Simulator-based Human Reliability Analysis using Bayesian Network: A Case Study on Situation Awareness in Engine Resources Management

A.M. Nizar, T. Miwa & M. Uchida
Kobe University, Kobe, Japan

ABSTRACT: Situational awareness (SA) is regarded as one of the important non-technical skills in constructing the seafarers' ability in daily decision-making and performing tasks, especially in Engine Resources Management (ERM). In maritime accidents that are mainly human error, insufficient SA is the specific factor that contributes to most of incidents. To quantitatively assess SA reliability, a Bayesian network of seafarers' performance in attaining SA in engine supervisory control is constructed. The adaptation of simulator data helped as combination along using subject matter expert input, which is a common practice in constructing human reliability analysis. Additionally, the simulator data can serve as the updating function when new data is observed. The result shows that the model can provide promising results as compared with expert expectation. Such kind of model can support the evaluation of the engine operation onboard, and mitigation can be provided to reduce the probability of human error.

1 INTRODUCTION

Regardless on doubt of the famous statement that said 80% of maritime incident are caused by the human error, the issue of human contribution to accidents is still remain important to study in the maritime safety area [1]. Departing from the traditional approach of studying human error for finding the lack of human limitation and blaming them, human factors study needs to be seen as the area where the human limitation is studied to improve working conditions. Thus, paradigm shift is needed to enhance the understanding of human capabilities, further, to facilitate the development of countermeasure and preventive strategy.

For the specific human performance factors that caused the most incident, situation awareness (SA) is prominent substituent that often mentioned. Human error in SA is labeled as "loss of SA". Incident analysis report that was conducted by Grech et. al revealed 71% of incidents in maritime operator are due to loss of SA [2]. Loss of SA will result in the operator failure to understand the operation condition and leading to the failure to take appropriate action. For thus, as the well concept that not only researcher and management define but also seafarer well understand and relate its importance, IMO already included it in crew resources management as one of the non-technical skill for seafarer to have [3,4].

1.1 *Human Reliability Analysis in Maritime Operation*

Practice to quantify human error through the process of human reliability analysis (HRA) has become the practice in other industry, from nuclear power plant through the air traffic control [5]. To uniform it, IMO with the formal safety assessment (FSA) guideline has also mentioned the HRA and various method of it in

maritime operation [6]. The HRA is not only useful in academic activity, but it was found to be benefit for stakeholders such ship owners and safety inspectors to identify and minimize the potential risk. The concept of hardware reliability analysis was also implemented in the HRA methods, including the hazard analysis and risk control stage. Seeing the HRA as similar with hardware analysis, it is also believed similar two combination of method can be fit to each other and combined to measure system reliability. Various HRA methods have been developed, also demonstrated in maritime operation cases. Cognitive reliability and error analysis (CREAM) is already applicated to derived the marine engineers performance reliability by combining Bayesian inference and fuzzy methods [7]. In seeing future projection, fuzzy methods also used combine with success likelihood index methods (SLIM) to demonstrate the autonomous operation regarding the human-machine interface [8]. Beyond onboard, the SLIM methods combined with system theoretic process analysis (STPA) are also used to analysis human-machine interaction in ship-to-ship LNG bunkering [9].

Most HRA aims to measure human error probability (HEP), that defined as an index to show the likeliness the human will conduct an error during the specific event. IMO defines it as the ratio of number of human error that have occurred, with the number of opportunities for human error [6]. In general, denominator for HEP is the number of chances the human conduct the error, compared to the hardware reliability where the denominator is the running time of equipment. This led to quantifying it by bottom-up approach to predict HEP by retrieving various data, mostly accident or incident report. The pitfall of employing only the failure database is the information only contain the number of failure event, without number of successful performance, where it is more close to assess the failure probability with empiric way [10]. Deciding the human error data from the accident report also has limitation since the number of accident reports is considered small compared to hardware failure data. This limitation often counters by including the expert judgement as the input, or the simulator experiment data.

In the lower factors, HRA can be included, combine, and consist of several performance shaping factors (PSF)[6]. PSF is often treated independently from one to the other. Several agree that PSF can be overlap, or its influence to each other should be considered [11]. The countermeasure of the dependency issue between PSF is by utilizing Bayesian network. The utilization of Bayesian network in HRA is increasing recently [5]. The Bayesian network allows us to analyze the likelihood of human error and identify the dependencies for complex modelling. It also came with the advantage of the ability to combine various data.

