Nuclear Engineering and Technology 52 (2020) 1697–1705

Contents lists available at ScienceDirect

Nuclear Engineering and Technology

journal homepage: www.elsevier.com/locate/net

Original Article

Considerations for generating meaningful HRA data: Lessons learned from HuREX data collection

Yochan Kim

Risk & Reliability Research Team, Korea Atomic Energy Research Institute, 111 Daedeok-daero 989 Beon-gil, Yuseong-gu, Daejeon, 34057, Republic of Korea

ARTICLE INFO

Article history:
Received 7 October 2019
Received in revised form
29 January 2020
Accepted 31 January 2020
Available online 5 February 2020

Keywords:
Data analytics
Human reliability analysis
HuREX framework
Lesson learned
Simulation data

ABSTRACT

To enhance the credibility of human reliability analysis, various kinds of data have been recently collected and analyzed. Although it is obvious that the quality of data is critical, the practices or considerations for securing data quality have not been sufficiently discussed. In this work, based on the experience of the recent human reliability data extraction projects, which produced more than fifty thousand data-points, we derive a number of issues to be considered for generating meaningful data. As a result, thirteen considerations are presented here as pertaining to the four different data extraction activities: preparation, collection, analysis, and application. Although the lessons were acquired from a single kind of data collection framework, it is believed that these results will guide researchers to consider important issues in the process of extracting data.

1. Introduction

After the significant extent that human errors contribute to system safety was revealed [1], various efforts have been made to reduce human errors in safety-critical systems. Human reliability analysis (HRA) is one approach among such efforts that (1) systematically analyzes human-machine systems, (2) predicts the possibility of human errors by integrating human failure mechanisms with system failure scenarios, and (3) facilitates communication between safety engineers and scientists in different areas using quantitative measures for improving systems in terms of human error. In the past several decades, multiple HRA techniques have been developed and successfully applied in diverse industries such as nuclear power plants (NPPs), chemical plants, air-traffic systems, and healthcare systems [2].

Despite the usefulness of HRA, the availability and quality of data supporting HRA techniques have been recognized as one of the most vulnerable aspects [3,4]. Data that have been practically employed in HRA techniques to date have largely relied on expert judgment or quite old sources, and rigorous treatments for the data have been insufficient to derive insights for human reliability [5].

There have been significant efforts to collect HRA data. For example, Kirwan et al. [6] developed the CORE (Computerized Operator Reliability and Error) database that includes human error probability (HEP) information obtained from different sources. Sträter and Bubb [7] accumulated CAHR (Connectionism Assessment of Human Reliability) data from boiling water reactors that contains HEPs and their relevant situations. The HERA (Human Event Repository and Analysis) database was generated from event investigations of NPPs [8]. Likewise, the GRS (Gesellschaft für Anlagen und Reaktorsicherheit) produced 37 HEPs from licensee event reports in German NPPs [9]. The HAMMLAB (Halden Reactor Project's Human-Machine Laboratory) conducted the experiments using full-scope simulators to extract human reliability information, including HEPs and contextual information, and compared the empirical findings with HRA results [10]. The US Nuclear Regulatory Commission (NRC) [11] also obtained empirical HEP distributions for four human failure events (HFEs) from training records of US operators. In more recent years, the SACADA (Scenario Authoring, Characterization, and Debriefing Application) and HuREX (Human Reliability data EXtraction) databases represent the briskest data collection activities [12,13]. These databases primarily collected HRA data from full-scope simulators. Based on observations of crew behaviors or interviews with licensed operators, human errors were identified and contextual information was obtained. Besides, the variety of experiments and analyses on human reliability and performance that has been conducted [14].

Despite several kinds of databases supporting HRAs have been proposed and implemented recently, considerations for enhancing

E-mail address: yochankim@kaeri.re.kr.

https://doi.org/10.1016/j.net.2020.01.034

the quality of data have not been sufficiently discussed yet [15]. The importance of data quality has been emphasized several times in various fields featuring data analytics [16,17]. Extraction of HRA data is necessary to have a deep understanding of the quality as much as or more than the extraction of data from other disciplines. It has been revealed that human errors are entangled with myriad contextual, situational, and organizational factors, plus the fact that there are different viewpoints and definitions of human errors and performance shaping factors (PSFs). Under such complex relations between human errors and PSFs and dissimilar definitions of human errors, it is difficult to convince that the data regarding human reliability or behavioral/cognitive characteristics can always be extracted consistently and reasonably in a way similar to the extraction of equipment failure data. This means that the quality of HRA data requires profound discussions for their preparation, collection, analysis, and application.

In this paper, we present key considerations to generate quality HRA data based on the experience of HRA data collection for the HuREX framework. Since the HuREX framework was developed to produce HRA data from training records using full-scope simulators, over 50,000 data points were generated from multiple different NPP systems. The considerations in this paper were established from discussion of the results of statistical analysis [18,19] and data comparison between HuREX and SACADA [20]. This paper assumes that HRA data will be used to provide empirical evidence for the prediction of HEPs, PSF effects, and recovery failure probabilities. The rest of this paper is organized as follows. Section 2 briefly introduces the HuREX framework and summarizes its data extraction activities. Section 3 provides the derived significant issues and recommendations to ensure meaningful HRA data. Section 4 appraises the previous data extraction activities against the considerations. The final section discusses the conclusions of this study and future research.

2. HuREX data extraction activity

2.1. HuREX framework

The HuREX framework was established to gather and analyze HRA data from NPP simulators [13,19]. The data is affordable from any kinds of experiments or training records from partial scale of simulators or replicas of NPP control rooms. Ham and Park also partially adopted the HuREX framework to analyze event investigation report [21]. Fig. 1 briefly describes the data collection and analysis process.

