actually lead them to that step.
- Step 6b of procedure ES-01 directs the operators to verify that charging is in service, and step 6c directs them to check RCP seal injection flow. Response not obtained (RNO) item c.3) for step 6c indicates that if a charging pump is not running, then steps should be taken to start the PDP. If it takes 8 minutes to get to this point, then there would be insufficient time remaining to establish seal-injection flow from the PDP.
- In procedure ES-01, it is not until step 14 is reached that the operators are asked whether a RCP is running. This is somewhat confusing, since the earlier RNO step addresses forcing water to the RCP seals gradually using the PDP.
- Overall, the operator interviews indicated that there was sufficient training and operator sensitivity to RCP seal cooling and RCP protection such that they would take the proper actions. The interview also indicated, however, that there was some confusion regarding procedural guidance for starting the PDP.
- In light of the potential confusion regarding the procedure, it was not clear that the operators would reach the step in procedure ES-01 to start the PDP within the ~8 minutes available before the RCP sealwater temperatures reached 230 °F.

This information was considered in addressing the potential failure to trip the RCPs and re-establish seal injection via the PDP pump.

*G8.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
Per the ATHEANA process, each of three experts provided a three-point distribution of the probability for this HFE. The results were then discussed to arrive at the following consensus distribution:

Mean HEP = 3E-02
1%-tile = 3E-03
99%-tile = 2E-01

The discussion that led to these values centered on the various contributing unsafe actions:

- Operators getting lost in the procedures;
- Operators failing to trip the RCPs; and
- Operators failing to start the PDP.

The analysts performing the expert elicitation stepped through each of the procedures and found them to be confusing and potentially misleading. Information from the operator interviews, however, indicated that available training and the sensitivity to the RCP seal cooling issue equip the operators with an adequate understanding of the need to secure the RCPs. The interviews also revealed, however, that there was some lack of clarity regarding the need to start the PDP, even though the procedure provided direction to do so. Overall, the analysts believed that the operators might, based on their training, not follow the procedures to the letter; this might help them to avoid becoming lost in confusing procedures. The operators are considered to be more likely to succeed than to fail, but failures would not be unexpected.

*G8.4.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | There are a lot of steps to go through. If the operators were diverted to other procedures or to other actions, the time available could be an issue. This was assessed to be good/fair as a PSF. | ND |
| **Time Pressure** | Time pressure was not cited as a major factor. | 0 |
| **Stress** | The workload was judged to be relatively light even though there are many steps that must be completed. The loss of charging was judged to be a relatively simple condition for operator response. Stress was not addressed as a significant factor. | N/P |
| **Scenario Complexity** | After reviewing the procedures, the sequencing of events that would be needed, and the familiarity with RCP seals compared with the lack of familiarity of the PDP, the analysts concurred that training seemed to provide adequate guidance for the diagnosis of this scenario. | 0 |
| **Indications of Conditions** | Indications were cited as being “not bad”, with plenty of indications available to the operators. | N/P |
| **Execution Complexity** | Relatively straightforward execution; no issues identified. | N/P |

| Factor | Comments | Influence |
|---|---|---|
| **Training** | Training in the response to a loss of CCW and seal injection is performed about once per year, and at least every two years. This was assessed to be "fair" with respect to the role of training as a PSF. | 0 |
| **Experience** | Experience was not addressed separate from training. | 0 |
| **Procedural Guidance** | The procedures themselves were considered to be very confusing and might cause the operators to be diverted into Addendum 4 of the annunciator response procedure. The operators might not be led to the instruction to trip the RCPs. | ND |
| **Human- Machine Interface** | The analysts judged that the control room employs control boards and displays standard to nuclear power plants, but not necessarily mimic displays found in many control rooms. The Quantitative Parameter Display System (QPDS) provides means to obtain a quick understanding of important plant parameters. | 0 |
| **Work Processes** | Not explicitly addressed outside team dynamics and communication. | 0 |
| **Communication** | During the simulator exercise observed by the analysts, crew communications were not always audible. The strategy on the part of the crew was for the supervisor to take the lead, read procedure steps aloud, direct activities, and request feedback from ROs on actions taken or parameters being monitored. | ND |
| **Team Dynamics** | The analysts observed that, in simulator exercises, the crew worked well together and there was a clear hierarchy of command. | N/P |
| **Other** | With respect to the availability of operating staff, it was noted that Units 1 & 2 have separate but identical control rooms. This means that if one unit is in an abnormal condition, substantially more than the minimum complement of operators can be available to assist.<br>With regard to the realism of diversions and deviations from accident sequences, the analysts believed that the deviations from procedures were helpful in this case and reflected the operators' training. If operators attempted to follow procedural directions verbatim, there could be problems. | N/P |

#### *G8.4.4 Comparison of Drivers to Empirical Data*

This section compares the drivers identified by the analysts for HFE 2A in the table above to those derived from crew performance.

- **Adequacy of time**. The SAIC analysis made using ATHEANA judged the adequacy of time to be a negative driver. The analysts cited the number of procedural steps that had to be completed and the potential to be diverted to other procedures, which would further detract from the available time. With respect to the response of the crews, time was also assessed to be a negative driver. It was noted, however, that with more training relevant to this scenario, there should have been adequate time to respond.
- **Time pressure**. Time pressure was not addressed as a separate factor with respect to crew response. The SAIC analysts judged time pressure to be non-driver.

- **Stress**. The SAIC analysts concluded that, even though there were many steps to be completed, the workload presented by this scenario was relatively light. The lack of stress was characterized as a nominal/positive driver. With respect to the assessment of crew response, stress was assessed not to be a driver, although some level of stress was observed among the crews.
- **Scenario complexity**. The SAIC analysis concluded that scenario complexity was not a driver, in light of the procedural guidance available, the familiarity of the operators with the RCP seal considerations (although not necessarily with the PDP). Observations of the crew response, on the other hand, indicated that the complexity of this scenario was a significant challenge because of all the things happening in close proximity in time. Based on crew response, this was assessed to be a main negative driver.
- **Indications of conditions**. The SAIC analysis concluded that plenty of indications were available to the operators, and that this was a nominal/positive driver. The crew responses indicated, however, that the availability of indications was a negative driver. This was particularly the case because of the failure of instrument bus 1201.
- **Execution complexity**. The SAIC analysts noted that the actions required were relatively straightforward, and cited execution complexity as a nominal/positive driver. Assessment of the crew responses indicated that this was a non-driver, although there could be some challenge to the operator depending on his level of experience.
- **Training and experience**. Training was assessed by SAIC to be a non-driver; it was judged to be conducted often enough to meet the basic needs of the operators. With regard to the crew response, training and experience were judged to constitute a main negative driver. Training related to the loss of RCP seal cooling is usually focused on scenarios involving a loss of offsite power, and not the types of failures encountered in the simulator exercises.
- **Procedural guidance**. Both the SAIC analysis and the assessment of crew response characterized the available procedural guidance as a negative driver. SAIC cited the potential to get diverted before reaching the step to trip the RCPs. The crew response indicated that the procedures relevant to this scenario were sufficiently complex that they presented a challenge to the operators to perform necessary actions in time.
- **Human-machine interface**. The SAIC analysts judged the human-machine interface to be a non-factor, weighing the lack of a mimic display against the digital display system used to convey important plant parameters. In the assessment of crew response, this aspect was characterized as nominal/positive, without further elaboration.
- **Team dynamics**. The SAIC analysts observed that the crew worked well together, and that there was a clear hierarchy of command. This was therefore a nominal/positive driver. It was assessed to be a non-driver based on observations of actual crew response. One instance of poor communication was noted, but otherwise team dynamics were adequate.
- **Other factors**. The SAIC team explicitly addressed several factors that were not evaluated for the actual crew responses. These included communication (assessed to be a potential negative driver owing to a tendency for some crew interactions to be inaudible); and the availability of crew members from the other unit to provide support (a nominal/positive driver).

*G8.4.5 Comparison of Qualitative Analysis to Empirical Data*
There were important differences between the SAIC assessment and the observations of crew response. These included especially training and experience and scenario complexity (both main negative drivers of crew response, but treated as non-drivers in the SAIC analysis). The SAIC analysis correctly identified the adequacy of time and procedural guidance as negative drivers.

*G8.4.6 Impact on HEP*
The SAIC analysis yielded a mean probability for the HFE of 0.03. With respect to the observations of the crews, none of whom successfully established flow from the PDP in time, this analysis would appear to have underestimated the impacts of some of the drivers, including the applicability of the operators' training and experience for this particular scenario and the complexity of the scenario.

## G8.5 HFE 3A

Human failure event (HFE) 3A considers failure of the operating crew to isolate the ruptured steam generator and control pressure below the SG PORV setpoint before the SG PORV opens. This precludes the potential that the PORV might stick open.

*G8.5.1 Summary of Qualitative Findings*
The HRA analysts considered the following potential contributions to the occurrence of the HFE:

- The possibility that failure to secure AFW could play a role.
- Failure to diagnose the SGTR and enter procedure E-30.

Aspects that relate to the consideration of these failures include the following:

- If nothing else is happening, the operators have indications from the main steam line radiation monitors, pressure indications, and trending (e.g., of SG levels). Therefore, significant information would be available to cue them to necessary actions.
- The time factor is beneficial even though there are many interim actions to take as part of the procedure.
- Procedures provide direction for SI initiation, isolation of any faulted SG, and then addressing the SGTR.

This information was considered in addressing the potential failure to isolate the ruptured SG and to control pressure to prevent opening of the SG PORV.

*G8.5.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
Per the ATHEANA process, each of three experts provided a three-point distribution of the probability for this HFE. The results were then discussed to arrive at the following consensus distribution:

Mean HEP = 3E-03
1%-tile = 1E-03
99%-tile = 6E-02

The SGTR scenario was considered to be well trained (at the plant in question, the training occurs every six months) and easy to recognize, even for novice operators. Therefore the cognitive part of the operator action is not considered difficult; the operators are more likely to omit something from the execution due to distraction. The time factor is also beneficial. Therefore the operator error in this case was judged to be in the realm of extremely unlikely.

*G8.5.3 Summary Table of Driving Factors*

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | If no other competing scenario is occurring, there is plenty of time available. There are 7 – 8 valves to close; these actions can be done in 10 minutes and are often trained on faster in the simulator. | N/P |
| **Time Pressure** | Time pressure was considered to be minimal. | N/P |
| **Stress** | Workload was considered to be relatively high, but combined with the lack of time pressure, stress was judged to be “medium”. | ND |
| **Scenario Complexity** | The cognitive part of the action is not difficult, but the operators are more likely to omit something from the execution portion due to distractions. Familiarity through training should provide good guidance to minimize this possibility. | N/P |
| **Indications of Conditions** | Radiation monitoring alarms and pressurizer level indications provide initial cues, and SG level without feedwater flow should provide ample guidance with respect to which is the ruptured SG that requires isolation and pressure control. | N/P |
| **Execution Complexity** | As noted for scenario complexity, the execution portion of this action includes a sufficiently large number of steps that there is the possibility that one or more could be omitted. | ND |
| **Training** | Training on response to a SGTR is performed about every six months. Together with the available procedural guidance, this is judged to be more than adequate. | N/P |
| **Experience** | Experience was not addressed separate from training. | 0 |
| **Procedural Guidance** | Available procedures and conditional information pages (CIPs) provide substantial guidance to allow conditions to be diagnosed. | N/P |
| **Human- Machine Interface** | The analysts judged that the control room employs control boards and displays standard to nuclear power plants, but not necessarily mimic displays found in many control rooms. The Quantitative Parameter Display System (QPDS) provides means to obtain a quick understanding of important plant parameters. | 0 |
| **Work Processes** | Not explicitly addressed outside team dynamics and communication. | 0 |
| **Communication** | During the simulator exercise observed by the analysts, crew communications were not always audible. The strategy on the part of the crew was for the supervisor to take the lead, read procedure steps aloud, direct activities, and request feedback from ROs on actions taken or parameters being monitored. | ND |
| **Team Dynamics** | The analysts observed that, in simulator exercises, the crew worked well together and there was a clear hierarchy of command. | N/P |

| Factor | Comments | Influence |
|---|---|---|
| **Other** | With respect to the availability of operating staff, it was noted that Units 1 & 2 have separate but identical control rooms. This means that if one unit is in an abnormal condition, substantially more than the minimum complement of operators can be available to assist. | N/P |

*G8.5.4 Comparison of Drivers to Empirical Data*

This section compares the drivers identified by the analysts for HFE 3A in the table above to those derived from crew performance.

- **Adequacy of time**. The SAIC analysis made using ATHEANA judged the adequacy of time to be a nominal/positive driver. The analysts noted that there is plenty of time, and the actions should not take more than about 10 minutes. The crew responses were characterized indicating that the adequacy of time was a non-factor. It would appear that the responses required a relatively small portion of the overall time available, so that time would seem to have been more than adequate with respect to the exercises as well.
- **Time pressure**. Time pressure was not addressed as a separate factor with respect to crew response. The SAIC analysts judged time pressure to be minimal, constituting a nominal/positive-driver.
- **Stress**. The SAIC analysts concluded that the relatively high workload, coupled with the lack of time pressure, constituted a negative driver with respect to stress. With respect to the assessment of crew response, stress was assessed not to be a driver.
- **Scenario complexity**. The SAIC analysis concluded that the scenario was relatively simple from a cognitive standpoint, and was therefore a nominal/positive driver. Observations of the crew response concluded that scenario complexity was a non-driver.
- **Indications of conditions**. The SAIC analysis judged the indications of conditions to comprise a nominal/positive driver, due to the availability of clear cues (radiation alarms, pressurizer level, etc.) with respect to existence of a SGTR. The observations of the crew responses confirmed this assessment.
- **Execution complexity**. The SAIC analysts noted that response required a large set of actions, such that execution complexity could be a negative driver. Assessment of the crew responses indicated that this was a non-driver.
- **Training and experience**. Training and experience was assessed by SAIC to be a nominal/positive driver because of the degree to which the SGTR is simulated. This was confirmed via the crew exercises.
- **Procedural guidance**. Both the SAIC analysis and the assessment of crew response characterized the available procedural guidance as a nominal/positive driver because the procedures supported response for this nominal scenario very well.
- **Human-machine interface**. The SAIC analysts judged the human-machine interface to be a non-factor, weighing the lack of a mimic display against the digital display system used to convey important plant parameters. In the assessment of crew response, this aspect was characterized as nominal/positive, without further elaboration.

- **Team dynamics**. The SAIC analysts observed that the crew worked well together, and that there was a clear hierarchy of command. This was therefore a nominal/positive driver. It was assessed to be a nominal/positive driver based on observations of actual crew response as well, with some examples of good teamwork noted.
- **Other factors**. The SAIC team explicitly addressed several factors that were not evaluated for the actual crew responses. These included communication (assessed to be a potential negative driver owing to a tendency for some crew interactions to be inaudible); and the availability of crew members from the other unit to provide support (a nominal/positive driver).

*G8.5.5 Comparison of Qualitative Analysis to Empirical Data*

This scenario was relatively straightforward, and the crews did not appear to be significantly challenged by it. The factors that could affect crew performance as assessed in the SAIC analysis were very close to the observations of the crew exercises.

*G8.5.6 Impact on HEP*

The SAIC analysis yielded a mean probability for the HFE of 0.003. No main negative drivers were noted in either the SAIC analysis or the evaluation of crew responses. Since all of the crews responded successfully well within the time available, a relatively low probability of failure appears to be justified.

## G8.6 Comparison Summary

*G8.6.1 Predictive Power*

The overall predictive power of the ATHEANA method as applied by the SAIC team in this study was judged to be fair. The SAIC team identified drivers and other qualitative elements that were generally consistent with the empirical observations for three of four HFEs. For the fourth HFE, the ATHEANA team substantially underestimated the impact of the negative drivers.

Similarly, for three of the four HFEs, the probabilities calculated by the ATHEANA team were consistent with the trend observed from the empirical observations. This was particularly the case for the two more straightforward actions (corresponding to HFEs 1A and 3A). Even for more difficult HFE 1C, the SAIC analysis was very consistent with the observed crew performance, if the analysis focuses on the outcome (avoiding challenging a SG PORV) rather than on the time criterion (action within 40 minutes). In that case, however, there were notable differences in the qualitative analysis as compared to the factors observed during the empirical studies. For HFE 2A, the SAIC team predicted a relatively low HEP; the analysis generally did not reflect the degree of challenge observed during the simulator exercises.

It might be noted that the SAIC applied ATHEANA in a manner that did not exercise some elements of the method. In particular, a focus of the ATHEANA method is on identifying error-forcing contexts (EFCs) and unsafe acts (UAs). In the case of the empirical studies, the EFCs were essentially pre-defined by the scenarios that were run. Beyond that, however, although SAIC broke some of the HFEs down into UAs, they did not attempt to quantify the UAs separately. Instead, they estimated probabilities for overall UAs corresponding to the HFEs themselves. It is not clear from the available information whether or to what extent the qualitative analysis and quantitative results might have been different had it been possible to follow through more fully with the ATHEANA process.

The table below summarizes the extent to which the ATHEANA analysis compares to the data collected from the simulator exercises for each of the four HFEs. As the table indicates (and as summarized above), the analyses were most consistent with the empirical data for HFEs 1A and 3A. The qualitative analysis diverges from the empirical observations for HFE 1C, and is not consistent with either the qualitative or quantitative observations for HFE 2A.

| HFE | HEP Ranking | | HEP Values | | Consistency with Empirical Data* | | |
|---|---|---|---|---|---|---|---|
| | Empirical | HRA | Empirical | HRA | Drivers | Op Story | Timing |
| 2A | 1 | 1 | 4/4<br>Very difficult | 0.03 | 2 | 2 | 3 |
| 1C | 2 | 1 | 3/4<br>Difficult | 0.2 | 3 | 3 | 2 |
| 1A | 3 | 3 | 0/4<br>Fairly difficult | 0.055 | 5 | 5 | 5 |
| 3A | 4 | 4 | 0/4<br>Easy | 0.003 | 4 | 5 | 5 |
| * 1= poor; 2= moderately poor; 3=fair; 4=moderately good; 5=good | | | | | | | |

G8.6.1.1 Qualitative Predictive Power in Terms of Drivers

Part of the ATHEANA process entails evaluating the scenario in the context of a set of PSFs, and to rank the quality of each PSF. These PSFs do not correspond directly to the factors used to identify drivers, but they can generally be mapped into these factors. Moreover, the rankings were used by the SAIC team to guide their judgments, but were not explicitly factored into the analyses.

As indicated in the table above, the SAIC ATHEANA team was quite successful in identifying the drivers for the two less difficult actions. One of these (HFE 3A, the failure to respond to a relatively uncomplicated SGTR) was generally very straightforward, with little in the way of potential deviations presenting themselves. For the other (HFE 1A, failure to initiate F&B cooling when the red path to FR-H1 was not satisfied), the potential difficulties in navigating the procedures were identified and taken into account. As in the exercises, however, available indications, coupled with substantial time, allowed a good chance for success.

The drivers were somewhat less closely defined for HFE 1C, although the SAIC team was successful in identifying several potential issues with implementing the action. The results reflected the difficulties presented for this scenario.

For HFE 2A, the SAIC team did not capture some of the main negative drivers. It is likely that the team was influenced by an expectation of success, given that procedural guidance was available (although it was judged to be potentially confusing) and that there was training for this action. For this action, consideration of possible deviation scenarios might have led to a result more consistent with the observed crew performance.

G8.6.1.2 Qualitative Predictive Power in Terms of Operational Expressions

The SAIC team was quite effective in identifying potential failure causes for the HFEs. Even for HFE 2A, for which the assessment did not match the empirical evidence well, the team identified the potential that the operators could get stuck at various points in the procedures and not reach the step to start the PDP within the available time. Based on their interviews with operators, the SAIC team judged that there was adequate understanding of the need to trip the RCPs, although there was less clarity with regard to starting the PDP. The conclusion reached was that there was a reasonable chance of failure, but a higher likelihood of success.

With regard to HFE 1C, the SAIC team recognized that there were priorities related to heat removal that could delay actions to limit pressure rise in the ruptured SG. These actions are in large measure what led three of the four crews to take more than 40 minutes to perform the actions directly related to avoiding challenges to the SG PORV. For two of these three crews, however, other actions to control the scenario still allowed SG pressure to remain below the SG PORV setpoint.

For HFE 1A, the team acknowledged the potential that the lack of clear path to FR-H1 could affect the operators' ability to reach the steps for F&B cooling. There was also a recognition, however, that other indications of lack of heat transfer would still provide compelling indications of the need for F&B cooling. Because the crews had substantial time, they were successful in all four of the simulator exercises. The qualitative and quantitative results were consistent with this as a moderately reliable action, with some potential for failure due to the misleading flow indicator.

For HFE 3A, the analysis did not identify specific challenges to the operating crew. The relatively straightforward nature of the response of the crews was consistent with the assessment made by the SAIC team.

G8.6.1.3 Quantitative Predictive Power

With respect to the quantitative estimates obtained by the SAIC ATHEANA team, the results appear to be consistent with the crew observations for the two less difficult HFEs, HFE 1A and 3A. The resulting values are arguably entirely consistent with the observations, although zero failures in four and three trials constitute very limited data sets. Nevertheless, the relative results for these two HFEs appear to be reasonable, given that there are qualitative reasons that the probability for HFE 1A would be higher than that for HFE 3A.

The quantification for HFE 2A (0.03) is not consistent with the empirical results in which four of four crews failed to achieve the intended outcome. The analysts in this case did not give sufficient weight to the qualitative factors they identified as potential drivers, and did not account for some factors that were main drivers in the crew responses.

