TESTING THE INTERNAL AT-POWER APPLICATION OF THE IDHEAS HRA METHOD

Huafei Liao[1], Stephanie Morrow[2], Gareth Parry[3], Dennis Bley[4], Lawrence Criscione[2], and Mary Presley[5]

[1] *Sandia National Laboratories, Albuquerque, NM, USA, 87185, hnliao@sandia.gov*
[2] *Nuclear Regulatory Commission, Washington, DC, USA, 20555, stephanie.morrow@nrc.gov; Lawrence.Criscione@nrc.gov*
[3] *Jensen Hughes, Walnut Creek, CA, USA, 95497, gparry@jensenhughes.com*
[4] *The WreathWood Group, Dublin, OH, USA, 43016, dbley@mac.com*
[5] *Electric Power Research Institute, Palo Alto, CA, USA, 94304, mpresley@epri.com*

The Integrated Human Event Analysis System (IDHEAS) internal at-power application (hereafter "IDHEAS AT-POWER") is a new human reliability analysis (HRA) method developed by the U.S. Nuclear Regulatory Commission (NRC) in collaboration with the Electric Power Research Institute (EPRI). It was designed for analyzing operator actions during internal, at-power nuclear power plant events. There is evidence that HRA results can vary substantially, depending on which method is used and how different analysts apply a particular method. IDHEAS AT-POWER was developed to provide a structured approach to the qualitative and quantitative analysis of human events. The purpose of the present study was to test IDHEAS AT-POWER and evaluate whether its guidance can be practically applied to produce consistent HRA results. The preliminary testing results of the method are presented in the paper.

I. INTRODUCTION

Human reliability analysis (HRA) uses systems engineering and behavioral science methods in order to render a complete description of the human contribution to risk and to identify ways to reduce that risk. It is used in the context of probabilistic risk assessment (PRA) to provide risk information regarding human performance to support risk-informed decisionmaking with respect to high-reliability industries. For example, risk information from HRAs is an important input to U.S. Nuclear Regulatory Commission (NRC) licensing and regulatory decisions.

Variability in HRA results is still a significant issue, even in state-of-the-art methodologies. The variability in HRA in turn contributes to uncertainty in PRA results. The sources of variability are most commonly attributed to the existence and use of different HRA methods that rely on different assumptions, human performance frameworks, quantification algorithms, and inconsistent implementation from analysts. This results in concerns over the robustness of HRA methods.

The Integrated Human Event Analysis System (IDHEAS) for human performance in internal at-power operation (hereafter "IDHEAS AT-POWER")[1] was developed as a joint effort between the U.S. NRC and Electric Power Research Institute (EPRI). Some of the objectives of developing this method were to enhance qualitative analysis, reduce analyst-to-analyst variability, and improve estimates of human error probabilities (HEPs) for internal at-power applications. The method includes the following key elements (see the full report for more information on the method).[1]

1. Description of the accident scenario and analysis of the human performance challenges in the scenario.
2. Identification and definition of the human failure events.
3. Feasibility analysis and time uncertainty analysis.
4. Task analysis and development of a Crew Response Diagram (CRD). The CRD is a graphical representation that identifies the critical tasks required for successful response, the expected crew response paths, and error recovery (correction) potential.
5. Implementation of the method's quantification model to estimate the HEPs. The quantification model provides a set of 15 crew failure modes (CFMs), a decision-tree (DT) for each CFM that includes the most relevant performance influencing factors (PIFs), and expert-estimated HEPs of every CFM under different combinations of the PIFs.
6. Model integration by analyzing dependencies between the HFEs in a PRA sequence, identifying and analyzing recovery actions, and documenting uncertainties in the HRA.

The purpose of the present study was to test IDHEAS AT-POWER and evaluate whether its guidance can be practically applied to produce consistent HRA results. It should be noted that due to the constraints on the study scope, some features of the method (e.g., time uncertainty analysis and dependency analysis) were not tested. The results presented in the paper are considered preliminary. More conclusive findings are expected through ongoing systematic and detailed analysis.

II. OVERVIEW OF STUDY METHODOLOGY

II.A. Testing Scenarios and Human Failure Events (HFEs)

The following three pressurized water reactor (PWR) scenarios were used in the testing. The first two scenarios were adapted from the U.S. HRA Empirical Study (hereafter "the U.S. Study")[2], with one human failure event (HFE) pre-defined for each. The third scenario was adapted from an actual event, for which three HFEs were defined.