Based on the characteristics of the system environment and the scope of data collection, information gathering templates (IGTs) are prepared, which include the data items or variables regarding crew contexts and responses that are potentially collected from simulator data including training records or experimental records. In the data extraction project, the three kinds of IGTs were developed: overview, response, and unsafe act. The overview IGT comprehends the basic information from simulations and the overall characteristics of scenarios, crews, and environments, e.g., simulation time, procedure progression information, simulated accidents and additional malfunctions, training level, and operating year of operators. The response IGT was designed to obtain the success or failure of task performance based on the procedures employed. With this IGT, the time to initiate a procedural step, procedural instructions, the types of operator tasks with related components and systems, and human error modes are evaluated. The unsafe act IGT includes detailed situational information when an operator error is observed. The data items in the unsafe IGT represent the following situational factors: familiarity and complexity of given tasks, clearness or structural issues in procedure instructions, communication, and recovery information.

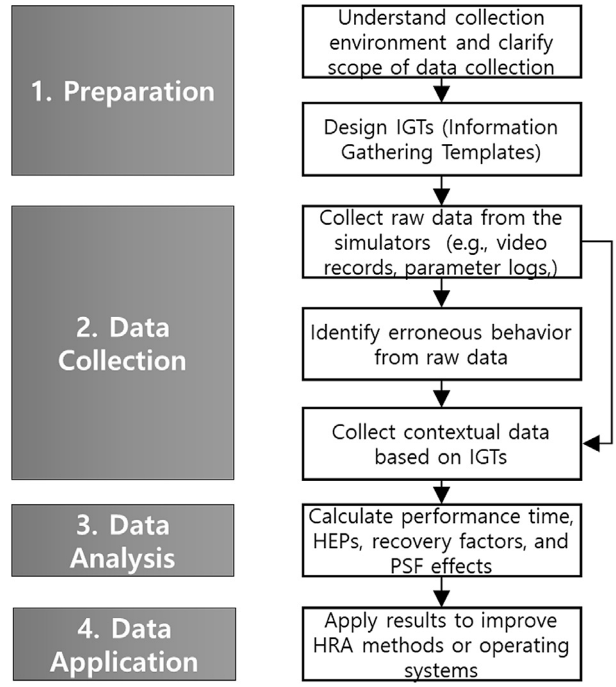

Fig. 1. Overall process of the HuREX framework (adapted from Ref. [18]).

During the raw data collection, video records, component manipulation logs, plant parameter logs, scenario information, and procedural path information are recorded (the first box of the data collection phase). From the raw data obtained, the tasks that were carried out by crews are selected and the erroneous behaviors are identified among the tasks (the second box of the phase). The erroneous behaviors in this analysis are determined when the following two conditions are met: (1) the operator behaviors have deviated from the procedure instructions and (2) the behaviors are causally related with negative consequences such as inappropriate equipment operation, inappropriate procedure transit, and inappropriate communication with external organizations. The tasks including erroneous behaviors were considered as failed tasks while the other tasks performed in the scenarios were seen as successful behaviors. Lastly, the contextual information is extracted based on the IGT forms developed in the previous phase (the third box of the data collection phase). All gathered information is stored in the database.

In the data analysis phase, quantitative information is estimated via several statistical treatments. For example, estimates of human performance time [22,23], HEPs [24], recovery failure probabilities [25], and PSF impacts on HEP [18,19] can be produced based on the obtained data. The human performance time was estimated by basic statistics such as mean and standard deviation or probability distribution fitting based on maximum likelihood estimation or Bayesian inferences. The HEPs and recovery failure probabilities are predicted using a Bayesian update with a non-informative Jeffrey prior. The PSF impacts were derived using two kinds of logistic regression analyses and stepwise variable selection.

Such quantitative information can be employed to update existing HRA methods or develop a new kind of HRA method. Analysis results also make it possible to identify areas of improvement for operating systems such as procedures, training, and interfaces. Therefore, during the last phase, the results of the previous phase are interpreted and reported with considerations of knowledge obtained from entire activities in the data extraction.

Besides, those results are compared or synthesized with other findings from empirical studies, expert opinions, or literature.

2.2. Summary of data extraction activity

Two separate HuREX data extraction projects have been performed for obtaining HRA data, as compared in Table 1. The first acquired the raw data from the full-scope simulators of a Westinghouse-type three-loop reactor and an OPR-1000 (Optimized Power Reactor 1000 MW). From 2 emergency scenarios and 12 abnormal scenarios, 223 simulations were observed. The primitive tasks of the procedures that the crew followed were analyzed from the observations. The tasks and human errors were distinguished according to the meaning and classification rules of human errors defined in the HuREX framework and over 10,000 data points were generated. In each data point, a binary state indicating the failure or success of a primitive task is entailed with the task types, error types, and the PSF levels for 26 variables. Table 2 shows the list of data items pertaining to each data point. Kim et al. [18,24] showed the results of statistical treatments using the data points.

The second data extraction project was the first project to generate a large HRA database for digitalized control rooms. The training simulator of an APR-1400 (Advanced Power Reactor 1400 MW) was employed to produce the raw data. A total of 168 simulations were recorded from 8 abnormal scenarios and 12 emergency scenarios in this project. The primitive tasks and operator performance of each task were analyzed like the first project, and as a result over 44,000 data points were produced. The information for the success/failure variable and over 50 contextual variables is contained in a data point. The results of time analysis using the obtained data were presented in Ref. [22], while HEP or PSF effect estimates are expected to be given in further studies.

3. Considerations for HRA data quality

This section describes the considerations for generating high-quality data based on the insights obtained from data extraction activities. For ease of understanding, we simply classify the considerations by the four data extraction processes: preparation, data collection, data analysis, and data report/application. However, it is important to note that these processes are performed iteratively, and the considerations are also correlated with each other. Fig. 2 summarizes the relations between the considerations reported in this paper. Some rounded boxes that overlap with each other imply that the considerations are shared specific sub-issues with each other. For example, *abstraction levels of operating tasks* can be seen as a special aspect of *consistent framework*. Some remarkable relations between considerations were also indicated by the arrows.