With respect to HFE 1C, the quantitative estimate developed by the SAIC team is somewhat optimistic relative to the experience of three of four crews failing to complete the action within 40 minutes. Only one of the four teams, however, failed to prevent lifting of the SG PORV. Given that time is not a direct determinant of reliability in the case of the ATHEANA application, the result could be judged to be very close to the empirical evidence on a functional basis.

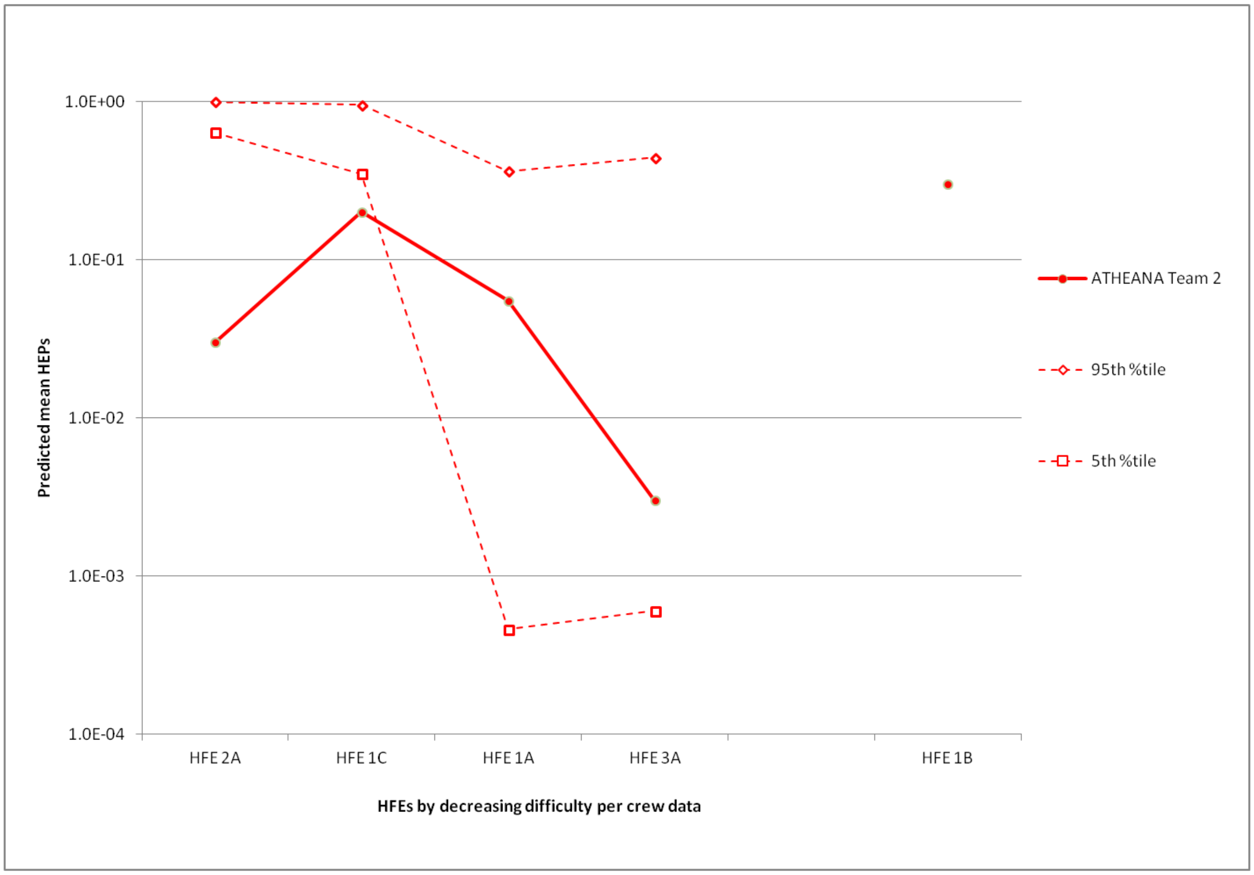



Figure 0.8 Predicted Mean HEPs with Uncertainty Bounds by SAIC HRA Team with ATHEANA

*G8.6.2 Traceability of the Qualitative Analysis and Quantification Process*

The ATHEANA method has a heavy emphasis on a good qualitative analysis, and the SAIC team followed the corresponding guidance to an appreciable extent. It was relatively easy to understand the thought processes behind their qualitative analysis. This was reflected in the descriptions that correspond to operational stories about how failure might come about in some cases, and in the comments regarding the ratings of the PSFs.

The translation of the qualitative information into quantitative estimates was somewhat less transparent. There is no formal model for quantification in ATHEANA that necessarily promotes reproducible results. In the case of the SAIC team's use of the method, the qualitative factors were discussed among three experts, and each suggested a three-point probability distribution. Each of the three inputs for each point were averaged to arrive at the overall three-point distribution for the HFE. A scale was used by the analysts to aid them is making the transition from an overall qualitative ranking to a probability estimate. Nevertheless, it is not easy to trace how the overall qualitative characterizations were defined, and how they were translated into probabilities.

*G8.6.3 Insights on Guidance*

Extensive guidance is available to aid analysts in applying the ATHEANA method. Much of this guidance was used effectively by the SAIC team in its analyses. One aspect of ATHEANA, however, is that the full implementation can in some cases be quite onerous. It would appear that the SAIC team made some decisions regarding portions of the method that would be omitted or performed in less detail than the guidance would suggest (e.g., in considering specific UAs). In addition, the available documentation suggests that the expert-elicitation process may have been less formal than the guidance would call for. For example, there is no evidence that a facilitator was involved, that great care was taken to control the introduction of bias or to calibrate the inputs. The quantification process appears to have been somewhat less formal, with each of three experts providing his or her inputs in a more holistic framework for the HFE. Moreover, the analysis stopped with the definition of the initial three points of a distribution (1%-tile, 99%-tile and most likely value), without following through by further developing the distribution. It is not possible to assess the impacts that would have resulted from following the available guidance more explicitly.

*G8.6.4 Insights for Error Reduction*

As noted in the section on insights related to “operational expressions”, the SAIC team identified opportunities to reduce the probability of failure for three of the four HFEs. These were tied to features of the response that could lead the operators to delay reaching appropriate steps in the procedures, or where (in the case of HFE 2A) there was a substantial lack of clarity in the procedure. These insights derived from the qualitative analysis were valuable, even where the assessment of drivers might not have been consistent with those exhibited in the crew exercises.

*G8.6.5 Remarks on Strengths and Weaknesses*

The extensive, formal qualitative analysis that is integral to the ATHEANA method provided the most useful results from the analyses as performed by the SAIC team. This was the case even though it appears that some simplifications were made to allow the team to use the method in a more practical manner. The issue for the SAIC team appeared to be in assessing the relative degree to which potential deficiencies they identified translated into possible contributions to failure, both qualitatively and quantitatively. ATHEANA presents a challenge to analysts in general because of the degree to which it relies on expert judgment in the quantification phase. This approach makes reproducibility of the results inherently more difficult. This challenge may have been magnified somewhat because the SAIC team was not able to follow the process as fully as the ATHEANA guidance might have called for.

## G.9 ATHEANA (NRC)

Note: The ATHEANA/NRC submittal provided the drivers and qualitative analysis in a discussion form, generally without identifying the drivers in terms of the PSFs defined in Table 3.1 for in the method assessments. This discussion describes specific expectations concerning crew behaviors, the responses, and the dynamics of the scenario. The results of the qualitative analysis, as reflected in this discussion and in the analyses of the HFE, make it clear that the scope of this analysis includes all of the PSFs used in the empirical study comparisons. The attribution of the drivers identified by the analysis team to PSFs and the assessment of the factor ratings provided in this summary was performed by the assessor with feedback from the HRA team.

### G9.1 HFE 1A

*G9.1.1 Summary of Qualitative Findings*

The analysis identified two main contributions:

a) take too long to work through the procedures for all explainable departures
b) fail for unexplained reason (with mean 5 E-3)

Contribution a) is quantitatively negligible. The quantification considered various departures from the nominal path and the variability in several steps and found that the time to establish F&B had a mean value of 12 minutes and a maximum value of 40 minutes, based on 100,000 Monte Carlo (MC) trials[8]. Given the 45-minute time window, this contribution has a mean value of 0.0.

Contribution b) reflects the HRA team's view that having time available (in the scenario) beyond the 99th percentile of the performance time, some crews may still fail.

The situation assessment is complicated by:

- AFWP appears to operate properly with full flow but feed to the SG does not occur due to the open AFW recirculation valve.
- SG levels drop very rapidly in this scenario and will be much lower at entry to E-0 than in usual training exercises, due to MFP trip, failure of SUFP, and AFWP, and additional shrink following reactor trip. This may delay recognition that no SGs are being fed. In addition, the rapid fall in SG level could cause some "confusion, delay, or jumping ahead in the procedures".
- The failure of the start-up feed pump (SUFP) may cause crews to try to restart; however, when the remaining two feed pumps trip, tripping the reactor becomes the only option so that this failure is not a major influencing factor.

The successful assessment is likely to be supported by the fact that all four SG levels are tracking identically, although only one AFWP is (apparently) providing feed. The trainer/operators from the plant indicated that the crews would recognize the open AFW recirculation valve, because this event has occurred at this plant and is included as part of regular simulator training.

[8] Distributions combined during the Monte Carlo trials were developed by the analysis team during the elicitation sessions, which usually provided estimates for 1st, 10th, 25th, 50th, 75th, 90th, and 99th percentiles.

In conclusion, while there are several complicating factors, the rapid decrease in SG levels is the primary negative factor. The analysis team also assumes that the open AFW recirculation valve would not be a major negative factor, due to the coverage of this event in training.

*G9.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
The distribution for the HEP is obtained with MC trials.

> Mean HEP = 5 E-3
> 1%-tile = 3E-5
> 99%-tile = 3E-2

A major part of the quantification deals with the quantification of "taking too much time", based on the scenario map and the elicited probabilities and time distributions. The dominant contribution b), is however based on the elicitation of phi-y, as indicated in sheet 'Branch Probabilities' of file 'Scen 1 LOFW then SGTR Rev 4-6.05.2011.xls' and file 'HFE-1A-1B Calc Sheet.pdf'. phi-y is the probability for "Unexplained failure to establish B&F in 45 minutes". The argument for the phi-y distributionis:

> "Even if there is substantial time beyond that required in the best estimate case and even beyond our 99th percentile time, allowing for non-optimal branches through the procedures and confusion, our team recognizes that some crews, on a bad day, could fail to complete the required actions. Delays on the order of 10 minutes or more would make failure possible. This branch allows for that possibility and is based on observing many crews over many years."

*G9.1.3 Summary Table of Driving Factors*

The team's documentation of the analysis does not address these factors explicitly. The comments and influence rating for each PSF are based on the assessors' reading of the provided analyses and supporting discussion.

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | The analysis describes the time window as "very ample time" and indicates that it is sufficient to allow for variability among the crews, in particular with respect to the key decisions and actions. These are expected to be reactor trip, branch to FR-H1 (from ES0.1), and completion of FR.H1 Step 12. | N/P |
| **Time Pressure** | Time pressure is not mentioned. The analysis estimates the probability of a reactor trip within 45 seconds, resulting in a time window of 45 minutes, to be 0.94. | N/P |
| **Stress** | Stress is not mentioned. | N/P |
| **Scenario Complexity** | The main complexity is identified as the very rapid drop in SG levels. Two other complicating factors are mentioned: 1) the AFWP with full flow indicated combined with the open AFW recirculation valve resulting in no feed to the SG and 2) the failure of the SUFP. However, both of these are not deemed to be major influencing factors. | ND |

| | | |
|---|---|---|
| **Indications of Conditions** | The indications are sufficient to allow a correct situation assessment. All SG levels are tracking identically while only one AFWP is running should contribute to the operators realizing that no SGs are being fed. | N/P |
| **Execution Complexity** | Difficulties in execution are not mentioned. | N/P |
| **Training** | The training of the open AFW recirculation valve as a part of regular training is identified as a key factor for correctly assessing that the sole functioning AFWP is not feeding the SG. | N/P |
| **Experience** | Experience is not addressed separately from training. | 0 |
| **Procedural Guidance** | No issues related to procedural guidance are identified: no problem areas were noted during elicitation discussions, which tracked path through EOP. | 0 |
| **Human- Machine Interface** | The HMI is not discussed. Implicitly, the ability to see all 4 SG levels and trends concurrently is important to correctly assessing the open AFW recirculation valve as a potential explanation of the observed indications. | 0 |
| **Work Processes** | No mention | 0 |
| **Communication** | No mention. | 0 |
| **Team Dynamics** | Elicitation noted that crews do weak or no briefings as they transition between EOPs. They brief later on. | 0 |
| **Other** | | 0 |

*G9.1.4 Comparison of Drivers to Empirical Data*

The HRA identified scenario complexity as the only negative driver for HFE 1A. The nominal/positive drivers include adequacy of time, which is described as "very ample time", indications of conditions, and training. In the empirical data, indications of conditions is also listed as a negative driver, although one underlying contribution is also covered by the scenario complexity (the underlying issue should not be double-counted in the comparison). The other contribution to the negative rating for indications of conditions in the empirical data refers to the FR-H1 entry criterion not being indicated in the plant computer.

The HRA identified no procedural issues for this HFE. The empirical data notes that the CSF status trees do not address misaligned valves (as a potential cause of Loss of Secondary Heat Sink: i.e. the misalignment of the AFW recirculation valve that led to AFW pump flow but no feed to SG B). In the observed crew performances, this characteristic of the CSF status tree led the crews to take longer to enter FR-H1 and to start B&F but all crews were successful.

*G9.1.5 Comparison of Qualitative Analysis to Empirical Data*

For HFE 1A, the qualitative analysis of the HRA was correct in the following points:

- the open AFW recirculation valve was not a major issue
- the behavior of the SG levels would allow the crews to recognize the AFW issue
- the adequacy of time would not be an issue

Concerning the diagnosis of the AFW problem, the qualitative analysis pointed to the identical behavior of the 4 SGs although one was expected to be fed. The observations showed that the discrepancy between a decreasing level in SG B and the expected AFW to SG B supported the diagnosis.

There was no evidence for the qualitative analysis prediction that the unusually rapid SG level decrease could lead to confusion or delay. While this prediction was not supported, it was not flagged as likely to lead to failure of the HFE; consequently the unsupported prediction should be assessed in light of the overall qualitative analysis result.

Overall, the team's qualitative analysis indicated that the crews would not have any major issues related to HFE 1A and that time would not be a problem. This is fully supported by the empirical data.

*G9.1.6 Impact on HEP*
The HEP is within the empirical bounds. The HEP is based on "unexplained failure" of the crew, which represents a residual failure mechanism.

**G9.2 HFE 1B**

*G9.2.1 Summary of Qualitative Findings*
HFE 1B differs from HFE 1A primarily in the time window available to the operators, based on whether they trip manually (1A) or the reactor trips automatically (1B, shorter time window). For HFE 1B, the qualitative analysis of the HRA team is essentially the same as for 1A. In summary, the scenario complexity factors are the initially unusually rapid SG level decrease, the masked failure of AFW flow, and the failure of the SUFP. These complicating factors are more critical for performance in view of the reduced time window.

*G9.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
The HEP distribution obtained in the ATHEANA analysis of HFE 1B is:

> Mean HEP = 3E-1
> The analysis did not directly address uncertainty in the HEP. Instead it estimated the uncertainty in the time to complete the actions:
> 1%-tile ≈ 2 minutes
> 99%-tile ≈ 33 minutes
> P(t > 13 minutes) = 0.30 is calculated from the full distribution.

Although not required to do so, the ATHEANA team evaluated the probability that the crews would not trip the reactor manually in less than 45 seconds (following the loss of MFPs 12 and 13) to be 0.06. This would lead to an automatic reactor trip.

The analysis notes that the response to HFE 1A and 1B are identical in terms of required actions and the time distributions for this response.

*G9.2.3 Summary Table of Driving Factors*
The HRA team stated that, "other than the shorter time available", the qualitative analyses for HFE 1B and 1A are identical.

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | 13 minutes is very little time. | MND |
| **Time Pressure** | The available time is not consistent with operator expectations. This implies lack of time pressure, lack of urgency. However, they will have already missed the 50% wide range SG level criterion for initiating F&B, so midway through FR H-1 it will be clear that actions should be carried out expeditiously. | ND |
| **Stress** | Stress is not mentioned. | N/P |
| **Scenario Complexity** | The complexity of the scenario due to three complicating factors 1) one AFWP delivering full flow that does not feed the SG due to an open AFW recirculation valve, 2) the very rapid fall of SG level, and 3) the SUFP failure. The rapidly falling SG level may delay the operators' realization that no SGs are being fed. In addition, this rapid fall could cause some confusion, delay, or jumping ahead in the procedures. The failure of the SUFP could result in an attempt to restart it. | ND |
| **Indications of Conditions** | The indications are sufficient to allow a correct situation assessment. 1) All SG levels are tracking identically while only one AFWP is running should contribute to the operators realizing that no SGs are being fed. | N/P |
| **Execution Complexity** | Difficulties in execution are not mentioned. | N/P |
| **Training** | The training of the open AFW recirculation valve as a part of regular training is identified as a key factor for correctly assessing that the sole functioning AFWP is not feeding the SG. | N/P |
| **Experience** | Experience is not addressed separately from training. | 0 |
| **Procedural Guidance** | | |
| **Human- Machine Interface** | The HMI is not discussed. Implicitly, the ability to see all 4 SG levels and trends concurrently is important to correctly assessing the open AFW recirculation valve as a potential explanation of the observed indications. | 0 |
| **Work Processes** | No mention | 0 |
| **Communication** | No mention. | 0 |
| **Team Dynamics** | Elicitation noted that crews do weak or no briefings as they transition between EOPs. They brief later on. | 0 |
| **Other** | | 0 |

*G9.2.4 Comparison of Drivers to Empirical Data*
N/A. This situation was not observed in the simulator study (all crews tripped manually).

*G9.2.5 Comparison of Qualitative Analysis to Empirical Data*
N/A. This situation was not observed in the simulator study (all crews tripped manually).

### *G9.2.6 Impact on HEP*

The HEP is based on the MC simulation of task duration for the different paths shown in the scenario map. The distributions estimated by the HRA team for the timing of the crew response elements are identical for both HFE 1A and 1B. The HRA team notes that the shorter time available in the case of scenario 1B may not be obvious to the crew; consequently, the same time distributions for the response of the crews are estimated for these HFEs.

## G9.3 HFE 1C

In scenario 1 (Total LOFW with induced SGTR), HFE 1C considers the failure of the crew to isolate the ruptured steam generator and control pressure below the setpoint for the SG pilot-operated relief valve (PORV). The conditions for this HFE arise given that the operators successfully establish F&B (modeled by HFE 1A). If the crews successfully initiate F&B, they will be able to establish AFW to one or several SGs. However, if they do so, an SGTR will occur in the first SG that is fed. HFE 1C models the crew's actions to isolate the affected SG and to control SG pressure.

### *G9.3.1 Summary of Qualitative Findings*

In its analysis, the HRA team identified the following factors and their underlying causes. For comparison, these causes are discussed relative to the conditions for HFE 3A, which is a basic SGTR scenario. In summary, the factors and causes are:

- Reduced time window (40 minutes vs. 3-4 hours for HFE 3A)
- Distraction: loss of all feedwater
- Confusion: open AFW recirc valve and running AFW pump and rapid drop in SG levels
- Loss of vigilance: second initiating event during recovery from first event [correction of analysis documentation regarding effect of loss of vigilance ("typographical" error)
- Masking/complexity: close proximity of SGTR to restoration of AFW
- Workload/competing tasks: need to complete FR.H-1 before addressing the SGTR
- Complexity: F&B flow path may trigger the Integrity of Critical Safety Functions

The dominant scenarios are paths 7, 2, 5, and 15 in the HFE 1C quantification map; these respectively contribute 0.30 (i.e. 30%), 0.27, 0.27, 0.12 of the HEP. (Path 6 contributes an additional 3%.) These paths are described next.

Path 7. A plant operator (PO) is dispatched to AFW recirculation valve during HFE 1A. SGTR is not noticed initially. The operators perform FR-H1 Steps 13-23 & 24, fail to jump to FR.P-1, instead cycling back to E-0 and not branching to E-3 in E-0 steps 13 or 22. At this point, the elapsed time has exceeded the TW.

Path 2. PO is dispatched to AFW recirculation valve. In contrast to path 7, the SGTR is immediately detected. The operators perform FR-H1 Steps 13-23 &24, fail to jump to FR.P-1, instead cycling back to E-0 and following E-0 to E-3. 48% of the performances on this path take more than 40 minutes.

Path 5. PO is dispatched. SGTR is not noticed immediately. Similar to Path 7 but they transfer to E-30 at E-0 Step 13. 48% of the performances on this path take more than 40 minutes.

Path 15. PO is not dispatched. The crew never restores AFW. Although the plant is cooling down and stable and the induced SGTR may be avoided, the path is defined as failure.

Note that no paths in which operators jump straight to E-30 are important, because they are very unlikely. Most of these also take longer to complete. Also, all paths involving failure to dispatch an operator to the AFW recirc valve early on, take substantially less time, but are unlikely, because of training on that failure.

*G9.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
The distribution for the HEP is obtained with MC trials.

