Extracting Human Reliability Information from Data Collected at Different Simulators: A Feasibility Test on Real Data

Salvatore Massaiu
OECD Halden Reactor Project, Halden, Norway

Abstract: This paper presents a feasibility test on extracting HRA-relevant information form data collected at different plant/simulators. Newly proposed methodologies for HRA simulator-data collection are trying to overcome the aggregation and generalization problems that stranded previous attempts at the creation of HRA data banks. Common to the different methodologies is that they insist on the need to precisely characterize the performance conditions. The difference is on the type of information they aim to collect, some focus on failure probabilities, other on situational influences on the performance of the join human-machine system. This paper investigates whether further information of use for HRA, like timing and performance variability, could be added to and extracted from such databases. The data used in this test derive from three simulator experiments. Two experiments were conducted at the Halden Human-Machine Laboratory, while the third at a training simulator at a U.S. nuclear power plant. All together 23 crews of licensed operators from four plants in two countries participated, and ten emergency scenarios were run. The test considers the data as a subset from a larger database, selected by a HRA user as relevant for the target application. The test shows that it is possible to extract three types of HRA-relevant data from records obtained at different simulators and plants: mean times of actions and diagnoses, response-time variability for critical actions, and standardized margins-to-failure information. This paper shows the feasibility of including and re-using traditional types of HRA data in newly proposed approaches to HRA database construction.

Keywords: HRA, Full-scale simulation, HRA Data, HRA database.

1. INTRODUCTION

During the eighties several simulator experiments were organized in order to provide empirical data for development and validation of Human Reliability Analysis (HRA) methods [1]. However, large-scale efforts directed at gathering HRA data were limited and mainly time reliability correlation data were acquired [2]. In the nineties the HRA discipline underwent a radical critique and new methods and practices were advanced, but without new systematic evaluations against empirical data. Efforts to build HRA databases were made at that time (e.g., the NUCLARR and HERA databases sponsored by the U.S. Nuclear Regulatory Commission) but are now discontinued. The result is that nowadays no publicly accessible data banks of HRA data exist.

There is general agreement on the need for collecting and sharing human performance data for probabilistic safety analysis (PSA) but no standardized HRA data collection approach exists. There are conceptual challenges relating to the generalizability of the data from the plant, scenario and task were they are acquired to other plants, and even to other scenarios and tasks at the same plant. There are disagreements on the classification of the observations (e.g., error types, characterization of the context). And there are practical impediments relating to proprietary and sensitivity issues.

These challenges are known and several international cooperative activities are addressing them in a renewed call for collection and exchange of HRA data. The Nuclear Energy Agency/Committee on the Safety of Nuclear Installations/R(2008)9 report [3] proposes a standardized HRA data collection approach for plant training simulators and provides a generic framework for collection and exchange of human performance data. The NRC has developed the Scenario Authoring, Characterization, and Debriefing Application (SACADA), a methodology and a software tool for creating a HRA database of human performance data from plants' routine training simulations [4,5], and a HRA Data Collection project is underway with a participating nuclear power plant (NPP). Electricité de France and the Institute for Energy Technology in Norway/Halden Reactor Project have started work on a classification system for Emergency Operating Systems (EOS), for allowing exchange and re-use of HRA data and information collected at different facilities [6].

This paper builds on these efforts by taking a look at real simulator data from different plant/studies and trying out ways of aggregating the data for possible reuse. The paper capitalizes on data from three dedicated HRA experiments that together constitute a limited but workable sample of observations, resembling a selected set of records from a larger HRA database. In this way reuse of data collected at different plants (the main challenge that have historically hindered previous efforts at creating HRA databases) is investigated.

2. TRADITIONAL HRA DATA COLLECTION

As soon as realistic simulations of control room operation became feasible, simulator studies became the natural arena for seeking nominal error probabilities (i.e., the probability that a given action will be performed erroneously when the task is not influenced by plant and situation specific factors). The idea was that from sets of scenarios containing the same human action, and from large samples of operators, it could be meaningful to at least count frequencies of the unsafe instances of the human actions. In this way, the reliability of the 'human component' could be estimated in basically the same manner as conventional reliability does for technical components.

Incorporating simulator data into probability safety analysis (PSA) and HRA in such a direct way proved unfeasible. When quantification was sought at the level of actions modeled in the PSA event trees, such as depressurizing the primary system or isolating a ruptured steam generator, two problems arose. First, it was noted that an extremely high number of sessions and crews (sample size) would be needed to observe these kinds of failures. For instance, Dougherty [7] refers to simulator studies performed in the late 1980's where, out of 1600 simulator opportunities, zero 'significant deviations' were recorded. Second, it was not trivial to define nominal conditions, whether similar actions observed in different scenarios and conditions could be combined for the calculation of the failure rates. Moray [8], who studied human performance for informing HRA, summarized the problem in the following way: "The attempt to find a single number is an attempt to establish a context-free universal fact about human performance. No such thing exists."