Bayesian network utilization in maritime operation is steadily increasing, either for HRA purpose or system reliability. It has wide application range from operator safety assessment to the evacuation training, including its application in offshore operation [12,13], ship collision [14], emergency situation [15], and ship engine operation [7]. The Bayesian network suggest PSF interaction and integrating different sources of information into the model, once the new information or data is exist, it can be updated easily to the model [11]. In the context of HRA, the Bayesian network provides the ability to contain and combine multiple types of information and data, including cognitive literature, operation events, statistical data, and expert judgment.

SA concept in human factors field is already applied in various work environments, include in maritime operation. In this study, loss in SA is considered as one of the factors that contribute to the human error event. HRA as the methods in quantifying human error is applied with adapting the engine plan simulator data combine with the subject matter experts. Further, Bayesian concept is employed to accommodate the dependency between the factors in contributing the human error.

2 MODEL CONSTRUCTION AND UPDATING

There are two terms involved in model construction and updating. In model construction, Bayesian network is applied to construct the model and calculate the probability of SA failure in the first place. While in model updating, the concept of Bayesian inference is used to recalculate the probability of SA failure by considering the new data from the simulator. The difference between the two methods is Bayesian network with its graphical methods does not necessarily imply the theoretical Bayesian inference. However, the Bayesian network in this study is called Bayesian since it employed similar rules for inferring the probability.

2.1 *Model construction*

The flow process of model construction and updating are shown in Figure 1. The Bayesian network in calculating human error mostly employed subject matter experts input as weighting factors, in this flow is to calculate nodes distance. Beyond that, subject matter experts also use variable control input to calculate condition probability distribution. Within this study, the role of subject matter expert is not removed, but instead reduced by substituting it with the result from simulator data, specifically in the stage of calculating the condition probability distribution. Bayesian network causal model uses directed acyclic graphs consisting of nodes and arcs. The node plays as the variable in the model, and the arcs denote the causal relationship between these variables. The nodes that the arcs point to are called child nodes, while the reference nodes are called parents nodes. A node with child node and no parent node is called root node. As shown in Figure 1, this part refers to the first and second stage. For nodes that are discrete, their effect on the child node can be quantitively expressed through a conditional probability distribution (CPD) that shows the influence of parent nodes. This part takes the three processes on the last section of model construction.
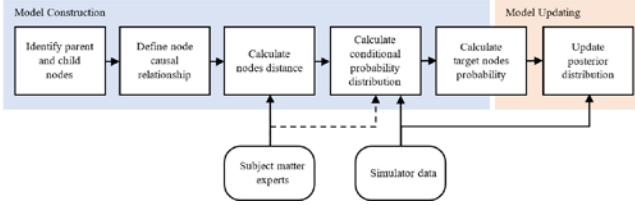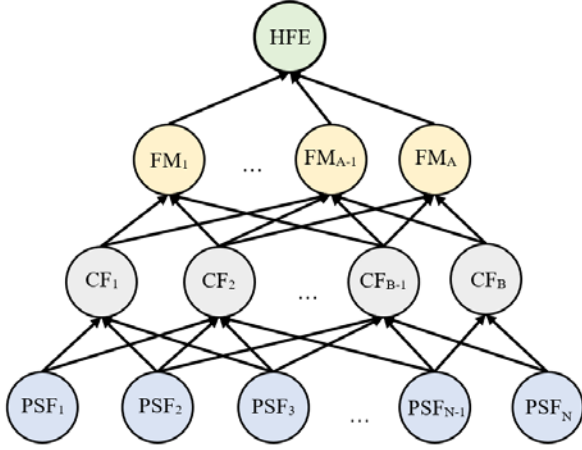
Figure 1. Flow of model construction and updating.