Two primary considerations were taken in selecting the scenarios. First, the scenarios should vary in complexity. For example, Scenario 1 was selected to represent a very straightforward scenario, whereas Scenario 2 was selected to represent a more complicated scenario. This allows for an evaluation of whether analysts' predictions using IDHEAS AT-POWER are sensitive to scenario differences and whether the predictions are adjusted accordingly. Complicated scenarios and HFEs are also important in that they can create a broad spectrum of performance issues to test an HRA method's capabilities to identify them. Second, we sought to test the capability of IDHEAS AT-POWER for event analysis applications, such as the Accident Sequence Precursor (ASP) program and Significance Determination Process (SDP) at the U.S. NRC. Therefore, Scenario 3 was developed based on an actual event in an NPP, with some scenario detail removed or summarized at a high level in order to allow analysts to (1) focus on the important aspects of the scenario that led to the event, and (2) understand the event sequence and associated contextual factors. Although plausible, some of the HFEs used in the testing are far more difficult than many of the HFEs in standard PRA scenarios.

II.A.1. Scenario 1: Standard Steam Generator Tube Rupture (SGTR)

While operating at power, a tube rupture occurs in a steam generator. The leak size is about 500 GPM at 100% power, which is sufficient to cause nearly immediate secondary radiation alarms and other abnormal indications/alarms. The operators have about 3 hours to isolate the ruptured SG and maintain Reactor Coolant System (RCS) pressure below the setpoint by cooling down the RCS.

- Definition of HFE 1: Failure to isolate the ruptured steam generator (SG) and control pressure below the SG PORV setpoint before SG PORV opening.

II.A.2. Scenario 2: Total Loss of Feedwater (LOFW) with Misleading Indicator

While operating at power, all main feedwater pumps are tripped and the start-up feed pump cannot start. Three of the four auxiliary feedwater (AFW) pumps fail after automatic start. The fourth AFW pump will start automatically and indicate full flow, but this flow will not reach the SGs because a recirculation valve is mis-positioned 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 is going into the SGs. The SG levels will go down, and the plant computer will not show a red path on the heatsink status tree because of the indicated flow from the running AFW pump. The crew will have to identify that the indication of AFW flow from the running AFW pump is false and establish Bleed and Feed (B&F) when the wide range (WR) level on any two SGs are less than 50%. All attempts to establish AFW before B&F initiation will fail.

- Definition of HFE 2: Failure to establish B&F within 45 minutes of the reactor trip, given that the operators initiate a manual reactor trip before an automatic reactor trip.

II.A.3. Scenario 3: Electrical Fault Causes Fire with Reactor Trip and Subsequent Loss of RCP Seal Injection and Cooling

While operating at power, an electrical feeder cable failure caused a fire, an automatic reactor trip, an automatic safety injection (SI), and other equipment malfunctions. Within the first minute of the initiating event, Reactor Coolant Pump (RCP) seal cooling (via Component Cooling Water (CCW)) is lost due to the closing of a CCW thermal barrier outlet isolation valve. Approximately 27 minutes into the event, the charging flow is diverted from the RCP seals to the RCS and RCP seal injection becomes inadequate because a Chemical and Volume Control (CVC) Valve fails open. As a result, the RCP seals begin to heat up and purge volume begins to empty. For successful recovery, operators need to re-open the CCW thermal barrier outlet isolation valve from the control room within approximately 19 minutes after all RCP seal cooling and injection are lost.

- Definition of HFE 3: Failure to restore CCW to the RCP thermal barrier heat exchangers by re-opening the CCW thermal barrier outlet isolation valve.

If the CCW thermal barrier outlet isolation valve is not opened in time or cannot be opened, operators have approximately 19 minutes from when seal cooling and injection are lost to trip the running RCPs and avoid catastrophic seal failure.

- Definition of HFE 4: Failure to trip the RCPs during a loss of all seal cooling and injection.

If operators fail to stop the RCPs and/or restore cooling prior to the seals being fully challenged, the seals will fail at the maximum leakage rate of 480 gpm per RCP. The RCP seal failure will lead to a small loss of coolant accident (SLOCA). The operators have at least two hours to initiate a cooldown and depressurization of the RCS prior to depletion of the RWST inventory to allow for the plant to be placed in shutdown cooling (SDC) using the residual heat removal (RHR) system.

- Definition of HFE 5: Failure to depressurize the RCS during a small loss of coolant accident (SLOCA).

