

# Overview and Preliminary Results of the U.S. Empirical HRA Study

Andreas Bye<sup>a\*</sup>, Vinh N. Dang<sup>b</sup>, John Forester<sup>c</sup>, Michael Hildebrandt<sup>a</sup>, Julie Marble<sup>d</sup>,  
Huafei Liao<sup>c</sup>, Erasmia Lois<sup>d</sup>

<sup>a</sup>OECD Halden Reactor Project, Institute for energy technology, IFE, Halden, Norway

<sup>b</sup>Paul Scherrer Institute, Villigen PSI, Switzerland

<sup>c</sup>Sandia National Laboratories, Albuquerque, NM, USA

<sup>d</sup>United States Nuclear Regulatory Commission, Washington, DC, USA

---

**Abstract:** Human reliability analysis (HRA), an important aspect of Probabilistic Risk Assessment (PRA), evaluates the contribution of human performance to risk. Work to improve HRA is motivated because it is a major contributor to variability in PRA results. This is due to the difficulties in predicting human action, the fact that different methods rely on different human performance frameworks and data, and that analysts may apply the methods inconsistently. In the International HRA Empirical Study, HRA predictions of different analysts and methods were compared to crew performance data at the Halden Reactor Project simulator facilities (HAMMLAB). This paper discusses the follow-up to this study that utilizes crew data from a US plant training simulator. A major objective of the US Empirical HRA Study was to test the consistency and accuracy of HRA predictions among different analyst teams using the same methods. At least 2 teams of analysts applied each method to predict the outcome of the scenarios. Another aim was to consolidate and extend the results so far obtained (in the HAMMLAB study) on the strengths and weaknesses of HRA methods, and to ascertain the extent to which they are replicated in a plant training simulator environment. This paper provides an overview of the study, its methodology, and the overall results of the empirical data and the method assessments. The differences between this study and the earlier International study are highlighted. A companion paper addresses specifically the intra-method comparisons and detailed results per HRA method based on these [1].

**Keywords:** HRA, PRA, Benchmarking, Crew Performance.

---

## 1. INTRODUCTION

Human reliability analysis (HRA) is an important aspect of PRA since it evaluates the contribution of human performance to risk, frequently identified as significant. However, HRA is a major contributor to the variability of PRA results. The application of different HRA methods that rely on different human performance frameworks and data, as well as inconsistent implementation from analysts, yield the most common sources of variability.

To address these variability issues, several international partners participated in the International HRA Empirical Study [2,3,4], in which HRA predictions of different analysts and methods were compared to observed crew performance data at the Halden Reactor Project HAMMLAB simulator facilities. The International HRA Empirical Study identified important strengths and weaknesses of the various HRA methods used in the study [5], 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, since only one of the methods examined in the International Study was applied by different analysis 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 was performed on a US nuclear power plant (NPP) simulator and is thus referred to as the US HRA Empirical Study, was to test the consistency and accuracy of HRA predictions among analyst teams using a given method.

The US Study also addressed two other potential limitations of the International Study. First, for logistical reasons, the HRA teams in the International Study 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 a NPP PRA. This type of information was provided to the HRA teams to the extent possible by the study team and the HRA teams were allowed to submit written questions that were answered by the study team and plant personnel as

needed. As a result, some of the HRA teams in the international study felt that this significantly limited their ability to perform an adequate HRA. In contrast, the HRA teams in the US study visited the reference plant and collected information relevant to performing their HRA as would normally be done in a PRA.

Further, there was some 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 US NPP crews. More specifically, some HRA teams in the International study thought that their expertise was more geared to understanding what US crews would do and that their US bias may have influenced their HRA method application. Thus, the US study served as a check against the effects of such bias on the results.

## **2. OVERVIEW OF METHODOLOGY**

The basic methodology of the US study was in many respects similar to the International HRA Empirical study, see [2,3,4,5]. This section emphasizes and discusses the differences between the studies.

