Constraints in Designing Simulator Scenarios and Identifying Human Failure Events for Testing HRA Methods
Abstract

A diversity of Human Reliability Analysis (HRA) methods are applied in Probabilistic Risk Assessments (PRAs) The Halden Reactor Project has, together with its member organizations, initiated a project to address a desire to test these methods against empirical evidence. In the initial phase of the project, specific accident scenarios have been chosen and human failure events (HFEs) have been identified. As part of these first trials, we are learning about some of the constraints, issues, and characteristics of the scenario designs that have to be dealt with. For instance, scenarios should be PRA/HRA relevant, sufficiently challenging to be useful tests of the HRA methods and yet be plausible, and also be feasible in a simulator setting. We are also learning about issues on how to define the HFEs of interest to ensure usefulness of the results to PRA/HRA. For instance, HFEs must be measurable and defined so that success of the action and failure of the action are clearly distinguishable. Process-based success criteria for individual HFEs other measures might be needed. This paper will discuss lessons learned about scenario design and the defining of HFEs to be able to progress toward testing HRA methods using simulations of accident scenarios.
Constraints in Designing Simulator Scenarios and Identifying Human Failure Events for Testing HRA Methods

Helena Broberg\textsuperscript{1}, Alan M. Kolaczkowski\textsuperscript{2}
\textsuperscript{1}OECD Halden Reactor Project, Halden, Norway
helena@hrp.no
\textsuperscript{2}Science Applications International Corporation, Lakewood, Colorado, USA
ALAN.M.KOLACZKOWSKI@saic.com

Abstract—A diversity of different Human Reliability Analysis (HRA) methods are applied in Probabilistic Risk Assessments (PRAs). The Halden Reactor Project has, together with its member organizations, initiated a project to address a desire to test these methods against empirical evidence. In the initial phase of the project, specific accident scenarios have been chosen and human failure events (HFEs) have been identified. As part of these first trials, we are learning about some of the constraints, issues, and characteristics of the scenario designs that have to be dealt with. For instance, scenarios should be PRA/HRA relevant, sufficiently challenging to be useful tests of the HRA methods and yet be plausible, and also be feasible in a simulator setting. We are also learning about issues on how to define the HFEs of interest to ensure usefulness of the results to PRA/HRA. For instance, HFEs must be measurable and defined so that success of the action and failure of the action are clearly distinguishable. Process-based success criteria can for example reflect on several HFEs, and to define success criteria for individual HFEs other measures might be needed. This paper will discuss lessons learned about scenario design and the defining of HFEs to be able to progress toward testing HRA methods using simulations of accident scenarios.

I. INTRODUCTION

A diversity of Human Reliability Analysis (HRA) methods are currently applied in Probabilistic Risk Assessments (PRAs). The Halden Reactor Project has together with its member organizations, and with particular support from the U.S. Nuclear Regulatory Commission (NRC)\textsuperscript{1}, initiated a project to address a desire to test many of these methods against empirical evidence. The focus of this empirical study of HRA methods is to compare outcomes predicted in HRA analyses with performance and it’s driving factors from simulator runs in the Halden man machine laboratory. This project is viewed as part of an overall goal to examine the validity of HRA method predictions. Lack of sufficient empirical validation remains a shortcoming of the current state of the art in HRA as applied to nuclear power plant events of interest. Having strong evidence that HRA methods provide predictions of human performance that actually correspond to observations of performance in real situations (based on past experience or in this case, from simulated events of interest in PRAs), will add to the validity of these predictions thus providing more confidence in HRA results when used by decision makers.

In the initial phase of the project, as a pilot exercise to design and improve further implementation of the project, two accident scenarios have been chosen and related human failure events (HFEs) have been identified. An international cadre of HRA teams using different methods, are predicting certain aspects of crew performance. This includes a summary of the most influencing factors on crew behavior, a qualitative assessment of difficulties associated with the actions of interest, and estimates of human error probabilities (HEPs) for the identified HFEs. The overall methodology for this initial assessment study, and HRA in the context of HRA testing with empirical data, are addressed in companion papers [1][2].