II.B. HRA Analyst Teams

Four HRA analyst teams (called Teams 1- 4 in the study) were recruited for the study, two from industry and two comprised of NRC analysts. Each team consisted of two or three analysts. An effort was made to control for levels of expertise and background across the teams and ensure that each HRA team had combined familiarity with nuclear power plant engineering, operations, PRA, and HRA. However, there were still some variations in expertise in the different areas across the teams. To avoid potential bias, the analysts did not have prior knowledge of the testing scenarios, and were not involved in the development of IDHEAS AT-POWER. Before they used IDHEAS AT-POWER for testing, all teams were trained in a two-day workshop to ensure that they fully understood the method and could use it independently. In addition, the project team also analyzed the scenarios with IDHEAS AT-POWER. They were treated as a fifth team (Team 5) and compared with the other teams.

II.C. HRA Inputs and Output Reporting

A prerequisite for HRA analyses in a 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, availability of cues, and potential distractions). The teams had the opportunity to interview the same designated operator to avoid introducing variability in the HRA predictions. The assumption is that the operator would provide the same information for questions, since the analysts were not able to perform other familiarization tasks, such as a plant visit, and observations of a crew in a simulator training scenario due to logistical considerations.

To ensure all of the analyst teams had access to the same information as a basis for their analyses, an information package was compiled by the project team for the analyst teams. The package provided the descriptions of the scenarios and the HFEs to be analyzed and the relevant procedures and other information that was necessary for the teams to perform their analyses. The analyst teams also had the opportunity to request clarifications or additional information from the project team. The questions from the individual teams and the answers from the project team were not 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 the information was decided to be germane enough that all HRA teams should have received it in the information package.

A template was provided to the analyst teams to document their analysis results, but the teams were allowed to document their analyses in other formats. The failure probabilities associated with DT paths were not provided to the analysts to avoid potential influence of the probabilities on the analyst teams. After the analyst teams submitted their analysis results, the final HEP of each HFE was calculated by the project team and then provided to the analyst teams for a reasonableness check.

Each team was also asked to complete a post-analysis questionnaire to provide additional information about their analysis of the HFEs and experience using the method.

II.D. Evaluation Criteria

The results of the testing were evaluated against the following criteria by comparing across analyst teams and with empirical data when available.

- Validity. In the current context, validity refers to whether IDHEAS AT-POWER can provide reasonable estimates about human reliability. The method should have sufficient guidance to enable analysts to identify the most likely mechanisms of failure that can lead to a given HFE, and consequently derive an appropriate HEP. Validity will be evaluated by comparing testing results against empirical data, when available, or the results of other HRA methods.
- Inter-analyst consistency. This criterion evaluates the extent to which different HRA analyst teams produce the same or similar results when using IDHEAS AT-POWER guidance. Previous benchmark studies have documented concerns with variability between analysts even when using the same method. Inter-analyst consistency will be evaluated for both the qualitative analysis and quantitative results using IDHEAS AT-POWER.
- Traceability. This criterion evaluates the extent to which IDHEAS AT-POWER documentation provides a clear link from the qualitative analysis to the quantitative inputs, and then to the HEPs. Traceability should be sufficient to allow for a third party to understand how the analysts derived HEPs for each HFE. Traceability is particularly critical in circumstances where there are disagreements in the HRA results, because it allows the analysts to identify where and why their results differ.
- Usability. This criterion evaluates the quality of the analysts' experience with using the IDHEAS AT-POWER method for HRA. For example, to what extent did analysts find IDHEAS AT-POWER guidance easy to use, how much time was required to analyze a scenario, and did the analysts encounter any difficulties with using the method?

- Utility. This criterion evaluates the extent to which IDHEAS AT-POWER provides useful information for decisionmaking. In addition to deriving HEPs to inform PRA, the qualitative outputs from HRA can be useful for understanding why human failure may occur, and thereby make changes to reduce opportunities for human error or increase opportunities for recovery.

III. PRELIMINARY ASSESSMENT OF TESTING RESULTS

The test produced a rich set of qualitative and quantitative data. Since one of the drivers for the development of IDHEAS AT-POWER was to reduce inter-analyst variability, the assessment focuses on the differences across the teams in both the qualitative analysis and quantitative results. In addition, comparison of IDHEAS AT-POWER predictions with the empirical data and other HRA methods' predictions produced in the U.S. Study[2] can provide valuable insights concerning IDHEAS AT-POWER and HRA in general.