The focus of both studies was on control room personnel actions required in the response to PRA initiating events. The HFEs were defined beforehand when the scenarios were designed. In order to compare predictions, one must ensure that the HRA teams are analyzing and predicting performance on the same human actions. The US study empirical data was collected on a US plant full-scope training simulator with a conventional control room of a PWR 4-loop Westinghouse reactor. Four crews of five licensed crew members participated. Three different scenarios were run. Scenario 1 was a total loss of feedwater (LOFW) followed by a steam generator tube rupture (SGTR). Scenario 2 was a loss of component cooling water (CCW) and reactor cooling pump (RCP) sealwater. Scenario 3 was an SGTR scenario without further complications. Collecting data in this environment was a little different from collecting data in the computerized HAMMLAB environment. The data were collected by observers both from the experimental team and from instructors at the plant taking notes, and from simulator logs, audio/video recordings and debriefing interviews. A difference between this study and the HAMMLAB study was that no audio/video could be removed from the plant site, so the experimental team was dependent on a well-prepared data collection process where notes, logs and questionnaires could be used without support from later video studies. The data were analyzed at the crew level first, before being aggregated over the crews on a HRA performance level that could be compared to the HRA predictions.

The HRA teams received a description of the scenarios and HFEs, as well as information on the relevant procedures, operator training, and conduct of operations. They also visited the plant to collect information, which was a major improvement over the International study. The HRA teams interviewed instructors about the scenarios, practices in the use of procedures, expected actions, training, the way the crews normally work in the control room and any other issues relevant for the analysis. The teams were also allowed to observe a general training session in the simulator, but it was a different scenario than the ones under analysis. This kind of plant visit and interviews/observation is considered good practice while doing HRA [6]. Nine HRA teams participated in the study and each team applied an HRA method to obtain predictions for the HFEs in the simulator scenarios. Two teams used ATHEANA, two teams used SPAR-H, two teams used ASEP, two teams used the HRA Calculator (with CBDT, HCR/ORE and THERP), and one team used methods that are used in the HRA Calculator, but did not use the actual software. Detailed descriptions of these methods including all the main references to each of the methods can be found in [2].

The predictions from the HRA teams were compared to the empirical data in several ways. Qualitative predictive power was assessed based on a comparison of predicted operational expressions and observed operational descriptions, as well as a comparison of predicted performance shaping factors and observed performance drivers. Quantitative predictive power was assessed to the extent possible in light of the small number of data points. For this, the predicted HEPs were compared to the level of difficulty of the HFEs in the empirical data. The latter was expressed by the number of failures of the HFE as well as an assessment of the difficulty for each HFE. This assessment was performed as a ranking based on number of failures, near-misses, operational difficulties experienced by the crews and a ranking done by unit supervisors. Aspects considered in the quantitative comparisons were: 1) potential optimism of the most difficult HFEs; 2) consistency of the ranking of the HFEs (by predicted HEP) with the reference difficulty ranking; 3) predicted HEPs relative to the confidence/uncertainty bounds of the reference data; 4) quantitative differentiation of the HFEs by HEP. In practice the quantitative comparisons were also used as a starting point for diving into

the qualitative issues: For example, why was an analysis optimistic for the most difficult HFE, or what was the underlying reason why an analysis did not manage to differentiate between the HEPs of the HFEs?

### 3. EMPIRICAL RESULTS INCLUDING SCENARIO AND HFE DESCRIPTIONS

A detailed description of the empirical results can be found in [7].

#### 3.1 Scenario 1 – Total Loss of Feedwater (LOFW) Followed by a Steam Generator Tube Rupture (SGTR)

**LOFW.** 2 minutes into the scenario all feedwater pumps tripped and the start-up feed pump could not be started. If the crew didn't trip manually, the reactor would trip on low SG level (20% NR) within 50-60 seconds. Auxiliary feedwater (AFW) received an automatic start order, but all four pumps failed to supply any feedwater due to pump failures (overspeed or shaft failures on three of the pumps). AFW pump 12 started and indicated full flow to the steam generator (SG), but because of a mis-positioned recirculation valve ("open" position), the water did not reach the B SG. There was no indication of the valve's position in the control room. The open recirculation valve masked the fact that no AFW at all was going in to the SGs. Criteria to start FR-H.1 (RESPONSE TO LOSS OF SECONDARY HEAT SINK) were met, but because of the indicated flow from AFW pump 12, the plant computer did not show a red path on the heatsink critical safety function status tree. The crew had to identify that the indication of AFW flow from pump 12 was false and decide to go to FR-H1.

*HFE 1A:* Failure to establish feed and bleed within 45 minutes of the reactor trip, given that the crews initiate a *manual reactor trip* before an automatic reactor trip.

*HFE 1B:* Failure to establish feed and bleed 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 feed and bleed include actuating Safety Injection and opening both of the pressurizer PORVs.