Figure 2. Bayesian network general construction

2.1.1 Identify nodes and causal relationship

The Bayesian network model has several advantages, such as the ability to clearly define causal factors. Figure 2 illustrates the causal factor mapping for the proposed model, which focuses on estimating the likelihood of human failure event (HFE) as the general objective of human factor analysis. Based on the simulator data, this study focuses on the Bayesian network depicted in Figure 2 in general and Figure 3 as the proposed model, which include three level of stages: failure mode, cognitive function, and performance shaping factors.
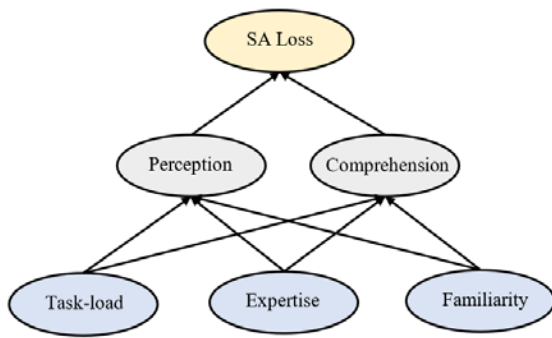

Figure 3. Bayesian network based on the simulator experiment

Failure Mode (FM). HFE can be described as many types as possible of activity or process during the work. Following the same term in hardware reliability, FM is constructed below the HFE to explain various types of supporting events that lead to the human error itself. This can range from processing information process, decision making process and taking an action process. In this study, SA loss is assigned as one of the FM that construct the HFE.

Cognitive Function (CF). FM are supported by various human mechanisms explained by CF nodes. CF is determined as a variable that can be examined and observed during the controllable environment such as simulator, but it is difficult to observe in work environment. As the simulator experiment design, CF is assigned as one or multiple dependent variables. As shown on the Figure 3, there are two CF nodes called perception and comprehension, that support the SA loss. Perception, as situation awareness definition, is the process consuming the cues from the environment, while the comprehension is the stage where this information from the cues combined with the working and long-term memory to create meaningful information for the work goal.

Performance Shaping Factors (PSF). The PSF are the lowest, or the root nodes of the modes, as other HRA methods also described. In relation to the simulator data, the PSF represents the independent variable, in this case task-load, expertise, and familiarity. Task-load refers to the complexity of the performed work, the expertise is determined by the skill and knowledge of the operator, and familiarity refers to experience with specific work environment hardware and scenario. These PSF nodes contribute to the SA loss by influencing the operator perception and comprehension.

2.1.2 Distance between nodes

The relative importance of each parent node in influencing the child nodes are established by considering the relative importance of one parent node compared to others. Røed suggested that this can be done using a weight w_i for each parent node i [12]. In this study, we employed analytical hierarchy process to conduct pairwise comparisons and determine the weight. We invited four experts from academia who have professional experience on board in engine department as described on Table 1. The experts provided their weighting answers individually.

Table 1 Profile of the expert as subject matter experts

No.	Professionals	Onboard	Duration
1	Academics	Chief Engineer	15 years
2	Academics	1st Engineer	5 years
3	Academics	Chief Engineer	14.5 years
4	Academics	Chief Engineer	10 years

The weighting process resulted in the construction of perception as cognitive function, where the task-load has a relative importance of w_T=0.26, expertise with w_E=0.41, and familiarity with w_F=0.33. Similarly, in the construction of comprehension, task-load had a relative importance of w_T=0.13, expertise with w_E=0.55, and familiarity with w_F=0.32. Lastly, the weight is also applied to determine the importance of each CF in construction FM. In this case, perception had w_P=0.14, while compression has w_C=0.86, in constructing the SA loss as the main goal HRA in this study.

$$D_j = \left| \sum_{i=1}^{n} D_{ij} W_i \right| \qquad D_j \in [0,1] \qquad (1)$$

The relative importance result is used to define the distance between the child node and its parent nodes. The probability of the child node in certain states should be assigned smaller if the parent nodes are in different states. Take the example of CF perception node, if its PSF task-load node is in the easy state, PSF expertise node in high state and PSF familiarity node in good state, the PSF node perception should have higher probability of being in the high state compared to medium and low states. Røed suggested the conditional probability can be measured using this distant methods [12]. Considering the direction of change its parent nodes, we applied the Li method that modified the equation into the absolute value, as mention in the equation above, to distribute the probability of the child node [16].