3.1. Preparation

Consistent framework. A consistent framework with which to collect data is the first important consideration. Consistency allows us (1) to continuously accumulate HRA data, (2) to clearly compare operator performance in different situations or environments, and (3) to reasonably conjecture the differences between the extracted data and data from other frameworks. The following decisions should be consistently made:

- How to define the tasks or human actions to be evaluated (e.g., success or failure) from the obtained raw data;
- How to classify the types of human behaviors that were evaluated as human errors or successful behaviors;
- How to determine the meaning of the PSFs or surrogate variables and their levels;
- How to designate an action changing the plant state that had been influenced by a previous human error, either as a recovery behavior, failed behavior, or successful behavior;
- How to store the inputted information, such as task performance and PSF variables, into a database.

To consistently manage the above issues, the related definitions and assumptions of the terminologies in the information to be collected should be explicitly addressed and applied. The assumptions or rules to determine the classes or types of behaviors, human errors, tasks, or contexts also need to be documented. When human errors or tasks are classified or any value is assigned to a PSF variable, a classification scheme that clearly distinguishes the classes or states and comprehensively describes the raw data is useful to generate insightful HRA data. In addition, definitions or classifications based on a theoretical basis of cognitive science or psychology are required, because the cognitive processes of human operators are analyzed during human reliability assessments. The US NRC [26] and Kim et al. [24], have presented significant examples of using cognitive literature for establishing such classification schemes.

Abstraction levels of operating tasks. In view of task classification, it is important to know that there exist different levels of tasks in terms of abstraction hierarchy that explain the goals–means relationships between entities [27]. Any task in a work domain can be explained with available means for achieving their goal. The means can also be seen as the other level of tasks that should be completed with a lower level of tasks. For example, Fig. 3 shows how tasks in NPP emergency situations can be described with different levels. In this hierarchy, it can be said that the higher levels of tasks can be achieved by the lower levels of tasks. At the highest level are the goals of system control as addressed by Corcoran et al. [28]. The second level tasks indicate the HFEs that can be defined in probabilistic safety assessment (PSA) scenarios or event trees, the third level tasks can be defined

Table 1
Summary of data extraction projects.

	1st extraction project	2nd extraction project
Reference plant	Westinghouse-type three loop plant and OPR-1000 (panel-based control room)	APR-1400 (computer-based control room)
Scenarios	2 emergency scenarios, 12 abnormal scenarios	12 emergency scenarios, 8 abnormal scenarios
Data source	Regular training records	Regular training records
Crew	Licensed operator working in commercial operations	Licensed operator working in commercial or trial operations
Operator roles in a crew	Five operators: shift supervisor (SS), reactor operator (RO), turbine operator (TO), electric operator (EO), and shift technical advisor (STA)	Five operators: SS, RO, TO, EO, and STA
Number of raw records	223	168
Number of data points	10,768	44,585

Table 2
Data items in 1st HuREX project (adapted from Ref. [18]).

Data items	Definition
Success/failure	Whether a human error occurred or not in the task
Task type	A Task type defined in HuREX task taxonomy
Error mode	Error of omission or error of commission
Training experience	Inclusion of scenario in the regular training program
Simulation mode	Type of simulated situation based on the reactor trip
Multiple initiating events	Whether inputted scenarios are single or multiple
Failed system/component	Existence of additional malfunctions in the system or component
Failed alarm/indicator	Existence of masked or failed indicators or alarms
Leadership of SS	Leadership style of shift supervisor
Cooperative attitude of BOs	Whether board operators showed active responses or communications during the simulations
Supervising level of STA	Whether the STA actively checked the operations of the systems
Independent checker	Independent review during significant system controls
Procedure compliance	How the shift supervisor gave directions based on the procedures
Overall communication strategy	Communication strategy that is frequently observed during the simulation
Time pressure	Urgency of the tasks determined based on the ongoing procedure and inputted scenario
Task familiarity	Whether the task can be experienced during the power reduction or raise
Contingency action	Inclusion of the task in the contingency action part when the emergency procedure is used
Type of state identification	Type of plant information required by the shift supervisor or obtained by the board operators
Note or caution	Inclusion of the task in the note of caution part of the procedure
Number of tasks	The number of tasks in the ongoing step
Number of manipulations	The number of manipulations to be controlled during performance of the ongoing step
Instruction contents	The content type to be instructed by the shift supervisor
Continuous action step	Inclusion of the task in the continuous action step
Confusing statement	Whether the procedure instruction related to the performed task is a negative form or includes an “or” condition
Multiple constraints	Existence of a parenthesized condition to be additionally checked in the procedure instruction
Clarity of decision-making criteria	Description of clear criteria in the procedure instruction to determine system status (e.g., pressure > 79 kg/cm^2)
Description of object	Description of component ID to be manipulated in the procedure
Specification of manipulation means	Whether the procedure instruction provides a means to control a target system
Diagnostic information clarity	Existence of indicator/alarm/display for diagnosis tasks

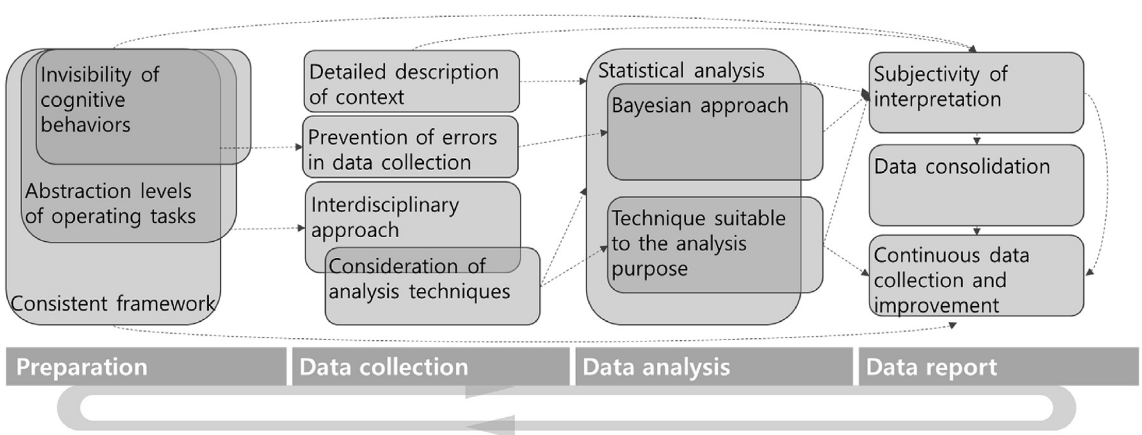

Fig. 2. A diagram depicting the relations between considerations.

by the training objective elements [12], and the fourth and fifth levels are defined by the task taxonomy of HuREX [24] and the cognitive processes in IDHEAS (Integrated Human Event Analysis System) [29], respectively.