The importance of the situation for human performance is recognized by several research directions in the fields of psychology and human factors, such as situated cognition [9], ecological psychology [10, 11] and cognitive system engineering [12,13], to name a few. Vicente, working within the Rasmussen, Goldstein and Hollnagel research tradition developed at the RISØ laboratory in Denmark, has termed the importance of the situation for human performance "context-conditioned variability" [14]. The impact of context-conditioned variability to HRA is that the parallel with component reliability where the context is more or less irrelevant and statistical data are available, cannot be fully maintained for the calculation of the reliability of the human component. HRA data are context laden, and therefore information about the context need to be captured.

One attempt at avoiding the sample and the context-conditioned variability problems in one single move was to collect human error probabilities for lower-level operator's tasks, i.e., tasks having a more defined context and higher failure probabilities (and therefore failure would be observed in smaller samples). This was the strategy followed by Swain and Guttman in THERP [15]: "Our general approach is to divide human behavior in a system into small units, find data (or use expert judgment to obtain appropriate estimates) that fit these subdivisions and then recombine them to derive estimates of error probabilities for the tasks performed in NPPs".

The fact that control room operation is performed by a team and the observation that most errors on small tasks were in fact corrected by other team members, or self corrected when the operators had enough time [16, 17], required the introduction of new concepts: recovery and dependency. Recovery has to consider the possibility that errors on one or several sub-tasks in a sequence could be detected and corrected in time (before an irreversible state is reached). Dependency accounts for the possibility that failing a task might influence the probability of failing another task later in the sequence. Recovery and dependence are in a sense ways of taking into account the team and dynamic aspects of accident operation without explicit models of team cognition and behavior.

Quantifying the building blocks of the tasks represented in the PSA, instead of these directly, could address the sample size problem, and data were provided to validate and extend decompositional methods like THERP. On the other hand, context-conditioned variability issues remained, and impaired the generalizability of the data. The standards of performance that define the failure of the basic actions refer to the given plant technology, the actual procedures, the organization of work, as well as other plant and situation specific aspects. This is an obstacle to direct generalizability of the failure data to other tasks, scenarios, and plants, with the additional drawback that the data might obsolete even at the same plant when the specific features are changed. In general, by collecting error data defined as behavior that diverges from plant-specific standards, specific contexts are imposed on the empirical material so that contextual differences need to be accounted for when re-using the data: for instance, if the procedures' steps are taken as standard of correct performance, data on the number of deviations form a step might be unusable if the step is modified or deleted. Even for error data on smaller bits of operator performance the context needed to be accounted for.

The following points summarize the traditional approaches to HRA data collection:
1. Data are collected on human errors for NPP tasks in nominal contexts.
2. 'Core damage' failures in nominal contexts are rarely observed in training or research simulators, therefore failure data for smaller tasks are more often obtained in the data collections.
3. The results cannot easily be generalized to different plants, scenarios and tasks, since the context of performance for the source data is not sufficiently and systematically described.

The same considerations apply to time reliability data, in fact the most collected type of traditional HRA data [2]. Response times can be collected for both high level tasks and for small units of behavior, and may be used to estimate probability of non-response curves, i.e., probabilities of human errors as a function of time available. The problems of generalizing to different tasks and contexts are similar for time reliability data as for human errors for required tasks.

3. CURRENT APPROACHES TO HRA DATA COLLECTION

Conscious of the limitations of the previous attempts to create HRA data banks, as well as the needs of second-generation HRA methods and emerging applications (e.g. HRA for not-at-power, external events) new approaches are being developed and tested for collecting, storing and exchanging HRA data. The SACADA [4], EOS [6] and the "Methodology for Conducting Simulator Experiments for HRA" developed at the Idaho National Lab [18], although from different theoretical perspectives and aiming at different end uses of the data in the HRA process, have all abandoned the notion of a nominal context. Instead they agree on the need to precisely characterize the performance conditions, in order to overcome the generalizability problems that have jammed the old attempts at the creation of HRA data banks.

This has important implications for the methodology used to generate and code human reliability data. Scenarios and contexts should be representative of the needs for PRA and the different performance conditions that operating crews may encounter. For subsequent use of the data it is therefore necessary to ensure that important information on the conditions of performance accompanies the data in order to prevent its potential aggregation with other data that reflect dissimilar conditions. This requires describing, with specificity, the characteristics of the join human-machine system (e.g. the plant, interfaces, procedures, crews, conduct of operations), and the kinds of cognitive and human performance demands present in the simulated scenarios that are used for data collection. It also requires such characteristics to be described consistently, following an approach that may be standardized, as well as applying associated, accepted models of human performance and accident operation. Characterizing plant, crews, scenarios and contextual conditions systematically is also an important part of understanding the data, not only the means of identifying and selecting data for future use and study.

The SACADA and EOS approaches to HRA data collection employs different ways to solve the limitation of classical HRA data where it comes to generalizing the data. The type of data and end-use considered by the two approaches are also different. SACADA collects success and failure data on tasks performed in simulator training (hence most actions are at a lower level of granularity compared to the Human Failure Events (HFEs) typically modeled in PSA event trees). Failure rates are obtained and re-used by matching tasks on their cognitive demand profile, i.e., a type of cognitive task analysis of the simulated tasks, assuming similar human error probabilities for similar cognitive demand profiles.