2.1.3 Conditional probability for child nodes

$$P_j = \frac{e^{-RD_j}}{\sum_{j=a}^{c} e^{-RD_j}} \qquad P_j \in [0,1] \tag{2}$$

To assign the probability distribution to each state on child nodes, the formula from Røed is used [12]. In the equation, the numerator is used to determine the probability of each state of the child node in the focus, while the denominator is normalization factor that all state of the child sum up to 1. The higher R index will make lower probability where the node in focus is state distant from the parent's states. While Røed uses the expert input to decide the R index, this study demonstrates how to reduce the uncertainty by using the simulator result. Approach using simulator data to decide R index has been demonstrated by Li in the nuclear power plant analysis [16]. In this study, we apply different approaches to decide the R index.

Table 2. Perception and comprehension sensitivity from simulator experiment

Variable	Perception, Comprehension							
Task Load	Easy				Complex			
Familiarity	Good		Bad		Good		Bad	
P1	1.40	0.54	0.42	0.42	1.81	1.38	0.95	0.00
P2	1.38	0.42	1.40	0.95	1.94	0.42	0.00	0.00
P3	2.35	1.38	1.38	0.42	1.40	0.95	-0.42	0.00
P4	2.77	0.95	0.95	0.42	2.35	0.00	0.86	0.00
P5	2.35	0.42	1.38	1.40	1.40	0.54	0.97	0.42
P6	2.77	1.94	1.93	0.00	2.77	0.42	2.77	0.54
P7	2.35	0.95	1.38	0.95	0.97	1.81	1.93	-0.54
P8	2.77	0.42	1.81	0.95	0.97	0.00	1.40	0.95
P9	1.38	1.38	1.94	0.95	1.94	1.81	2.35	0.97
P10	1.40	0.54	0.95	0.42	1.81	0.00	-0.54	-0.95
P11	1.38	0.00	1.81	0.00	0.97	0.00	1.38	0.00
P12	1.40	0.42	1.38	0.95	1.81	0.42	0.54	0.00
P13	1.81	-0.54	0.97	-0.42	0.97	-0.42	-0.43	0.00
P14	2.77	0.00	0.95	-0.42	0.54	0.54	0.97	0.00
P15	1.94	0.00	1.94	-0.42	0.95	-0.42	0.86	0.00
P16	2.35	0.42	1.38	0.54	2.35	0.42	0.97	-0.95

The simulator result is explained in Table 2. The simulator experiment is designated with two task-load states: easy where the simulation is under ocean going scenario, and complex for entering-port scenario where stand-by engine procedure must be conducted. Familiarity has a bad state where the first measurement is taken, and good when the repeated measurement is taken. The sensitivity of perception and comprehension level of situation awareness measured by freeze-probe methods under signal detection theory (SDT) [17,18]. The higher number on it means the ability of each participant to discriminate between the false alarm and correct rejection of the questioned parameter is better. The participants P1 until P8 are categorized with expertise state high, and P9 until P16 categorized with expertise state low. Within this data, two thresholds are applied to categorize the measurement result into three states: low, mediate, and high. Important to note that the average mediate state should have the larger number of distributions.

Table 3. Simulator result distribution for perception

Parent or CF	Pr (CF | Parent State)							
Task-load	Easy				Complex			
Experience	High		Low		High		Low	
Familiarity	Good	Bad	Good	Bad	Good	Bad	Good	Bad
Percep- High	0.75	0.00	0.25	0.00	0.25	0.13	0.13	0.13
tion Mediate	0.25	0.80	0.75	1.00	0.75	0.50	0.75	0.38
Low	0.00	0.13	0.00	0.00	0.00	0.38	0.13	0.50

Table 4. Simulator result distribution for comprehension

Parent or CF	Pr (CF | Parent State)							
Task-load	Easy				Complex			
Experience	High		Low		High		Low	
Familiarity	Good	Bad	Good	Bad	Good	Bad	Good	Bad
Compre- High	0.13	0.13	0.00	0.00	0.13	0.00	0.13	0.00
hension Mediate	0.88	0.75	0.50	0.50	0.63	0.38	0.38	0.13
Low	0.00	0.13	0.50	0.50	0.25	0.63	0.50	0.88

The assigned conditional probability retrieved from the simulator result was then compared with the conditional probability from the Bayesian network with R value varied from 0 to 5. We simply assign RMSE function as stated below to decide the comparison between probability from simulator data e and the probability from Bayesian model f with scenario i and the state j. The m and n explains the numbers of scenario and number states respectively. The R value was then decided based on the comparison which has the lower RMSE. In this case, R=3.15 and R=1.22 are assigned for calculating with Equation 2 to calculate the CPD for perception and comprehension nodes. The result of each CPD is shown in Table 5 and Table 6 for the perception node and comprehension node, respectively.