While this project is only in its initial phase, observations can already be made about the considerations that need to be addressed to be able to design meaningful simulated scenarios and to address corresponding HFEs worthy of being useful toward validation of HRA methods. This paper addresses these early observations.

II. BACKGROUND

The scenarios used for the initial phase of the HRA empirical test project were originally designed for a different HRA study: PSF/masking. The goal of the PSF/masking study is to gather information about how performance shaping factors (PSFs), and in particular scenarios that have characteristics that mask (i.e., hide) other important failures that require operator attention, affect performance in HRA relevant tasks. In late 2006, 14 crews of licensed operators from a nuclear plant with pressurized water reactors, participated in a series of scenarios in the Halden man machine laboratory. While not specifically designed for the HRA empirical test project, two of the scenarios from the PSF/masking study were judged suitable for the pilot phase of the HRA empirical test study. These include a classic steam generator tube rupture event and a steamline break with a subsequent steam generator tube rupture. In the second scenario, the plant response is initially driven by the steamline break, and it takes time for the cues of the subsequent steam

\textsuperscript{1}The opinions expressed in this paper are those of the authors and not of the U.S. Nuclear Regulatory Commission.
generator tube rupture to be evident so that the operators detect
and respond to the tube rupture. Using these two scenarios for
the pilot phase of the HRA empirical test project has led to
some lessons learned for designing simulated PRA relevant
scenarios and identifying potentially important HFES for the
purposes of examining (via the simulations) and predicting (via
the use of HRA methods) operator performance.

When conducting simulator studies there are different
approaches to scenario design, depending on the purpose of the
study. One approach, commonly used in usability testing, is to
design scenarios to encompass specific features of a human
system interface or software. The scenarios include typical
tasks that the user would perform in the course of his or her work
activities. With this approach, scenarios are treated discretely;
users can perform a series of possibly unrelated tasks.

Another approach is to emphasize the whole experience by
crafting realistic scenarios and let them progress as naturally as
they can. This way you can also gain insights about the overall
experience, and not just specific features. There are many
advantages with this approach: work processes and each
situation will be more realistic when scenarios are allowed to
progress without interruption. You may also observe things that
may evolve over time, and get a better understanding for
decision making and other processes in the control room. Even
though the goal is to let the scenario develop as naturally as
possible, manipulations to direct the crews towards certain
actions may be needed, in addition to the initiating event. As a
simulated scenario designer, you have a lower degree of control
of how the scenario will evolve in practice, and thus scenarios
may progress in unforeseen ways as influenced by the
operators. If the actions of interest take place late in the
scenario, in specific parts of the event tree, it may be difficult to
compare the results of multiple simulated events. The timing of
these scenarios are also difficult, which can be problematic
when you only have a certain amount of time available to run
the simulation. Nevertheless, given the interest of examining the
overall operator response to the chosen steam generator tube
rupture events, including multiple HFES that might be typically
modeled in a PRA for such an accident, scenarios designed with
this second approach has been used for the pilot phase of the
empirical study of HRA methods.

III. OBSERVATIONS REGARDING SCENARIO DESIGN AND
DEFINING HFES OF INTEREST

Our experience thus far with the empirical study of HRA
methods, as well as with other simulations conducted at the
Halden Reactor Project, suggests that the following are important
elements of designing scenarios and selecting HFES for HRA study purposes.

- Identify What We Want to Test
- Identify Risk Significant Human Actions in PRA
- Sort Out Actions that Cannot be Simulated
- Define HFES
- Add Manipulations
- Choose or Create Tools to Collect Empirical Data and
  Record HRA Predictions

A. Identify What We Want to Test

Scenarios have to be designed to allow for examination of
what it is we want to test or otherwise gain information about.
For instance, the main objective of the PSF/masking project is
to investigate relationships between PSFs and measurable
performance of HRA relevant tasks particularly under
conditions involving masking; i.e., when one problem/task is
partly hidden from the operators by other problems or other
information. To be able to find out about how PSFs influence
performance we first need to identify HRA relevant tasks and
then find ways to manipulate the PSFs in large part through the
scenario design including masking effects.