**SGTR.** After the crews had started bleed and feed (B&F), they could establish AFW flow to the SGs by closing the recirculation valve and/or cross-connecting the flow from the running AFW pump to the other SGs. If the crews tried to establish AFW before initiating B&F, the scenario called for delaying this possibility. Moreover, in the scenario, establishing AFW flow led immediately to a tube rupture in the first SG that was fed (note that the tube rupture was masked by AFW flow to the SG, as long as it is fed.) The leak size of the ruptured tube was about 500 GPM at 100% power. There was initially no secondary radiation because there was only a minimum steam flow. The blow down (BD) and sampling were secured because of the SI. By the time the crews had filled the SG(s) enough to exit FR-H1, they could have problems with the RCS integrity status tree and be forced to enter procedure FR-P1 (RESPONSE TO IMMINENT PRESSURIZED THERMAL SHOCK CONDITION), which delayed the possibility to go to the SGTR procedure E-30 (STEAM GENERATOR TUBE RUPTURE).

*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 40 minutes. The actions include isolating the ruptured SG (close feedwater and main stream isolation valves) and maintaining SG pressure below the setpoint by cooling down the RCS (cooling the secondary system by dumping steam and depressurizing the RCS).

#### *Performance of Human Failure Events 1A and 1C*

*HFE 1A:* 0/4 crews failed. All crews manually tripped the reactor. All crews established bleed and feed well within 45 minutes of the reactor trip.

*HFE 1B:* no data. There is no data for HFE 1B. Because all crews tripped manually, none faced this condition (none had an automatic reactor trip).

*HFE 1C:* 3/4 crews failed (within the 40 minute time frame). Three crews (Q, S, T) isolated the ruptured SG and controlled pressure below the SG PORV setpoint and avoided SG PORV opening. Thus, three crews succeeded from a plant perspective, but only 1 crew accomplished the desired actions within the 40 minute time (S). One crew (R) isolated the ruptured SG according to procedure E-30 but failed to reduce the RCS pressure and the SG PORV did open approximately 40 minutes after the SGTR (Table 1).

Table 1. Performance on HFE-1C

Crew	Reach 100% SG level (WR)	Open SG PORVs	Isolate ruptured SG and control pressure within 40 minute time frame
Q	Yes	No	No
R	Yes	Yes	No
S	No	No	Yes
T	No	No	No

*Performance drivers in the scenario*

The difficulty in the scenario was identifying that all AFW was lost because a mispositioned recirculation valve on the AFW pump 12 (feeding SG B) was open, indicating flow that in reality did not go into the SG. Lack of FW or AFW is an entry criterion to procedure FR-H1, which will guide to start B&F. All crews detected that the level in SG B continued to decrease even though AFW flow was indicated, and they questioned if they actually had flow to the SG. Three of the four crews suspected that the recirculation valve was open. The main performance shaping factors for HFE 1A was thus the scenario complexity, the indication of conditions and procedure guidance, since the critical safety functions status trees did not address misaligned valves. For HFE 1C the adequacy of time was an additional driver, since they had quite little time to diagnosed the SGTR and work through the procedures. The procedural guidance itself was not good for this HFE, since they were held up in higher priority procedures regarding the LOFW and bleed and feed actions.

**3.2 Scenario 2 – Loss Of CCW and RCP Sealwater**

The B train was out of service and made CCW pump B unavailable. 2 minutes into the scenario the distribution panel 1201 failed with the consequences that the following controlling channels had to be taken in manual control: A and B SGs, PRZ level control, Rod control, Nuclear Instrumentation (NIS), PRZ pressure control. The failing distribution panel was unrelated to the loss of CCW and sealwater but increased the complexity of the scenario and masked the status of CCW and sealwater by keeping the crew busy due to the number of alarms.

The feedwater regulation valve on SG A could not be operated manually and remained fully open, feeding the SG. If the crew did not trip the reactor, there would be an automatic turbine trip on high SG level (87%), which caused a reactor trip. Because of the loss of 1201, one AFW pump did not start. When the reactor tripped, bus E1C had a bus lockout due to a bus fault, and CCW pump A tripped at the same time. There were 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).

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. Procedure 0POP04-RC-0002 and procedure ES-01 both have guidance to start the Positive Displacement Pump (PDP) when CCW and seal injection is lost. However, it can only be started if the RCP seal temperatures are below 230F, and reaching these procedure steps takes some time. In accordance with the Westinghouse RCP vender manual, if seal inlet or seal inlet bearing reach 230F, the potential for seal damage is too great to risk placing seal injection in service.