The EOS method focuses on reusing information in rich qualitative task analyses. The plant and simulation settings are systematically profiled in terms of the characteristics of the join human-machine system, i.e. the ensemble of crew, interface, and procedures, by describing the system according to standardized categories, e.g., the crew composition, the prescribed way of working, the communication style, the procedures type. The EOS will then help assessing the similarity and differences of the systems (plants) and consequently the relevance of the empirical material to the target application. The EOS approach is more focused on reusing qualitative information for the purposes of second-generation HRA methods. The EOS approach also considers other HRA analyses as data (knowledge) that can be reused, as they contain information and reasoning about possible system failure elaborated by other experts.

In these approaches types of information that are still widely used by HRA are left out. One example is timing information. Another is procedure progression variability. In this paper we concentrate on showing how timing information and performance variability information could also be collected in HRA simulator experiments and reused via data-banks, by benefitting form the context-of-performance characterizations provided, e.g., by the SACADA and EOS methods.

2. TEST DATA

The data analyzed in this paper derive from experiments that were originally arranged for evaluating aspects of HRA practice, including assessing the predictive capabilities of commonly used HRA methods [19], intra-method consistency [20], and for investigating teamwork and procedure following in complex emergency operation [20, 21]. The studies provided useful insights to the specific research questions addressed, but it became apparent that there was no agreed methodology to extend the use of the collected data beyond the specific scopes that motivated their generation, as for instance recording the results in central data bank. This is a typical situation when it comes to the use of research simulators for collecting HRA data. It is in fact not uncommon for individual plants to use the results of simulator observations to support aspects of their plant-specific PSA. Yet, collected data are rarely, if ever, used to verify or validate HRA methods, or to improve the HRA discipline at large.

The data used in this test derive from three simulator experiments. Two experiments were conducted at the Halden Human-Machine Simulator, while the third at a training simulator at a U.S. nuclear power plant. All together 23 crews of licensed operators from four plants in two countries participated, and ten emergency scenarios were run. Table 1 below shows the details of the dataset used, including the experiment name and references, the simulated plant type, the number of crews in the experiment, and the number of design basis scenarios run. All crews used similar versions of the emergency operation guidelines developed by Westinghouse Owner Group. The test imagines that the data are a subset from a larger database, created according to the prescripts of modern approaches to HRA data base creation, and that therefore it was possible for the HRA user to select a sample that matched the target application.

Table 1: Sources of the data used

Experiment name	Ref. #	Simulated plant	# of crews	# of scenarios
PSF/Masking	[19]	Framatome PWR 900 MW	14	4
US training simulator	[20]	Westinghouse PWR	4	3
HRA-2011	[21, 22]	Westinghouse PWR	5	3

4. RESULTS

Three types of information have been extracted from the data set. The first type of information is about mean times for diagnosis and actions. The second type of information is response time distributions for critical tasks, i.e., tasks that impact the plant process, and likely the likelihood of subsequent Human Failure Events in the event sequences. The third type of information is margin-to-failure scores, a measure of human performance variability to normalize different tasks outcomes on a common scale.

4.1. Timing of Actions

Time is a factor of uppermost importance for HRA. Some approaches use time as a surrogate cause of failure and allow the time factor to incorporate the effects of most, if not all, drivers of performance. For these approaches the estimation of the timing of actions in a scenario is essential and it is not limited, as for all other HRAs, to the correct estimation of the maximum allowable times from thermohydraulic calculations (i.e. the times when irreversible state arise and the required operators' actions are not longer useful). Typically, these methods have to differentiate between the time required for the actions to be completed, the time available for formulating the correct diagnosis of the situation, the time available for recovering errors, and the time for eventual delays. All these timings are connected to each other, but in most time-reliability methods the single most important estimation is the diagnosis time, since it is often the case that the HEPs are derived from time response curves for diagnosis of abnormal events in the control room, or outside.

The HRA analysts calculate the timing of action based on the qualitative information obtained for the task analysis. An important source for timing determinations is the estimated entry and progressions in the emergency procedures set. Scenario specific factors also play a role, for instance the combination of events might create cumbersome progressions in the procedures or cause reasons for delaying important actions (e.g. tripping the reactor). HRA data collected at research and training simulator can constitute a valuable source for timing information determinations.

These points can be illustrated by reference to the International Empirical Study, where various HRA analyst where asked to predict the outcome of simulated emergency scenarios [19]. Tables 2 and 3 below summarizes the different timing estimates made for two HFEs in two steam generator tube rupture scenarios, and the assumed impact for task success. The first HFE (1A) is the identification and isolation of the ruptured steam generator in a scenario without added complications (base scenario), given a maximum available time defined by study's assessment group in 20 minutes (from rupture to isolation). The second HFE (1B) is again the identification and isolation of the ruptured steam generator, but in this case in a scenario complicated by a steam line break concurrent with the tube break and by the lack of radiation indications (partly due to the steam line isolation that follows the steam line break and partly to failures in other radiation sensors). The study defined the maximum available time for isolating the ruptured SG in this scenario in 25 minutes.