The meanings of tasks in a specific level (as well as their successes and failures) are dissimilar from those in other levels. Because human reliability or the effects of situational factors on it can differ according to task level [30], data collectors should clearly define the levels of tasks and discuss how the results from the collected data can be interpreted after the analysis phase. The abstraction hierarchy of tasks such as Fig. 3 will be useful in guiding which tasks are to be selected for observation and analysis during human reliability data generation. In many cases, a higher-level task is achieved by carrying out multiple tasks in a lower level, and so the observation of tasks in a low abstraction level for data collection is beneficial to secure more data points. In addition, some tasks in the highest levels can be achieved by entirely different sets of low-level tasks; hence, it can be tricky to derive generic information from the performance data of higher-level tasks. On the other hand, in the data collection using the low-level task, we observed the following issues. First, the success of an HFE-level task does not always require the successful performance of all procedural instructions. During the simulation observation, there are some cases where an error in a detailed task made by an operator did not affect the overall task performance in the higher levels of tasks. Therefore, when performance or reliability in low-level tasks should be associated with HFE-level reliability, it is important to identify the tasks that are significant to the higher task levels. Second, as many errors in low-level tasks can be implicitly or explicitly recovered during all periods of high-level task performance, a way to capture such recovery behaviors should be prepared before the data collection phase. Lastly, it was observed that,

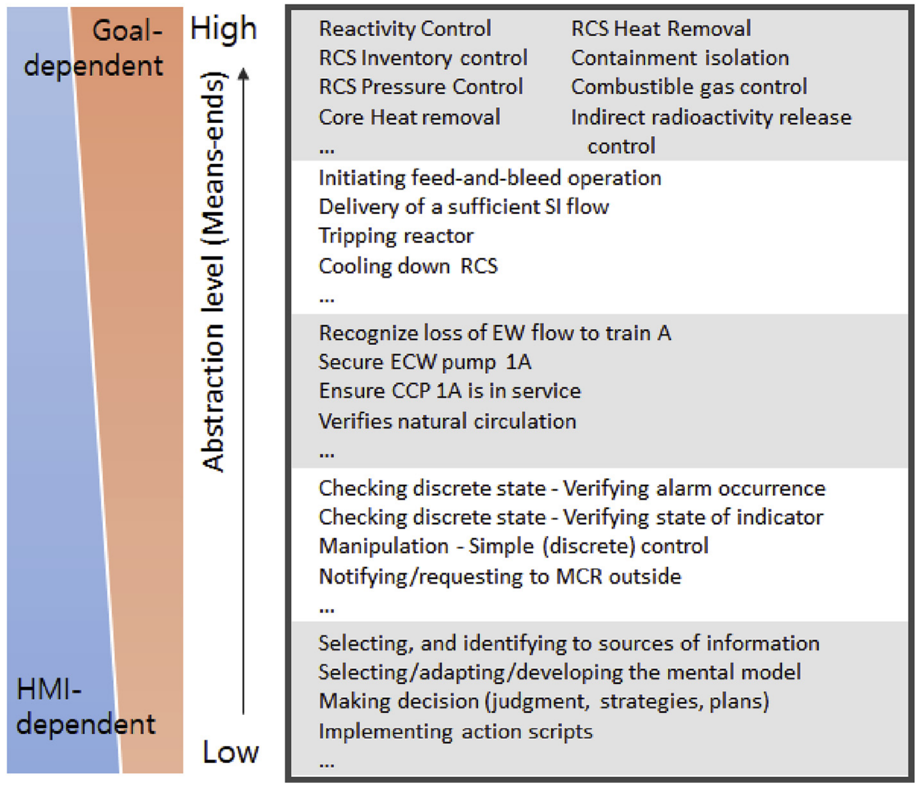

Fig. 3. Abstraction hierarchy of emergency tasks in NPPs (HMI: human-machine interface).

during training, operators tended to be somewhat negligent in low-level tasks that were not directly influential on plant safety. Data analysts thus need to consider the effects of these actions in interpreting results.

Invisibility of cognitive behaviors. Because many HRA methods have a goal to identify cognitive failure mechanisms, data items that have been alluded to in cognitive science literature might be included in data collection templates. However, capturing the cognitive characteristics of human operators usually requires subjective evaluation by data collectors. In fact, subjectivity can exist during the whole process of data extraction. The subjective evaluation should be carefully dealt with in any other process such as the determination of PSF levels in order to reduce the variability of data collection. But, it should be noted that identification of error mechanisms demands deep knowledge and experience of cognitive engineering and a consensus among experts may not be reached in some cases. For example, to identify the cognitive mechanism behind a wrong report in delivering plant information, a data collector may suspect that it is caused by a misperception of the parameter indicators while the other collector thinks that it is related with a mistake in speech [31]. In this situation, the data collectors cannot perfectly deduce the main reason of the reporting error without interviewing the human operator. In addition, even if the human operator can give information to the collectors about the cognitive cause of the error, the information might sometimes be weakly informative because an operator seldom makes an error while consciously aware of the specific error type. During the interview for identification of the root-cause of an observed human error, the interviewee often infers a plausible cause, but do not remember the exact cause of her/his error.