Mean HEP = 4.2E-2
1%-tile ≈ 8E-3
99%-tile ≈ 1E-1

An important insight of this analysis is the priority of FR-H1 with respect to the SGTR, coupled with the possible dispatch of a PO to the AFW recirculation valve. Due to the scenario definition, if the crew does not dispatch a PO, it may establish F&B and complete more of FR-H1 prior to restoring AFW. In this case, the crew has somewhat more time to deal with the SGTR, which is not induced until the AFW is restored.

*G9.3.3 Summary Table of Driving Factors*
The comments are based on the assessors' reading of the provided analyses. For HFE 1C, the correspondence between the ATHEANA analysis and these factors is fairly straightforward.

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | The SGTR may not be detectable within the time window due to the masking effect of the AFW flow, which is restored shortly before the SGTR.<br>The priority of FR-H1 and the need to complete this procedure before dealing with the SGTR may use part of the available time. (The completion of this procedure uses crew resources.) | ND |
| **Time Pressure** | Note: as described, the issue regarding the priority of FR-H1 noted in the analysis could result in time pressure, for cases in which the operator is aware of the SGTR. | 0 |
| **Stress** | The loss of vigilance following recovery from the first event may result in a suboptimally low level of vigilance / stress. | ND |
| **Scenario Complexity** | Some confusion may have arisen due to the open AFW recirculation valve in the prior phase of the scenario (HFE 1A). Here the principal complexity results from masking (Indications of Conditions).<br>An additional complexity element is that the F&B flow path may trigger the Integrity CSF. | ND |
| **Indications of Conditions** | The SGTR follows closely the restoration of AFW. The restored AFW flow will also lead to an SG level rise and may consequently mask the SGTR. It may not be until the SGs fil up and AFW is throttled that the operators will note the uncontrolled SG level. This may occur past the 40-min. criterion. | ND |

| | | |
|---|---|---|
| **Execution Complexity** | Difficulties in execution are not mentioned. | N/P |
| **Training** | Training is not mentioned as a driving factor. | 0 |
| **Experience** | Experience is not addressed separate from training. | 0 |
| **Procedural Guidance** | The priority of FR.H-1, which addresses a Critical Safety Function, may result in delay, even if operators notice the SGTR immediately. Note: the HRA team notes that the priorities of the procedures are appropriate, rather than a defect of the procedural guidance. Nevertheless, in this scenario, these priorities have a negative impact on the HFE. | ND |
| **Human- Machine Interface** | No mention. | 0 |
| **Work Processes** | No mention | 0 |
| **Communication** | The HRA team notes that it found little evidence for timely, formal briefings, which could have helped in this scenario. | 0 |
| **Team Dynamics** | (See related issue under PSF communication.) | 0 |
| **Other** | | 0 |

*G9.3.4 Comparison of Drivers to Empirical Data*

The HRA analysis for HFE 1C identified all negative drivers in the empirical data: adequacy of time, scenario complexity, indications of conditions, and procedural guidance.

Additionally, it identified a suboptimally low level of vigilance/stress as a potential negative driver, as a result of the crew successfully recovering from the first event (the LOFW preceding the induced SGTR). There was no empirical evidence for this reduced vigilance.

*G9.3.5 Comparison of Qualitative Analysis to Empirical Data*

In the HRA's qualitative analysis for HFE 1C, three of the four dominant paths (7, 2, and 5) relate to not transferring to the SGTR procedure and to delays in transferring to E-30. Path 15 is an HFE failure, based on not dispatching a local operator to the AFW recirculation valve, which has the consequence that AFW is not restored and the induced SGTR does not occur.

The analysis of this HFE defines "functional success" paths in which the operators isolate the SG and control pressure to avoid lifting of the SG PORVs by following instructions in FR-H1 and FR-P1. These paths are not included as dominant contributions to the HFE. The observed crew performances (empirical data) show, however, that remaining in these ERGs, in some cases with criteria assessment failures within these ERGs led to failure of the HFE.

The empirical data shows diverse contributing causes and failure modes of the HFE that do not match the predictions of the qualitative analysis. The qualitative analysis and corresponding scenario map prepared by the HRA team show that entering FR-P.1 would lead to success, with no probability that the crews would fail to complete the actions within the time window. Of the three observed failures, one crew (R) failed to reduce pressure, one crew (Q) reduced RCS pressure late, and one crew (T) isolated the SGTR late.

- Crew (R) failed to reduce pressure at FR-H1 step 28 through a misinterpretation of the FR-H1 procedure (active loop definition). Crew R failed to identify the SGTR due to the masking effect of the AFW flow to the ruptured SG. They completed and exited FR-H1 (without stopping the last HHSI pump in Step 28) and during a crew brief planned to go through E-10 to cooldown. They diagnosed the SGTR during the brief on exit of FR-H1 and transferred a few minutes later to E-30 from the E-10 CIP. In E-30, they failed to connect the decreasing RCS pressure with the lifting of the SG PORV; they attributed the decreasing RCS pressure to a primary leak and transferred consequently to EC31 (tube rupture combined with primary leak). The SG PORV release was stopped by the RCS cooldown in EC-31.

- Crew (Q) reduced RCS pressure late. Crew Q (correctly) remained in FR-H1. They identified and isolated the ruptured SG while in FR-H1 (but could not transfer to E-30 due to the priority of FR-H1). The crew failed to stop the last HHSI pump while in FR-H1 within the TW because they followed the FR.H1 emphasis on improving subcooling first (43' after the end of the 40' TW).
- Crew (T) isolated the SGTR late. Crew T was caught up in FR-H1 and FR-P.1 and isolated the ruptured SG late (approximately 90' after the SGTR, i.e. 50' after the TW). It should be noted that Crew T avoided lifting of the SG PORV.

In terms of the more detailed contributors to HFE failure, the qualitative analysis correctly identified

- masking of indications of a ruptured SG due to AFW flow to the ruptured SG
- workload and the competing tasks, in particular, driven by the priority of the FR-H1 procedure

However, it missed the delayed appearance of radiation indications.

*G9.3.6 Impact on HEP*

The HEP for HFE 1C is driven by the sum of paths 7, 2, 5, in which the time to reach the ruptured SG isolation steps exceeds the TW. The duration of the paths is estimated by MC simulation of the durations of the tasks along the path, based on task duration distributions elicited from the HRA analysis team.

The predicted mean HEP is approximately a factor of 10 below the 5th percentile empirical bound (4.2E-2 vs. 3.5E-1).

## G9.4 HFE 2A

HFE 2A addresses failure of the operators to trip the RCPs and start the positive displacement pump (PDP) to prevent a RCP seal LOCA following the total loss of RCP seal cooling.

*G9.4.1 Summary of Qualitative Findings*

The analysis assumes two parallel paths, either of which leads to success if completed in time. The first path proceeds through the EOPs, through the first 6 steps of ES-0.1. The second path involves "short-cutting" through the EOPs recognizing the loss of CCW and charging and the loss of seal injection. This short-cutting is justified as the use of skill-ofthe-craft or by using the AOP for Loss of CCW. The crew must continue to proceed through the EOPs, but if they recognize the loss of CCW, they could (and should) immediately trip the RCPs and respond to the loss of seal injection. The HEP results from the minimum estimated time to complete the steps along the two success paths, if either pathway (t3 + t5) or (t2 + t4 + t5) is less than 7-9 minutes.

The analysis predicts that the two paths, i.e. proceeding through the EOPs and short-cutting, would be carried out in parallel (rather than as alternatives).

The most influencing factors are time available and competing demands for the operators' attention. The analysis notes that if "nothing else is going on" the AOP to trip the RCPs would apply when cooling and seal injection to the running pumps are lost. In this event, the loss of the AC bus draws the operators' attention to recover or take manual control of the affected equipment and the additional failures lead to a reactor trip require the operators to enter E-0. In addition, having two operators instead of the usual three makes it less likely that they notice the loss of CCW, charging, and the coming seal LOCA. If they do notice, tripping the RCPs and starting the PDP will be quick and easy to accomplish.

*G9.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
The distribution for the HEP is obtained by considering the time variability in the success paths using MC simulation:

Mean HEP = 0.075

The analysis did not directly address uncertainty in the HEP. Instead it estimated the uncertainty in the time to complete the actions:

1%-tile = 0.3 minutes
99%-tile = 12 minutes

Because the success criteria were 7-9 minutes, it calculated:

$P(t > 7 \text{ minutes}) = 0.10$ and $P(t > 9 \text{ minutes}) = 0.05$

The mean HEP is obtained from the average of P at 7 and 9 minutes, = 0.075

[Note: for the direct path through the EOPs, it calculated that $P(t > 7 \text{ minutes}) = 0.30$ and $P(t > 9 \text{ minutes}) = 0.11$]

A branching probability for the two parallel paths, phi-2, was initially elicited but was not used in the final quantification.

*G9.4.3 Summary Table of Driving Factors*
Factors for which no comments are shown in this table were not discussed as driving factors.

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Expeditious but methodical progress through the EOPs,, which is required given the reactor trip, leads to a required time close to the available time.<br>The reduced crew size mentioned under 'Team Dynamics' may be related to this factor as well. | MND |
| **Time Pressure** | | |
| **Stress** | | |

| | | |
|---|---|---|
| **Scenario Complexity** | The initial loss of a vital AC bus draws the operators' attention to recovering manual control of the affected equipment. The additional failures leading to a reactor trip require entry into E-0. | ND |
| **Indications of Conditions** | | |
| **Execution Complexity** | The manual actions are easy and quick to accomplish. | N/P |
| **Training** | With respect to the loss of DP 1201 (the vital AC bus), the analysis notes that the problems are well-known and the response is well trained. "It just takes time". | |
| **Experience** | | |
| **Procedural Guidance** | The AOP for Loss of CCW would be straightforward but in the scenario, the initial loss of an AC bus and subsequent failures leading to reactor trip require the operators to follow the AOP for loss of DP 1202, E-0 then ES 0.1, as well as notice the loss of CCW and respond to it. The EOPs eventually address the RCP issues, but not early on. | ND |
| **Human- Machine Interface** | | |
| **Work Processes** | | |
| **Communication** | | |
| **Team Dynamics** | The fact that only two ROs are available instead of the usual three make it less likely that they noptice the loss of CCW, charging, and the coming seal LOCA. | ND |
| **Other** | | |

*G9.4.4 Comparison of Drivers to Empirical Data*

For HFE 2A, the qualitative analysis correctly identified the following drivers:

- scenario complexity, with the loss of the vital AC bus and corresponding need to manually control the affected equipment and the reactor trip requiring entry into E-0 (rather than dealing with the Loss of CCW alarm directly)
- adequacy of time (the main negative driver), based on the expected time required to perform the procedural steps being close to the TW
- procedural guidance, based on the need to follow E-0 as a result of the reactor trip.

*G9.4.5 Comparison of Qualitative Analysis to Empirical Data*

The qualitative analysis correctly identified that

- the crew would be carrying out the EOPs in parallel with an AOP and/or knowledge-/training-based response to the Loss of CCW
- time would be critical if the EOPs were followed methodically
- and that the bus failure would result in significant complexity and workload.

The analysis failed to identify that the condition to be below 230 °F in order to restore seal cooling would block the operators from starting the PDP. In addition, it assumed that the tripping of the RCPs would be closely coupled with the starting of the PDPs.

*G9.4.6 Impact on HEP*
The predicted mean HEP for HFE 2A is a factor of 10 below the 5th percentile empirical bound.

It is unclear to the assessor whether in the modeling of the parallel paths (EOPs and knowledge-/training-based response), the elicitation accounted for interference between the paths. (In other words, did the estimated distributions for the durations of activities in one path account for the crew following the other path in parallel?).

**G9.5 HFE 3A**
HFE 3A addresses the failure of crew to isolate the ruptured steam generator and control pressure below the SG PORV setpoint before SG PORV opening.

*G9.5.1 Summary of Qualitative Findings*
The analysis identified two main contributions:

a) take too long to work through the procedures for all explainable departures
b) fail for unexplained reason (with a mean of 2E-5)

Contribution a) is quantitatively negligible (mean = 0.0). The quantification considered various departures from the nominal path and the variability in several steps and found that the time to establish F&B had a mean value of 40 minutes and a maximum value of 90 minutes, based on 50 000 Monte Carlo trials[9]. Given the 2-3 hour time window, this contribution has a mean value of 0.0.

Contribution b) reflects the HRA team's view that while the chance of requiring more than 2 hours is nil (negligible), some crews may still fail.

*G9.5.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)*
The distribution for the HEP is:

Mean HEP = 2E-5
1%-tile ≈ 2E-7
99%-tile ≈ 7E-5

[9] Distributions combined during the Monte Carlo trials were developed by the analysis team during the elicitation sessions, which usually provided estimates for 1st, 10th, 25th, 50th, 75th, 90th, and 99th percentiles.

*G9.5.3 Summary Table of Driving Factors*
Factors for which no comments are shown in this table were not discussed as driving factors.

| Factor | Comments | Influence |
|---|---|---|
| **Adequacy of Time** | Time is ample (no crews are expected to require more than 2 hours) and the time window is 2-3 hours.<br>Time is expected to be sufficient for recoveries because the EOP/ERG provides “additional flags” to send the operators back to the E-0 – E-30 sequence and the time available is much longer than the time required. | N/P |
| **Time Pressure** | | |
| **Stress** | | |
| **Scenario Complexity** | | |
| **Indications of Conditions** | | |
| **Execution Complexity** | | |
| **Training** | Perfect match of training, given basic SGTR. | N/P |
| **Experience** | | |
| **Procedural Guidance** | Perfect match of procedures. | N/P |
| **Human- Machine Interface** | | |
| **Work Processes** | | |
| **Communication** | | |
| **Team Dynamics** | | |
| **Other** | | |

*G9.5.4 Comparison of Drivers to Empirical Data*
For HFE 3A, the training, procedural guidance, and time match the basic SGTR scenario. In addition, there is ample time.

*G9.5.5 Comparison of Qualitative Analysis to Empirical Data*
The qualitative analysis found that there were no negative factors and ample time. The main failure contribution is an “unexplained” or residual failure.

The performance of the crews fully supports this qualitative analysis.

*G9.5.6 Impact on HEP*
The mean HEP of 2E-5 is dominated by the “unexplained” or residual failure contribution. It should be noted that the residual failure contribution is different for HFE 3A and HFE 1A. The documentation of the analysis does not discuss the basis for the team’s judgment of the appropriate distribution.

The mean HEP is an order of magnitude below the 5%tile bound of the empirical HEP.

## G9.6 Comparison Summary

### *G9.6.1 Predictive Power*

#### G9.6.1.1 Qualitative Predictive Power in Terms of Drivers

At the level of driving factors, the ATHEANA (NRC) analyses of the HFEs (excepting HFE 1B, for which there is no empirical reference) were quite successful in identifying nearly all of the negative drivers.

#### G9.6.1.2 Qualitative Predictive Power in Terms of Operational Expressions

In terms of operational expressions and the overall qualitative analysis (comparison of qualitative analyses), the predictive analyses were less successful.

For HFEs 1A (F&B in LOFW) and 3A (basic SGTR), the overall prediction is that the crews would not have any major issues. This was fully supported by the empirical data.

For HFE 1C (isolation and pressure reduction in induced SGTR subsequent to LOFW and F&B), the predictions correctly identified that the performance of FR-H1 could delay the completion of the required tasks. On the other hand, the predictions failed to identify the issues with the RCS pressure reduction steps in FR-H1. Moreover, the predictions showed that paths in which the crews entered FR-P.1 would lead to success, with no contributions to failure at all. In the data, 3 crews entered FR-P.1; two crews did not complete all required tasks while one crew failed to complete the required tasks within the time available as defined by the success criteria.

For HFE 3A, the predictions coupled the tripping of the RCPs to prevent seal LOCA on loss of CCW strongly with the starting of the PDPs to restore seal flow, i.e. the predictions tended to assume if they successfully trip the RCPs, they successfully start the PDPs (“if they [the crews] do notice, tripping the RCPs and starting the PDPs will be quick and easy to accomplish”). All crews tripped the RCPs but failed to start the PDP (at all). The underlying issue with the 230F criterion for PDP start was not identified.

#### G9.6.1.3 Quantitative Predictive Power

The predicted ranking of the HEPs from the ATHEANA NRC analyses matches the empirical qualitative difficulty ranking very well. The two most difficult HFEs have comparable HEPs while the HEP difference between 3A and 1A is also well supported.

As can be seen in Figure 0.9, the HEPs for the two most difficult HEPs are about one order of magnitude under the 5th percentile bound. The HFE 1A ((third most difficult) is well within the bounds. The mean HEP for HFE 3A (the least difficult HFE) of 2E-5 is over an order of magnitude below the 5th percentile bound of 6E-4. It should be noted however, that the empirical HEP for HFE 3A is relatively high for an action for which no performance issues at all could be identified. The mean HEP is 1.25E-1, driven by the weak evidence of observing only 3 crew performances, all of which were successes. The mean value of 2E-5 predicted by the HRA team nevertheless seems quite low (the predicted upper bound is 7E-5 while the lower bound is 2E-7), even in light of the 2-3 hour time window.

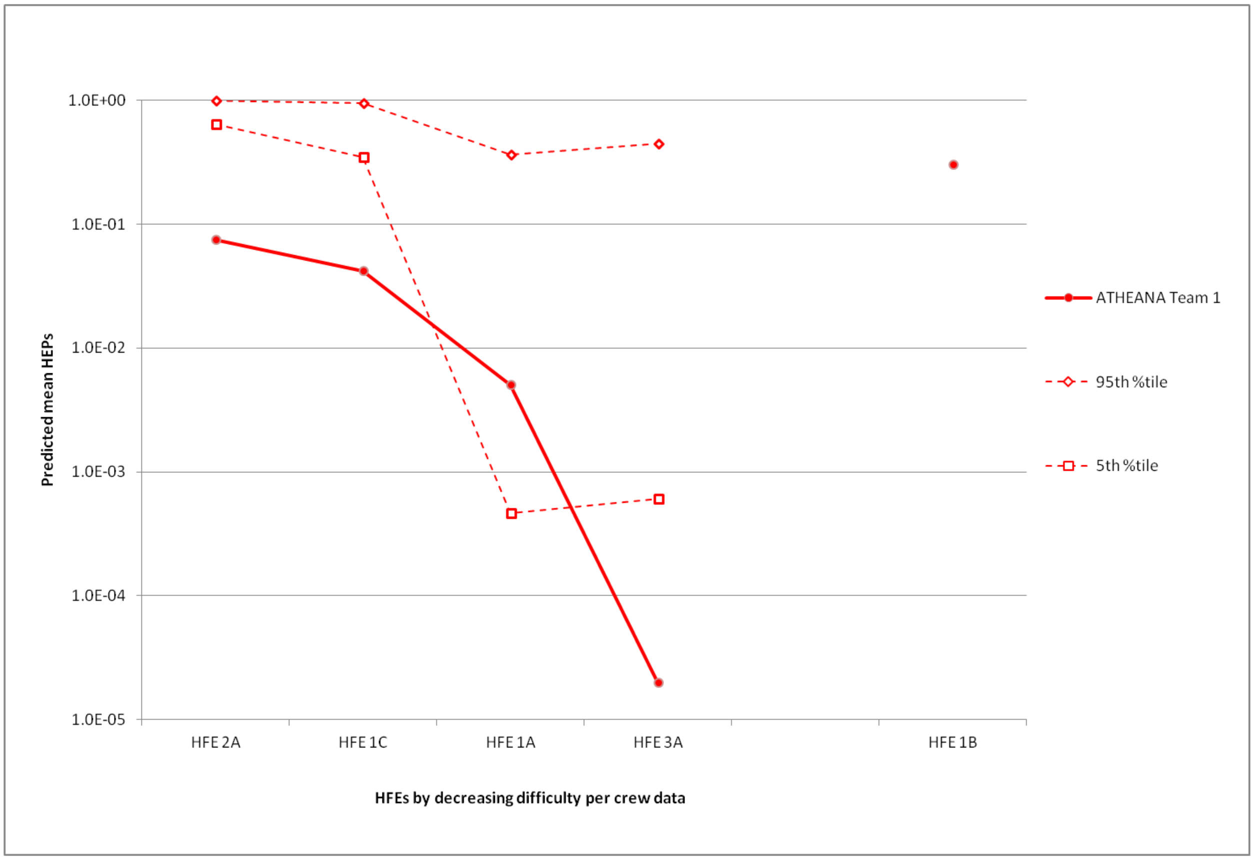



Figure 0.9 Predicted Mean HEPs with Uncertainty Bounds by the NRC HRA Team with ATHEANA

*G9.6.2 Assessment of Guidance and Traceability*

The detailed scenario map developed by this HRA team captures a comprehensive set of potential crew responses, in terms of decisions or branches that may or may not be taken as a result of deliberate decisions of the crew or "error" as well as of the timing of decision and execution subtasks. The development of such a scenario map is not inconsistent with the ATHEANA guidance but has not been used in previously documented ATHEANA analyses. The HRA team notes that its assessment "was performed using the ATHEANA method as documented in NUREG-1624 Rev. 1 (2000), with some modifications that have been developed in the years since its publication."

As such, it can be seen as the innovative application of an analysis approach. The scenario map could potentially represent a highly detailed implementation of the ATHEANA qualitative analysis steps of identifying the expected crew response and corresponding deviation scenarios.

The traceability of the numerical HEP results is very good. The HEP for each HFE is the result of Monte Carlo simulation of all paths in the scenario map, which uses as input the distributions of the branch probabilities and timing distributions. These distributions are mainly based on expert judgment.

According to the HRA team, the elicitation of the branch probabilities and timing (duration) parameters for the scenario map for each HFE followed the structured elicitation process described in the “ATHEANA Users Guide,” NUREG-1880, 2007.