Since differences in the qualitative analysis have a direct impact on the quantitative results, the quantitative predictions for each HFE are examined first in the study to direct the assessment of the qualitative analysis. By tracing back from the observed variability in the quantitative results to its sources, we can understand how the variability arises in the analysis process, whether the variability is caused by the method or by the use of the method, pinpoint areas of concern, and identify ways in which the method might be improved.

III.A. Expected HEPs

After the analyst teams completed their analyses using IDHEAS AT-POWER, but before the mean HEPs predicted by IDHEAS AT-POWER were calculated and provided back to the teams, each analyst team was asked to provide an expected mean HEP for each HFE based on their qualitative analyses. The expected HEPs can be considered subjective, quantitative expectations of operator performance. As shown in Fig. 1, no team was more consistently conservative or optimistic than others. All teams agreed that HFE 1 was the least difficult event, and the expected HEPs for this event were within one order of magnitude of each other. For HFEs 2, 3 and 5, the variability across teams was more than one order of magnitude. Teams 1, 4 and 5 were fairly consistent in their expectations across the HFEs. Furthermore, they expected comparable HEPs between HFEs 2 and 5.

The variability in the expected HEPs suggests that the teams had different perceptions of how the crew challenges (i.e. operational problems), scenario dynamics, and contextual factors may influence performance. It was observed that the analysts sometimes made different assumptions when they had insufficient information, different expectations of how operators would behave, and different interpretations of the information obtained from operator interviews.

III.B. Overall Quantitative Results Predicted Using IDHEAS AT-POWER

The predicted mean HEPs of all HFEs by the five HRA analyst teams are presented in Fig. 2.

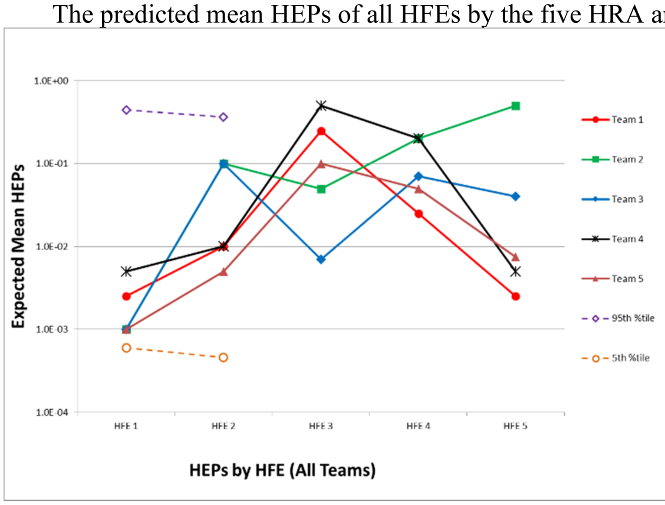

Fig. 1. Expected Mean HEPs by HFE

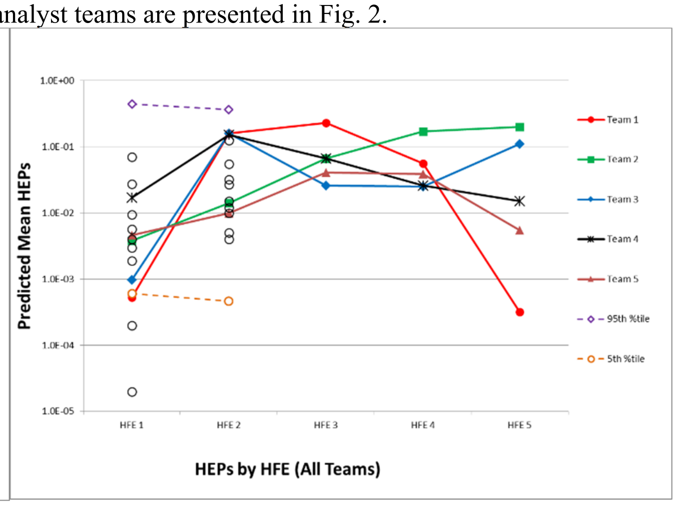

Fig. 2. Predicted Mean HEPs by HFE

III.B.1. Comparison to U.S. Study Data

The points connected by dotted lines represent the 5th and 95th percentile uncertainty bounds for HFEs 1 and 2. The bounds were derived through a Bayesian update with the Jeffrey's prior of the crew failure rates from the simulations performed in the U.S. Study.[2] As can be seen in Fig. 2, the HEP predictions of HFEs 1 and 2 fall within the 5th and 95th percentile uncertainty bounds, with one HEP for HFE 1 slightly below the 5th percentile. Although this is a large uncertainty interval, based as it is on a small sample with no failures, it provides some evidence that the teams did not significantly under- or over-estimate the two HFEs. The HEPs predicted using IDHEAS seem to converge with predictions from the U.S. Study, which are represented by the open circles in Fig. 2, and fall in a relatively smaller range.