To input information on human cognitive characteristics, (1) a specific guideline should be written, and/or (2) training or drills for data collectors for the evaluation should be performed before the collection phase. Alternatively, a taxonomy of PSFs or human errors based on human behaviorism can be prepared, and objectively observable data items can be used for data collection.

3.2. Data collection

Interdisciplinary approach to human error identification and PSF rating. The data generation and analysis processes for HRA basically belong to data science, which is a multidisciplinary activity of computer science, statistics, and domain knowledge [32]. On the other hand, identifying human error information and significant PSFs from raw data (such as simulation records or experimental data) can also be seen as a sort of root-cause analysis or human error investigation process, which requires multidisciplinary reviews in general [33]. Given that even HRA applications are expected to be performed with multidisciplinary teams [34,35], it is not surprising that HRA data should be collected with an interdisciplinary approach. Based on the HuREX data collection experience, the following areas of expertise are expected to generally apply to HRA data: (1) safety/reliability engineering, (2) cognitive science, (3) NPP system dynamics, (4) data analytics, and (5) computer software development and database management. For example, understanding the NPP system dynamics is essential to evaluate whether certain operator behaviors are safe or unsafe. The data collectors sometimes need to consult experienced plant operators for comprehending the operator's behaviors. On the other hand, the data should be collected with consideration of the data structure. Because multiple experts can participate in the data collection activity and it is required to merge data in different kinds of forms (e.g., IGTs of HuREX), the relations between data forms should be clearly determined and the database key or index should be consistently used.

Consideration of analysis techniques. Without anticipation of the data analysis process, data collection can produce very limited information to be analyzed. It is important to generate appropriate data after a thorough review of which particular analytical methods can be used and which contents or structures of data are needed for the methods. For example, Kim et al. [5] alluded that generalized linear models such as a logistic regression that could be employed for estimating impacts of the PSFs require the contextual data items not only when crews failed certain tasks but also when they successfully performed them. Because, certain PSF states that appear to have impacts on human failures are sometimes more closely related with human success cases. Consequently, if the PSF data for success cases are not gathered, the negative influences of those PSF states are likely to be overestimated. Actually, the detailed contextual data in the original HuREX framework were obtained only when the human errors were observed. To statistically model the quantitative relation between PSFs and HEPs, we additionally generated detailed contextual information on the successful behaviors.

Prevention of errors in data collection. Unlike industrial big data such as manufacturing records and monitoring data of automated machinery, most HRA data is generated manually because the selection of PSF levels or the identification of human errors still requires entangled multidisciplinary knowledge. However, manual data collection has a weakness in that data quality can be degraded by human errors or violations by the collectors. To minimize bias or flaws in manually collected data, the following strategies can be employed: (1) human error identification by data collectors having the least conflict of interest with the results, (2) peer-reviews of the collected data, (3) establishment and application of explicit rules for determining information regarding human errors and contexts, and (4) development of support systems, such as a user interface-based data generation system and a data integrity-verification program.

Detailed description of context. Many data frameworks typically attempt to collect PSF information using items or variables predetermined during the preparation phase. However, as mentioned previously, the entire data extraction process is iterative, or more specifically, data analysis sometimes requires the gathering of additional PSF information. For example, in the course of data analysis, a number of operator errors are observed at a particular procedural step, which may reveal a structural problem that the related data item was not properly defined in the preparation phase. In this situation, new PSF variables should be defined, and data for the new variables should be collected. To facilitate the identification of significant PSF variables, detailed contexts including procedures, human-machine interfaces, and communications should be well documented during the collection phase.

3.3. Data analysis

Statistical analysis. Statistical techniques are essential to derive significant information supporting HRA. For example, due to the scarcity of HRA data, some researchers have attempted or proposed to simply compare HEPs in a certain context with other HEPs in another context for estimating the effects of the context [36–38]. However, to rigorously scrutinize the influence of contextual factors, it is important to statistically infer whether the difference between the two HEPs is actually due to chance or due to certain factors. It is therefore desirable to use developed statistical measures, such as confidence intervals or p-values, for testing hypotheses on the significance of factors [18].

Multivariate analysis techniques, which simultaneously examine multiple variables, are more effective than univariate techniques to compare PSF variables and then select the significant variables among them. In the multivariate logistic regression in Refs. [5,18], there were cases in which a particular factor that had been proven as significant to human error by univariate analysis was not selected as a significant factor in the multivariate analysis. This was because the importance of the variable was comparatively negligible, with other factors better explaining human reliability.

Bayesian approach. It is hard to say that collected HRA data are perfectly precise, even when a range of systems or guidelines for high quality data is implemented in collection activities. In addition, since data points in the tens or hundreds of thousands are often required to observe a human error, it is often difficult to ensure a sufficient amount of data. Along these lines, Bayesian inference techniques are useful to derive meaningful insights when data is insufficient and/or imprecise. In addition, Bayesian inference can be used to propagate uncertainties of the estimates through simulations or combine the results of data analysis with other kinds of data sources, such as expert judgments and prior knowledge from literature. Bayesian approaches have been widely employed in PSA [39], and Kim et al. [19] also recently introduced a Bayesian technique for estimating the relations between the human reliability and the PSFs.

Technique suitable to the analysis purpose. An analysis technique that is able to produce results in conformity with the given HRA application should be used. For example, most HRA methods usually assess HFEs using intuitive models such as decision trees, simple Bayesian networks, or multiplier tables for calculating their HEPs with PSF states. These models are beneficial to support practitioners in easily tracing their evaluation process and verifying its reasonableness following human reliability assessment. To update these existing HRA methods or develop new easy-to-use techniques, quantitative information generated in the analysis phase should not only have a high precision or prediction rate of human reliability, but also be applicable to simple HEP quantification models. For example, in practical HRA applications, complex network-based algorithms, such as deep neural networks, may not be attractive for modeling the relation between PSFs and human reliability, recovery factors, or dependencies for existing HRA strategies. Because the large-scale neural networks may not be easy for the HRA practitioners to interpret, making it difficult to trace or understand the results of their analyses, an additional technique aiding interpretation of practitioners could be required [40].