The dominant contribution to failure for some of the HFEs was a residual failure probability if the failure to complete the actions required according to the success criteria and within the time available was negligible (the probability of requiring more time than the available time was found to be negligible in the Monte Carlo simulation). The distribution for this residual failure probability was HFE-specific but guidance for determining this distribution was not shown.

*G9.6.3 Insights for Error Reduction*

The detailed mapping of the potential crew responses in the scenario map provides substantial insights for error reduction. The scenario map makes the HRA model of performance on the task addressed by an HFE very transparent.

*G9.6.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses*

The ATHEANA analyses performed by this analysis team were based on a detailed, comprehensive development of a large number of possible crew responses represented by a scenario map. This map considered numerous possible paths through the procedures, additional rule-based and knowledge-based paths consistent with the conduct of operations as assessed by the HRA analysis team, variability in the timing of the subtasks, as well as its effect on the plant conditions that a crew would face. This detailed analysis allowed the analysis team to identify a most of the performance issues that were observed in the crew performances in the simulator.

In quantitative terms, the difficulting ranking of the HFEs as well as the differentiation in the HEPs were both very good. Although the comparison of predicted HEPs vs reference HEPs were assigned a low weight in the assessment, it is worth noting that the analyses produced mean values below 1E-1 for the two most difficult actions. It is unclear whether these HEPs were due the omission of critical factors or performance issues or instead due to inaccuracies in the distributions elicited by expert judgment for the scenario map input probabilities and task durations.

The comprehensive scenario maps, which are a prediction of a set of very specific paths to success and failure that are considered in the analysis, support the identification of concrete insights / measures for error reduction.

Some of the characteristics of this implementation of ATHEANA include:

- the development of a model representing a comprehensive range of possible crew responses and contributors to failure in scenario- and task-specific terms
- a very large analysis effort to develop this range of responses, represented as a scenario map, and elicit the branch probabilities and durations needed to quantify the map
- the capability to consider a wide range of factors that may affect branching probabilities and durations, coupled with the difficulty to calibrate the expert judgments concerning the impact of these factors
- the capability to propagate uncertainties in all of the input probability and timing distributions to the HEP distribution

# APPENDIX H INTRA-METHOD COMPARISONS

This appendix describes the results of the intra-method comparisons performed to separate analyst effects from method effects to the extent possible and identify the factors that contribute to variability in HRA results for each method. This analysis is done HFE by HFE and the results are summarized across HFEs in Chapter 7, the Intra-Method Comparison chapter in the main body of this report.

## H1 ASEP (Team 1 and Team 2)

### H1.1 HFE 1A

*H1.1.1 Discrepancies in Operational Stories*

As discussed in Section 0, the two teams showed discrepancies on procedural paths operators would take to respond to the scenario, which in turn led to the discrepancies on their estimation of execution time. In addition, although both teams recognized the cues to initiate F&B (F&B), Team 1 considered that the misleading indications and poor procedural guidance might delay operators' diagnosis (see discussion in Section 0).

Overall, the operational stories by both teams were not completely consistent with each other or the empirical data. It should be noted that it appears that Team 1's somewhat better prediction on some aspects of the scenario may be attributed to their more detailed qualitative analysis that seems to go beyond the ASEP methodology.

*H1.1.2 Discrepancies in Driving Factors*

The two teams showed discrepancies on the following three driving factors:

- Adequacy of time. Team 1 predicted the factor of adequacy of time to be a negative driver, whereas Team 2 predicted it to be a nominal/positive driver. Team 1 expected that the crews would be likely to enter other procedures before FRH1 (e.g., FRH5 "Low Steam Generator Level" and F003 "Heat Sink Critical Safety Function Status Tree", but the crews did not actually enter these specific procedures) due to a yellow path on critical safety function status caused by the misleading AFW indication, and therefore there would be less time available for diagnosis and execution. It should be noted that Team 1 recognized that there were two procedural opportunities in FRH5 to diagnose flow diversion, and they assumed in their analysis that operators would make the correct diagnosis at the first opportunity and then quickly transfer to FRH1.

  In contrast, Team 2 assumed that the crews' experience and training would lead them to enter FRH1 directly upon appropriate cues, after entering EO00. The crew data indicated that the crews entered EO00 and ES01 before they entered FRH1. Team 1 explicitly considered these two procedures, but it is not clear based on Team 2's analysis whether the team assumed that ES01 would be entered. Team 2 may have factored entering these procedures into the 5-minute delay as prescribed in ASEP Table 8-1 Item 5.a.

  Team 2 estimated the overall time to perform F&B initiation actions in FRH1 to be two minutes, in contrast, Team 1 team showed some conservatism by estimating two minutes for each action in FRH1. Regardless, Team 1 identified that adequacy of time would be an important driver; while the data suggests that there was plenty of time available.

- Indications of conditions. Team1 concluded that the factor of indications of conditions was a main negative driver, as they predicted that the misleading indications of AFW flow and initial safety function status in HFE 1A would reduce crews' available diagnosis time. In contrast, Team 2 concluded that the factor was not a driver. While the results indicated that the indications of conditions did have an impact on the crews performance, it did not seem to have a large impact since all crews were successful. It seems that Team 2 tended to focus only on whether the cues for required actions would appear, and did not weight heavily how other indications would negatively affect crews' performance. For example, Team 2 recognized that crews would initiate F&B by looking at the steam generator level indication, but it was not clear in their analysis that they considered that the misleading flow indication due to a mis-positioned recirculation valve would delay crew's diagnosis.

- Procedural guidance. Team 1 predicted that the factor of procedural guidance was a negative driver; as they concluded that the symptom based EOPs did not provide guidance to promptly identify the mis-positioned recirculation valve. In contrast, Team 2 predicted that the factor was either a nominal driver or not a driver at all. It appears that Team 2 tended to only consider whether the required post-diagnosis actions were covered in the procedure, and did not address whether the procedures would help crews' timely diagnosis, whereas Team 1 also took into account the impact of the procedural guidance on diagnosis.

*H1.1.3 Discrepancies in Quantitative Analysis*

Given that Team 1 identified more negative drivers than Team 2 for HFE 1A, it may be understandable to see that the HEP estimated by the Team 1 team for those HFEs is larger than that obtained by Team 2. Nevertheless, the HEPs obtained by both teams were small, which is consistent with the finding that none crews failed this scenario.

If we look at the diagnosis and execution HEPs separately, the execution HEP estimated by the Team 1 team are one order of magnitude larger than that by Team 2. Team 2 used ASEP Table 8-5 to obtain a total HEP for F&B initiation steps in FRH1, assuming these steps to be dependent. The Team 1 team used THERP guidance and estimated HEPs for each individual F&B initiation step in FRH1. The different approaches resulted in numerical differences. In addition, the Team 1 team estimated an HEP for EO00 and ES01 actions, whereas it is not clear how those actions were treated in Team 2's analysis (see discussion in Section 0). It should be noted that the procedural steps in F003 and FRH5 were considered an integral part of calculated diagnosis by Team 1, hence the steps did not appear to impact the execution HEP.

For diagnosis HEPs, although both teams used the Nominal Diagnosis Model (ASEP Figure 8-1), the HEP estimated by Team 1 was larger than that by Team 2. As discussed in Section 0, this may be largely due to the Team 1 assumption of more procedural steps to work through and conservatism in estimating execution time (thus leaving less time for diagnosis). In addition, Team 2 selected the lower bound diagnosis HEP as their interviews with operators indicated that operators would be aware of the need for F&B under this scenario. In contrast, Team 1 selected the nominal diagnosis HEP as they predicted that the misleading indications would increase diagnosis difficulty.

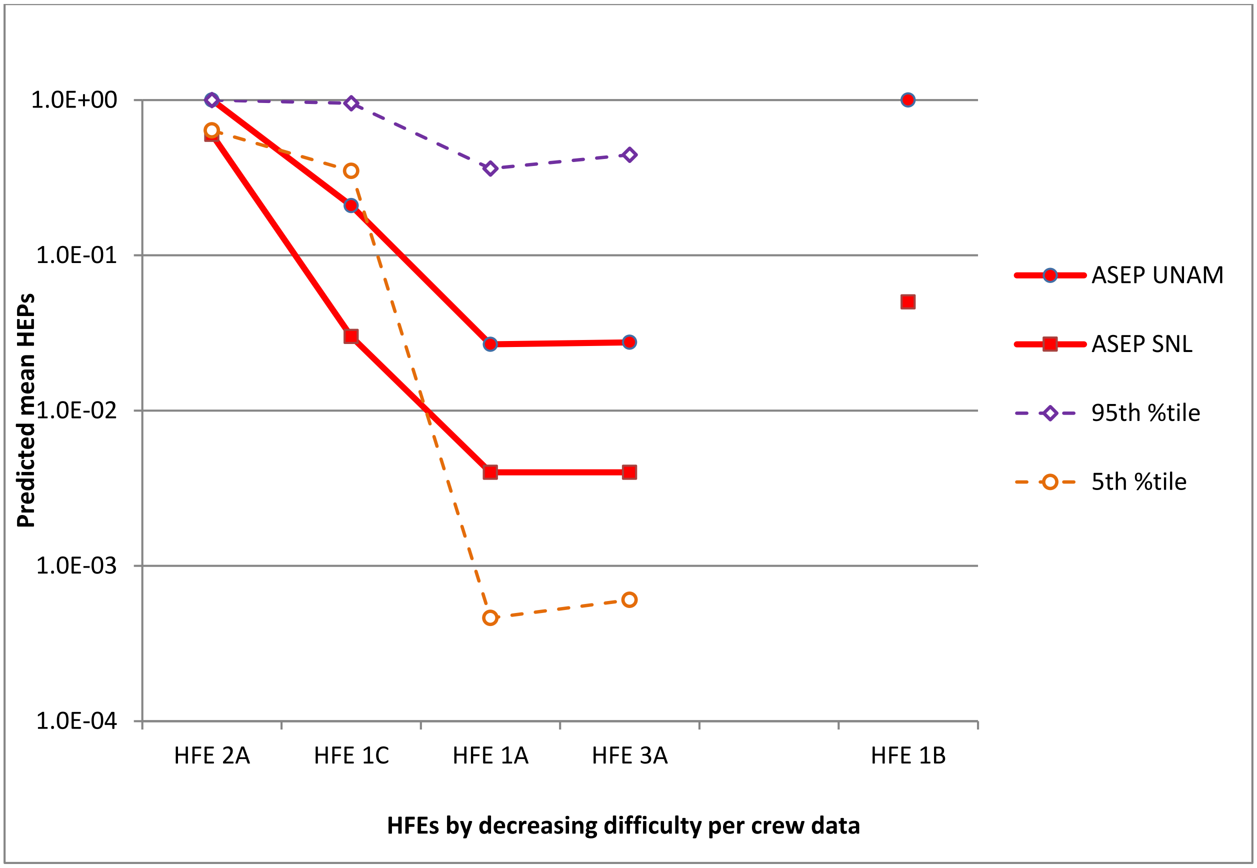



Figure 0.1 Predicted Mean HEPs with Uncertainty Bounds by HRA Teams with ASEP

*H1.1.4 Summary Table of Driving Factors*

| Factor | Empirical | HRA | | Comments |
|---|---|---|---|---|
| | | Team 1 | Team 2 | |
| Adequacy of Time | N/P | ND | N/P | |
| Stress | 0 | N/P | N/P | |
| Scenario Complexity | ND | MND | N/A | |
| Indications of Conditions | ND | MND | 0 | |
| Execution Complexity | N/P | N/P | N/P | |
| Training | N/P | N/P | N/P | |
| Procedural Guidance | ND | ND | N/P | |

## H1.2 HFE 1B

*H1.2.1 Discrepancies in Operational Stories*

Since the procedural paths predicted by both teams in HFE 1B were the same as those in HFE 1A, the two teams showed the same discrepancies on procedural paths and execution time in HFE 1B as those in HFE 1A (see discussion in Sections 0 and 0). As a result, Team 2 estimated the available diagnosis time to be five minutes, whereas Team 1 claimed that there was no time available for diagnosis. Furthermore, based on operator interviews that indicated that operators would be aware of the need for F&B under the scenario, Team 2 selected the lower bound diagnosis HEP, which led to a much smaller HEP compared to that by Team1 (see more discussion in Section 0).

Similar to HFE 1A, although both teams recognized the cues to initiate F&B, Team 1 considered the misleading indications might delay operators' diagnosis (see discussion in Section 0).

*H1.2.2. Discrepancies in Driving Factors*
The two teams showed discrepancies on the driving factor of indications of conditions:

- Indications of conditions. Similar to HFE 1A, Team 1 concluded that the factor of indications of conditions was a main negative driver, as they predicted that the misleading indications of AFW flow and initial safety function status in HFE 1B would reduce crews' available diagnosis time. In contrast, Team 2 concluded that the factor was not a driver. It seems that Team 2 tended to focus only on whether the cues for required actions would appear, and did not weight heavily how other indications would negatively affect crews' performance. For example, Team 2 recognized that crews would initiate F&B by looking at the steam generator level indication, but it was not clear in their analysis that they considered that the misleading flow indication due to a mis-positioned recirculation valve would delay crew's diagnosis.

*H1.2.3 Discrepancies in Quantitative Analysis*
The HEP estimated by Team 2 (0.05) is much smaller in magnitude than that by Team 1 (1.0). Such a difference can lead to opposite conclusions. With the Team 2 estimation, it can be concluded that operators are not likely to fail the scenario; however, Team 1's estimation suggests that operators are most likely to fail. Due to lack of empirical data, it is impossible to check which team's prediction is more accurate.

Both teams' HEPs are diagnosis dominated, and both teams used the Nominal Diagnosis Model (ASEP Figure 8-1) in estimating diagnosis HEPs. The discrepancy in HEPs can be attributed to their prediction of procedural paths and estimation of execution time. Team 1 arrived at a diagnosis HEP of 1.0 as they claimed that there was no time available for diagnosis. In contrast, Team 2 obtained a nominal diagnosis HEP of 0.2 with the lower and upper bounds being 0.05 and 1.0, respectively. The nominal HEP is not substantially different from 1.0, however, Team 2 selected the lower bound as the operator interviews indicated that operators would be aware of the need for F&B under the scenario.

*H1.2.4 Summary Table of Driving Factors*

| **Factor** | **Empirical** | **HRA** | | **Comments** |
|---|---|---|---|---|
| | | **Team 1** | **Team 2** | |
| Adequacy of Time | | MND | MND | |
| Stress | | N/P | N/P | |
| Scenario Complexity | | MND | N/A | |
| Indications of Conditions | | MND | 0 | |
| Execution Complexity | | 0 | N/P | |
| Training | | N/P | N/P | |
| Procedural Guidance | | N/P | N/P | |

### H1.3 HFE 1C

*H1.3.1 Discrepancies in Operational Stories*

The two teams showed discrepancies on how they treated operators' actions earlier in HFE 1A (starting SI) (see discussion in Section 0). Although both teams recognized that the cues for SGTR would be difficult to detect initially, Team 1 also considered the FRH1 steps to stop SI (HH SI pumps), and thus obtained a HEP that was more consistent with the actual crew failure rate.

*H1.3.2 Discrepancies in Driving Factors*

The two teams showed minor discrepancies on the following driving factors:

- Adequacy of time. Although both teams identified adequacy of time as a negative driver for HFE 1C, relatively more detailed qualitative analysis by Team 1 made them understand that the crews would be delayed in getting through FRH1. For HFE 1C, it was crucial for the crews to follow FRH1 steps to stop SI (HH SI pumps) to maintain SG pressure below the setpoint and avoid opening the PORVs, which means that the crews' responses in HFE 1C would be affected by their actions earlier in HFE 1A (starting SI). Team 1 recognized the context of SI (e.g., that it had been started), and explicitly considered the FRH1 steps in their analysis. Hence, the team estimated the available diagnosis time to be 5 minutes. In contrast, although Team 2 recognized that the crews would be in FRH1 initially, they claimed that dealing with a SGTR should be independent of the previous operator actions, and thus did not consider FRH1 steps in their analysis. As a result, Team 2 estimated the available diagnosis time to be 25 to 30 minutes, leading to a more optimistic HEP and less recognition of the delays in responding.

- Indications of conditions. Both teams identified indications of conditions as a negative driver for HFE 1C as they both recognized that the cues for SGTR would be difficult to detect initially. However, Team 2 seemed to believe that this was an important factor, and thus decided to select the upper bound diagnosis HEP. In contrast, Team 1 team seemed to believe that operators training and experience would dilute the effect of this factor, and thus selected the lower bound diagnosis.

*H1.3.3 Discrepancies in Quantitative Analysis*

Although the two teams were, at least to a large degree, consistent on their qualitative analyses for HFE 1C, the HEP estimated by Team 1 was one order of magnitude larger than those by Team 2.

If we look at the diagnosis and execution HEPs separately, the execution HEP estimated by Team 1 is one order of magnitude larger than that by Team 2. Team 2 used ASEP Table 8-5 to obtain a total HEP for SG isolation steps in EO30, assuming these steps to be dependent. Team 1 used THERP guidance and estimated HEPs for each SG isolation step in EO30. The different approaches resulted in numerical differences. It should be noted that although Team 2 did not specify the SG isolation procedural steps, it appeared that Team 1's analysis covered more steps. Another reason for the difference between the execution HEPs is that unlike Team 2, Team 1 considered FRH1 steps to terminate SI (see discussion in Section 0).

For diagnosis HEPs, both teams used the Nominal Diagnosis Model (ASEP Figure 8-1), and the diagnosis HEP estimated by Team 1 is also one order of magnitude larger than that by Team 2. This can be partly attributed to the differences in available diagnosis time estimation due to the different treatments of FRH1 steps to terminate SI (see Section 0). Team 2 estimated the available diagnosis time to be 25 to 30 minutes, whereas Team 1predicted it to be five minutes.

Another reason is, as discussed in Section 0, how the two teams weighed in the factor of indication of conditions.

*H1.3.4 Summary Table of Driving Factors*

| Factor | Empirical | HRA | | Comments |
|---|---|---|---|---|
| | | Team 1 | Team 2 | |
| Adequacy of Time | ND | MND | ND | |
| Stress | 0 | N/P | N/P | |
| Scenario Complexity | MND | N/P | N/A | |
| Indications of Conditions | ND | ND | MND | |
| Execution Complexity | N/P | N/P | N/P | |
| Training | N/P | N/P | N/P | |
| Procedural Guidance | ND | N/P | N/P | |

### H1.4 HFE 2A

*H1.4.1 Discrepancies in Operational Stories*

Compared to Team 2, the Team 1 analysis appeared to have somewhat better predictions of actual crew behavior. This may be due to Team 1's more detailed qualitative analysis, (particularly with respect to factors that could influence diagnosis), which went beyond the basic ASEP guidance. In addition, the two teams seemed to have obtained different information from interviews with operators regarding crews' knowledge and experience about the impacts of the loss of CCW (see more discussion in Section 0).

*H1.4.2 Discrepancies in Driving Factors*

The two teams showed discrepancies on the following driving factors:

- Indications of conditions. Team 1 concluded that the factor of indications of conditions was a negative driver, as they predicted that the unreliable indications due to a bus failure in HFE 2A would complicate the scenario and then degrade crews' diagnosis. In contrast, Team 2 concluded that the indications of conditions factor was not a driver. It seems that Team 2 tended to focus only on whether the cues for required actions would appear, and did not weight heavily how other indications would negatively affect crews' performance for diagnosis. For example, Team 2 noted that the loss of RCP cooling and injection would result in a high RCP thermal barrier alarm, and realized that multiple failures caused by the bus failure would lead to dynamic execution by requiring crew actions to control multiple functions. However, the emphasis seemed to be put on the nature of task execution, rather than the context in which many unreliable indications (e.g., various alarms caused by the failures of Distribution Panel 1201 and Bus E1C) competed for the crews' attention and affected their diagnosis.

- Procedural guidance. Team 1 predicted that the factor of procedural guidance was a negative driver, as they concluded that the symptom based EOPs did not provide guidance to properly address alarms. For example, the team noted that the procedures with high priorities in HFE 2A did not provide instructions for alarm prioritization to start the PDP. In contrast, Team 2 predicted that the factor was either a nominal driver or not a driver at all.

- Training.  Team 1 predicted training to be a main negative driver, but Team 2 predicted it to be a nominal/positive driver.  The discrepancy can be attributed to discrepancies in the results (or at least their interpretation) of their interviews with plant instructors.  The Team 1 team came away with the perspective that the crews were not readily aware of the full impact of a total loss of CCW on RCPs and the immediate actions they needed to take.  However, Team 2 inferred otherwise.

*H1.4.3  Discrepancies in Quantitative Analysis*
Given that Team 1 identified more negative drivers than Team 2, it may be understandable to see that the HEP estimated by Team 1 (1.0) is larger than that obtained by Team 2 (0.6). Nevertheless, the two teams' HEPs are not substantially different.

It should be noted that the two teams seemed to have obtained different information (or at least different impressions) from interviews with operators regarding crews' knowledge and experience about the impacts of the loss of CCW.  The information obtained by Team 2 caused them to select the nominal diagnosis HEP, whereas the information obtained by Team 1 led them to select the upper bound diagnosis HEP.

*H1.4.4  Summary Table of Driving Factors*

| **Factor** | **Empirical** | **HRA** | | **Comments** |
|---|---|---|---|---|
| | | **Team 1** | **Team 2** | |
| Adequacy of Time | ND | MND | MND | |
| Stress | 0 | MND | ND | |
| Scenario Complexity | MND | MND | N/A | |
| Indications of Conditions | ND | ND | 0 | |
| Execution Complexity | 0 | ND | ND | |
| Training | MND | MND | N/P | |
| Procedural Guidance | ND | ND | 0 | |

**H1.5  HFE 3A**
*H1.5.1  Discrepancies in Operational Stories*
Teams 1 and 2 had generally consistent operational stories, which were also consistent with the empirical data.