*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 seal water inlet or lower seal water bearing temperatures are greater than 230F (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 230F is about 7-9 minutes from loss of CCW.

### Performance of Human Failure Event 2A

HFE 2A: All 4 crews failed to trip the RCPs and start the Positive Displacement Pump (PDP) before RCP seal water temperatures were greater than 230 degrees.

Table 2. 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

All crews stopped the RCPs. Three crews stopped them before the RCP seal temperatures were above 230 degrees, and one crew a minute after. All crews exceeded the 1 minute criterion in POP4-RC2 for stopping the RCPs. None of the crews started the PDP.

Factors that made the crew detect the loss of all CCW and seal injection late were the complication of the scenario (main negative driver). There were many on-going actions after the 1201 failure and the reactor trip. Better board scans and checking the alarms could have made the crews identify the loss of CCW and seal injection earlier. Three of the four crews stopped the RCPs from knowledge and one crew from procedure. The crews that stopped the RCPs from knowledge were helped by training. The crews were late in stopping the RCPs because of the complication of the scenario. None of the four crews started the PDP. The crews did not find procedure guidance to start the PDP (negative driver), or it took them too long to go through the procedures. The start of PDP is for example in procedure ES-01 but they did not reach the step in time. E-0 and ES-01 were the highest priority procedures in the scenario. The other procedure that could help them, POP4-RC2 was not the highest priority procedure and the ROs that handled the procedure lacked experience and training in working in this procedure. In addition, the complication of the scenario made the task difficult. They identified the situation late and had problems recognizing the urgency. The crews did not have training on this specific scenario. They were used to train on loss of CCW and sealwater only in loss of offsite power scenarios, and therefore did not expect a loss of CCW and sealwater in another scenario. One crew expressed that the seals were not a concern, which may be a training problem. Therefore, training and experience was identified as another main negative driver.

### 3.3 Scenario 3 – Steam Generator Tube Rupture (SGTR)

Scenario 3 was a standard Steam Generator Tube Rupture scenario without added complications. 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.

*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: isolating the ruptured SG (feedwater and main steam isolation valves closed) and maintaining RCS pressure below the setpoint by cooling down the RCS (cooling the secondary by dumping steam and depressurizing the RCS).

### Performance of Human Failure Event 3A

All 3 participating crews succeeded on the HFE (Crew S did not participate in this scenario due to a simulator problem). The crews had no major difficulties in this scenario. Factors that supported their performance were indication of conditions, procedures, training and teamwork.

### 3.4 Difficulty Ranking of HFEs

For the ranking of HFE difficulty, three data sources were considered. A difficulty ranking of the HFEs was made by three of the four Unit Supervisors (US) that participated in the study (column “US rank” in Table 3). The three analysts made an independent difficulty rating based on the review of crew performance and an

in-depth analysis of the scenario complications (column “difficulty”). Finally, the number of HFE failures were considered (column “failure rate”). Table 3 below shows that all these judgments converge.

Table 3. Evaluated difficulty of the HFEs

HFE	Task	US rank	Failure rate	Difficulty
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	1 / 4 (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 crew being well trained on similar scenarios (the recirculation valve open).

#### 4. OVERALL RESULTS OF COMPARISON BETWEEN EMPIRICAL DATA AND HRA METHOD PREDICTIONS

The HEPs predicted by each method are plotted in Figure 1 alongside the Bayesian uncertainty bounds derived from the crew data. On the horizontal axis, the HFEs are ordered by their difficulty ranking in the empirical data (see above). It should be noted that HFE 1B was not ranked since there was no empirical data for this HFE. Some preliminary findings can be drawn from these curves. (Note that the findings in this paper are preliminary because the project is on-going and the full result report is not yet finished).

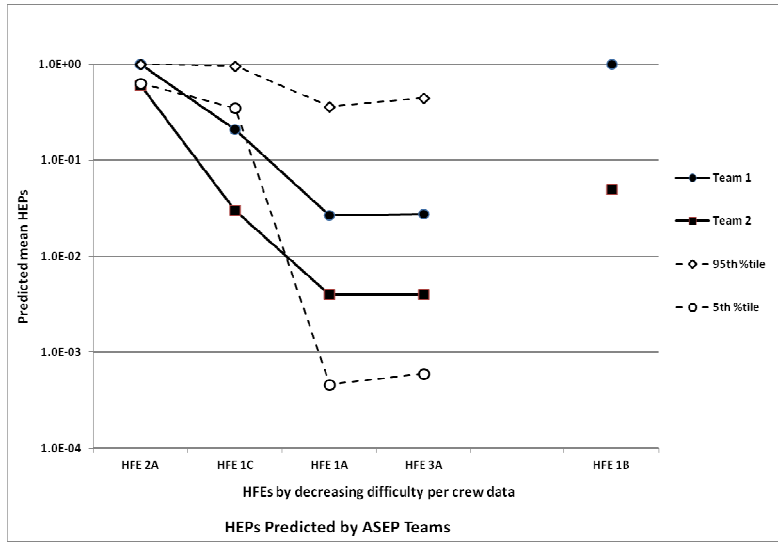
##### 4.1 Ranking of HFEs was reasonable for most methods