III.B.2. Variability in Predicted HEPs

No team predicted consistently conservative or optimistic HEPs compared to other teams. The differentiation across HFEs varied across teams. For Team 1, there is a difference of approximately three orders of magnitude between the highest and the lowest HEPs. In contrast, the HEPs of all HFEs of Teams 4 and 5 fall within a range of approximately one order of magnitude.

Similar to the expected HEPs, there was also variability in the predicted HEPs. The variability of the HEPs for HFEs 2, 3, and 4 is approximately one order of magnitude, whereas there is significantly more variability among the HEPs for HFEs 1 and 5. Although all the HRA teams identified HFE 1 to be the easiest HFE, this HFE showed variability of approximately two orders of magnitude, with five predictions exhibiting a pattern of three clusters. One reason for Team 4 being significantly higher than the other four teams is that the one of the dominating CFMs identified by Team 4 was deemed unlikely by other teams. The disagreement on PIF ratings and treatment of recovery also contributed the variability.

For HFE 2, Teams 1, 3 and 4 obtained approximately the same HEP. A closer look at their analyses revealed that the five teams agreed on the same dominating CFM. However, their choices of three adjacent paths of the corresponding DT caused the HEPs of Teams 2 and 5 to be about one order of magnitude lower than those of Teams 1, 3 and 4. It should be noted that Team 4 misinterpreted some PIFs associated with the CFM.

The HEPs of HFE 3 fall in a relatively narrow range. One result that is worth noting is that Teams 2 and 4 obtained exactly the same HEP with the selection of exactly the same set of CFMs and DT paths. Team 1's relatively higher HEP was caused by double counting the effect of unavailable cues with two CFMs. Without double counting, Team 1's HEP would have been closer to those of other teams.

The HEPs of HFE 4 also fall in a relatively narrow range, and Teams 3 and 4 produced approximately the same HEP. However, the teams did not agree on the dominating CFMs.

For HFE 5, the HEPs significantly deviate from each other, leading to approximately three orders of magnitude difference, which echoes the disagreement of the teams in the difficulty levels and the expected HEPs of the HFE.

III.C. Observations from Qualitative Analysis

The qualitative analyses showed varying degrees of inter-analyst consistency across HFEs. Some teams produced qualitative analyses with richer, more detailed content, which seemed to help the teams understand scenario conditions and procedural interactions. In general, the qualitative analyses appear to have the ability to lead to an understanding of the scenario dynamics and performance drivers, and accommodate analysts' knowledge to characterize scenario complexity.

Overall, there tended to be relatively less variability in the qualitative analyses of less difficult HFEs (e.g., HFE 1) and HFEs (e.g., HFEs 2 and 5) that had well-defined procedural success paths compared to HFEs for which procedural paths were not clear (e.g., HFEs 3 and 4) due to uncertainty regarding how operators' would respond to potentially ambiguous procedures and plant parameters. There also tended to be less variability in HFEs that analysts were familiar with than those that analysts were not. For HFEs 1 and 2, there is a good consensus in the identified procedural paths and performance drivers, which are also consistent with the simulator data collected in the U.S. Study.[2] For HFEs 3 and 4, there is less agreement on main performance drivers, and the teams sometimes made different judgments and assumptions on which procedure would be entered and how it would be entered. For HFE 5, although the teams had a consensus on the procedural path and the critical tasks, there is considerable disagreement on main performance drivers as evidenced in the selection of CFMs and PIF ratings.

The teams differed in how they modeled execution tasks. In IDHEAS AT-POWER, an execution task is intended to be addressed in an integral manner rather than by assessing each of the individual sub-tasks. For example, in an SGTR scenario, the operator would need to perform RCS cooldown and depressurization and then control RCS pressure after the ruptured SG is isolated. These tasks should be treated and quantified holistically as one task rather than separately as some teams did.