*H1.5.2  Discrepancies in Driving Factors*
The two teams had consistent predictions on driving factors, which were also consistent with the empirical data.

*H1.5.3  Discrepancies in Quantitative Analysis*
Although the two teams had consistent qualitative analyses, the HEP by Team 1 (0.0275) was one order of magnitude larger than that by Team 2 (0.004).

For diagnosis HEPs, although both teams recognized that operators would be familiar with the cues associated with the scenario, Team 2 selected the nominal HEP, whereas Team 1 selected the lower bound HEP.

Both teams agreed that the HEPs were execution dominated, and Team 1 obtained a larger execution HEP than Team 2.  Team 2 used ASEP Table 8-5 to obtain a total HEP for SG isolation steps in EO30, assuming these steps to be dependent. Team 1 used THERP guidance and estimated HEPs for each SG isolation step in EO30. The different approaches resulted in numerical differences. It should be noted that although Team 2 did not specify the SG isolation procedural steps, it appeared that Team 1's analysis covered more steps.

*H1.5.4 Summary Table of Driving Factors*

| **Factor** | **Empirical** | **HRA** | | **Comments** |
|---|---|---|---|---|
| | | **Team 1** | **Team 2** | |
| Adequacy of Time | 0 | N/P | N/P | |
| Stress | 0 | N/P | N/P | |
| Scenario Complexity | 0 | N/P | N/P | |
| Indications of Conditions | N/P | N/P | N/P | |
| Execution Complexity | 0 | N/P | N/P | |
| Training | N/P | N/P | N/P | |
| Procedural Guidance | N/P | N/P | 0 | |

## H2. HRA Calculator

Two teams – NRC and Scientech – performed an analysis of the five HFEs using the EPRI HRA Calculator. A third team – NRI – performed the analysis using a hybrid CBDT & ASEP method. This intra-method comparison will compare the qualitative and quantitative performance of all three teams. It is important to note that the CBDT & ASEP method differs considerably from the EPRI HRA Calculator and thus the NRI analysis cannot be directly compared to the other two teams in all aspects of the analysis.

In the NRI approach, the cognitive contribution to the HEP is modeled as the sum of the *identification* failure and the *diagnosis/delay* failure. In this approach, CBDT is used to quantify *identification* failure and time reliability curves from THERP are used to quantify the *diagnosis/delay* failure. The EPRI HRA Calculator quantifies the cognitive contribution to the HEP by using the maximum value between CBDT and HCR/ORE (a different time reliability correlation). For this comparison, the CBDT values can be directly compared. The time reliability components cannot be compared except to determine if timing was the main driving factor (time reliability values dominate the analysis). In cases where timing is the main driving factor, the timing itself can be directly compared.

For the execution component, the two approaches use disparate quantification methods and cannot be compared. NRI uses ASEP to quantify execution errors, while the EPRI HRA Calculator uses THERP tables. Comparisons in the execution component will be limited to the qualitative input.

### H2.1 HFE 1A

*H2.1.1 Discrepancies in Qualitative Analysis*

There were a few notable discrepancies in the expected progression of the scenario, and these discrepancies in the qualitative analysis translated directly into discrepancies in the quantitative analysis. The qualitative assessment varied in level of documentation, so these discrepancies are best discussed in the quantitative analysis section below.

The operational story in the qualitative analysis was fairly consistent between each assessment and the actual data. The crew figured out that AFW was lost (as expected) and that led them to FRH1 where F&B was started without difficulty. The only noteworthy difference is that the NRI assessment did not correspond as well to the empirical data because they failed to identify in the qualitative analysis that the open recirculation valve would complicate the scenario. This difference in the qualitative analysis is what drove that team to miss some of the important drivers for HFE 1A discussed in the following sections.

*H2.1.2 Discrepancies in Quantitative Analysis*

At the highest level, there were discrepancies between how each analysis team defined the logic of the HFE and broke the HFE into subcomponents for quantification. The table below summarizes the subcomponents quantified. The major differences are that the NRC team did not credit execution recovery and the Scientech team identified an additional failure mode of ‘failure to enter FRH1.’

| NRC | Scientech | NRI |
|---|---|---|
| | Cognitive: Failure to enter FRH1 | |
| Cognitive: Failure to establish F&B in 45 minutes | Cognitive: Failure to establish F&B in 45 minutes | *Cognitive: Failure to establish F&B |
| Execution: Actuate SI | Execution: Actuate SI | Execution: Actuate SI |
| Execution: Open PORV | Execution: Open PORV | Execution: Open PORV |
| *No recovery credited* | Recovery: Verify RCS Feed Path | Recovery: Verify RCS Feed Path |
| | Recovery: Verify Adequate bleed path | Recovery: Verify Adequate bleed path |

*Note: In this hybrid method identification and diagnosis are quantified using the *combination* of CBDT and a time-reliability correlation; in the EPRI Calculator identification and diagnosis are quantified as one cognitive step using *either* CBDT or a time-reliability correlation.

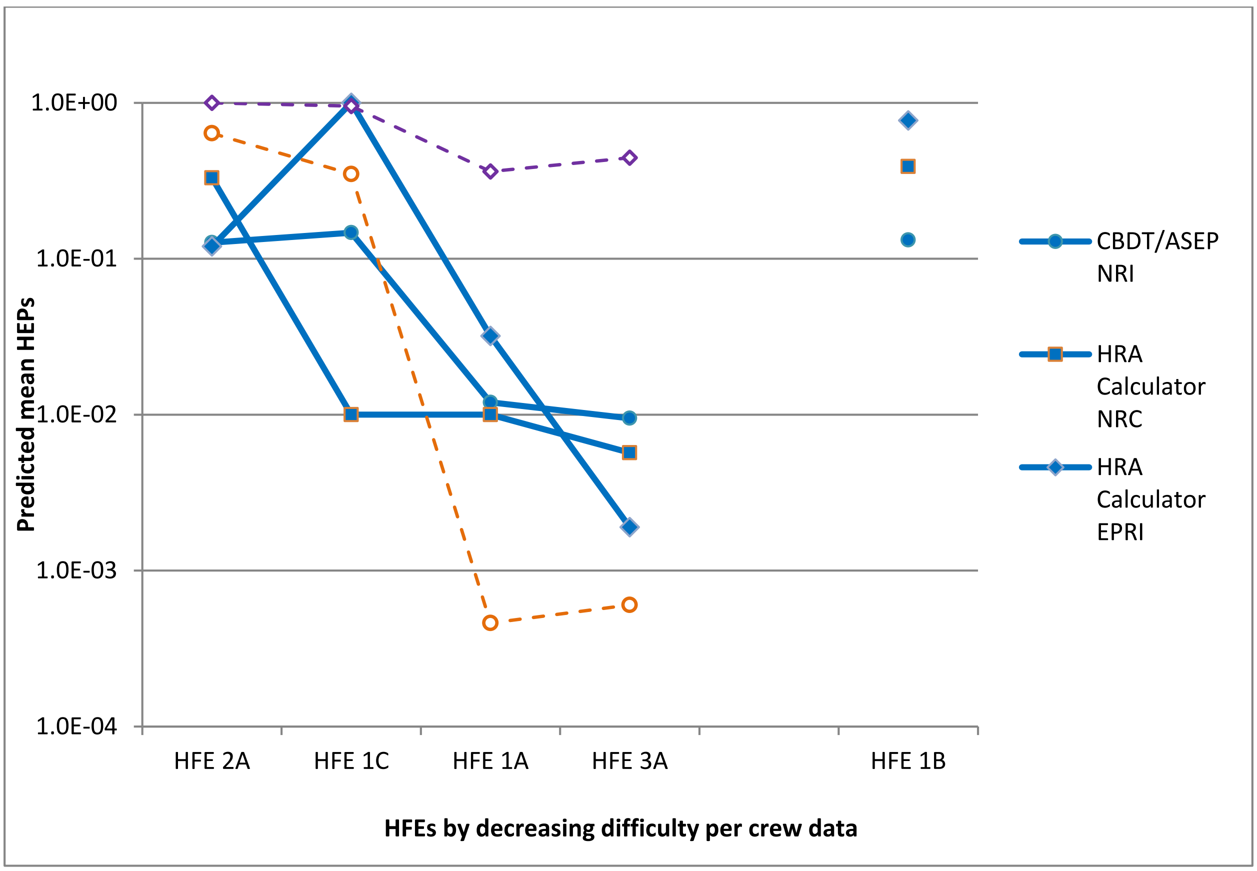



Figure 0.2 Predicted Mean HEPs with Uncertainty Bounds by HRA Teams with HRA Calculator

H2.1.2.1 Cognitive Failure

For the cognitive failure all three teams found that the contribution to the HEP from the CBDT method dominated over the time reliability method(s), and so they all used CDBT to quantify the cognitive portion of failure to establish F&B with the following results:

| NRC | Scientech | NRI* |
|---|---|---|
| 6.2E-03 | 3.1E-02<br>1.6E-04 | 3.8E-03 |

*Note: this is only the identification portion. The diagnosis portion was quantified using a time reliability correlation (contribution of 5.0E-04). This differs from the Calculator, so is not examined here.

Amongst the three analyses, there was some consistency in the identified drivers for the cognitive failure and some differences. The table below describes the identified drivers from the decision trees and their assessment for each method. In the table below, red indicates a major driver, yellow a minor driver and grey is negligible.

| Failure Mechanism | Scientech Base | Final* | NRC | NRI |
|---|---|---|---|---|
| Availability of information | 5.0E-2 | 2.5E-2 | | |
| Information misleading | 1.0E-2 | 5.0E-3 | 3.0E-3 | |
| Skip a step in procedure | 3.0E-3 | 1.6E-4 | 3.0E-3 | 3.0E-3 |
| Failure of attention | 1.5E-2 | 9.6E-4 | 1.5E-4 | 4.5E-4 |
| Misinterpret decision logic | | | | 3.0E-4 |

*Final value includes credit for recovery; NRC and NRI did not credit cognitive recovery.

The major discrepancies in the cognitive assessment are as follows:

- Information misleading
  - *Cue not as stated*: In the scenario there is a false AFW flow indication due to an open recirculation valve. This false indication leads the heat sink tree to indicate a yellow path, instead of the red path that would lead the operators to FRH1. However, there are plenty of other indications that would eventually lead the operators to F&B. Each of the three analysts dealt with this differently. The Scientech team separated the decision to enter FRH1 from the rest of the process to establish F&B (given successful entry into FRH1); for failure to enter FRH1 they choose the "all cues *not* as stated" branch. The NRC team quantified failure to establish F&B as one cognitive step, and modelled the step as "all cues *not* as stated." NRI also quantified failure to establish F&B as one cognitive step, but they modelled the step as "all cues as stated." NRI did not discuss the misindication of the AFW flow in their assessment, but from the audio transcript of the operator interview, it seems the analyst concluded that the misindication of the recirculation valve position would not have a significant effect on the crew's actions because the main indications were the SG levels, not the red path on the CSFST.
  - *Warning of Differences*: While Scientech and NRC both modelled some form of "all cues *not* as stated", the major difference between their assessments of this failure mechanism

was that NRC assessed that the procedures provided a warning of differences between the expected behaviour and actual behaviour (leading to a branch HEP value of 3.0E-03) and Scientech's assessment that there was no such warning (leading to a branch HEP value of 1E-02). In actuality there is no obvious warning in the procedures that the cues may be different. However the crew is trained in recognizing an open recirculation valve (training, though is handled as another branch point). The NRC justifies their decision stating "False AFW flow indication would delay the operators' diagnosis on knowing the complete loss of heat sink. However, the operators would perform F&B when criteria reached. The equal reduction of all SGs' levels would eventually lead operators to learn that all feed water were not available." While the NRC team's decision may be an oversight, it seems like it was a purposeful application of analyst judgement to the trees in order to reflect the teams understanding that although there is misleading information, there are sufficient cues that will lead them down the right path. Although there is no justification in the NRI analysis, this same analyst judgement may be why that team chose the "all cues as stated" branch. Scientech appeared to strictly adhere to the trees in the sense of not adjusting to account for conditions not directly addressed by the trees, but they may seem to account for the additional cues, etc. via recovery factors (see next bullet).

- Cognitive recovery: Only the Scientech team chose to credit cognitive recovery (e.g., self-review, extra crew, etc.). Crediting recovery made a significant difference for Scientech; it reduced one cognitive HEP from 7.8E-02 to 3.1E-02 and the other HEP from 3.0E-03 to 1.6E-04. All three methods assessed the same branch for the "Skip a Step in the Procedure" tree (brach c of $P_{ce}$), leading to a probability of 3.0E-3. The recovery factor Scientech applied reduced the HEP to 1.6E-04, removing that failure mechanism from the list of drivers.
    - NRI noted that the method allows for crediting cognitive recovery, however, in practice, they usually use the recovery factors for the cognitive part in very specific situations (scenarios with very long time windows, very simple and well-known scenarios, etc) which is not the case here.

Other discrepancies, although not major (i.e., do not lead to large differences in numerical results), demonstrate sources of variability in applying the CBDT method. These discrepancies include:

- Workload: For tree $P_{cb}$ (Failure of attention), the analysts had to assess workload. Scientech chose high workload, the NRC chose low and NRI used the average of high and low (which is allowed by the hybrid method, but not in the Calculator).

- Procedure Logic: For tree $P_{cg}$ (Misinterpretation of decision logic), the analysts had to assess the logic in the procedure. The NRI team chose a branch that indicated there was an AND or OR in the procedure (citing step 9 of FRH1, which indeed has an OR statement). The main steps to successfully perform F&B are FRH1 steps 10-13, which do not contain AND or OR logic. Scientech and the NRC did not indicate there was any AND or OR logic in any of the critical parts of the procedures they expected the crew to follow, and therefore took a different path through the decision tree. Specifically, the Scientech analysis determined that step 2 in FRH1 would direct the crew to step 10 (steps to start F&B). In this progression they do not encounter any AND or OR logic (i.e., step 9 of FRH1). This progression is also consistent with that described in the NRC operator interview.

- Parameter Monitoring: Also for tree Pcb, the NRC and NRI were consistent in deciding that the parameter was monitored and alarmed. In the NRC documentation it was not clear which parameter they were basing their judgements on. The NRI analysis specified that "conservatively, 'monitor' was chosen since operators should monitor Critical Safety Functions and SG level during the scenario," but "all essential indications are placed on the front panels and are alarmed." Scientech divided the HFE into two cognitive mechanisms – failure to enter FRH1 and failure to start F&B once in FRH1. The first failure mechanism was consistent with the NRC/NRI in designation as a "monitored" parameter, but differed in indicating that it was not alarmed. Here the parameter being monitored is the CSFSTs, but "a red path on the critical safety function trees is not alarmed from the control room." For the second failure mechanism, Scientech decided the parameter was a "check" parameter and that it was also not alarmed. In this case it seems the parameter is referring to the SG WR level.

#### H2.1.2.2 Execution Failure

While the high level execution steps are consistent between the NRC and Scientech assessments, there is much variance in how those steps were quantified. The two execution steps were 1) Actuate SI and 2) Open both PORVs. The final execution HEP was 4.3E-03 for the NRC and 1.1E-03 (2.0E-02 unrecovered) for Scientech.

In the analysis of these two steps, there were three major discrepancies:

1. For both steps each team quantified the probability of an error of omission. However, for both steps the Scientech team chose table 20-7b item 2, while the NRC team chose item 1 in the same table (factor of 3 difference between two entries):

    a. Item 1: Omission of item when procedures with checkoff provisions are correctly used. Short list, <= 10 items.
    b. Item 2: Omission of item when procedures with checkoff provisions are correctly used. Long list, >10 items.

2. In addition to the error of omission for each step, Scientech also modelled the appropriate error of commission. However, the contribution due to these EOCs was negligible.

3. The NRC did not credit recovery. Scientech (and also NRI) credited the following checks that would lead to recovery: 1) verify RCS feed path and 2) verify adequate RCS bleed path.

H2.1.2.3 Comparison to Empirical Data

The table below provides a breakdown of the HEPs for each assessment:

| HEP | NRI | NRC | Scientech | Experimental |
|---|---|---|---|---|
| **Cognitive (CBDT)** | 3.8E-03<br>[4.3E-03]* | 6.2E-03 | 3.1E-02 | -- |
| **Execution** | 7.7E-03** | 4.3E-03 | 1.1E-03 | -- |
| **Total** | **1.2E-2** | **1.0E-02** | **3.2-02** | **0/4 fail** |

*value in brackets is the total cognitive HEP, which includes the time-reliability curve “diagnosis” value from ASEP, which does not correspond to the Calculator method.
**Quantified by ASEP, does not correspond to the Calculator method

The final HEPs were quite consistent between the three groups, however, there is a difference in contribution due to cognitive errors versus execution errors. For the NRC team, 62% of the HEP was cognitive, while for Scientech, 97% of the HEP was cognitive. While the experimental data yielded no failures, there were some insights that support a mid-range HEP with reasonable contribution from both cognitive and execution error: several crews had cognitive difficulty, including one crew who did not know that the recirculation valve was open (cognitive) and another crew skipped a step in the procedure (execution) which they were able to recover from.

*H2.1.3 Comparison of Drivers to Empirical Data*

At the high level, the drivers indicated by the three teams matched the empirical data fairly well; although the specific operational expressions were not consistently predicted by all teams.

The empirical data concluded that this scenario was moderately complex as the cues did not match the scenario. This maps to the “Scenario Complexity,” “Procedural Guidance” and “Indication of Conditions” PSFs. All three assessments concluded that the “Scenario Complexity” and “Procedural Guidance” were negative drivers. Two of the assessments identified these drivers for the correct reasons (complexity due to mismatch between procedure and cue [all cues not as stated]). The NRI assessment identified these drivers for other reasons which did not correspond to the empirical data (e.g., procedural guidance was a negative driver because it was assessed to contain difficult procedure logic and had poor graphical distinction for the relevant steps). Likewise, Scientech and NRC assessments identified “indication of conditions” as a driver, but the NRI assessment did not. Two of the three assessments, then, predicted the main drivers for the correct reason. Nevertheless, it should be noted that all crews were successful on 1A and therefore these drivers apparently did not have large affects.

Finally, all three assessments found the same operational expression of “skip a step in procedure.” This failure mechanism was manifested in the empirical data: one crew skipped a step (failed to stop RCPs before starting F&B) but recovered from it. Only one assessment (Scientech) credited recovery for this failure mechanism.

Two of the three teams mentioned that the workload would be elevated prior to entry to FRH1 because the crew will be busy trying to check and restore flow to the SGs. This workload was a minor numerical driver. The empirical data confirms that the crews wanted to verify they didn’t have AFW to SG B, and at least 3 crews called a PO to check the valve. There was evidence this increased workload did cause some teams to delay initiation of F&B due to task prioritization (crew members trying to start the AFW pumps to avoid need for F&B). However, there is no evidence that this increased workload adversely impacted their success in the identification and diagnosis of the condition.

All three teams also identified training (on the open recirculation valve) as a positive driver in the qualitative/quantitative analysis; this is consistent with the empirical data. Training factor only translated to an impact on the HEP for the Scientech analysis.

Each of the methods also identified other drivers which were not supported by the empirical data, but these were not major drivers. Stress was the only other driver that was consistently applied by the analyses. There was no evidence of stress in the empirical data.

| Factor | Empirical | NRI | NRC | Scientech |
|---|---|---|---|---|
| Scenario Complexity | ND | ND | MND | MND |
| Procedural Guidance | ND | ND | MND | MND |
| Indications of Conditions | ND | N/P | MND | MND |
| Stress | 0 | ND | ND | ND |
| Adequacy of Time | N/P | ND | N/P | N/P |
| Execution Complexity | N/P | ND | N/P | N/P |
| Other: Workload | | N/P | N/A | ND |
| Training/Experience | N/P | N/P | 0 | N/P |
| Human-Machine Interface | N/P | N/P | N/P | N/P |
| Work Processes | N/A | N/P | N/P | N/P |
| Communication | N/A | N/P | N/P | N/P |
| Time Pressure | ---- | N/A | N/A | N/A |
| Team Dynamics | N/A | 0 | N/A | N/A |

### H.2.2 HFE 1B

As timing dominated this analysis, the NRI results are not comparable to the other teams using the EPRI HRA Calculator and will not be considered in the comparison for this HFE. Also, empirical data was not available for this scenario.

*H2.2.1 Discrepancies in Qualitative Analysis*

In the qualitative analysis Scientech noted that the PSFs for this scenario would be timing, high stress, high workload and cognitive complexity. The NRC identified the same PSFs with the exception of workload, and they categorized the cognitive complexity as moderate rather than high.

The two analyses defined a similar progression through the scenario, with small differences in timing for establishing F&B within 13 minutes:

| Action | NRC | Scientech |
|---|---|---|
| T_delay | 1 min | 7 min* |
| T_1/2 | 8 min | 1 min |
| T_m | 2 min | 5 min |
| **Total** | **11 min** | **13 min** |

*This is broken down in the analysis as 5 minutes for the STA to get to the CR and receive the cue for loss of AFW and 2 minutes to start FRH1.