The objective of the empirical study of HRA methods is to
come the predictions from different HRA methods (such as
the failure rate and the PSFs and plant conditions having the
most influence on the predicted failure rate) against similar
measures from empirical data in the form of observed operator
performance in simulated scenarios. Further, results among the
different HRA methods are also to be compared. To do this we
need scenarios with HRA relevant tasks where multiple PSFs
could potentially influence operator performance during the
conduct of the scenarios and those most important could be
observed or otherwise learned from subsequent techniques such
as post-scenario operator interviews. Additionally, if one wants
to truly test the HEPs from the HRA methods, the scenarios
must be sufficiently challenging to likely induce observable
failures of interest. And yet, this must be done without
designing scenarios that are too unrealistic in order to induce
such failures.

B. Identify Risk Significant Human Actions in PRA

To create scenarios relevant for PRA/HRA one must
correspondingly identify human actions that are of interest and
potentially risk significant. The scenarios need to then be
designed such that the identified human actions will be a part of
the simulated scenario(s) with an opportunity for the related
HFES to be manifested in the operators’ performance. As was
done in this first phase of the empirical study of HRA methods,
the actions and related HFES of interest will typically come
from existing PRA/HRA information. This information is
likely to be in the form of the modeled human events as
depicted in event trees and fault trees, or from the PRA solution
cut sets or sequence results expressions.

C. Sort Out Actions that Cannot be Simulated

Sorting out actions that cannot be simulated involves one of
the major constraints of creating scenarios to be simulated for
studying events of interest to HRA. The scenarios must be
feasible in a simulator setting. From the cut set list of scenarios
with high frequency we must abandon those that take too long
(e.g., actions involving the change to recirculation cooling) and
actions that are not performed in the control room (e.g., align
city water locally).

Further, the simulator itself might also have some limitation
of what can be simulated. In this pilot empirical study of HRA
methods we used operators from one plant and the simulator of
another plant, and therefore had to disregard all scenarios that
included a lot of actions on equipment that was different in the
plants. Our list of viable actions to be simulated became much shorter because of this concern.

Hence, practical considerations of what can be simulated, how long the simulations may take, and how accurately events can be simulated all play a role in designing scenarios to be used in such HRA studies.

D. Define HFEs

Our limited experience thus far indicates that one of the most important aspects of defining the HFEs of interest is to as precisely as possible, define what is meant by “failure” of the human action. These failure definitions need to relate to that used in the PRA/HRA but also be observable in the simulated scenario. For instance, failure for some actions might result in a near immediate plant state change such as the onset of core damage will then occur, or the plant pressure will then rise uncontrollably. Some failure definitions, however, may not have such a near immediate and clearly undesired plant state change, but instead have a lesser effect such as simply delaying the overall timing of the operators’ response or affect the evolution of the scenario (e.g., system pressure ends up at a level that is not preferred) but not necessarily the final outcome of interest (e.g., core damage will occur). Yet others may, because of the way the HFEs are defined, involve multiple actions or steps to be carried out and yet if one of the steps is not performed correctly, the overall human action response may still be successful from the PRA point of view (hence, is failure to perform that one step really a failure?). Hence, defining “failure” is not necessarily straightforward.

Yet another set of factors may have to be considered in defining the HFE. This is with regard to expectations of the HRA method to be applied. For instance, many methods use a time reliability function to estimate the diagnosis failure probability portion of the HEP. Use of this function requires that there be a definitive time established for when the desired action must be performed (typically referred to as the allowable time). Then on the basis of the time it is estimated to take to actually implement the action, the analyst can back out a time by which the diagnosis must occur. Use of the time reliability function for that diagnosis time yields the diagnosis failure probability. To use such a method, the HFE must also be defined in terms of time, whether directly or as can be inferred from plant thermal-hydraulic expectations for the scenario.