Differences arose among CRDs in terms of how tasks were grouped together, or whether a task was explicitly considered due to low likelihood of failure. For example, all teams seemed to agree that identification of the ruptured SG was not likely to fail in HFE 1, but they treated the task differently. The first approach was to model the task explicitly, the second was to treat the task as an assumed success and not model it, and the third approach was to implicitly model the task by combining it with the identification of the SGTR. However, the different approaches did not seem to have a significant impact on the quantitative results.

Differences were observed in analysts' assumptions when there was uncertainty about operator actions. In HFEs 3 and 4, some operator actions depended on plant parameters or operators' interpretations of the procedures. Analysts had to make assumptions about those actions, which resulted in variations in the analysts' CRD construction and CFM selection.

IV. DISCUSSION & CONCLUSIONS

IV.A. Validity

In general, IDHEAS AT-POWER appears to have the capability to capture a broad range of failure modes, contextual conditions, and influences on behavior associated with the difficult HFEs and complex scenarios in the study. For HFEs 1 and 2, the HEPs predicted with the method were not inconsistent with the simulator data, and converge with other HRA methods' predictions in the U.S. Study. No team produced consistently optimistic or conservative HEPs. The differentiation in predicted HEPs varied across teams, ranging from one order of magnitude to three orders of magnitude. The teams' expected HEPs, in most cases, are consistent with the predicted HEPs, which suggests that the method is capable of factoring analysts' understanding of a scenario into the analysis process and produce an HEP that closely agrees with analysts' expectations.

IV.B. Inter-Analyst Consistency

The study revealed some degree of variability in both the qualitative and quantitative analyses, which seems to vary across the HFEs. Particularly, there is considerable variability in execution task and recovery modeling, which suggests that future versions of the method guidance and training should devote more attention to ensuring consistent use of the CFMs for execution tasks and modeling of recovery actions. Although some teams produced the same or similar HEPs for a given HFE, there was not always a consensus on CFMs and PIFs. As a result, our preliminary evaluation does not suggest strong evidence of inter-analyst consistency. Due to good traceability, the sources of the variability can be identified and are summarized below.

IV.B.1. CRD Construction

It is not surprising to see variations in the CRD when the analyst teams chose different procedural paths to model crew success. This seemed to occur when there was no clear procedural path, or when analysts had different expectations of typical operators. When the teams chose the same procedural path, they sometimes decomposed the HFEs differently (e.g., different task groupings) and/or identified different critical tasks. Some teams decomposed execution tasks into a lower level, which is not necessarily consistent with the method guidance as it was intended to be applied. Nevertheless, the difference in task decomposition alone did not seem to have a significant impact on the final HEPs in the study.

Another source for variability in the CRD was recovery modeling, which has a potential impact on the quantitative analysis results. One reason for the variability might be that the guidance was not very clear on how to follow a failure path from a critical task through the procedures to identify possible recovery paths.

IV.B.2. Analyst Judgment in Addressing Scenario Complexity and Uncertainty

It was observed in the study that the analysts relied on their judgment to address scenario complexity and uncertainty, which can lead to different understanding of a scenario in terms of procedural paths and scenario dynamics. The analyst judgment effect seems to be magnified for scenarios that are not defined in standard PRAs. When there is lack of familiarity with a scenario, analysts tend to rely on their prior knowledge of procedures, operations, and other scenarios and expectations of typical operator performance. As discussed in Section III.C, HFEs 3 and 4 were complicated by the uncertainty about operators' choices of procedural paths either due to unclear procedural guidance or due to the analysts' uncertainty of the values and trends of plant parameters used for situation assessment. The analysts dealt with the uncertainty by making assumptions to simplify their analyses. The method and its guidance provided insufficient help to the analysts to yield consistent results.

IV.B.3. CFM and PIF Evaluation

The selection of CFMs is based on an assessment of the activities needed to perform a critical task; hence analysts' understanding of the scenario dynamics will influence their decisions on potential failure modes. When the guidance or training is not sufficient, analysts have to rely on their judgment in interpretation of CFM scopes and applicability for a particular situation. In addition, IDHEAS AT-POWER employs a selection-out rather than selection-in philosophy for identifying applicable CFMs. Under this philosophy, each CFM is considered potentially applicable to a critical task unless the inapplicability of a CFM can be justified. Although analysts may disagree on the applicability of a CFM, the analysts who screen in the CFM may eventually select a DT path that yields a small failure probability with most or all associated PIFs being in nominal conditions. This situation may have minimal effect on the final HEP when the failure probability is negligible compared to other dominating CFMs. However, the effect can result in a material impact on the final HEP when there is not a clear dominating CFM. Since this study, guidance has been added on renormalization of the lowest failure probabilities to a prechosen lower bound, which can help resolve this issue.