There are several reasons for the differences in timing between the two analyses. One difference is how the cue is defined and what is included in that definition. This method does not give clear guidance on how to define the cues. T_delay is evaluated at 7 minutes in the Scientech analysis because the *cue to start F&B* was defined as step 2 in FRH1. For the NRC analysis t_delay was evaluated at 1 minute because that is the time it takes for the AFW to kick in and fail, which is the *cue to enter FRH1*. The t_1/2 values then differ according: for Scientech it is simply performing step 2 and transferring to step 10 of FRH1, which takes 1 minute. For the NRC, it includes diagnosing the LOFW, getting into FRH1, and getting to step 10. The difference in cue definition is because Scientech split entry into FRH1 as a separate HFE (see quantitative analysis).

Manipulation time, however, differed substantially between methods. Both teams seem to base manipulation time on completion of steps 10-13 of FRH1. However, the NRC team reduced the manipulation time under normal conditions (which was assessed at 5 minutes) to 2 minutes because of *high* time pressure: knowing they have a short time frame will lead the crew to cool down as close to the max cool down rate as possible. Scientech, however, noted that time pressure would be *low* because the crew is not aware that this is time critical action – they do not train to reach F&B criteria with in a specific time, therefore, the operators will not be rushing through the procedures to meet the PRA defined success criteria.

*H2.2.2 Discrepancies in Quantitative Analysis*

At the highest level, there were discrepancies between how each analysis team defined the logic of the HFE and broke the HFE into subcomponents for quantification. The table below summarizes the subcomponents quantified. The major differences are that the NRC team did not credit execution recovery and the Scientech team identified an additional failure mode of 'failure to enter FRH1.'

| NRC | Scientech |
|---|---|
| | Cognitive: Failure to enter FRH1 |
| Cognitive: Failure to establish F&B in 13 minutes | Cognitive: Failure to establish F&B in 13 minutes |
| Execution: Actuate SI | Execution: Actuate SI |
| Execution: Open PORV | Execution: Open PORV |
| *No recovery credited* | Recovery: Verify RCS Feed Path |
| | Recovery: Verify Adequate bleed path |

The table below summarizes the results of the quantitative analysis for both teams:

| HEP | NRC | Scientech |
|---|---|---|
| **Cognitive (HCR/ORE)** | 3.9E-01 | 7.5E-01 |
| **Execution** | 4.3E-03 | 1.1E-03 |
| **Total** | **3.9E-01** | **7.5E-01** |

H2.2.2.1 Cognitive Failure

In this analysis, time was the driving factor, so both teams used HCR/ORE to quantify the cognitive contribution to the HEP. As discussed in the qualitative analysis section, the analyses differed in their timing assessments. The difference in timing, however, had only a moderate impact on the analysis (HFE of 3.9E-01versus 5.0E-01 for failure to start F&B within 13 minutes).

The main difference between the two analyses here was that Scientech identified and factored into their quantification an additional cognitive step/failure mechanism – failure to enter FRH1 – whereas NRC quantified the entire HFE as one cognitive step. This accounts for a difference in HEP of 5.0E-01 and 7.5E-01.

Another difference between the two assessments was the CP value selected; Scientech used and NRC used CP3. Numerically, this was only a minor difference between the assessment: a change in CP designation would only change the NRC value by 10% (lower) and would not change the results of the Scientech analysis at all. The justification for each team's choice is as follows:

- CP1 is the category for the "DO immediately" actions. If Wide range SG level is less than 50% the operators will perform F&B. They will not need to monitor the cues for F&B while in FRH1 because these conditions will already exist before the operators enter FRH1.
- CP3 is defined as a response following an event that gives rise to a primary cue that has to be achieved *before* some plant parameter reaches a critical value. This critical value can be regarded as a soft prompt or secondary cue. In this case, the primary cue is the reactor trip and recognition of no AFW flow. The actions for F&B then need to be performed before core damage.

H2.2.2.2 Execution Failure

For both analyses, the contribution due to execution error was negligible. The analysis for the execution portion was identical to that for HFE 1A.

**H2.3 HFE 1C**

*H2.3.1 Discrepancies in Qualitative Analysis*

The operational story forming the basis of the qualitative analysis for the three teams for this HFE is difficult to compare. The NRC analysis misinterpreted the HFE, assuming the SGTR would happen after the crew had exited FRH1. Thus, the timing and qualitative assessment of PSFs are not necessarily applicable. The documentation for NRI was not detailed enough to determine the operational story. However, two items were addressed in qualitative assessments that are comparable against the empirical data: 1) affect of masking; and 2) available time.

Amongst the three teams, there were differing assessments on the importance of the masking effect. NRI indicated that the most probable failure mechanism is *wrong* diagnosis due to the initial “masking effect”. The NRC team indicated that masking might make diagnosis difficult, but the operators are monitoring the SG levels closely and the diagnosis should happen quickly nonetheless. Similarly, Scientech did not believe that the masking effect would be of importance. The empirical data did find “masking” was a negative performance driver, *delaying* diagnosis of the SGTR; although, no team failed to diagnose it.

Both NRI and Scientech also noted that adequacy of time would be an issue. The difference between the two assessments was that NRI indicated that, due to the short time window, the crew *may* not be able to perform the required actions on time; whereas Scientech indicated that they *would* not be able to perform the require actions on time. The empirical data supports the notion that there was not adequate time; only one crew made the time criteria and they barely made it (37 minutes out of a 40 minute time frame).

Scientech was fairly clear in their description of the operational story. The major elements of this analysis – that was not reflected in the other two assessments – were:

- the structure of the procedures would not allow the operators to leave FRH1 to address the SGTR [the analysis appeared to misinterpret the phrase “active loop”, so they thought the crews would be stuck in FRH1]
- crew could not exit FRH1 without permission from management, and could not isolate the SG in parallel because it conflicts with FRH1
- the operator’s training/belief in the adequacy of the procedures would cause them to fail.

The Scientech assessment was incorrect in its assessment that the SGTR could not be addressed with the procedures as they were, it was correct in predicting that the degraded performance would manifest itself in a confusion in how to address the SGTR while in the FR procedures. The NRI analysis seemed to imply that the teams would use E-30 to address the tube rupture, presumably transferring from FRH1 to E-30 and then performing the steps in E-30 to isolate the SG and maintain RCS pressure below the set point. In reality, the crews found the SGTR challenging because they felt “stuck” in the FR procedures. In the empirical data, isolation of the SG and maintaining the RCS pressure below setpoint were done in a variety of ways – none of the operational stories accounted for the observed crew variability in scenario progression. Only two crews used the E-30 procedure at all; of the two crews that used E-30, one was directed there via the procedures (FRH1-> FRP1-> ES-11-> E-30) nearly two hours into the scenario and the other started it from knowledge.

One interesting insight from the operational story defined by Scientech was that the operational story described that the crew would get caught in a loop with step 28 of FRH1, where they could not secure the last HHSI pump because the AFW flow will not be adequate to sufficiently lower the temperature. This description is, in essence, the same thing that happened to the crew which misinterpreted the phrase “active loop”. In this case the HRA assessment team also neglected to understand the implication of the phrase “active loop”, and that there was a possible success path.

*H2.3.2 Discrepancies in Quantitative Analysis*
Based on the timing assessment performed as part of the qualitative analysis, the Scientech team set the HEP to 1.0, and cannot be compared to the other analyses any further.

H2.3.2.1 Cognitive Failure

Both the NRC and NRI assessments used CBDT to quantify cognitive failure, resulting in very disparate values: NRC's assessment of the trees resulted in a cognitive HEP of 6.0E-03 while the CBDT contribution in the NRI assessment was 8.1E-02. In this quantification there were two driving differences:

- Accurate Indications: While the NRC team acknowledged that masking might make diagnosis more difficult, for the trees (Pca and Pcd) which ask if the indications were accurate they choose the "yes" branch because operators are monitoring the SG levels closely and the indications are technically correct. The NRI team decided "conservatively" that because of the masking effect, some cues and parameter values are not accurate/as stated in procedures. In Pca (Availability of Information), this resulted in a difference of HEP between 5.0E-02 and "negligible"; for Pcd (Information Misleading) this was a difference of 1.0E-02 and "negligible." Masking was one of the negative drivers in the empirical study for its effect on delaying diagnosis in a time constrained scenario, and its effect on scenario complexity.

- Parallel Procedures: In the NRI analysis, the assessment team concluded that the SGTR would be dealt with in parallel with FRH1. The NRC analysis assumed the crew would be out of FRH1, only working through the SGTR procedure. This difference caused the NRI to determine the workload to be "high", but the NRC to select "Low" in $P_{cb}$ (Failure of Attention. This resulted in a difference of HEP between 3.0E-03 and 1.5E-02. This difference also shows up in tree $P_{ce}$ (Skip a Step in Procedure), where NRI choose the multiple procedure branch and NRC choose the single procedure branch. This resulted in a difference in HEP of 3.0E-03 versus 6.0E-03. Working items in parallel was a negative driver in the empirical data because it contributed to the scenario complexity.

Another discrepancy, although not major (i.e., does not lead to a large difference in numerical results), was $P_{cg}$ (Misinterpret Decision Logic). NRI determined that E-30 had AND or OR logic in the procedure (step 2); NRC did not indicate that this logic appears in the relevant procedure. Step 2 does indeed contain OR logic. This leads to a difference between 3.0E-04 and "negligible" HEP.

H2.3.2.2 Execution Failure

While the numerical results for execution failure cannot be compared between NRC and NRI, there was one notable discrepancy between the two analyses. The NRC quantified two execution steps: 1) Failure to identify ruptured SG based on uncontrolled level rising (step 2 of E-30); and 2) Isolate SG (step 3 of E-30). The NRI quantified two different execution steps: 1) Isolate SG (step 3 of E-30) and 2) Maintain SG pressure below set point (step 4 of E-30).

H2.3.2.3 Comparison with Empirical Data

The HEPs for this HFE were: 1.0E-02, 1.47E-01 and 1.0. The empirical data shows that only one of the four teams met the time criteria; however, three of the four teams were ultimately able to prevent opening the PORVs. The 1.0E-2 value seems optimistic when compared to the empirical data; however, this was the assessment that misinterpreted the HFE and thus missed some of the cognitive difficulties presented by being in FRH1. The other two HFEs are a more reasonable fit with the data. The 1.0 value seems to be a bit of an overestimate, however, it is worth noting that this method gives no credit for knowledge-driven activities, and the empirical crew that did succeed did so via a knowledge-driven path.

*H2.3.3 Comparison of Drivers*

Using CBDT, both the NRC and NRI assessments correctly identified Scenario Complexity, Indications of Conditions and Procedural Guidance as negative drivers. While these were all negative drivers in the empirical data, the reasons varied. Because the Scientech analysis quantified this HFE based on feasibility, not all the PSFs below were evaluated as part of their analysis. The PSFs identified here are only the ones that caused the HEP to be set to 1.0.

For NRI, scenario complexity was a negative driver because workload was high (multiple procedures) and there was a mismatch between the available cues and indications (masking). For the NRC, scenario complexity was also a driver because the plant did not respond as expected (because of the mismatch); however, in their analysis this was only a negative driver as far as it increased the stress level, making execution more error prone. For scenario complexity, the NRI rationale for identifying this as a negative driver was consistent with the empirical data, but the level of impact (main negative driver versus negative driver) was not. The NRC rationale was not consistent with the empirical data.

Again, for NRI, indications of conditions was a negative driver because the masking effect would make diagnosis more difficult. This is consistent with the empirical data. The Scientech analysis did not consider the masking effect to be an important driver; the NRC analysis only considered the masking effect in the qualitative analysis – it was not a driver in the quantitative assessment. For both NRI and NRC, indications of conditions was also a negative driver because of failure of attention due to the saliency of the cues (Pcb; monitor cue, not alarmed); this was not supported by the empirical data.

Both NRI and NRC considered procedural guidance as a negative driver because of procedure readability issues (e.g., lack of graphical distinction between procedural steps). This was not supported by the empirical data. The Scientech analysis, however, identified procedural guidance as a negative driver because of the difficulty entering E-30 given they were held up in a higher procedure. This is consistent with empirical data. However, Scientech found this to be the main negative driver, whereas the empirical data indicated that the scenario complexity was the main driver.

The NRI and Scientech assessments found adequacy of time as a negative driver. For Scientech, it was considered a substantial driver, whereas the NRI analysis classified this as a slightly negative driver. In the empirical data adequacy of time was an important driver. Three of the four teams correctly performed the actions, but only one team did it in the 40 minutes timeframe.

| Factor | Empirical | NRI | NRC | Scientech |
| --- | --- | --- | --- | --- |
| **Scenario Complexity** | MND | ND | ND | 0 |
| **Procedural Guidance** | ND | ND | ND | MND |
| **Indications of Conditions** | ND | ND | ND | 0 |
| **Adequacy of Time** | ND | ND | N/P | ND |
| **Stress** | 0 | ND | ND | 0 |
| **Execution Complexity** | N/P | ND | N/P | 0 |
| **Human- Machine Interface** | N/P | N/P | ND | 0 |
| **Training/Experience** | N/P | N/P | 0 | 0 |
| **Work Processes** | 0 | N/P | N/P | 0 |
| **Communication** | 0 | N/P | N/P | 0 |
| **Time Pressure** | --- | n/a | n/a | n/a |
| **Team Dynamics** | 0 | 0 | n/a | 0 |

## H2.4 HFE 2A

### *H2.4.1 Discrepancies in Qualitative Analysis*

As confirmed by the experimental data, all three teams determined this scenario to be a time limited, complicated evolution. While they agreed this would be a complicated scenario, the basic operational story between the three analyses differed.

It is useful to look at this HFE in two parts: 1) identify loss of CCW and sealwater and trip RCPs and 2) start the PDP. For the first portion, the analyses are fairly consistent in their operational story: the reactor trips, the crew performs the immediate actions (Steps 1-4 of E0) and then take a step back to look at the system. At that point there will be clear cues that the CCW has been lost (e.g., rising RCP seal temperature), so the operators will trip the RCP immediately. Operators are trained to monitor the RCPs and trip them if necessary. The NRI and NRC[10] analyses assume the operators will use RC-02 (step 3) to trip the pumps, but the Scientech analysis states that "Operator training develops an instinct that will cause an operator to trip the RCPs without referring to a procedure in the interest of time." The data shows that one group tripped the RCPs using RC-02, while the rest ordered the trip from knowledge when they detected the loss of CCW and sealwater. Only two of the three crews tripped the RCPs within one minute of detecting the loss; both of these crews tripped from knowledge.

Timing was determined by all three analyses, and confirmed by the empirical data, to be a key factor. The plant criteria (Step 3 in RC-02) states that affected RCPs should be stopped within one minute of loss of RCP seal injection flow and thermal barrier cooling. As predicted by both the NRC and Scientech timelines, and confirmed by the empirical data, none of the teams could make this criteria based on the diagnosis times alone. NRI did not provide timing data in their documentation.

[10] Based on the operator interview insights, NRC analysis indicates that both paths are possible, but in their summary, only discuss the use of RC-02.

| Action | Experimental | NRC | Scientech |
|---|---|---|---|
| **Diagnose Loss of CCW** | 4-6 min | 5 min | 3 min |
| **Trip RCPs** | ~1-2.5 min | 1 min* | 1 min |

*the 1 minute was the manipulation time given to "respond to Loss of CCW". It is not clear from the analysis if this is just tripping the RCPs or is intended to include both tripping the RCPs and starting the PDP.

The second part of this HFE is starting the PDP. It seems that the NRC analysis assumed the operators would use RC-02 (step 11) to stop the PDP; although the analysis does not specifically mention starting the PDP, nor was it discussed in the operator interview. The NRI and Scientech analyses assumed the operators would stop the PDP via step 6_c3 of ES-01. Both are possible procedural paths to stop the PDP, but neither were demonstrated as viable in the empirical data because time was too short.

Again, NRI did not provide specific timing information, but timing analysis was provided by the NRC and Scientech. As mentioned before, the NRC analysis determined it would take 5 minutes to diagnose the loss of CCW and then 1 minute to respond to it. [Note that the NRC analysis did not specify that the manipulation time of 1 minute included starting the PDP.] The Scientech analysis determined that it would take 6 minutes to get to the necessary procedural step to start the PDP and 2.5 minutes to start the PDP. These timing analysis were overly optimistic. Three of the four crews started RC-02, but were unable to get through the procedures fast enough to start the PDP in time. Two of these crews did not start RC-02 until 1-2 minutes after the RCP seal temperature exceeded 230 °F; the third crew did not even trip the RCPs until a minute after the seal temperatures exceeded 230 degrees. The fourth crew did not use RC-02, but stayed in ES-01. They were also unsuccessful in starting the PDP; they did not get to step 6 of ES-01 until ~2 minutes after the RCP seal temperature exceeded 230 °F . None of the teams were able to start the PDP, so there is no data on how long that step took.

In addition to the operational story, there were other elements of the qualitative analysis that were relevant; these are discussed in the section on driving factors.

*H2.4.2 Discrepancies in Quantitative Analysis*

For all three assessments, time-limited cognitive error (HCR/ORE for Scientech/NRC and THERP time-reliability correlation for NRI) dominated the quantitative analysis and are in fair agreement across the board:

| HEP | NRI | NRC | Scientech | Experimental |
|---|---|---|---|---|
| **Cognitive (HCR/ORE or TRC)** | 1.0E-01 [1.1E-01]* | 2.8E-01 | 1.1E-01*** [1.0] | -- |
| **Execution** | 1.5E-02** | 5.2E-03 | 7.8E-03 | -- |
| **Total** | **1.3E-1** | **2.8E-01** | **1.2-01** | **4/4 fail** |

*value in brackets is the total cognitive error value from ASEP, which includes "identification"; this portion does not correspond to the Calculator method.
**Quantified by ASEP, does not correspond to the Calculator method
***Based on a 9 minute time frame; value in parenthesis is the 7 minute value.

#### H2.4.2.1 Cognitive Failure

Overall, the cognitive failure HEPs for the three analyses were very similar. The biggest discrepancy between the analyses was the difference in timing. This is discussed in detail in the qualitative analysis section above. For HCR/ORE, the important part of timing is the ratio of the time available after the cue (T_sw – T_delay – T_m) to the diagnosis time (T_1/2).

Another difference in the analyses was that Scientech divided their HEP into two separate HFEs – failure to trip RCPs and failure to start PDP – applying both a 7 minute and 9 minute time frame to each. The 7 minute time frame resulted in an HEP of 1.0, whereas the 9 minute time frame resulted in a 1.1E-01. The NRC analysis quantified this HEP as one cognitive step, using an 8 minute time frame, resulting in a value of 2.8E-01 [Note: using a 9 minutes time frame yields a result of 2.1E-01]. In reality, the time frame available (between loss of CCW and RCP seal temperature >230 °F) ranged from 7 to 8.5 minutes. 8.5 minutes was the minimum time required by the Scientech analysis; six minutes was the minimum time required for NRC analysis. In this case the empirical data showed 4/4 failures, which is consistent with the generally high HEP values. Given the actual timing and the fact that no team had sufficient time to complete the actions, the empirical data is most consistent with the Scientech assessment.

As mentioned previously, Scientech divided their HEP into two separate HFEs – failure to trip RCPs and failure to start PDP. The relative contributions of the two cognitive failure modes (for the 9 minutes time frame) 4.2E-03 and 1.1E-01. This is also consistent with the empirical data considering that all teams tripped the RCPs (although one team did so after the seal temperature limit was exceeded), but none of the teams were able to start the PDP.

#### H2.4.2.2 Execution Failure

While the execution failure was a negligible contributor to both all three analyses, there was one difference in the quantification that is worth pointing out. Like the cognitive portion, the two analyses divided the HFE differently:

| NRC | Scientech | NRI |
|---|---|---|
| Execution: Trip RCPs | | Execution: Trip RCPs |
| Execution: Start PDP | Execution: Start PDP | Execution: Start PDP |
| *No recovery credited* | Execution: Close PDP recirculation valve | *No recovery credited* |
| | Recovery: Monitor sealwater temperatures | |

### *H2.4.3 Comparison of Drivers*

As seen in the discussion on the quantitative analysis, the cognitive contribution significantly dominated this HEP for all three analyses. NRC and Scientech used HCR/ORE to quantify the cognitive contribution. In this method, the only factors considered are 1) timing; and 2) cue type (CP designation). Because of this method of quantification, driving factors other than timing cannot be parsed out easily – they are either nominal or they are accounted for by adjusting the timeline. While both analyses discussed important factors in their qualitative analysis, neither analysis specifically discussed if/how the timelines were adjusted for these factors. This is why, in the table below, the main negative driver for both the NRC and Scientech analyses are "Adequacy of Time" (a driver which is supported by the empirical data). Likewise, the dominating factor for NRI was the cognitive portion which was also quantified using a time-reliability correlation. Therefore, “Adequacy of Time” was a major driver for their numerical results, too. Other items identified as negative drivers for the three teams in the table below are very minor contributors as they mainly influence the execution value.

From the qualitative analysis perspective there was a big difference in how the teams did in matching the drivers:

Both NRC and Scientech indicated that operators would be trained to monitor and trip the RCPs. Scientech elaborated that the cues and indications clearly direct operators into the correct procedures and the crew is not expected to mis-diagnose the scenario. NRI, however, indicated that the most probable failure mechanism would to be *late* diagnosis. The empirical data found that, contrary to Scientech and NRC’s qualitative assessment, monitoring of control boards and acknowledging alarms was less than adequate [Work Process]. Late diagnosis was supported by the empirical data.