3.4. Data report and application

Subjectivity of interpretation. Even though reducing subjectivity is beneficial in producing transparent results, it is also noticed that the interpretation of the analysis results often requires the subjective opinions of data analysts. If any PSF variable is revealed as significant to a type of human reliability, it is desirable to discuss which specific reasons or features can affect the cognitive activities of operators. In addition, how the obtained results can be generalized to predict the effects of abstract PSFs on the HEPs can be also deliberated. On the other hand, when some variables are correlated with each other, it is critical to identify which variables are actually meaningful. For example, for a set of abnormal situation data collected from 'A' simulator and a set of emergency scenario data gathered from 'B' simulator, the situation modes and simulator designs are often correlated. In this case, data analyst judgment can influence the interpretation. To derive factual conclusions from the analysis, it is recommended that data analyzers should collaborate with data collectors or carefully review the description of the relevant situations (refer to the consideration, *detailed description of context.*). Otherwise, more data should be collected to clearly resolve this issue.

Table 3
The evaluation summary of the two HuREX projects on the basis of the considerations.

<table>
<tr><th>Considerations</th><th>1st HuREX project</th><th>2nd HuREX project</th></tr>
<tr><td rowspan="2">Consistent framework</td><td colspan="2">· The criteria discriminating success and failure were developed based on the definition of an unsafe act. The criteria were consistently applied by educating the data collectors with several workshops.
· The task and error taxonomies were developed taking into account the procedure sentences, cognitive models, existing error types, and simulator characteristics [24]. The classifications were implemented into the database management systems.
· Spreadsheet-based databases were developed and managed.</td></tr>
<tr><td>· For clarifying subtle issues in the identification or classification of human errors and recovery actions, additional rules were established.</td><td>· The developed rules were elaborated and adjusted using the characteristics of the APR1400 simulator (e.g., use of the computer-based procedure).</td></tr>
<tr><td>Abstraction levels of operating tasks</td><td colspan="2">· The procedure sentence-based primitive tasks were defined (the second level tasks from the bottom in the hierarchy of Fig. 3). This task level is more concrete than the general HFE in HRA practice; hence, the model to link the results of the projects with the HRA applications should be developed.</td></tr>
<tr><td>Invisibility of cognitive behaviors</td><td colspan="2">· The tasks distinguishable from audio-video records and procedure sentences were used.</td></tr>
<tr><td rowspan="2">Interdisciplinary approach to human error identification and PSF rating</td><td colspan="2">· HRA experts, statisticians, cognitive scientist, software/database developer participated in this project.</td></tr>
<tr><td>–</td><td>· The plant operators and training instructors were regularly interviewed for explaining plant dynamics.</td></tr>
<tr><td>Consideration of analysis techniques</td><td colspan="2">· The data that can be used for average HEPs or recovery failure probabilities were mainly obtained (e.g., occurrences of success and failure cases)
· To estimate PSF effects, the contextual information for both successes and failures were additionally generated.</td></tr>
<tr><td rowspan="2">Prevention of errors in data collection</td><td colspan="2">· The human reliability data was generated by the 3rd party researchers.
· Database containing properties of procedure sentences were developed and used.
· Spreadsheet-based data (i.e., IGTs) were obtained and managed.</td></tr>
<tr><td>–</td><td>· Softwares to aid observe multiple videos simultaneously and to generate IGT data along with procedure sentences were developed and employed.
· Software to check synthetic errors were developed and employed.
· Monthly workshops were held for peer-reviewing the data collection.</td></tr>
<tr><td>Detailed description of context</td><td>· The procedure sentences related to human errors were documented.</td><td>· The procedure sentences, significant conversations, and consequences related to human errors were documented.</td></tr>
<tr><td>Statistical analysis</td><td>· Several statistical criteria such as p-value, confidence interval (for maximum likelihood estimation), credible interval (for Bayesian analysis), and Bayesian information criterion were used to test the hypotheses or quantify the parametric uncertainty.</td><td>· Similar criteria used in the first project will be employed.</td></tr>
<tr><td>Bayesian approach</td><td>· Bayesian inference based on Jeffrey's prior was applied to the estimation of HEPs.
· The PSF multipliers were also predicted by a Bayesian inference [19].</td><td>· Bayesian inference is planned to be used for estimation of HEPs, recovery failure probabilities, PSF effect, and so on.</td></tr>
<tr><td>Technique suitable to the analysis purpose</td><td>· Logistics regression models that allow estimates that can represent the multiplicative HRA models were used.</td><td>· A logistic regression technique or decision tree algorithm can be used.</td></tr>
<tr><td>Subjectivity of interpretation</td><td>· The data analyzer and data collector collaborated with to understand the statistical results.</td><td>–</td></tr>
<tr><td>Data consolidation</td><td>· The estimates for PSF multipliers were combined with the multipliers in existing HRA methods [19].
· The HEPs from this project were compared with those obtained from the Micro tasks [41].
· The statistical results were compared with the results of the second extraction project.</td><td>The expert knowledge from a formal elicitation was obtained and is expected to be combined [42].</td></tr>
<tr><td>Continuous data collection and system improvement</td><td>· This project provides the motivation for the second extraction project.</td><td>· Some improvements regarding procedure expressions were reported to the operating company.</td></tr>
</table>

Data consolidation. No HRA data can be asserted as perfect. Moreover, results derived from HRA data cannot be considered as generic, because the operating culture in terms of national/organizational culture and system design including plant dynamics, automation levels, and interface designs are too various. To derive generic and robust insights from collected data, data comparison or consolidation with results from other data sources or expert judgments is highly important. To do so, it is necessary to record the characteristics of the operational environment and the basic assumptions of each data collection. For example, the HuREX data obtained to date, as based on regular training records, contains operator responses in 30–40 min simulations because of the particular training policies of the operating company. In this case, it is hard to accurately simulate situations regarding long-term operation or recovery behaviors with longer time margins.