In our pilot empirical study of HRA methods, we encountered this difficulty of defining the HFEs. In the case of a steam generator tube rupture event, there are specific categories of actions expected to be taken as directed by the procedures and training. These include, for instance, recognition and isolation of the affected steam generator, forced cooldown of the reactor coolant system, depressurization of the system, and termination of safety injection when appropriate. These actions collectively stop the leakage past the tubes, significantly decrease the radiation being released to the secondary side of the plant, and allow for control of the plant to eventually achieve safe shutdown. Failure of any or even all of these actions does not necessarily cause core damage. On the other hand, it is highly desirable that these actions be conducted expeditiously in order to stop the primary-to-secondary leakage and decrease the radiation release. In such a case, what does it mean to fail to isolate the steam generator or to fail to cooldown the system? Should these actions be done by a certain time? If so, how does the operator know what those time expectations are when they are not explicitly provided in the procedures or in training? In this case, and in light of the fact that many HRA methods need to know a time by which the action needs to be taken as discussed above concerning time reliability curves, we decided to use somewhat artificial definitions of failure of these actions based on very loosely defined training expectations. These definitions imposed a time by which the actions should be taken, as well as what precise steps had to be taken to implement the desired action (e.g., close main steam line isolation valves as part of the overall isolation action). We could then observe whether or not the actions were performed in the desired times and if all the critical steps were indeed carried out.

The point here is that identifying the human actions to be addressed and particularly defining the corresponding HFEs, is not always as straightforward as it may seem. It requires considerable thought to ensure the intent of the failures from a PRA perspective remains intact, and yet the failures can be observed and measured as objectively as possible.

In the pilot empirical study of HRA methods, HFEs were defined after the simulator scenarios were run, since these scenarios were originally designed for another study. We believe that the definitions of the HFEs should be written earlier in the process. When deciding on how to collect data, the HFEs should already be defined to ensure that they can be measured in a simulator setting.

E. Add Manipulations

In order to accomplish the desired goals of a study, it may be necessary to manipulate the scenarios in ways that are necessary to be able to predict, observe, and measure the performance effects of interest. For example, in the PSF/masking study used for the pilot phase of the empirical study of HRA methods, the goal was to observe the impact of PSFs and especially masking on performance, to investigate if and how this impact should be accounted for in HRA. To be able to observe the effects of masking, we needed to have masking in the scenarios be present and not present so that performance differences due to the masking could be readily measured. Hence, and as done for this work, two versions of the overall scenario type (in this case, a steam generator tube rupture) were designed and simulated. We labeled the purposely manipulated case involving a masking effect as the “complex” scenario and when no manipulations (i.e., masking) was included, we called the scenario the “base” scenario. For the pilot study, the base steam generator tube rupture scenario was an uncomplicated tube rupture with all the classic indications of such a rupture. The complex steam generator tube rupture involved a steam line breach nearly immediately followed by a tube rupture. The steam line breach drives the initial plant response thereby masking the signs of a tube rupture and the automatic isolation of the steam line causes most of the secondary radiation indications to remain normal – removing a classic sign of a tube rupture (i.e., high secondary radiation). The one remaining radiation measurement from steam generator sampling was failed, leaving the operators to diagnose
the tube rupture on other symptoms such as steam generator level and reactor coolant system pressure.

As part of adding purposeful manipulations in order to meet the goals of the study, a balance should be maintained between having the scenarios sufficiently challenging to likely cause differences in operator performance or induce the potential to observe the failures of interest, and yet not so contrived so that the resulting scenario is unrealistic or absurd; i.e., the scenario must be plausible. Put another way, the scenarios must be reasonably sensitive to capturing differences in human performance, but they should be sufficiently realistic that there is a logical success-path and yet allowing for a possibility to fail.