Scientech also noted that complexity of task itself would *not* be difficult, but, time was short and the workload would be high due to lost distribution panel and use of multiple procedures. NRI said that there would be difficulty due to the dynamic scenario with unusual combination of events (with respect to what they’ve been trained on) and short time window. The empirical data showed that many thinks happening at the same time made it difficult to detect the priority items, the crew was not accustomed to this unusual progression and, though the procedural guidance was there, crews had problems reaching the critical items in time [Scenario Complexity, Training, Procedural Guidance].

| Factor | Empirical | NRI | NRC | Scientech |
|---|---|---|---|---|
| **Scenario Complexity** | MND | ND | 0 | 0 |
| **Training/Experience** | MND | ND | N/P | N/P |
| **Adequacy of Time** | ND | ND | MND | MND |
| **Work Processes** | ND | N/P | N/P | N/P |
| **Procedural Guidance** | ND | ND | 0 | 0 |
| **Indications of Conditions** | ND | N/P | 0 | 0 |
| **Stress** | 0 | ND | 0 | ND |
| **Execution Complexity** | 0 | ND | N/P | N/P |
| **Other** | n/a | N/P | ND | ND |
| **Human- Machine Interface** | N/P | N/P | N/P | N/P |
| **Team Dynamics** | 0 | 0 | N/P | N/P |
| **Communication** | 0 | N/P | 0 | 0 |
| **Time Pressure** | n/a | n/a | n/a | n/a |

### H2.5 HFE 3A

#### *H2.5.1 Discrepancies in Qualitative Analysis*

This was a fairly easy scenario for the operators, and all three analyses reflected this. All three teams recognized that the operators would have little trouble performing the necessary actions within the long time frame because operators are well trained on SGTRs (design basis scenario). The cognitive complexity was assessed to be low, and no other difficulties expected (PSFs optimal). This matches the empirical data which saw no failures and no negative drivers in the operators' performance.

*H2.5.2 Discrepancies in Quantitative Analysis*
For all three assessments, the final HEP was found to be fairly low, which is consistent with the empirical data which saw no failures and no negative drivers in the operators' performance. The results of the quantitative analyses are summarized below:

| HEP | NRI | NRC | Scientech | Experimental |
|---|---|---|---|---|
| **Cognitive (CBDT)** | 3.5E-03 | 3.2E-03 | 2.7E-05** | -- |
| **Execution** | 6.0E-03* | 2.6E-03 | 1.8E-03 | -- |
| **Total** | **9.5E-3** | **5.8E-03** | **1.9-03** | **0/0 fail** |

*Quantified by ASEP, does not correspond to the Calculator method
**Assessed value, cognitive contribution capped at 1E-04

H2.5.2.1 Cognitive Failure

For all three analyses, CBDT dominated the cognitive contribution. The results from the three assessments overall were very similar. The Scientech analysis saw a significant numerical discrepancy between their analysis and the NRI and NRC analyses. This numerical discrepancy came primarily from two sources: 1) they dissected the HFE into three separate cognitive failure mechanisms, each quantified separately; and 2) they credited cognitive recovery. Scientech divided the HFE into the following three failure mechanisms: A) Failure to diagnose SGTR, B) Failure to isolate SGs, and C) Failure to maintain RCS pressure below the setpoint. Each of these three mechanisms were quantified using CBDT, and the only non-negligible mechanism found in each assessment was "Skip a Step in Procedure", which had a value of 3.0E-3 ($P_{cd}$[e]). This is the same mechanism that dominated the NRC and NRI CBDT analyses. However, since Scientech had three instances, the non-recovery contribution would have been 9.0E-3; triple that of the NRC/NRI analyses. But, unlike NRC and NRI, Scientech applied a cognitive recovery factor to each of these failure mechanisms, taking credit for self review over the long time frame. This reduced their cognitive HEP from 9.0E-3 to 2.7E-5 (then capped at 1.0E-4). So, while the driving failure mechanism for cognitive error is the same for the three analyses, there is a substantial difference in the end HEP. The empirical data did not see this failure mechanism, but that was expected by the teams given the low HEP.

One other discrepancy, though quite minor (3.0E-4 vs. "negligible"), between the NRI analysis and the NRC and Scientech analyses was the assessment of the "Misinterpret Decision Logic" tree ($P_{cg}$). The NRI analysis assessed that there was AND or OR logic in the procedure, citing step 2 of E-30. There in fact is an OR statement in step 2 of E-30. The NRC and Scientech assessment of that tree either did not recognize that logic or decided it was not important.

H2.5.2.2 Execution Failure

While the execution contribution amongst the analyses – particularly between the NRC and Scientech analysis which are both using THERP – are fairly consistent, there significant discrepancies in how that portion of the HFE was divided (see table below). The other major discrepancies between the NRC and Scientech analyses were that, Scientech credited recovery steps, and, for nearly each execution step, they also modelled both an error of omission (table

20-7) and an error of commission (table 20-11/12), whereas the NRC analysis only modelled the error of omission.

*H2.5.3 Comparison of Drivers*

The empirical data saw no negative drivers. As evidenced by the qualitative assessments and the low resultant HEPs, the negative drivers found by the NRI, NRC and Scientech analyses were not strong drivers. In fact, the designation “ND” and “MND” were only based on the relative numerical contribution; the overall numerical contribution was quite low. Scientech, however, noted that “Execution Complexity” would be a driver just based on the large number of steps that had to be performed. For “Indications and Conditions”, the NRC team considered it a “negative driver” because in the CBDT it is a “monitor” not a “check” parameter – this leads to a numerical contribution of 1.5E-04. Likewise, “Procedural Guidance” was a driver because a major failure mechanism was skipping a step in the procedure (cognitive and execution) because of the checklist nature and the fact that the steps were not graphically distinct.

For this HFE, the positive drivers are just as important to examine. The rationale behind positive values for “Training and Experience”, “Indications and Conditions” and “Procedural Guidance” for the empirical data was that this was a frequently trained scenario, the cues were clear for the scenario and the procedures supported the base case well. The qualitative analyses for the three teams support these conclusions.

| NRC | Scientech | NRI |
|---|---|---|
| Execution: Identify Ruptured SG | | |
| Execution: Isolate ruptured SG | | Execution: Isolate ruptured SG |
| *No recovery credited* | | Execution: Maintain RCS pressure below setpoint |
| | Execution: Adjust ruptured SGs PORV controller setpoint to between 1260 PSIG and 1265 PSIG | *No recovery credited [team noted that this was a conservatism in the analysis]* |
| | Execution: Check SG 1D - RUPTURED | |
| | Execution: Close ruptured SGs MSIVs and MSIBs | |
| | Execution: Stop AFW flow | |
| | Execution: Determine required core exit temperature | |
| | Execution: Block low steamline pressure SI | |
| | Execution: Place steam dump ‘INTLK SEL’ switches to BYPASS INTERLOCK | |
| | Execution: Dump steam to condenser from intact SG(s) at maximum rate | |
| | Execution: Stop RCS cooldown | |
| | Execution: Place group ‘C’ pressurizer heater control switch to PULL TO LOCK | |
| | Execution: Place all other pressurizer heater group control switches to OFF | |
| | Execution: Initiate maximum pressuizer spray | |
| | Execution: Normal spray valves CLOSED | |
| | Execution: Auxiliary spray valves CLOSED | |
| | Recovery: NR level – GREATER THAN 14% in ruptured SG | |
| | Recovery: Core exit T/Cs LESS THAN required core exit temperature | |
| | Recovery: Maintain core exit T/Cs LESS THAN required temperature | |
| | Recovery: Continue to depressurize until cooldown conditions are met | |

The one positive driver in the empirical data that the qualitative analyses did not explicitly describe as a positive factor was “Team Dynamics”.

| Factor | Empirical | NRI | NRC | Scientech |
|---|---|---|---|---|
| **Procedural Guidance** | N/P | ND | MND | ND |
| **Execution Complexity** | 0 | 0 | N/P | MND |
| **Scenario Complexity** | 0 | ND | N/A | N/P |
| **Indications of Conditions** | N/P | N/P | ND | 0 |
| **Training/Experience** | N/P | N/P | N/P | N/P |
| **Work Processes** | N/P | N/P | N/P | N/P |
| **Human- Machine Interface** | N/P | N/P | N/A | 0 |
| **Communication** | 0 | N/P | N/P | N/P |
| **Adequacy of Time** | 0 | N/P | N/P | N/P |
| **Stress** | 0 | 0 | N/P | N/P |
| **Other** | N/A | N/P | N/A | N/A |
| **Team Dynamics** | 0 | 0 | N/A | N/A |

## H3 SPAR-H

The figure below provides a graph of the numerical results from each team plotted alongside the results of the empirical study.

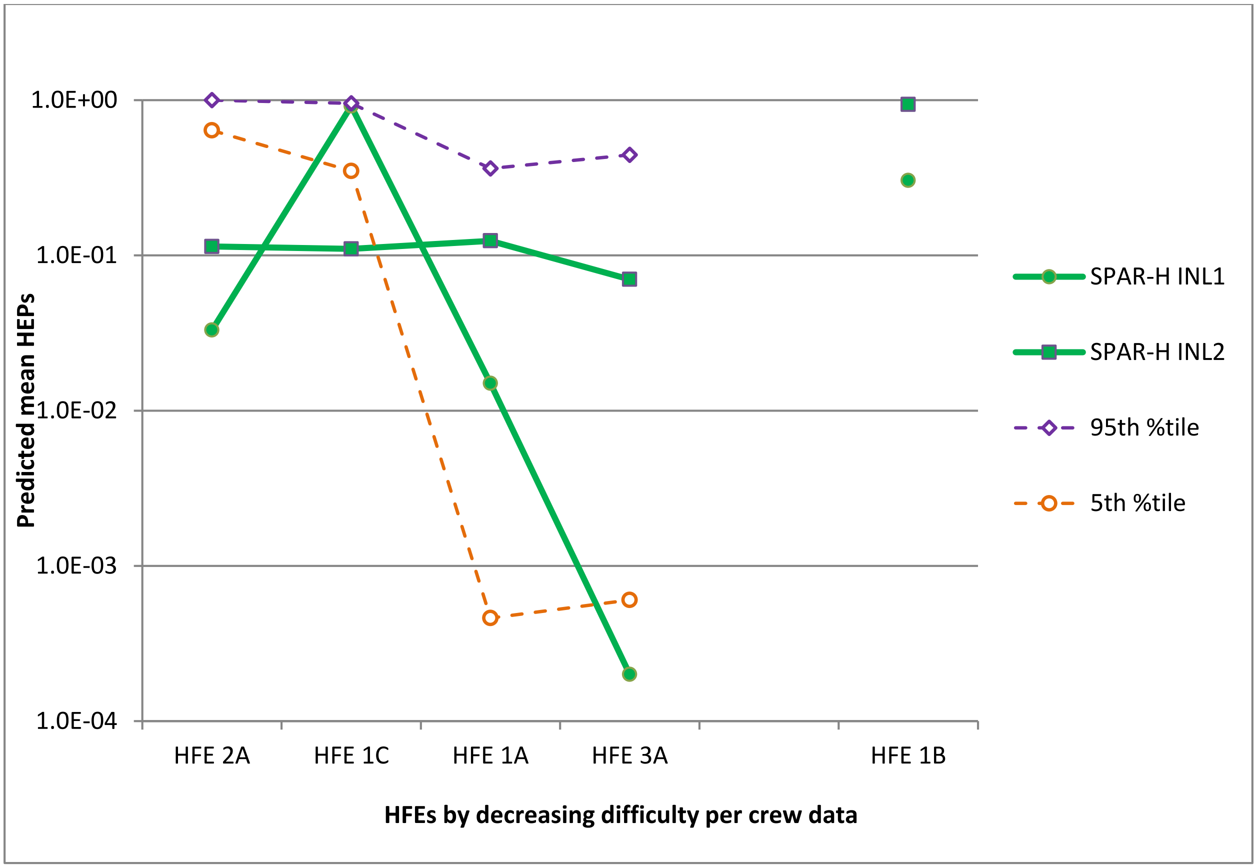



Figure 0.3 Predicted Mean HEPs with Uncertainty Bounds by HRA Teams with SPAR-H

### H3.1 HFE 1A

*H3.1.1 Qualitative Analysis*

In the INL 1 team analysis, Scenario 1A was modeled as combination diagnosis/action event but not decomposed further. In contrast, the INL 2 analysis decomposed the scenario into 6 basic events. Basic events were individually determined to be either diagnosis, action or combination events. The INL 1 analysis of Scenario 1A did not discuss transitions in procedures, although they discussed some details of the developing scenarios. In contrast, transitions in the procedures were a significant point in the INL 2 analysis and served to distinguish basic events. Regardless of the decomposition, HFE1A(1), failure to recognize that AFW 12 pump is not flowing to the SG was the dominant event in the INL 2 analysis due to the effect of the missing condition indicator. In addition, the symptom based procedures used attributed a positive effect on the HEP in 2 basic events.

Also, the main negative driver for both analyses was the misleading condition indicator. However, the INL 1 team analysis accounted for it under Complexity (a factor of 2), while INL 2 team accounted for it under misleading indicator in Ergonomics (a factor of 50). The INL 1 team analysis also rated Ergonomics poor (a factor of 10) due to lacking CR status indication of a potentially important recirculation valve. They accounted for the misleading AFW 12 pump indicator under Complexity, and specifically stated that this was not assigned under Ergonomics in order not to double-count effects.

In the INL 2 team SPAR-H analysis the main negative driver as the PSFs indication of conditions and a secondary negative driver was scenario complexity. The empirical data also identified these as negative drivers in this scenario. Both SPAR-H analyses indicated that procedures would have a nominal or positive effect, while the empirical data indicated that Procedural guidance was a negative driver because the Critical Safety Function Status Trees do not address the misaligned valves. However, the INL 2 team SPAR-H analysts indicate that the recognition of the loss of feedwater is a knowledge based decision (implying that there is no procedural indication); the effect of the lack of procedural guidance is included in the complexity PSF.

In the empirical data, the critical negative PSFs were Scenario complexity, Indication of conditions, and Procedural guidance. Both SPAR-H analyses indicated the effects of Scenario complexity and Indication of conditions quantitatively. However, while procedural guidance was called out in the qualitative analysis by the INL 2 team SPAR-H analysis, it was not captured as a driver in the quantitative analysis. The INL 1 team SPAR-H analysis did not address procedures. Both SPAR-H analyses indicated that Adequacy of Time was a positive factor as did the empirical data.

The INL 2 team SPAR-H analysis did not apply quantitative effects from work procedures (which could be combined under the SPAR-H PSF ‘work processes’), but the analysts state in their analysis that the cues for AFW are reviewed by multiple team members, so it is likely that the misleading indication will be questioned. They note that this type of team process is not typically included in a SPAR-H analysis, and therefore it is not reflected in the quantitative analysis or PSFs. The INL 1 team analysis attributed a positive effect for work processes because they noted good communication among the crews.

*H3.1.2 Quantitative Analysis*

The HEP estimates provided by the two SPAR-H analyses for scenario 1A were not easily comparable, and differed by a factor of 10. The INL 2 team SPAR-H PSF assumed a factor of 50 on Ergonomics (due to the missing indicator), while INL 1 team attributed a factor of 10, thus INL 2 team HEP was much higher. Other differences in the HEP estimates between the two analyses seem to be due to the decomposition of PSFs to basic events. Finally, the INL 1 team analysis included a negative effect of stress, while INL 2 team did not, and the INL 2 team analysis did not include the effect of PSFs on the Action portion of the HEP, possibly due to the decomposition of the scenario into basic events. The table below shows the difference in PSF assignments. Values indicate that they were applied to the action portion of the PSF. Values separated by commas in the INL 2 team analysis indicate that the PSF that value on a different basic event.

| PSF Type | INL 1 team HFE1a (diag/act) | INL 2 team HFE 1A |
|---|---|---|
| **Time** | .1/1 | .1/.01 |
| **Stress** | 2/2 | 0 |
| **Complexity** | 2/2 | 2, 2 |
| **Experience** | .5/.5 | .5 |
| **Procedures** | 1/1 | .5, .5 |
| **Ergonomics** | 10/10 | 50 |
| **Fitness** | 1/1 | 0 |
| **Work Processes** | .8/.5 | 0 |

### H3.2 HFE 1B

The two SPAR-H analyses based their initial analyses of scenario 1B on the analyses above. The differences noted about remained, but included further differences noted here. The INL 2 team SPAR-H analysis of scenario 1B assumed that if the crew allowed the reactor to trip automatically, it was indicative of extremely poor work processes. Their analysis identified that a significant amount of the available time would be taken up before the automatic trip occurred. The analysts further assumed because recognition of total loss of feedwater was a knowledge based task, which combined with the missing condition indicator for the flow from the AFW pump would mean that there would be insufficient time to recognize that water from pump 12 was not reaching the steam generators, and thus a failure probability of 1.0. The INL 1 team SPAR-H analysis also indicated that the automatic trip of the reactor would reduce available time; INL 1 team assumed that automatic trip would yield nominal time to recognize the failure rather than insufficient time. This accounts for the difference between the HEPs estimated.

### H3.3 HFE 1C

*H3.3.1 Qualitative Differences*

The INL 2 team SPAR-H analysis assumed that the crew could get stuck in a 'do-loop' and in FR-H1 steps 28 and 29, due to the need to obtain an adequate subcooling margin to move on. For the assessment of available time, the analysts assumed that the crew would cycle through steps 28 and 29 three times before meeting the RNO criteria of step 29(a), which directs them to transfer to E-10.

In contrast, the INL 1 team SPAR-H analysis did not focus on the procedures, but on the strategy that operators might use to control the plant. They state that…" *not taking care of isolating the SG right away may represent a strategy that is not in error. Thus, the HFE is expected to be higher and is represented as such in the SPAR-H analysis. Alternately, lack of SGTR response may be due to lowered awareness; the analysis team identified that focus upon task completion for F&B could result in diminished situation awareness regarding additional operation concerns such as SGTR isolation."*

The INL 1 team assumed that the crew's use of procedures would be positive, but did indicate that procedures could have the crew ignore the SGTR for some time. In addition, the INL 1 team SPAR-H analysis indicated that the range for the PSF complexity should be higher than allowed by SPAR-H. They did not assess the misleading indicator as a missing indicator (which would have given a factor of 50) because there are alternative indicators. Given that the HEP was already at .9, they did not feel it was necessary to factor in a large multiplier to attract attention to the problem. This would indicate a difference in the usage of the SPAR-H method between the two teams.

*H3.3.2 Quantitative Differences*

The INL 1 team SPAR-H analysis identified Stress, Complexity and Indication of Conditions as the main negative drivers; all of which were factored at their highest weights. The INL 1 team SPAR-H analysis did not identify Available time as a driver. In comparison, the INL 2 team SPAR-H analysis identified Available Time, Time Pressure as the main negative drivers and Execution Complexity (for the F&B) as a negative driver.

The primary quantitative differences between the two analyses derived from the factors being maximized in the INL 1 team SPAR-H analysis and different drivers being selected but not maximized in the INL 2 team SPAR-H analysis. The INL 2 team SPAR-H analysis included the following drivers on performance in the quantification: Time Pressure/Available time (as identified in the empirical data), Stress, Scenario Complexity and Execution Complexity and Procedural guidance. The INL 2 team SPAR-H qualitative analysis also indicated that unfamiliarity (which would be factored in terms of experience or training), Team factors, and Work Processes could affect crew performance. However, these factors were not included in the quantification, perhaps because the analysts, while they were aware they could affect performance, could not be certain that they would be present without further crew information.

*H3.3.3 Comparison to Empirical data*

In the empirical data, the main negative driver was considered to be the scenario complexity, since the tube rupture was initially masked by the AFW flow to the ruptured SG. It was also noted that when the tube rupture occurred, the crews already had an emergency situation and were working in FR-H1.

These issues are well described by the INL 1 SPAR-H team when explaining the complexity PSF ("preoccupation with the F&B", etc), even though they don't go into details on which procedures that were in use and exactly how the flow was masking the tube rupture. Indication of conditions was thus also noted in the empirical data as a negative driver, as also identified by the SPAR-H analysis. Procedural guidance was noted as a negative driver in the data, since it was not easy to manage the SG isolation while they were "stuck" in the procedure for F&B (FR-H1 (and FR-P1)). The INL 1 SPAR-H team notes this goal conflict several places, even in a comment on procedures: *"procedure following may have the crew ignore isolating the SG for quite a while".* However, they did not change the weight for the procedures PSF, but kept it nominal. The last negative driver in the data was the adequacy of time due to the 40 minutes period set up as a goal in the HFE. Time was considered nominal in the INL 1 team SPAR-H analysis.

The INL 2 team SPAR-H analysis also identified PSFs in their qualitative analysis that were not included in the quantitative analysis. In the INL 2 team SPAR-H analysis, the analysts were accurate that performance would be affected by the misleading condition indicator. Crews did not perceive the rising SG until AFW was stopped. The difficulty in transitioning from FR-H1 to E-30 was also identified in the qualitative analysis, as was the utility of the Conditional Indications Page (FoldOut page). Crews also indicated that they felt run down by the long complicated scenario, the prediction of this was the basis for the Stress PSF.

### H3.4 HFE 2A

*H3.4.1 Qualitative Analysis*

The INL 1 team SPAR-H analysis predicted the HFE 2A to be a moderately complex scenario. Two negative drivers were identified: Complexity and stress.

In the empirical data, it was clear that all the crews were late in detecting the loss of CCW and sealwater, causing them to trip the RCPs late and also gave them too little time to start the PDP before the RCP sealwater temperature reached 230 degrees, then the crew simply could not move through the procedures fast enough. The main reason for them being late was that they lacked training on the specific type of scenario, the situation was very complex and they had a problem recognizing the urgency. The INL 1 SPAR-H team does not question the combination of these factors and does not analyze in detail how some of these issues impact performance. The team stated that they relied on interviews when assessing the HFE. There might be two reasons for the mismatch based on this: That the interviews paied too little attention to the details, or that the self-assessment of the interviewees was not sufficient.