Continuous data collection and system improvement. The ultimate goal of HRA is to improve systems and reduce human errors through quantification. We specify here that data collection is also very useful for identifying system improvements. Data collection offers another opportunity to find system problems that HRAs have not easily captured from the observation of operator hesitations, conversations, and non-verbal communications. If improvements are repeatedly identified and the HRA method is continuously updated through the continuous HRA data collection, system safety will be more effectively secured.

4. Evaluations of HuREX activities against considerations

The efforts in the two HuRex projects for handling the considerations proposed in this paper are encapsulated in Table 3. Overall, both data extraction projects utilized the HuREX to collect the data with the consistent definitions and taxonomies. The crew errors were assessed based on a detailed level of the task, and observable behaviors were selected as the target evaluation actions of human errors reflecting the invisibility of cognitive behavior. In the second project, more detailed rules than that of the first project were used to maintain consistency. In addition, more experts participated in the second project, and periodical workshops and interviews were conducted to verify data quality. The data collection for APR1400 control rooms has recently been completed and statistical analyses and applications have not been carried out yet, but many statistical results are expected in the future.

From the perspective of the lessons learned, the following issues of the HuREX framework were found. First, the task level mainly concerned in these projects is consistent but very concrete than the HFE in many HRA applications. It is required to develop a model for plausibly associating the results of extraction with the HRA applications. Second, for the effects of PSF, more data needs to be collected to minimize subjectivity in interpretation. Third, some values of HEPs from the HuREX were compared with the HEPs estimated by other simulator experiments [41]. Likewise, the various comparison efforts between the estimates should be encouraged. In addition, how to synthesize the findings or data from different data sources should be discussed in the future. Lastly, ongoing data extraction needs to be continued to improve system safety and derive general insights into human reliability. For example, it can be attempted to compare domestic data with data from various countries or compare the reliability for current APR1400 operators with the reliability after the accumulation of more operating experience.

5. Discussion and conclusion

It is clear that HRA data is of great importance in improving both the quality of human error quantitative analysis as well as the safety of human-machine systems. In this paper, we presented a number of considerations for ensuring high-quality HRA data following the lessons learned from the HuREX data extraction experience. Because the lessons were acquired from only one data collection framework, some recommendations may not exactly align with other collection frameworks. Despite this, it is believed that these results will guide researchers to consider important issues in the process of extracting data. We plan to analyze the recently collected HuREX data with the considerations conceived in this study. The results will then be compared with the data from HAMMLAB experiments, the SACADA database, and other experiments.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the Nuclear Research & Development Program of the National Research Foundation of Korea grant, funded by the Korean government, Ministry of Science, ICT & Future Planning (grant number 2017M2A8A4015291).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.net.2020.01.034.

References

[1] B. Kirwan, A Guide to Practical Human Reliability Assessment, Taylor & Francis, London, 1994.
[2] J. Bell, J. Holroyd, Review of Human Reliability Assessment Methods, Health and Safety Laboratory, United Kingdom, 2009.
[3] E. Hollnagel, Y. Fujita, The Fukushima disaster – systemic failures as the lack of resilience, Nucl. Eng. Technol. 45 (2013) 13–20.
[4] E.M. Dougherty Jr., Human Reliability Analysis - where shouldst thou turn? Reliab. Eng. Syst. Saf. 29 (1990) 283–299.
[5] Y. Kim, J. Park, W. Jung, I. Jang, P.H. Seong, A statistical approach to estimating effects of performance shaping factors on human error probabilities of soft controls, Reliab. Eng. Syst. Saf. 142 (2015) 378–387.
[6] B. Kirwan, G. Basra, S.E. Taylor-Adams, Core-data: a computerized human error database for human reliability support, in: Proceedings on IEEE Sixth Annual Human Factors Meeting, 1997.
[7] O. Sträter, H. Bubb, Assessment of human reliability based on evaluation of plant experience: requirements and implementation, Reliab. Eng. Syst. Saf. 63 (1999) 199–219.
[8] B. Hallbert, R. Boring, D. Gertman, D. Dudenhoeffer, A. Whaley, J. Marble, J. Joe, E. Lois, Human Event Repository and Analysis (HERA) System – Overview, NUREG/CR-6903, vol 1, US Nuclear Regulatory Commission, Washington DC, 2006.
[9] W. Preischl, M. Hellmich, Human error probabilities from operational experience of German nuclear power plants, Reliab. Eng. Syst. Saf. 109 (2013) 150–159.
[10] US Nuclear Regulatory Commission, The International HRA Empirical Study: Lessons Learned from Comparing HRA Methods Predictions to HAMMLAB Simulator Data, US Nuclear Regulatory Commission, Washington DC, 2014. NUREG-2127.
[11] US Nuclear Regulatory Commission, The U.S. HRA Empirical Study – Assessment of HRA Method Predictions against Operating Crew Performance on a U.S. Nuclear Power Plant Simulator, NUREG-2156, US Nuclear Regulatory Commission, Washington DC, 2016.
[12] Y.J. Chang, D. Bley, L. Criscione, B. Kirwan, A. Mosleh, T. Madary, R. Nowell, R. Richards, E.M. Roth, S. Sieben, A. Zoulis, The SACADA database for human reliability and human performance, Reliab. Eng. Syst. Saf. 125 (2014) 117–133.
[13] W. Jung, J. Park, Y. Kim, S.Y. Choi, S. Kim, HuREX-A framework of HRA data collection from simulators in nuclear power plnats, Reliab. Eng. Syst. Saf. (2018), https://doi.org/10.1016/j.ress.2018.07.036.
[14] J. Park, D. Lee, W. Jung, J. Kim, An experimental investigation on relationship between PSFs and operator performances in the digital main control room, Ann. Nucl. Energy 101 (2017) 58–68.
[15] A. Bye, A. Wright, M. Reid, Human reliability data for modern control rooms, a survey, in: Asian Symposium on Risk Assessment and Management 2019, Gyeongju, Korea, 2019. September 30 – October 2.
[16] R.Y. Wang, D.M. Strong, Beyond accuracy: what data quality means to data consumers, J. Manag. Inf. Syst. 12 (4) (1996) 5–33.
[17] L. Cai, Y. Zhu, The challenges of data quality and data quality assessment in the big data era, Data Sci. J. 14 (2015).
[18] Y. Kim, J. Park, W. Jung, S.Y. Choi, S. Kim, Estimating the quantitative relation between PSFs and HEPs from full-scope simulator data, Reliab. Eng. Syst. Saf. 173 (2018) 12–22.
[19] Y. Kim, J. Park, Incorporating prior knowledge with simulation data to estimate PSF multipliers using bayesian logistic regression, Reliab. Eng. Syst. Saf. 189 (2019) 210–217.
[20] Y. Kim, J. Park, Characteristics of the HuREX framework as a tool for HRA data, in: PSAM14 Workshop on Collecting HRA Data, September 16, 2018. Los Angeles, CA.
[21] D.H. Ham, J. Park, Use of a big data analysis technique for extracting HRA data from event investigation reports based on the Safety-II concept, Reliab. Eng. Syst. Saf. 194 (2020) 106232.
[22] Y. Kim, J. Choi, J. Park, W. Jung, S.J. Lee, Estimating diagnosis time of emergency situations in digitalized control rooms, in: Proceedings of the AHFE 2018 International Conference on Human Error, Reliability, Resilience, and Performance, July 21-25, 2018. Florida, USA.
[23] Y. Kim, J. Kim, J. Park, S.Y. Choi, S. Kim, W. Jung, H.E. Kim, S.K. Shin, An HRA Method for Digital Main Control Rooms – Part I: Estimating the Failure Probability of Timely Performance, KAERI, 2019. KAERI/TR-7607/2019.
[24] Y. Kim, J. Park, W. Jung, A classification scheme of erroneous behaviors for human error probability estimations based on simulator data, Reliab. Eng. Syst. Saf. 163 (2017) 1–13.
[25] Y. Kim, J. Park, S. Kim, S.Y. Choi, W. Jung, Estimating recovery failure probabilities in off-normal situations from full-scope simulator data, in: Proceedings of the KNS 2016 Autumn Meeting, October 27-28, 2016. Gyeongju, Korea.
[26] US Nuclear Regulatory Commission, Cognitive Basis for Human Reliability