The ranking of HFEs is rather important for HRA, since HRA is a tool not only to find probabilities of human errors for use in PRA, but also to identify important events and prioritize where fixes and improvements are needed in the plant. Viewing the curves in Figure 1, most methods identified the correct ranking (decreasing curves) with only a few exceptions in terms of the relation between 2A and 1C. Note that the HFEs 2A, 1C and 1A were all rather difficult events. While plausible, they comprised far more difficulties for the crews than many of the HFEs in standard PRA scenarios. HFE-3A was a standard SGTR scenario with no further complications, and it was shown in the empirical data by the difficulty ranking to be the easiest. All the HRA teams identified this HFE to be the easiest HFE, with the lowest HEP. However, there was quite some variability in the HEPs for this HFE *between* the different methods and teams, see section 4.5. The reasons for this variability are not yet fully explored and we are looking into this.

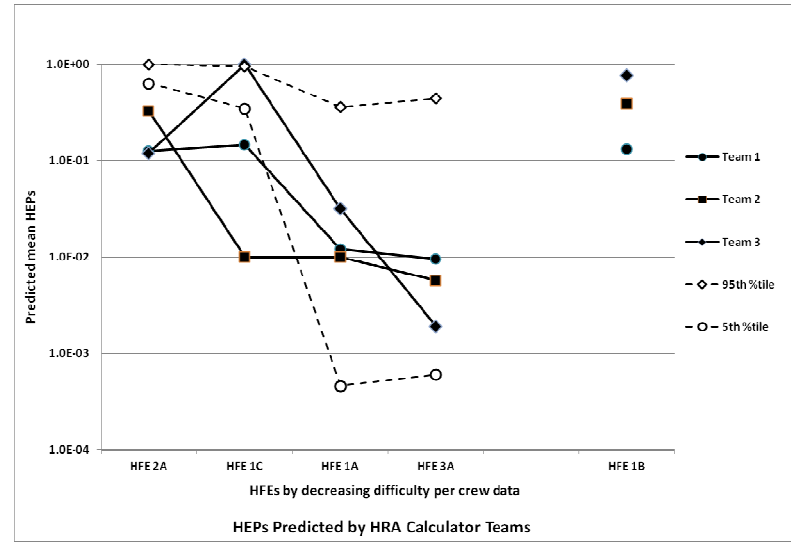
##### 4.2 For most HFEs, there is one order of magnitude difference across teams using a given method

Looking at the HEPs in Figure 1, there is generally about one order of magnitude or less difference across teams within each method. This is the case for most of the HFEs, except HFE-3A for SPAR-H and ATHEANA, and the reasons for these differences are briefly discussed in [1] and are being further investigated. For HFE-1C, it also seems that there is a larger difference for the HRA Calculator teams, but the value for one of the teams was apparently based on an incorrect interpretation of this HFE’s definition.