The INL 2 team SPAR-H analysis identified one success path with one critical subtask, and also identified one recovery path with two critical subtasks. They also identified time and complexity as the main negative drivers.

Regarding the CRT modeling, they identified other failure paths as well, but one failure path is not analyzed since there is no chance of success. This seems to be a strange reason. Should not difficult or impossible failure paths be included in the analysis? Another question relates to the inclusion of the recovery paths in the modeling of this HFE. The INL 2 team SPAR-H analysis accurately indicates that identifying loss of CCW in the midst of all the other alarms is very difficult for the crew and uses up significant time. The analysts indicate that there is a chance for recovery from failing to detect loss of CCW if the crew works rapidly once the RCP stator alarm occurs. However, it is questionable whether they should give credit to recovery paths like this when the time frame is only 7 minutes.

Both teams consider the procedures PSF to be nominal. SPAR-H as such normally doesn’t consider the fit between the procedures and the situation in detail, rather it only asks for whether the general quality of the procedures are good, typically whether procedures are symptom-oriented or diagnostic or not.

*H3.4.2 Quantitative Analysis*

The INL 1 team SPAR-H analysis did not manage to assess the combined effect of the drivers, resulting in an optimistic HEP for this HFE that was the most difficult one with four of four crews failing. Especially, the lack of time for the crews was not identified. The crews were late in diagnosing the event and the time required to move through the procedures was not adequate. The SPAR-H team estimated 5 minutes to complete, while 7-9 minutes were available. A 40% margin is nominal in SPAR-H. A question on the method is whether this margin should still be used in such short time intervals of a few minutes.

The INL 2 team SPAR-H analysis modeled recovery in this short time interval HFE. This leads to dependent sub-events, multiplying the first basic event with the HEPs of the recovery events. The first sub-event had a HEP of 0.51, while the final HEP was 0.14, an optimistic HEP given the empirical data of four of four crews failing. This could have been higher if other failure paths had been included or recovery had been judged differently in the CRT modeling.

### H3.5 Scenario 3A

*H3.5.1 Qualitative Analysis*

This scenario was designed to be a standard ("vanilla") PSA scenario, and it turned out to be as well. It is often trained and everything ran as expected following well prepared procedures and good indications. The INL 1 team SPAR-H analysis predicted this, and it seems that for standard scenarios the need to go into details in scenario developments in order to find out the difficulties for the crews is less prominent than with more difficult scenarios.

The INL 2 SPAR-H team conducted a good qualitative analysis of this HFE, breaking down the HFE into sub-events. The breakdown is based on the analysts' ability to perform the qualitative analysis and the technique that they use to define the basic events. In this case, the use of CRTs to define basic events lead to very detailed analysis of the scenarios, and examination of the role of the procedures in the events. There was again one disturbing comment about a failure path that was not included in the quantification since there was zero opportunity for success.

*H3.5.2 Quantitative Analysis*

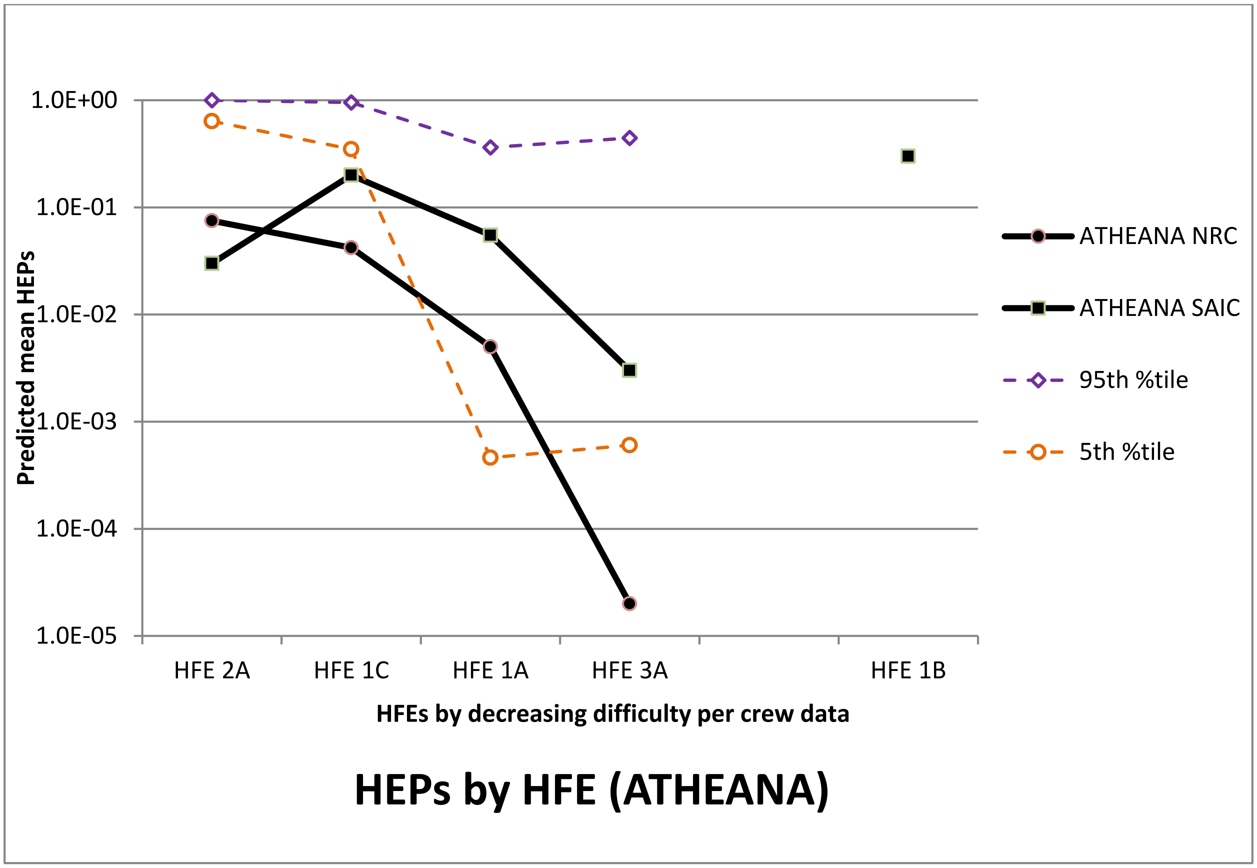
The INL 1 team SPAR-H analysis predicted correctly this HFE to be the easiest one.

The INL 2 team SPAR-H analysis disturbingly enough did not manage to discriminate the value of the HEP of this HFE sufficiently from the other HFEs, in spite of a good qualitative analysis. In the empirical data this was shown to be the easiest HFE, a well-trained one without any complications or failures. It seems that adding up the HEPs for the basic events in this case lead to a rather high HEP for the HFE 3A. The analysts did not give credit for time, training &procedures, or experience which would further reduce the HEP.

## H4 ATHEANA Intra-Method Comparison

Two assessment teams – SAIC and NRC – utilized the ATHEANA HRA method to predict the results in this empirical study. Figure 1 provides a graph of the numerical results from each team plotted alongside the results of the empirical study.

Figure 1: HEP Predictions using ATHEANA vs. Empirical Data




### H4.1 HFE 1A

Overall the SAIC and NRC assessments for HFE 1A characterized the scenario in a manner generally consistent with the responses observed in the simulator exercise. Time was more than adequate, and no main negative drivers were cited, which is consistent with a scenario in which all crews completed the required response well within the time available. While the depth and approach of the qualitative assessment differed significantly between the SAIC and NRC assessment teams, both assessments were fairly consistent in the key points of the qualitative analysis. The table below provides a breakdown of the quantitative results:

| HEP | SAIC | NRC | Experimental |
|---|---|---|---|
| **Total** | **5.5E-2** | **5.0E-03** | **0/4 fail** |

Both analyses came up with two or more failure modes in the initial qualitative analysis, but ended up quantifying only one scenario. For SAIC, this scenario was the failure to initiate F&B cooling, conditional on failing to recognize and correct the diversion of AFW flow through the recirculation path. The mean value was assessed at 5.5E-02, with an upper bound of 1.0E-1.

The upper bound was based on consensus opinion that operator could be stuck in another procedure causing late time pressure. The NRC analysts assessed all explainable departures as having a negligible contribution to the HFE, so the scenario they quantified was failure for “unexplainable” reasons. They assessed this HEP, based on their team’s experience with unanticipated failures, to have a mean value of 5.0E-03. While both teams are within the bounds of what can be deduced from the experimental data, the NRC value is nearly an order of magnitude lower than that assessed by SAIC.

The primary driver behind this discrepancy seems to be the weight the respective teams placed on the effect of the missing indication on the HFE. Both identified that the false indication for the AFW flow existed, but also acknowledged that there were sufficient cues and training to allow the operators to succeed. The NRC team noted that the operators were trained specifically on the open recirculation valve based on an event that had previously occurred at the plant. The SAIC team indicated“operators are probably not particularly well trained on the recirculation valve.” However, they counted training as a positive factor because they expected the operators to be well trained on lack of heat sink. This training, in conjunction with the indications for loss of heat sink, would allow the operators to be successful. The empirical data suggested that the specific training on open recirculation valve was an important positive driver.

Both teams also identified time as more than adequate; their timing estimates were fairly consistent with each other and with the empirical data: the SAIC team provided a point estimate of 16.5 minutes to start F&B, NRC provided a distribution with a mean value of 12 minutes (with a min and max time of roughly 2 and 40 minutes, respectively) and the empirical data showed a range of times roughly from 10 to 17.5 minutes.

In addition to the above factors, the NRC team noted that if there were confusion or delay, the primary cause would be the rapid fall in SG level. There was no evidence for the qualitative analysis prediction that the unusually rapid SG level decrease could lead to confusion or delay. While this prediction was not supported, it was not flagged as likely to lead to failure of the HFE and the unsupported prediction should be assessed in light of the overall qualitative analysis result

The SIAC team noted that, during the simulator exercise observed by the analysts, crew communications were not always audible. This was considered a negative driver in their analysis. Below is a table comparing the drivers between the analyses and the empirical data:

| Factor | Empirical | SAIC | NRC |
|---|---|---|---|
| **Procedural Guidance** | ND | ND | 0 |
| **Scenario Complexity** | ND | 0 | ND |
| **Indications of Conditions** | ND | ND | N/P |
| **Adequacy of Time** | N/P | N/P | N/P |
| **Execution Complexity** | N/P | N/P | N/P |
| **Training/Experience** | N/P | N/P | N/P |
| **Human- Machine Interface** | N/P | 0 | 0 |
| **Time Pressure** | ---- | N/P | N/P |
| **Stress** | 0 | 0 | N/P |
| **Team Dynamics** | N/A | N/P | 0 |
| **Communication** | N/A | ND | 0 |
| **Work Processes** | N/A | 0 | 0 |

While the SAIC team judged the response (i.e., the initiation of F&B cooling) may have been quicker if there had not been a misleading indication of AFW status, the analysts judged scenario complexity to be a non-factor. Although they acknowledged the misleading flow indication, they noted that the criteria for establishing F&B cooling were tied to SG levels, and that there was limited complexity associated with this set of parameters. Similarly, the NRC team acknowledged that the procedure did not explicitly address the misaligned valve, there are sufficient indications to diagnose the loss of heat sink, and they determined the procedures to be sufficient to lead the operators to success.

**H4.2 HFE 1B**

Empirical data was not available for this scenario.

**H4.3 HFE 1C**

The operational story forming the basis of the qualitative analysis for the two teams for this HFE is difficult to compare as there are multiple procedural and non-procedural success paths available. This variety was also reflected in the empirical data.

The SAIC analysis lays out a timeline with an estimated 263 minutes from tube rupture to end of the scenario; this seems to be at odds with the success criteria of 40 minutes outlined in the HFE definition. Procedurally, at step 33 of FR-H1 the operators will transfer to ES11, which will eventually lead to success. Alternatively, if the SGTR has been diagnosed, the operators can take actions in parallel with FR-H1 as long as those actions do not interfere with FR-H1. The operator may choose to look at procedure E-30 (steam generator tube rupture) to perform these actions. The main difficulty in this scenario, according to the SAIC analysis, is diagnosis of the SGTR, which is complicated by lack of clear indication of the SGTR (including lack of initial radiation alarms and masking), distraction (by focus on FR-H1) and crew utilization (bulk of the diagnosis burden is left to the secondary side operator).

The NRC analysis focuses on functional success via step 13 of FR-P1 (FR-P1 reached via orange path on CSF Tree once FR-H1 is exited); timely diagnosis of the SGTR is not necessary for success. If the operators make it into FR-P1, they will be successful within the given time frame; the team assessed that 93% of the time, operators will make it to FR-P1. The failure mechanisms all revolved around failure to enter FR-P1. In these cases, the operators must diagnose the SGTR and address it by going to E0 and eventually getting to E-30 or deviating from the EOPs and going directly to E-30. In these cases it is a race against time, with a small chance for diagnosis and the actions to be completed in the given time frame. The NRC team did not consider actions taken in parallel with FR-H1.

The SAIC analysis produced a mean value of 2.0E-01, while the NRC analysis found a mean of 4.2E-02. The experimental results found that 3 of the 4 teams failed to perform the necessary actions within the 40 minute time frame. The resultant HEPs from both analyses fall below the HEP range supported by the empirical results. While the operational story of the NRC generally fit the empirical data, the resulting HEP was significantly lower than is supported by the data. This is noteworthy because their analysis considered 100% success (based on time estimates) of meeting the criteria if the operators correctly entered FR-P1; 2 of the 3 failures seen in the empirical data are from teams which did get to FR-P1, but were still unable to meet the time criteria.

The table below provides a summary of the drivers for the analyses as compared to the empirical results:

| Factor | Empirical | SAIC | NRC |
|---|---|---|---|
| Scenario Complexity | MND | ND | ND |
| Procedural Guidance | ND | ND | ND |
| Indications of Conditions | ND | ND | ND |
| Adequacy of Time | ND | N/P | ND |
| Stress | 0 | ND | ND |
| Training/Experience | N/P | ND | 0 |
| Execution Complexity | N/P | N/P | N/P |
| Human- Machine Interface | N/P | 0 | 0 |
| Communication | 0 | ND | 0 |
| Time Pressure | --- | N/P | 0 |
| Team Dynamics | 0 | N/P | 0 |
| Work Processes | 0 | 0 | 0 |

### H4.4 HFE 2A

Both assessments did a moderate to good job in matching the qualitative analysis to the empirical data in determining the area of difficulties, however, both teams significantly underestimated the effect of those difficulties and thus, the resultant HEP. The table below provides a breakdown of the quantitative results:

| HEP | SAIC | NRC | Experimental |
|---|---|---|---|
| **Total** | **7.5E-2** | **3.0E-02** | **4/4 fail** |

There are generally two success paths for this HFE: one is through step 6 of ES01 and the other is to diagnose the Loss of CCW and address the RCP and start the PDP in parallel with ES01 via either skill-of-craft actions or the Loss of CCW AOP 0POP04-RC-002. The NRC analyzed both paths and determined they would be performed in parallel, basing their HEP estimates on the quickest success path. The SAIC analysis seemed to focus on the success path via ES01, but did mention that operators were trained to address Loss of CCW early. The empirical data showed that the operators did indeed perform ES01 in parallel with addressing the RCP based on skill-of-craft knowledge or AOP RC002 upon recognition of the Loss of CCW. The empirical data also showed that the operators did not start the PDP as part of the seal protection actions performed in parallel with ES01.

For both analyses, timing was determined to be a driving factor. For the procedural path via ES01, the NRC analysis estimated a mean time of 6.2 minutes, with $P(T > 7 \text{ minutes}) = 0.3$ and $P(T > 9 \text{ minutes}) = 0.11$. The SAIC analysis gave a point estimate of roughly 8 minutes to complete the necessary actions for success. Based on the empirical data both teams underestimated the time necessary to complete the necessary actions via ES01.

For completing the necessary actions outside ES01, the NRC team estimated a mean time 3.9 minutes, with $P(T > 7 \text{ minutes}) = 0.1$ and $P(T > 9 \text{ minutes}) = 0.05$. This was based on the understanding that operators were well trained and starting the PDP was closely coupled with stopping the RCP. The SAIC analysis, however, noted that, while there was good training on stopping the RCP, there seemed to be some confusion on the need to start the PDP. This is supported by the empirical data which saw teams stop the RCP outside of ES01, but not addressing the PDP.

Overall, both teams thought this to be a straight forward scenario where timing was an issue and procedures were not completely clear. The empirical data showed that the scenario was more complex than the analysts predicted. The table below provides a summary of the drivers for the analyses as compared to the empirical results:

| Factor | Empirical | SAIC | NRC |
|---|---|---|---|
| **Scenario Complexity** | MND | 0 | ND |
| **Training/Experience** | MND | 0 | 0 |
| **Adequacy of Time** | ND | ND | MND |
| **Procedural Guidance** | ND | ND | ND |
| **Work Processes** | ND | 0 | 0 |
| **Indications of Conditions** | ND | N/P | 0 |
| **Stress** | 0 | N/P | 0 |
| **Execution Complexity** | 0 | N/P | N/P |
| **Human- Machine Interface** | N/P | 0 | 0 |
| **Team Dynamics** | 0 | N/P | ND |
| **Communication** | 0 | ND | 0 |
| **Time Pressure** | n/a | 0 | 0 |

### H4.5 HFE 3A

This was a straightforward scenario with no major complicating factors. The empirical data saw no negative drivers, and both assessments are fairly consistent with the empirical data in this respect. The SAIC analysis noted a few negative drivers (i.e., many steps equated to execution complexity and stress, and communications were sometimes inaudible as observed in the simulator), but none were judged to be significant. The table below provides a breakdown of the quantitative results:

| HEP | SAIC | NRC | Experimental |
|---|---|---|---|
| **Total** | **3.0E-3** | **2.0E-05** | **0/3 fail** |

The NRC analysis noted that the main failure mechanism was failure for “inexplicable” reasons or random failures. Similarly, the SAIC analysis noted that if failures occurred, they were likely to be in execution, but there was ample time for recovery. Like the NRC, they judged this failure to be “in the realm of extremely unlikely.” While both assessments had similar qualitative assessments, the quantitative vales were nearly two orders of magnitude different. The empirical data saw no failures.

NRC FORM 335
(12-2010)
NRCMD 3.7

U.S. NUCLEAR REGULATORY COMMISSION

# BIBLIOGRAPHIC DATA SHEET

*(See instructions on the reverse)*

1. REPORT NUMBER
(Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.)
NUREG-2156

2. TITLE AND SUBTITLE
The U.S. HRA Empirical Study -
Assessment of HRA Method Predictions against Operating Crew Performance on a U.S. Nuclear Power Plant Simulator

3. DATE REPORT PUBLISHED

| MONTH | YEAR |
|---|---|
| June | 2016 |

4. FIN OR GRANT NUMBER

5. AUTHOR(S)
John Forester, Huafei Liao, Vinh N. Dang, Andreas Bye, Erasmia Lois, Mary Presley, Julie Marble, Rod Nowell, Helena Broberg, Michael Hildebrandt, Bruce Hallbert, Tommy Morgan

6. TYPE OF REPORT
Technical

7. PERIOD COVERED (Inclusive Dates)

8. PERFORMING ORGANIZATION - NAME AND ADDRESS (If NRC, provide Division, Office or Region, U. S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)
Huafei Liao, P.O. Box 5800 MS 0748
Sandia National Laboratories
Albuquerque, NM 87185

9. SPONSORING ORGANIZATION - NAME AND ADDRESS (If NRC, type "Same as above", if contractor, provide NRC Division, Office or Region, U. S. Nuclear Regulatory Commission, and mailing address.)
Division of Risk Analysis
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

10. SUPPLEMENTARY NOTES

11. ABSTRACT (200 words or less)
This report documents the U.S. Human Reliability Analysis (HRA) Empirical Study (referred to as the U.S. study in the report), which is a large systematic data collection effort supported by the U.S. Nuclear Regulatory Commission with participation of organizations from five countries representing industry, regulators, and the research community. The objective of the U.S. study was to improve the insights developed from the International HRA Empirical Study [1–4] (referred to as the International Study) and address the limitations of that study.

Similar to the International Study, the U.S. study evaluated the performance of different HRA methods by comparing method predictions to actual crew performance in simulated accident scenarios conducted in a U.S. nuclear power plant (NPP) simulator. There was significant agreement in the findings and conclusions between the International and U.S. studies in terms of the strengths and weaknesses of the HRA methods evaluated in both studies and in the overall findings about HRA and the identified needed improvements.

12. KEY WORDS/DESCRIPTORS (List words or phrases that will assist researchers in locating the report.)
Human reliability analysis, simulator experiment, operator performance

13. AVAILABILITY STATEMENT
unlimited

14. SECURITY CLASSIFICATION
*(This Page)*
unclassified
*(This Report)*
unclassified

15. NUMBER OF PAGES

16. PRICE

Printed on recycled paper

Federal Recycling Program

**UNITED STATES**
**NUCLEAR REGULATORY COMMISSION**
WASHINGTON, DC 20555-0001

OFFICIAL BUSINESS

Blog @NRCgov You Tube

NUREG-2156 The U.S. HRA Empirical Study: Assessment of HRA Method Predictions against Operating Crew Performance on a U.S. Nuclear Power Plant Simulator June 2016