$$RMSE = \sqrt{\frac{1}{mn} \sum_{i}^{m} \sum_{j=1}^{n} \left(e_{ij} - f_{ij} \right)^2} \tag{3}$$

Table 5. Perception CPD

Parent or CF	Pr (CF | Parent State)							
Task-load	Easy				Complex			
Experience	High		Low		High		Low	
Familiarity	Good	Bad	Good	Bad	Good	Bad	Good	Bad
Percep- High	0.80	0.34	0.23	0.10	0.44	0.13	0.11	0.03
tion Mediate	0.17	0.55	0.64	0.46	0.46	0.64	0.55	0.17
Low	0.03	0.11	0.13	0.44	0.10	0.23	0.34	0.80

Table 6. Comprehension CPD

Parent or CF		Pr (CF | Parent State)							
Task-load		Easy				Complex			
Experience		High		Low		High		Low	
Familiarity		Good	Bad	Good	Bad	Good	Bad	Good	Bad
Compre-hension	High	0.54	0.36	0.25	0.19	0.46	0.29	0.23	0.16
	Mediate	0.30	0.42	0.46	0.35	0.35	0.46	0.42	0.30
	Low	0.16	0.23	0.29	0.46	0.19	0.25	0.36	0.54

2.1.4 Calculating target node probability

The SA loss as the failure mode in the proposed Bayesian model in this study is binary state; means either the condition meets the true or false state. While the node in the cognitive function and performance shaping factor level is constructed by three states, the following step is necessary to align it with the failure mode level. The condition probability distribution in the cognitive function is assigned as equation below.

$$P_j = P_{basis} \sum_{i=1}^{n} w_i \sum_{k=a}^{f} P_{ik} Q_{ik} \tag{4}$$

The conditional probability of failure mode node P_j is calculated based on the probability of each parent cognitive function node P_{ik} with states $k=a,b,c$. The weighted value w_i is retrieved from subject-matter experts like the previous step in weighting the PSF nodes. Q_{ik} is the corresponding adjustment for the P_{basis}. We follow the practice by Li to use P_{basis}=0.01 as the basic probability of error in SA, and 100-fold as the number of compromise adjustment. This configuration will made the adjustment factor for parent nodes state to have Q_{ik}=0.01, Q_{ik}=1, Q_{ik}=100 for parent node high, mediate, and low respectively.

Given the example, when the Task-load is Complex, Experience is high, and familiarity is bad, the probability distribution of perception is 0.13, 0.64, and 0.23 from Table 5, and the probability distribution comprehension is 0.29, 0.46, and 0.25. Where the weighting can be retrieved from the previous explanation of subject matter experts. The probability of SA loss given this condition can be calculated as follow:

$$P_{fail} = P_{basis} \left(0.14 \times \left(0.13 \times 0.01 + 0.64 \times 1 + 0.23 \times 100 \right) + 0.86 \times \left(0.29 \times 0.01 + 0.46 \times 1 + 0.25 \times 100 \right) \right) =$$
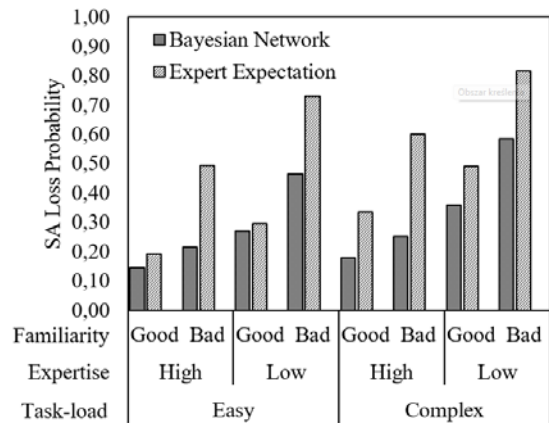$$= 0.253$$

Figure 4. SA loss probability comparison from Bayesian network and expert expectation

The result from the Bayesian network modelling is shown in Figure 4. The comparison includes the expert expectation that was measured using the free scale on the paper. The expert given the condition of simulator result and asked how likely the participant will fail in attaining the SA in each scenario. The validation is not with an aim to validate until the level of unit, but the tendency of the pattern comparison. It can be accepted the Bayesian network result is follow the expert expectation pattern at the most, except in interception of scenario familiarity is good and expertise is low, the Bayesian network result have tendency to have higher probability in these scenarios.