Note that depending on the desires of the study, if observing failures is a requirement of the study needs, one has to make the scenarios more difficult. In achieving some level of plausibility for the scenario, the scenario should not be so complicated that its likelihood of occurrence is an extremely low frequency. Besides that, the scenario should not be so difficult that it is likely that most or all of the crews will simply give up, stating that the scenario makes no sense or provides no means to succeed. This balance is an important part of ensuring that the scenario design, and particularly any additional manipulations of the scenarios, do not become unrealistic or prevent useful operator performance information to be gathered because the crews may give up.

This balance was achieved in the base and complex steam generator tube rupture scenarios used in the pilot empirical study of HRA methods. To investigate the nature of how and when crews fail, and succeed, scenarios that have sufficient difficulty so that some crews will fail and some succeed will provide useful data. As important as creating scenarios so difficult that failures can be observed, it is desired that not all crews fail on a specific task. One advantage of simulator studies, compared with for example event reports, is that success mechanisms can be studies as well as failure mechanisms. When designing the scenarios used in the pilot empirical study of HRA methods, the goal was to observe both failures and successes. Through the use of both design basis accidents and more challenging scenarios, we were able to meet this goal and did not have any crews give up on the scenarios including the complex scenario. Previous Halden studies relating to HRA have used this model [3].

When taking a first look at the data from the simulator experiment we see that some crews that succeed in the base scenario fail on several tasks in the complex scenario. This gives us the possibility to investigate what PSFs lead to a failure, and what lead to a success. Using both base and complex scenarios, we believe, can give insights as well in possible relationships among PSFs.

F. Choose or Create Tools to Collect Empirical Data and Record HRA Predictions

Tools to collect data should be based on the HFE definitions and on the PSFs that we wish to study. This will ensure, before the simulator runs, that we can measure performance and PSFs. For simulator runs with the purpose to test HRA methods we must also make sure that we collect data that can be compared with the predictions. This includes describing both what happened and why it happened.

Most of the data for examining crew performance was collected during the simulator runs. We logged selected process data and recorded audio and video. In addition, a scheme for evaluating performance, developed in Halden, was used: the operator performance assessment system, OPAS [4]. In OPAS important actions are predefined and during the simulator run we collect time stamps for these actions. For this experiment, OPAS was used a bit differently than its original form. Instead of counting only positive scores, we timed several types of actions (e.g., procedure transfers), and also gave a performance score for each main task in the scenario, corresponding to the HFEs. Our experience is that both OPAS and logs are important tools to get an overview of crew performance and how the scenario developed.

During the simulator runs, data on PSFs was also collected. Below is a table of PSFs evaluated during the experiment, and how the data was collected. These PSFs are a selection from PSFs defined in NUREG-1792 [5]. Some PSFs, such as Environment and Available Staffing, are not particularly relevant for in-control room actions, and were therefore not evaluated in this simulator study. Other PSFs were estimated by an expert or by investigating the operators’ background, and not collected during the experiment.

<table>
<thead>
<tr>
<th>Performance Shaping Factor</th>
<th>Data Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training and Experience</td>
<td>Interview</td>
</tr>
<tr>
<td>Procedures and Administrative Controls</td>
<td>Interview Operator rating</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>Interview Operator rating</td>
</tr>
<tr>
<td>Complexity</td>
<td>Interview Observer rating Operator rating</td>
</tr>
<tr>
<td>Workload/Time</td>
<td>Interview rating</td>
</tr>
<tr>
<td>Pressure/Stress</td>
<td>Interview rating</td>
</tr>
<tr>
<td>Team/Crew dynamics</td>
<td>Interview rating</td>
</tr>
<tr>
<td>Human-System Interface</td>
<td>Interview rating</td>
</tr>
<tr>
<td>Communication</td>
<td>Interview rating</td>
</tr>
</tbody>
</table>

A thorough evaluation of PSFs and failure mechanisms is performed after the experiment, using audio/video recordings and logs. The data collected during the experiment, in interviews and questionnaires, can support this analysis. Our impression so far in analyzing the scenarios is that important insights can come from listening to how the crews reason when making decisions.