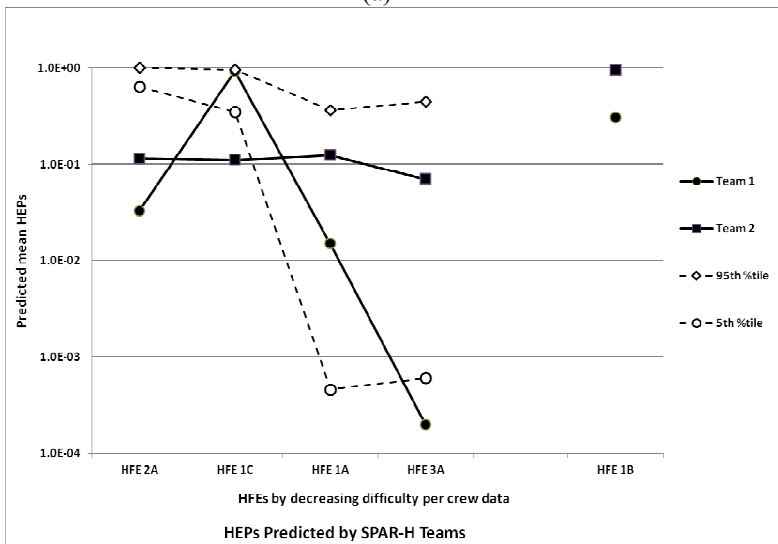
This could be considered a positive finding from an HRA perspective in the sense that greater degrees of variability have been found in other studies, especially taking into account that the observed convergence applied to the three most difficult HFEs and the notable variability in the crew performance on each of the HFEs. There were observed differences in crew performance not only in timing but also in some of the sequences of actions or paths through the procedures. For the difficult HFEs, this complicates the work for HRA analysts, even though one attempts to model average behavior, it is not so easy to do this when the variability in plausible behavior is so large, as seen in the observations in this study.



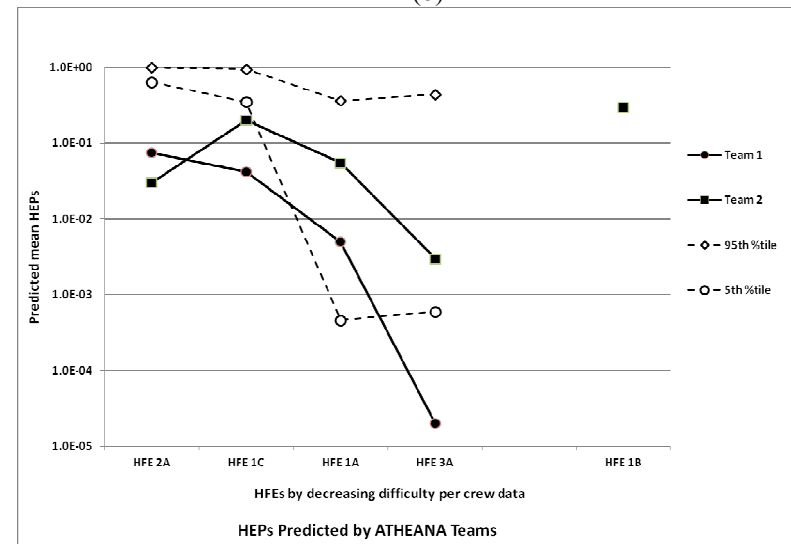
(a)



(b)



(c)



(d)

Figure 1. Predicted Mean HEPs by HRA Methods with Bayesian Uncertainty Bounds

HFE-1B is not much discussed in this study, since there were no observations in the empirical data for this HFE (i.e., all the crews tripped manually before an automatic trip would have occurred, so the HFE was not relevant). Nevertheless, comparing the predictions of HFE-1B shows that this was even more uniform than the others. Eight of the nine predicted HEPs by the different teams were well within one 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.

### **4.3 Some methods seem to be more consistent than other methods**

This study cannot prove quantitatively more or less consistency or accuracy for one or the other methods, with only two or three data points (HRA teams) per method. However, a review of the figures suggest that ASEP and ATHEANA yield somewhat more consistent quantitative results, across the two analysis teams that used each of these methods. Consistency here refers to ranking of the set of HFEs as well as to the differences in the HEPs for a given HFE. As discussed above, one of the HRA Calculator teams (with the lowest value) seems to have misunderstood the HFE definition for 1C, so these teams seem to produce relatively consistent results as well. Although statistically weak due to the small set of HFEs and data, such observations led to further questions and analysis. Are the predictions of these teams in fact consistent when the qualitative predictions (performance issues, failure mode, dominant factors and failure “mechanisms”) from these analyses are compared? In other words, do the two ASEP and the two ATHEANA analysis teams (respectively) obtain consistent quantitative results based on the same underlying reasons? These intra-method comparisons are discussed in more detail in the companion paper [1] and in the study report that is under preparation. Although the consistency of the quantitative results is generally encouraging, the detailed analysis that considers the qualitative predictions nevertheless found significant differences in the identified basis for the HEPs. Thus, the common results may have been somewhat serendipitous and raises potential questions about the reliability of the methods. These differences may also be important for safety management because they would be expected to lead to different safety insights and recommendations for potential improvements.