2.2 Model updating

$$P(H|D) = P(H) \frac{P(D|H)}{P(D)} \tag{5}$$

In this model updating method, the calculated probability for each combination of performance-shaping factor, cognitive function, and failure mode will be re-calculated given the condition if the new simulator data exists. The method is based on the Bayesian inference as explained in the equation below. The aim is to update the posterior distribution $P(H|D)$ with the prior known information distribution $P(H)$ with collected data from likelihood model $P(D|H)$ that normalized with probability of distribution $P(D)$ The Bayesian inference able to use every time the new data exists, means the posterior data from one modelling process became prior data for the next modelling study or stage.

2.2.1 Prior distribution construction

Regarding SA loss as the one factor for human error event, it can be explained using the binomial distribution explained by equation below. The distribution expresses the uncertainty about the number of failures x occurred in the given condition of demands n with the parameter probability of failure p. This parameter p is uncertain that sometimes derived from expert judgment or data. Groth et al suggest the p can be retrieved by using the Bayesian inference [19].

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x} \tag{6}$$

We used the same approach as the previous method in deciding the p by using the beta distribution. Probability density function as shown in below equation express the beta function $B(\alpha, \beta)$ with function to normalize the distribution. From with explanation from Groth et al, the distribution can be conjugated with binomial distribution, where posterior distribution parameters α_{post} and β_{post} are assigned using equation below [19]. The next step in using Bayesian inference is with assigning the value of α and β.

$$f\left(p;\alpha,\beta\right) = \frac{p^{a-1}\left(1-p\right)^{\beta-1}}{B\left(\alpha,\beta\right)} \qquad (7)$$

$$a_{post} = a_{prior} + x$$

$$\beta_{post} = \beta_{prior} + n - x \qquad (8)$$

To specify the prior distribution p_0, we applied the constrained non-information (CNI) distribution [19] As shown in the following equation, the a is estimated as the number of failure contained in the prior distribution, and the denominator ($\alpha+\beta$) is considered as the number of demands. The beta distribution is constructed by $\alpha=0.5$ and β derived from the equation constraint. The extract prior distribution p_0 from the new existing simulator data is shown in Table 7.

$$E\left(Beta\left(\alpha,\beta\right)\right) = \frac{\alpha}{\alpha+\beta} = E\left(p_0\right) \qquad (9)$$

Table 7. Prior distribution (p_0)

Expertise	Task-load	Familiarity	$E(p_0)$	α	β	Prior Distribution
High	Easy	Good	0.145	0.5	2.9	p_0, Beta (0.5, 2.9)
		Bad	0.215	0.5	1.8	p_0, Beta (0.5, 1.8)
	Complex	Good	0.179	0.5	2.3	p_0, Beta (0.5, 2.3)
		Bad	0.253	0.5	1.5	p_0, Beta (0.5, 1.5)
Low	Easy	Good	0.269	0.5	1.4	p_0, Beta (0.5, 1.4)
		Bad	0.465	0.5	0.6	p_0, Beta (0.5, 0.6)
	Complex	Good	0.357	0.5	0.9	p_0, Beta (0.5, 0.9)
		Bad	0.584	0.5	0.4	p_0, Beta (0.5, 0.4)

2.2.2 Posterior distribution updating

Table 8. Simulator result distribution

Variable	Perception, Comprehension							
Task Load	Easy				Complex			
Familiarity	Good		Bad		Good		Bad	
P17	1.81	1.38	1.81	0.54	0.95	0.00	2.35	0.95
P18	1.81	0.42	1.40	0.00	0.43	0.00	0.95	0.00
P19	2.35	0.95	2.35	0.95	1.40	0.54	1.40	0.42
P20	1.94	0.54	1.81	0.54	1.40	0.00	2.35	0.54
P21	1.81	0.43	1.94	1.38	1.40	0.97	1.81	0.97
P22	2.35	1.38	2.35	2.35	1.40	0.97	1.38	1.38
P23	1.81	2.35	1.38	0.42	2.35	0.42	0.00	1.40
P24	2.35	1.38	2.35	0.95	2.35	1.38	1.81	0.42