Analysis, NUREG-2114, US Nuclear Regulatory Commission, Washington DC, 2016.
[27] K.J. Vicente, J. Rasmussen, Ecological interface design: theoretical foundations, in: IEEE. T. Syst. Man. Cyb., vol 22, 1992, pp. 589–606, 4.
[28] W.R. Corcoran, N.J. Porter, J.F. Church, M.T. Cross, W.M. Guinn, The critical safety functions and plant operation, Nucl. Technol. 55 (3) (1981) 690–712.
[29] US Nuclear Regulatory Commission, An Integrated Human Event Analysis System (IDHEAS) for Nuclear Power Plant Internal Events At-Power Application, vol 1, US Nuclear Regulatory Commission, Washington DC, 2017. NUREG-2199.
[30] E.M. Hickling, J.E. Bowie, Applicability of human reliability assessment methods to human–computer interfaces, Cognit. Technol. Work 15 (1) (2013) 19–27.
[31] Y. Kim, J. Park, Suggestions of HRA method improvement for the practical assessment of human reliability, J. Ergon. Soc. Korea. 37 (3) (2018) 229–241.
[32] M. Barber, Data science concepts you need to know! Part 1. Towards Data Science, 14, 2018. Webpage published January, https://towardsdatascience.com/introduction-to-statistics-e9d72d818745.
[33] US Nuclear Regulatory Commission, Integrated Safety Analysis Guidance Document, NUREG-1513, US Nuclear Regulatory Commission, Washington DC, 2001.
[34] J. Silva-Martinez, Human systems integration: process to help minimize human errors, a systems engineering perspective for human space exploration missions, Reach. Out. 2 (2016) 8–23.
[35] US Nuclear Regulatory Commission, Good Practices for Implementing Human Reliability Analysis, NUREG-1792, US Nuclear Regulatory Commission, Washington DC, 2005.
[36] A.R. Kim, J. Park, Y. Kim, J. Kim, P.H. Seong, Quantification of performance shaping factors (PSFs)'weightings for human reliability analysis (HRA) of low power and shutdown (LPSD) operations, Ann. Nucl. Energy 101 (2017) 375–382.
[37] B. Kirwan, The development of a nuclear chemical plant human reliability management approach: HRMS and JHEDI, Reliab. Eng. Syst. Saf. 56 (2) (1997) 107–133.
[38] Y.J. Chang, C. Franklin, L. Criscione, J. Xing, Example use of the SACADA data to inform HRA, in: The Enlarged Halden Programme Group Meeting, Fornebu, Norway, 2016.
[39] D. Kelly, C. Smith, Bayesian Inference for Probabilistic Risk Assessment: A Practitioner's Guidebook, Springer, New York, 2011.
[40] D. Gunning, Explainable Artificial Intelligence (Xai), Defense Advanced Research Projects Agency (DARPA), 2017. https://www.darpa.mil/attachments/XAIProgramUpdate.pdf.
[41] S. Massaiu, A. Fernandes, The Reliability of Identification Tasks in Digital and Analogue Interfaces: A Re-analysis of Two Micro-tasks Studies, Halden reactor project, 2019. HWR 1218.
[42] Y. Kim, Expert elicitation for estimating PSF effects on HEPs in computer-based control rooms, in: 29th European Safety and Reliability Conference, September 22 - 26, 2019. Hannover, Germany.