Table 2: Timing estimates for HFE1A: ruptured steam generator isolation

Team	Method	Delay	Diagnosis	Required	Available	Impact of time to failure
NRC	ASEP-THERP	7		13	20	Time as main HEP driver.
EPRI	CBDT+THERP	7			20	From RX trip, but main error cause is missing the first transition step to E-3. Some time for recovery.
INL	SPAR-H		10		20	Based on 10 minutes to reach E-3. Sufficient time if Rx is tripped manually.
NRC	SPAR-H		8-10	13-15	20	The remaining 5 to 7 minutes to manually trip the reactor are sufficient.
PSI	CESA				20	Limited time for recovery. But failure mainly due to random execution errors.
NRI	DT+ASEP				20	Time not a limiting factor for the HFE.
EDF	MERMOS				20	Lack of urgency main factor for failure. 20 min criterion seen as arbitrary.
Ringhals	HEART			16-18	20	Time from trip to isolation. Time shortage is the main HFE difficulty.
IRSN	PANAME		13-15	18-20	20	Time is tight and the main HEP driver for an easy diagnosis.
VTT	B-THERP		10-12	13-15	20	Time sufficient for HFE success.
NRC	ATHEANA	11		18-26	20	The crews might wait up to 11 min. to trip the reactor and then the time will bee too short. The allowed 20 min. considered as arbitrarily defined.
KAERI	K-HRA	5	11		20	Delay: 5 minutes to trip the reactor. Time is tight and the main driver.
NRI	CREAM				20	Adequate time but with small margin.

Table 3: Timing estimates for HFE1B: ruptured steam generator isolation in a scenario with complications

Team	Method	Delay	Diagnosis	Required	Available	Impact of time to failure
NRC	ASEP-THERP		9-12		25	Time sufficient to reach a transition to the right procedure.
EPRI	CBDT+THERP		18-20		25	Time minor contributor, even if no time for recovery.
INL	SPAR-H	>5		20	25	Time insufficient, HEP = 1.
NRC	SPAR-H				25	Limited time and limited time to recover, but rated nominal.
PSI	CESA				25	Short time but not a determining factor.
NRI	DT+ASEP				25	Short time (and unrealistic), but not main factor.
EDF	MERMOS				25	Lack of urgency as main factor together with short available time.
Ringhals	HEART				25	Shortage of time as second most important factor.
IRSN	PANAME		15-20	20-25	25	Time is short and one of two main factors.
VTT	B-THERP		15	18-25	25	Time available close to time required and other negative PSFs determine a high HEP for the HFE.
NRC	ATHEANA	15	6	18-26	25	No diagnosis in the first 15 minutes (delay) due in part to masking. The crews may simply run out of time to meet the “arbitrary” 25-minute time frame.
KAERI	K-HRA	15			25	Delay: no signs of SGTR for 15 min after Rx trip. Time is the most important factor for HFE failure.
NRI	CREAM				25	Delayed interpretation one of failure types identified.

The table shows that although the analysts received the same information regarding the scenarios, the HFEs, the crews, the simulated plant, and even detailed printouts of plant status parameters at different points in the scenarios, there was variation in the timing estimates. Furthermore, the timing evaluation had in many cases a significant impact on the quantification of the HEPs, and not only for methods that employed time-reliability curves.

4.2. Average Times

Observing simulator sessions is a good HRA practice and would have obviously benefited the analyses. Table 4 below, reporting the timing information observed in the data, shows that many analysts' time estimates were realistic, but some were not and in many cases strongly influenced quantification. For instance, there were no significant delays in tripping the reactor in the base scenario, as by some assumed. Also, it took on average double as long in the complex scenario to enter the SG isolation procedure compared to the base scenario. This information would have resulted in very different HEP estimates for some analyses (e.g. K-HRA and ATHEANA in the base scenario).

Table 4: Time responses in SGTR scenarios

	Base N=14		Complex N=14		U.S. crews N=8		All N=36
SGTR to:	**Mean**	**Max**	**Mean**	**Max**	**Mean**	**Max**	**Mean**
RX trip	02:23	06:29	At SGTR	-	02:11	03:39	**02:19**
E-0	At trip	07:06	At trip	01:11	At trip	-	**At trip**
Stop AFW to ruptured SG	06:44	09:41	13:27	22:42	02:52	04:43	**08:59**
Enter E-3	10:15	16:08	20:46	40:32	12:58	18:12	**15:18**
SG isolated	15:32	21:29	26:54	45:27	21:13	33:42	**21:34**

The table also reports the corresponding timing information regarding U.S. crews that run a base SGTR scenario in their training simulator and other U.S. crews that run a multiple SGTR scenario in the Halden Man Machine Laboratory (HAMMLAB). The times are in line with those observed with Swedish crews, considered the differences in scenario specifics, simulated plants, and emergency procedures details. The table also shows the averages for the entire sample: if these data where provided with standardized context of performance information they could be used as generic timing data (conservative in this case, as over half the scenarios are SGTR with complications) for cases where the analysts do not possess specific plant/scenario data, if comparable in terms of context-of-performance characteristics to target plant.