The differing quantitative results for the SPAR-H applications are particularly interesting and appear to be due to very different applications of the method. Team 1 used SPAR-H in the way most commonly used, by judging the PSFs on the HFE level after making a qualitative analysis as thorough as they thought necessary. Alternatively, Team 2 utilized Crew Response Trees (CRTs) to determine the decision points and basic sub-events for the analysis based on the break points in the procedures. A CRT represents a detailed decomposition of actions, events, and potential procedural branching points that could lead to a failure of the HFE. This is not a normal part of applying SPAR-H and it is unclear at this point to what extent the CRT exercise contributed to a lack of differentiation among the HEPs in the Team 1 results. Some hypotheses for the results from this way of using SPAR-H are discussed in [1]. For example, there might be an effect of the total number of PSFs used on the various sub-events that comes into play. SPAR-H utilizes a scaling factor when more than three PSFs are applied for one HFE. There is no guidance for how this should be handled in the case where one HFE is decomposed into several sub-events. Another plausible explanation is that the PSF levels are simply misjudged. We are investigating the reasons or combination of reasons for these results. These two ways of modeling the human actions and HFEs for the use with SPAR-H may or may not resemble two different methods. This is further investigated in the study.

### **4.4 Except ASEP, all other teams underestimated HFE 2A**

Plausible reasons for some of the HRA analyses underestimating HFE 2A are probably insufficient qualitative scenario analysis combined with or in addition to an over-reliance on answers from instructors. In the case of one HRA team, they 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 time. 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. The empirical results showed that the



crews did not manage the action. After the complex situation lead 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. So a better interview technique and a self-standing detailed analysis for making a good qualitative scenario analysis may help here; analyzing complexity and procedure progress in a better way.

#### **4.5 All teams agreed that HFE 3A was the easiest, but there was significant variability in the HEPs for this HFE**

As described above, HFE 3A is the isolation of the ruptured steam generator in a basic SGTR scenario. Qualitatively, all analysis teams concluded that this HFE was the easiest of the set. In addition, most teams identified no particular challenges or negative factors that could lead to a potential failure of this HFE. An exception was SPAR-H Team 2 that identified complexity as a performance driver for this HFE and this led to an apparent overestimation of the HEP and contributed to the estimated HEPs for this HFE differing by three orders of magnitude. Some additional reasons for this are discussed in [1]. However, it appears that HRA methods may not have a common baseline: there is no consensus across the methods on the appropriate failure probability for “easy”, practically routine actions. In other words, the methods do not produce the same value for an operator action in an abnormal/emergency scenario under optimal conditions. This observation is also related to the issue of defining human performance limits [8].

In the earlier International HRA Empirical Study [2,3,4], the assessment group as well as the HRA analysis teams noted that the HRA methods were not precise in defining the baseline conditions, for instance, those for which the “best” performance shaping factor ratings applied. Consequently, this issue relates both to the assumptions regarding what constitute “optimal” performance condition as well as the failure probabilities appropriate for (consistently defined) optimal conditions. While this issue may be less critical in terms of its effect on the ranking of HFEs, this issue does cause problems when overall risk figures-of-merit or failure probabilities for similar HFEs in different PSA studies are considered.

## **5. CONCLUSIONS**

The experience of conducting an HRA data collection at a plant training simulator has produced valuable methodological and empirical insights for the researchers and for the participating plant, and is considered by all participating parties as a promising research strategy for the future.

Variability in the crew performance and especially variability in how the crews performed the difficult scenarios was found in this study and confirmed the findings from the international study. Difficult accident scenarios create a variation in crew performance since the procedures are often not directly written for these situations, although they cover them in general. Also, small changes in the early development of the scenario, often induced by actions by the crews, may create large changes in the context for the crews at a later stage in the scenario.

Regarding the issue of a plant visit, we saw that the HRA teams used this for creating instrumental input to their analyses. However, there was variability in how the HRA teams performed interviews and thereby which results they got out of them. Differences in the qualitative analyses led to different predictions by the teams, regardless of the HRA method that they used.

Plants can learn much in performing these kinds of studies running very difficult scenarios. One typical example comes to mind in this study, and that is that the most difficult HFE, 2A, was seen to be a problem for the crews in performing the required actions within the time available, as shown in chapter 3. This was taken seriously in the plant, since it is an important event. After these runs, the plant has focused parts of its training on this scenario and HFE, and has considerably improved the operator performance on this HFE. This also reflects the paradox and difficulty for HRA and PRA in adjusting for the training PSF. To which extent is the training sufficient, and to which extent is it intermingled with other PSFs? Further discussion on training as a PSF, partly based on this study, is done in [9].