Another source of variability was due to differences in the selection of PIFs. Analysts indicated that they had difficulty in interpreting the guidance for selecting PIF levels and answering the associated reference questions. Inconsistencies in the modeling of recovery and selection of recovery PIFs also contributed to variability in the results. It should be noted that when a recovery opportunity is identified for a critical task in the CRD, it will be credited via the Recovery Potential PIF if the selected CFM(s) has this PIF. The teams were not always in agreement regarding identification of recovery opportunities in the CRD, and were not always consistent in ensuring that when a recovery was modeled in the CRD the associated Recovery Potential PIF was credited, and vice versa.

IV.C. Traceability

The study indicates that traceability is a strength of IDHEAS AT-POWER. The method elements and analysis framework made it easy to pinpoint the sources of differences between method applications by different analysts. However, it should be noted that the ability to trace analysts' decision process on CFMs and PIF ratings is, to some extent, a function of the quality of analysts' documentation of their decision process. In particular, good documentation will be necessary to allow traceability if analysts base the PIF ratings on information that is not covered by the reference questions provided for PIF assessment, or analysts attempt to stretch the method to address CFMs or PIFs that are not adequately addressed in the method. In addition to improvement of the guidance for CFM and PIF assessment, emphasis on the importance of documentation can help further enhance traceability.

IV.D. Usability

The analyst teams indicated that IDHEAS AT-POWER was not easy to learn and extensive resources were needed to use the method. Several factors impacted the method's usability. First, it was time-consuming to develop a CRD and document the analysis. In particular, analysts reported that the documentation can be long, and there was sometimes unnecessary redundancy when documenting details. On average, it took most teams about 38 hours to analyze one HFE, which is considered much higher than the time needed with other HRA methods currently used by the NRC and industry. Even when accounting for time saved from repeated use of the method, analysts' still expected a typical analysis to take, on average, 16 to 24 hours.

Analysts reported a significant learning curve when it came to selecting CFMs and PIFs. Understanding the intent of each CFM and PIF is a skill that requires practice and training, and can be challenging for beginners. However, analysts believed that it would take less time as users become more familiar with the CFMs. In some cases, evaluation of CFMs and PIFs was difficult either due to inadequate guidance or because subjective judgment was needed.

It was also clear from analysts' feedback that using the method requires significant operations input for determining recovery paths and timing. As a result, effective use of the method can be, to some degree, dependent on analysts' experience in HRA and expertise in human performance and plant operations.

IV.E. Utility

Rich insights for error reduction can be derived from IDHEAS AT-POWER results. The analyst teams noted that the analysis process aided the development of a thorough understanding of the scenario. In addition, the CFM and PIF assessment helped analysts consider a broad range of failure modes and influencing factors. The insights on the performance issues and associated contextual factors are useful for training, plant design, and risk management decisionmaking.

IV.F. Additional Observations

In addition to the evaluation criteria described above, the project team had the following additional observations as a result of testing the IDHEAS AT-POWER guidance.

IV.F.1. Sensitivity of Binary Decision Trees

As discussed in Section III.B.2, although the five analyst teams agreed on the same dominating CFM for HFE 2, their choices of three adjacent paths of the same DT caused the HEPs of Teams 2 and 5 to be about one order of magnitude lower than those of Teams 1, 3 and 4. Therefore, the binary nature of the decision trees can cause the method to be quite sensitive to analysts' choices of PIF levels, and becomes a source of inter-analyst variability.

IV.F.2. Subjective Judgment in Method Application

When analyzing HFEs 3 and 4, the analyst teams encountered difficulty with performing the task analysis because there was not a single well-defined procedural path that would lead the crew to success. In most cases, the operators' choice of procedural paths is not between the right path and the wrong one, because each path can lead to success. However, one option may not be optimal because it decreases the likelihood of success (i.e., increases the conditional HEP). This situation is not captured by the CFMs and not addressed in the method guidance. All the analyst teams assumed and analyzed only one of the possible procedural paths. A possible resolution to this gap may be to analyze all the possible procedural paths on a case-by-case basis. The HEP for a particular path could then be assessed as a conditional failure probability given the relative probability of each case. Expert judgement would be needed to assess relative probability and derive a final HEP using this approach.