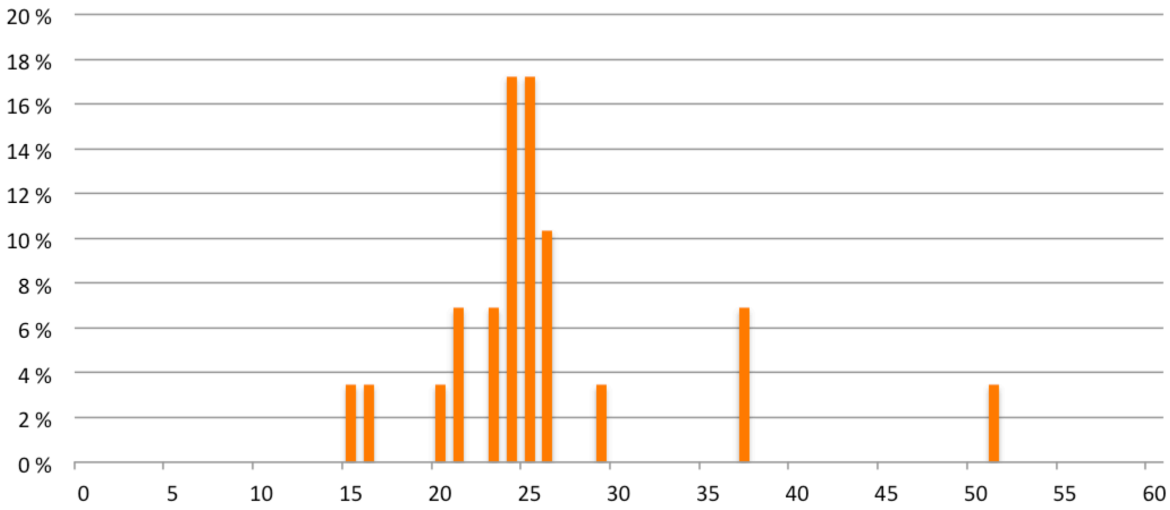
For methods that use time reliability curves, one of the most important information is the time available for diagnosis, as this can be associated with an HEP – once other contextual considerations are made (e.g. type of event, familiarity). In the case of SGTR events, the diagnosis is formally completed by entering the tube rupture isolation procedure (E-3 for Westinghouse plants). The complete sample of 36 SGTR observations across the three studies indicates an average time of about 15 minutes to enter this procedure from the initial cues associated with the tube rupture (i.e. radiation alarms or automatic reactor trip in these simulations). Again, an analyst working on a SGTR scenario for a Westinghouse PWR plant could benefit from access to such empirical information.

4.3 Response-Time Variability

A second type of data important in many HRA applications is response-time variability. For a given task in a given scenario and for a given plant, an extensive data collection could result in empirically informed non-response time distributions. If a data sample of the same size collected at different plants was available the same could not be done easily, as different plant PRA criteria, procedural criteria and other plant/simulation specific aspects would likely strongly reduce the relevant sample. However, the data from different plants can provide very useful information to HRA analyses at other plants.

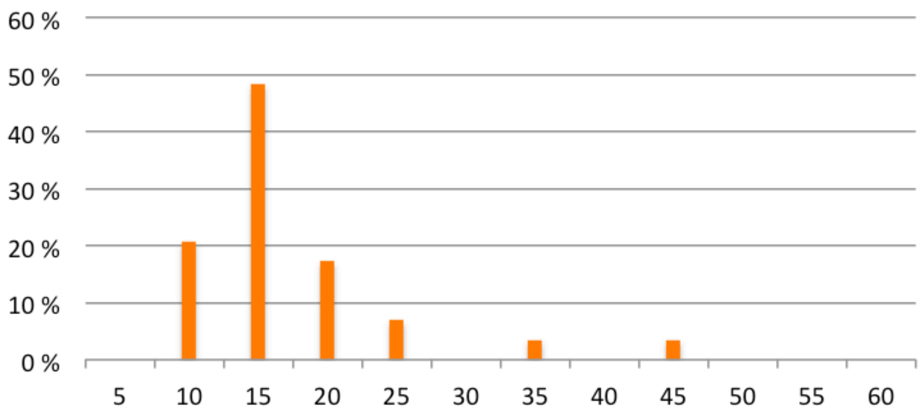
Figure 1 below is based on data from the three studies collected on total loss of feedwater (TLFW) scenarios. In this event the crews have to trip the reactor as early as possible after feedwater flow to the steam generators is lost, in order to keep secondary water inventory longer in the steam generators, and hence increase the time available before the occurrence of fuel damage. This action is critical as it has direct consequences on the evolution of the plant process and likely impact the failure probabilities of the HFEs in the event sequence. Figure 1 shows the crew variability for this action and that significant delays are possible. In the specific case, the crew that tripped the reactor 51 seconds after total LOFW had less than 30 minutes available to establish Feed and Bleed before core damage, compared to about 90 minutes for the fastest crews.