An experiment of this size (14 crews) is too small for robust quantitative analyses like calculating Human Error Probabilities, but a very large data set to analyze qualitatively. One challenge in analyzing the data is how to get an overview of the phenomena observed in the runs, not only for performance but also for PSFs and common problems.

The experience from observing the simulator runs is helpful for selecting what to analyze further when reviewing DVDs. A lesson learned for the next phase in the HRA empirical test
study is to collect more qualitative data during the runs in a form that is easily accessible. This could be done by adding comment fields to the expert ratings, and actions in OPAS. Letting the process expert make a written summary after each run is another idea.

IV. CONCLUSIONS REGARDING THE MOST IMPORTANT LESSONS LEARNED

When conducting a simulator study for testing HRA methods you should be very clear about what you want to test; i.e., the objectives of the study. This will determine the overall scope of the study and provide input into the scenario design.

We recommend using existing/available information such as from a PRA to guide which scenarios and corresponding HFEs are worthy of being examined. They should be potentially risk important so that expensive simulator time is properly prioritized and not used on scenarios or HFEs that almost assuredly would not be risk significant even if the human failure rate were quite high.

Certain constraints cannot be avoided when performing simulator studies. They limit the actions in the PRA that can be studied in a simulator setting. Examples are:

- Human actions that don’t take place in the control room
- Time constraints for simulations
- Simulator model capacity
- Compromise between realism and usefulness of the study when you choose the level of difficulty in the scenario

To learn about human reliability and especially the effects of PSFs, we believe it is of value to have well-designed manipulations of scenarios that to the extent possible, adjust the influences of interest in ways that will allow the examination of the effects of such influences on human performance. A useful way to do this is to use the concept of designing both base and complex cases of the same scenario and to then generalize from the results, to improve HRA methods.

Care is required in defining the HFEs of interest. They need to be defined in ways that both correspond to the human failures of interest in PRAs and can be observed and measured in a simulator environment. Be aware that the HRA methods being tested or evaluated may also influence how the HFEs should be defined so that the HRA methods can actually be used to analyze the HFEs. We believe that measuring performance with HFEs based on completed actions and times are valuable in analyzing effects of PSFs. HFEs should be defined in the scenario design process, before creating tools for data collection.

Care must also be taken in ensuring consistency between the performance data to be collected and the form of the predictions from the HRA methods. This is not a trivial task and we are still learning from our initial pilot project. Suffice it to say that what can be observed or measured may not always correspond to what comes out of HRA method results. We are finding that aids have to be created (such as operator interviews) in order to bridge the gap between what can be directly observed or measured in a simulator setting and what is predicted about human performance using a HRA method.

It can be very time consuming analyzing the data from simulator runs. We believe that tools that facilitate the overview, like OPAS, are useful. We also believe that more easily accessible qualitative data, like comments (expert comments were recorded on audio, but not summarized), would be useful. To understand what leads to failure/success, a more detailed analysis is needed.

ACKNOWLEDGMENT

The authors are grateful to Ronald Boring, Idaho National Laboratory, and to Andreas Bye and Michael Hildebrandt, OECD Halden Reactor Project, for their contributions to the paper and to Per Øivind Braarud and Salvatore Massaiu, OECD Halden Reactor Project, for their work in the area of simulator scenario development for HRA purposes as a source of many of the ideas discussed in this paper. The authors gratefully acknowledge the guidance and discussions provided by Vinh Dang, Paul Scherrer Institute, Erasmia Lois, U.S. Nuclear Regulatory Commission and John Forester, Sandia National Laboratories.

REFERENCES