## Acknowledgements

The authors gratefully acknowledge the contributions of Helena Broberg and Salvatore Massaiu, the Halden Reactor Project, Bruce Hallbert and Tommy Morgan, Idaho National Laboratory, and Amy D'Agostino, USNRC for major parts of the experimental work done in the project. The work of the nine HRA teams has of course been of invaluable importance, as was that of the additional assessment team members, Alysia Bone, USNRC, Katrina Groth, Sandia National Laboratories, and Stuart Lewis, Electric Power Research Institute. Very special thanks goes to the US nuclear power plant that supported the study with their training simulator, operating crews, instructor support in designing the scenarios, and multiple staff supporting the data collection and analysis. The plant support is obviously a major contribution to supporting the improvement of HRA and the safety of nuclear power plants.

This study is a collaborative effort of the Joint Programme of the OECD Halden Reactor Project, the U.S. Nuclear Regulatory Commission (USNRC), the Swiss Federal Nuclear Inspectorate (DIS-Vertrag Nr. 82610) and the U.S. Electric Power Research Institute. In addition, parts of this work were performed at Sandia National Laboratories and Idaho National Laboratory (INL) with funding from the USNRC. Sandia is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000. INL is a multiprogram laboratory operated by Battelle Energy Alliance LLC, for the United States Department of Energy under Contract DE-AC07-05ID14517. The opinions expressed in this paper are those of the authors and not those of the USNRC or of the authors' organizations.

## References

- [1] Marble J, Liao H, Forester J, Bye A, Dang VN, Presley M, Lois E. Results and Insights Derived from the Intra-Method Comparisons of the US Empirical HRA Study. Proc. PSAM11&ESREL2012, June 25-29, 2012, Helsinki, Finland.
- [2] Lois E, Dang VN, Forester J, Broberg H, Massaiu S, Hildebrandt M, Braarud PØ, Parry G, Julius J, Boring R, Männistö I, and Bye A. International HRA Empirical Study—Phase 1 Report: Description of Overall Approach and Pilot Phase Results from Comparing HRA Methods to Simulator Data. HWR-844, OECD Halden Reactor Project, Halden, Norway and NUREG/IA-0216, Vol. 1. US Nuclear Regulatory Commission, Washington, DC, 2009.
- [3] Bye A, Lois E, Dang VN, Parry G, Forester J, Massaiu S, Boring R, Braarud PØ, Broberg H, Julius J, Männistö I, and Nelson P. International HRA Empirical Study—Phase 2 Report: Results from Comparing HRA Method Predictions to Simulator Data from SGTR Scenarios. HWR-915, OECD Halden Reactor Project, Halden, Norway and NUREG/IA-0216, Vol. 2. US Nuclear Regulatory Commission, Washington, DC, 2011.
- [4] Dang VN, Forester J, Boring R, Broberg H, Massaiu S, Julius J, Männistö I, Nelson P, Lois E, and Bye A. International HRA Empirical Study—Phase 3 Report: Results from Comparing HRA Method Predictions to Simulator Data on LOFW Scenarios. HWR-951, OECD Halden Reactor Project, Halden, Norway, 2011. To be issued as NUREG/IA-0216 Vol. 3.
- [5] Forester J, Lois E, Dang VN, Bye A and Boring RL. Conclusions on Human Reliability Analysis (HRA) Methods from the International HRA Empirical Study. Proc. PSAM11&ESREL2012, June 25-29, 2012, Helsinki, Finland.
- [6] Kolaczowski A, Forester J, Lois E, Cooper S. Good Practices for Implementing Human Reliability Analysis (HRA). NUREG-1792, U.S. Nuclear Regulatory Commission, Washington DC, USA, 2005.
- [7] Broberg H, Hildebrandt M, Nowell R. Results from the 2010 HRA Data Collection at a US PWR Training Simulator. HWR-981, OECD Halden Reactor Project, Halden, Norway, 2011.
- [8] Kirwan B, Umbers I, Edmunds J, Gibson H. Quantifying the Unimaginable – the Case for Human Performance Limiting Values. Proc. 8th Int. Conf. on Probabilistic Safety Assessment and Management (PSAM9), 18-23 May 2008, Hong Kong, China.
- [9] Skjerve AB, Bye A. Towards guidance for assessment of "training" as a performance-shaping factor in human reliability analysis. Proc. PSAM11&ESREL2012, June 25-29, 2012, Helsinki, Finland.