Figure 1: Seconds from TLFW to reactor trip (percent of crews) - N=29



Another example, from the same total loss of feedwater scenarios is depicted in figure 2. After the crews trip the reactor, they will be directed to the relevant emergency guidelines for the event. One of the steps in the heat sink procedure will instruct the crews to stop the reactor cooling pumps (RCPs). Also this action has consequences for the plant process and the available time to core damage: the earlier the RCPs are stopped the longer the time available. Figure 2 shows that two of 29 crews significantly delayed this action, thereby consuming available time to avoid core damage.

Figure 2: Minutes from reactor trip to RCPs stop (percent of crews) - N=29



Response-time distributions of the type presented here cannot obviously be used to uncritically derive HEPs or to obtain non-response time curves. They are nonetheless a valid source of empirical information, and the value of the data is not limited to analysts that do not have access to direct observation of operator performance for their analyses. For instance, the presence in the data of significant delays and their relative frequency cannot be disregarded by the fact that similar delays have not be observed at the analyzed plant. There is in fact no guarantee that the performance conditions and task demands the crews have been trained for and evaluated on in their simulator sessions were fully representative of the conditions encompassed by a PSA scenario, and of those that could arise in a real accident.

A critical aspect is therefore that the database allows access to the individual data records in order to determine the circumstances surrounding the delays (e.g. plant conditions, procedures, crew composition, training) thus allowing the user to assess the relevance to the analyzed context. Likewise, the user should be allowed full sorting and filtering access of the data, in order to obtain the most relevant sub-sample for further analysis. In this respect, coding the observations according to the EOS and SACADA classifications, would not only provide the data required for these methodologies, but also facilitate the re-use of timing information in new contexts.

4.4. Margins to Failure

A proposal for generalizing HRA data from one plant/scenario to different plants/scenarios is to score human performance on different tasks on a common scale. The notion of success/failure is already a normalized measure of human performance. However, such a binary characterization limits the ability to draw useful insights from even substantial amounts of observed trials. For example, it is possible that all crews in a simulation session complete the needed actions to mitigate a design basis event. This would indicate only successes on the HFEs and no conclusion regarding the relative difficulty of the tasks could be obtained. However, performance quality variability typically exists, and the information indicating that some crews succeed with substantial 'margin' to spare while others did not, is lost by the dichotomous measure.

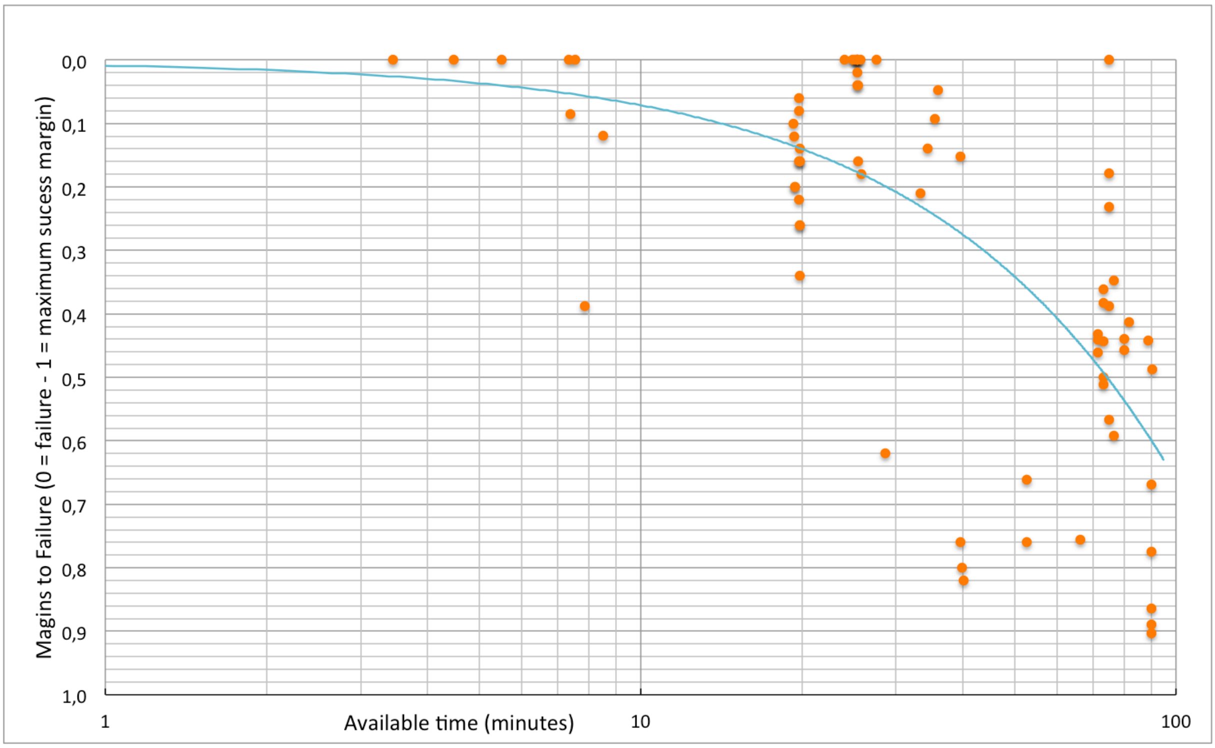
For this reason Hallbert and al. [18] propose a continuous measure to characterize performance that capture issues such as available or remaining margin to failure and variability in performance amongst crews. The approach is called the "limit state concept" and includes a data analysis technique that normalizes the raw performance measurement in terms of the limit state. "This means that the raw performance measures are adjusted using their relationship to the limit state as a notionally common scale" [18, p. 40].