The process updating the prior distribution with new simulator data can be obtained by utilizing Equation 8. The aims is calculating the posterior distribution p_1 based on the prior distribution p_0 and the new distribution parameters α_{post} and β_{post}. For thus, number of failures x occurred in the given condition of demands n need to be defined. Table 8 illustrates the additional simulator result from eight participants with expertise level assumed to be high. In this step, a different approach is used to highlight the participant who has perception or comprehension equal or below 0 are categorized as loss in SA. Based on the categorization, as also shown on Table 9, there are eight trials for each scenario combination, then number of opportunities can be assigned $n=8$, and the number of failure x is assigned in each scenario combination. The α_{post} and β_{post} is retrieved using Equation 8, and the updated SA loss for selected scenario.

Table 9. Posterior distribution for (p_1)

Expertise	Task-load	Familiarity	Data (x/n)	α	β	Posterior Distribution	$E(p_1)$
High	Easy	Good	0/8	0.5	10.9	p_1, β (0.5, 10.9)	0.044
		Bad	1/8	1.5	8.8	p_1, β (1.5, 8.8)	0.145
	Complex	Good	3/8	3.5	7.3	p_1, β (3.5, 7.3)	0.324
		Bad	2/8	2.5	6.4	p_1, β (2.5, 6.4)	0.282

3 DISCUSSION

HRA methods to measure and predict human error in maritime operation are already developed with various methods. The Bayesian network and Bayesian inference is used in this study to demonstrate method which use the simulator data for the probability distribution calculation. The Bayesian network has the advantage of treating the dependencies of PSF, which often treat independent of each other in recent HRA methods. Further, Bayesian inference concept in the second stage demonstrated another possibility to update the probability distribution of human error if the new data from the simulator exists. This has an advantage since there is no necessity to reconstruct the model.

In the first stage of model construction, the three-node level is introduced. FM were introduced as the possible process or activity that support event led to the human error. Situation awareness is introduced as single FM in this study. CF is introducing as human mechanism in construct the activity, its assigned as dependent variable in simulator experiment. Perception and comprehension were introduced in constructing situation awareness. Last, PSF is introduced as the lowest node in the model. In relation to the simulator experiment, the PSF represents the independent variable. Three PSF are assigned in the model: task-load, expertise, and familiarity. The proposed methods to replace the expert judgment in deciding R-value are demonstrated to reduce the subjective expert judgment uncertainty. However, the input from expert judgment is still mandatory to put the weighting factors between the nodes in the same level node. The second stage demonstrates updating the prior distribution with the Bayesian inference methods. Here the additional simulator data is used to recalculate it into posterior distribution.

The human error probability as the output from the model was compared to the expert expectation for each scenario combination. It is observed that Bayesian network results follow the same pattern as the expert expectation input. However, several comparisons such the scenario with familiarity is good and expertise is low, is have higher evaluation from the Bayesian Network. Thus, the Bayesian network in this model still lacks sensitivity, especially in the scenario which has close result of human error probability.

The general construction of HRA with Bayesian network offers the flexibility to cover more PSF into the model. However, considering the PSF that can be observed in the simulator will limit to detect all PSF exist in other studies. This is the coming limitation of the Bayesian network in HRA context. Thus, the incomplete representation cannot explain the complex

relationship between variables. The second limitation from employing simulator data is, during the session, the participant is well now their performance is being observed. This implies the participants put more effort during the simulation. This must be noted since human error calculation is aimed to measure the error during the normal condition.

Work onboard a ship is divided into big portions of navigation and engine operation work. Defining general PSF that includes the two areas is the future challenge that must be considered. The remaining challenge in HRA is the various definition of the denominator in probability, the number of opportunities for human error which remain wide interpretation for each method. An approach to combine the result of HRA needs an adjustment method to tackle this challenge. Similar to the reliability in hardware, such study may become reference under the IMO guideline to include in the formal safety assessment.

REFERENCE

[1] Wróbel K. Searching for the origins of the myth: 80% human error impact on maritime safety. Reliability Engineering & System Safety 2021;216:107942. https://doi.org/10.1016/j.ress.2021.107942.