The guidance for selecting applicable CFMs and determining PIF ratings rely, to some extent, on subjective descriptions rather than being based on concrete, objective, and measurable criteria. This may hinder the utility of the guidance, as subjectivity in analysts' interpretation of the guidance may become a potential source of inter-analyst variability. This is exemplified in the study by (1) confusion and difficulty in interpretation of CFM scopes and PIF level definitions, and (2) considerable disagreement in judgment on whether the specific aspect of workload applicable to a CFM is high or low, whether a procedure is complex or simple, and whether an execution task is straightforward or not. Some of the disagreement had a significant impact on the final HEP.

IV.F.3. Treatment of Execution

Execution tasks are treated in an integral manner and categorized as two generic categories: simple and complex. It was observed that analysts disagreed on categorization of a particular execution task based on their interpretation of the guidance. Analysts also expressed frustration over not being able to differentiate between execution tasks that both met the definition for "complex," but that clearly had very different levels of complexity. In addition, although IDHEAS AT-POWER has a focus on the cognitive aspects of operator behavior, the holistic treatment of execution means that analysts are not encouraged to explicitly address cognitive demands during the execution of a response. This might lead to a failure to identify important PIFs and result in under-estimation of HEPs. The treatment of execution could be improved with additional execution CFMs, or additional guidance and examples for simple and complex actions.

IV.G. Conclusions

Although there were some instances where the method was not applied consistently with method guidance, the teams generally followed the analysis process and applied the method as intended. Overall, the method provides a structured qualitative analysis framework and self-consistent quantification approach. In general, it has the capability to capture a broad range of failure modes, contextual conditions, and influences on behavior associated with the difficult HFEs and complex scenarios in the study. In addition, it is capable of factoring analysts' understanding of a scenario into the analysis process and produce an HEP that closely agrees with analysts' expectations. The use of CRDs, CFMs, and DTs promotes good traceability, which, with good documentation, enabled us to easily pinpoint differences between analysts. The insights on the performance issues and associated contextual conditions that the method can provide indicates its utility.

Analysts reported that applying the method was resource intensive, both in terms of the analysis and documentation requirements. Significant operations knowledge was needed to use the method, and a substantial learning curve was reported for understanding how to apply the various CFMs and PIFs. It was observed that analysts' knowledge of operations and scenarios played a role in helping them understand scenario dynamics. Teams with relatively less knowledge tended to rely more on the designated operator to identify procedural paths. Significant inter-analyst variability was observed in some cases. Even when teams derived similar final HEPs, there were often variations in the chosen CFMs and PIF levels.

There were several areas where the test indicates that improvements to the guidance may be helpful, including selection of the appropriate level of a PIF, selecting one or more procedural paths and CFMs in the face of uncertainty, and the depth of documentation needed to support the qualitative analysis. Like with other methods, analysts tended to rely on their judgment when faced with challenging scenarios that went beyond the method's guidance and training. Additional guidance may be needed to assist analysts with using their judgment in a more consistent manner. Note that these conclusions are preliminary based on an initial review of the testing results and feedback from the analyst teams. Analysis of the testing results are ongoing.

ACKNOWLEDGMENTS

The authors would like to thank the analysts who participated in the testing. Their participation and feedback provided invaluable insights into the practical use of the IDHEAS AT-POWER method.

This report was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product, or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights. The views expressed in this paper are not necessarily those of the U.S. Nuclear Regulatory Commission.

REFERENCES

1. Xing, J., Parry, G., Presley, M., Forester, J., Hendrickson, S., & Dang, V. (2016). An Integrated Human Event Analysis System (IDHEAS) for NPP Internal At-Power Application. NUREG-2199, Vol.1, In Press. Washington, D.C.: U.S. Nuclear Regulatory Commission (ADAMS: ML16147A358).
2. Forester, J., Liao, H., Dang, V.N., Bye, A., Lois, E., Presley, M., Marble, J., Nowell, R., Broberg, H., Hildebrandt, M., Hallbert, B., & Morgan, T. (2016). The US HRA Empirical Study – Assessment of HRA Method Predictions against Operating Crew Performance on a US Nuclear Power Plant Simulator. NUREG-2156. U.S. Nuclear Regulatory Commission, Washington, DC (ADAMS ML13225A516).