Figure 3 below, show the application of this data analysis approach to performance data from the three studies considered in this paper. The crew performances of four HFEs in three different design basis events have been analyzed according to the limit state concept. The limit states (i.e. success criteria for the HFEs) are defined, with some specificity and uncertainty, by the PRA of the design basis accidents. One HFE was "establish Feed and Bleed before core damage" in total loss of feedwater events. The second HFE was "isolate the ruptured steam generator before it overfills" in SGTR scenarios. The third and forth HFEs are in a RCP Seal LOCA scenario (one requiring to stop the RCPs and the other to start the positive displacement pumps before irreversible damage states).

The 73 crew performance observations of the four HFEs are scored on a scale ranging from 0 margins (i.e., failure) to 1 (i.e., maximum success margin possible), based on the timing of their actions in the scenario, the status of plant parameters at the time of actions, and the limit states (success criteria) provided by the PSA model. For instance, in the case of the HFE "Feed and Bleed", the success margin would represent the percentage of remaining available time before core damage at the time Feed and Bleed was established, based on the time the crew tripped the reactor. In this scale 0 would correspond to the action executed after the core began to melt, and 1 the action was executed at the same time as reactor trip (100 percent time available). A curve relating time form reactor trip to core damage (obtained form the PSA) is adopted for calculating the available time. Similar calculations are made for the other HFEs.

Figure 3 shows the margin to failure scores obtained by the crews relative to the time available for the four HFEs. The interpolated polynomial curve resembles typical "probability of non-response" curves for the considered time region, although the upper and lower bounds would likely be larger here, as failures and near failures were seen also with extensive time available. It should be stressed, however, that the Y-axis here represents the margin to failure scores on a linear scale and not human error probabilities on a logarithmic scale.

Figure 3: Margins to failure by available time - N=73



This approach permits insights to be drawn regarding three properties related to performance reliability:

1. Whether the actions meet the success criteria for the defined HFE;
2. The amount of margin available between task performance and the limit state for the task as defined by the HFE;
3. Variability among crews in performing the task(s).

The measure would also permits estimation of measures of central tendency and calculating statistics that can be used in reliability analysis.

5. CONCLUSIONS

Several international cooperative activities are addressing the need of collecting and sharing human performance data for probabilistic safety analysis (PSA). New methodologies for data collection, classification and storage have been proposed, and some are being tested. Common to the different methodologies is that they insist on the need to precisely characterize the performance conditions, in order to overcome the aggregation and generalization problems that stranded previous attempts at the creation of HRA data banks.

The paper follows on such developments and assumes that by means of these classification systems a HRA user has extracted a sample of relevant data from the database. The test investigates whether a HRA user could extract further relevant information, beyond the primary uses assumed by the methodologies. The data used in this trial are real simulator observations from three dedicated HRA experiments performed with three different simulators, and with crews from four different plants in U.S. and Europe.

The first type of information that is extracted is about timing of actions. The paper shows that average response times from the entire sample could be used as generic timing data, provided the data match the target application. Given the nature of the data, with over half the records from design basis scenarios with extra complications, such mean values could be assumed to be conservative estimates.

The second type of information is response time distributions for critical tasks, i.e., tasks that impact the plant process, and likely the likelihood of subsequent HFEs in the event sequences. This information does not allow direct derivation of HEPs or non-response time curves. Rather, it is a source of empirical information about crew variability and of the possibility and relative frequency of significant delays that need to be taken into account in the qualitative and quantitative HFE modeling. This information cannot be disregarded even by analysts that have collected performance observations on-site, since there is no guarantee that the performance conditions and task demands the crew have been trained for and evaluated on in their simulator sessions were fully representative of the conditions encompassed by a PSA scenario, and of those that could arise in a real accident.

The third type of information extracted applies a methodology for scoring human performance on different tasks on a common scale, on the lines of the "limit state" concept for generalizing HRA data. This approach permits insights to be drawn regarding three properties related to performance reliability: (1) whether the actions meet the success criteria for the defined HFE; (2) the amount of margin available between the task performance and the failure criteria for the task as defined by the HFE; and (3) variability among crews in performing the task(s). The measure would also permits estimation of measures of central tendency and calculating statistics that can be used in reliability analysis.

A critical assumption for this test is that the database allows access to the individual data records in order to determine the circumstances surrounding performance (e.g. event description, plant conditions, procedures, crew composition, training) thus allowing the user to assess the relevance to the analyzed context. Likewise, the user should be allowed full sorting and filtering access of the data, in order to obtain the most relevant sub-sample for analysis.