[2] Grech MR, Horberry T, Smith A. Human Error in Maritime Operations: Analyses of Accident Reports Using the Leximancer Tool. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2002;46:1718–21. https://doi.org/10.1177/154193120204601906.

[3] Cordon JR, Mestre JM, Walliser J. Human factors in seafaring: The role of situation awareness. Safety Science 2017;93:256–65. https://doi.org/10.1016/j.ssci.2016.12.018.

[4] International Maritime Organization. International convention on standards of training, certification and watchkeeping for seafarers. London: International Maritime Organization; 2011.

[5] Patriarca R. Human reliability analysis: Exploring the intellectual structure of a research field. Reliability Engineering and System Safety 2020;203:27. https://doi.org/10.1016/j.ress.2020.107102.

[6] International Maritime Organization. Revised Guidelines For Formal Safety Assessment (FSA) For Use in The IMO Rule-making Process. 2018.

[7] Yang ZL, Bonsall S, Wall A, Wang J, Usman M. A modified CREAM to human reliability quantification in marine engineering. Ocean Engineering 2013;58:293–303. https://doi.org/10.1016/j.oceaneng.2012.11.003.

[8] Liu J, Aydin M, Akyuz E, Arslan O, Uflaz E, Kurt RE, et al. Prediction of human–machine interface (HMI) operational errors for maritime autonomous surface ships (MASS). J Mar Sci Technol 2021. https://doi.org/10.1007/s00773-021-00834-w.

[9] Ahn SI, Kurt RE, Turan O. The hybrid method combined STPA and SLIM to assess the reliability of the human interaction system to the emergency shutdown system of LNG ship-to-ship bunkering. Ocean Engineering 2022;265:112643. https://doi.org/10.1016/j.oceaneng.2022.112643.

[10] Mkrtchyan L, Podofillini L, Dang VN. Bayesian belief networks for human reliability analysis: A review of applications and gaps. Reliability Engineering & System Safety 2015;139:1–16. https://doi.org/10.1016/j.ress.2015.02.006.

[11] Park J, Jung W, Kim J. Inter-relationships between performance shaping factors for human reliability analysis of nuclear power plants. Nuclear Engineering and Technology 2020;52:87–100. https://doi.org/10.1016/j.net.2019.07.004.

[12] Røed W, Mosleh A, Vinnem JE, Aven T. On the use of the hybrid causal logic method in offshore risk analysis. Reliability Engineering & System Safety 2009;94:445–55. https://doi.org/10.1016/j.ress.2008.04.003.

[13] Vinnem JE, Bye R, Gran BA, Kongsvik T, Nyheim OM, Okstad EH, et al. Risk modelling of maintenance work on major process equipment on offshore petroleum installations. Journal of Loss Prevention in the Process Industries 2012;25:274–92. https://doi.org/10.1016/j.jlp.2011.11.001.

[14] Aydin M, Akyuz E, Turan O, Arslan O. Validation of risk analysis for ship collision in narrow waters by using fuzzy Bayesian networks approach. Ocean Engineering 2021;231:108973. https://doi.org/10.1016/j.oceaneng.2021.108973.

[15] Musharraf M, Bradbury-Squires D, Khan F, Veitch B, MacKinnon S, Imtiaz S. A virtual experimental technique for data collection for a Bayesian network approach to human reliability analysis. Reliability Engineering & System Safety 2014;132:1–8. https://doi.org/10.1016/j.ress.2014.06.016.

[16] Li P, Zhang L, Dai L, Li X-F. Study on operator's SA reliability in digital NPPs. Part 3: A quantitative assessment method. Annals of Nuclear Energy 2017;109:82–91. https://doi.org/10.1016/j.anucene.2017.05.019.

[17] Nizar AM, Miwa T, Uchida M. Measurement of situation awareness in engine control room: approach for non-technical skill assessment in engine resource management. WMU J Marit Affairs 2022;21:401–19. https://doi.org/10.1007/s13437-022-00270-6.

[18] Hogg DN, Folles⊘ K, Strand-Volden F, Torralba B. Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. Ergonomics 1995;38:2394–413. https://doi.org/10.1080/00140139508925275.

[19] Groth KM, Smith CL, Swiler LP. A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods. Reliability Engineering & System Safety 2014;128:32–40. https://doi.org/10.1016/j.ress.2014.03.010.