This paper shows that timing and standardized performance data coded according to the new classification systems for HRA simulator data collection could not only provide the data required for these methodologies, but also facilitate the extraction and re-use of other types of information of critical importance to HRA, like timing, margins-to-failure, and crew variability information. This data can be naturally collected at research simulators, while raw data collected at training simulators could be later analyzed, formatted, and stored in a publicly available data bank.

References

[1] A. J. Spurgin, "*Human Reliability Assessment Theory and Practice*," CRC Press, 2009.

[2] D. I. Gertman and H. S. Blackman, "*Human Reliability and Safety Analysis Data Handbook*," John Wiley & Sons, 1993.

[3] Nuclear Energy Agency/Committee On The Safety Of Nuclear Installations. "*HRA Data and Recommended Actions to Support the Collection and Exchange of HRA Data*," NEA, 2008, Paris.

[4] Y. J Chang and E. Lois, "*Overview of the NRC's HRA Data Program and Current Activities*", in "Proceedings of the 11th International Probabilistic Safety Assessment and Management Conference & The Annual European Safety and Reliability Conference", 25–29 June 2012, Finland.

[5] L. Criscione, S.-H. Shen, R. Nowell, R. Egli, Y. J. Chang and T. Koonce. "*Overview of Licensed Operator Simulator Training Data and Use for HRA*", in Proceedings of the 11th International Probabilistic Safety Assessment and Management Conference & The Annual European Safety and Reliability Conference. 25–29 June 2012, Finland.

[6] S. Massaiu, P. O. Braarud, and P. Le Bot. "*Including organizational and teamwork factors in HRA: the EOS approach*", in proceedings of the Enlarged Halden Program Group, 2013, Norway.

[7] E. M. Dougherty, "*Human reliability analysis--where shouldst thou turn?*", Reliability Engineering & System Safety, 29, pp. 283–299, (1990).
[8] N. Moray. "*Dougherty's dilemma and the one-sidedness of human reliability analysis (HRA)*", Reliability Engineering and System Safety, 29, pp. 337-344, (1990).
[9] P. Robbins, and M. Aydede. "*The Cambridge handbook of situated cognition*," Cambridge University Press, 2008.
[10] R. Shaw and J. Bransford, "*Perceiving, Acting, and Knowing: Toward an Ecological Psychology*," Lawrence Erlbaum, 1977.
[11] J. J. Gibson. "*The Ecological Approach To Visual Perception*," Psychology Press, 1986.
[12] E. Hollnagel and D. D. Woods. "*Cognitive systems engineering: New wine in new bottles*", International Journal of Human-Computer Studies, 51(2), pp. 339–356, (1983).
[13] J. Rasmussen, A. M. Pejtersen and L. P. Goodstein, "*Cognitive systems engineering*," Wiley, 1994.
[14] K. J. Vicente, "*Cognitive work analysis: toward safe, productive, and healthy computer-based work*," Routledge, 1999.
[15] A. D. Swain and H. E. Guttman, "*Handbook of human reliability analysis with emphasis on nuclear power plant applications*," Sandia National Labs-NUREG CR-1278, Nuclear Regulatory Commission, 1983, Washington DC.
[16] D. D. Woods. "*Some results on operator performance in emergency events*", In D. Whitfiled (Ed.), "*Ergonomic Problems in Process Operations*," pp. 21–32, Institute of Chemical Engineers Symposium, 1984, Birmingham.
[17] L. Norros and P. Sammatti, "*Nuclear power plant operator errors during simulator training*," VTT Technical Research Centre of Finland Helsinki, 1986, Finland.
[18] B. Hallbert, T. Morgan, J. Hugo and J. Oxstrand. "*Simulator Data Sharing for HRA: Tenets & Methodology for Conducting Simulator Experiments for HRA - Data Generation and Analysis Focused on the Limit-State Concept*," NRC Technical report INL/EXT-12-26327 (Draft), Nuclear Regulatory Agency, 2012, Washington DC.
[19] J. Forester, V. N. Dang, A. Bye, E. Lois, S. Massaiu, H. Broberg, P. Ø. Braarud, R. Boring, I. Männistö, H. Liao, J. Julius, G. Parry and P. Nelson, "*The International HRA Empirical Study – Final Report – Lessons Learned from Comparing HRA Methods Predictions to HAMMLAB Simulator Data*", HPR-373, OECD Halden Reactor Project, 2013, Norway.
[20] J. Marble, H. Liao, M. Presley, J. Forester, A. Bye, V. Dang and E. Lois. "*Results and Insights Derived from the Intra-Method Comparisons of the US HRA Empirical Study*", in proceeding of 11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference 2012 (PSAM11 ESREL 2012), (2012).
[21] S. Massaiu, and L. Holmgren, "*Complex diagnosis of multiple steam generators leaks: Preliminary results from the HRA-2011 experiment*", in proceedings of the Enlarged Halden Program Group, 2013, Norway.
[22] S. Massaiu, and L. Holmgren, "*An Empirical Investigation of Team Decision-Making with Emergency Procedures*", paper presented at the American Nuclear Society's 2013 Risk Management Embedded Topical Meeting: Risk Management for Complex Socio-technical Systems, (2013).