Enhanced Hazard Analysis and Risk Assessment for Human-in-the-Loop Systems

David Kaber and **Maryam Zahabi**, North Carolina State University, Raleigh

Objective: The objective of this study was to enhance the existing system hazard analysis (SHA) technique by introducing the concepts of human and automation reliability quantification as well as fuzzy classification of system risks. These enhancements led to formulation of a new overall system risk-reliability score.

Background: Many system safety analysis methods focus on individual physical component failure. Some human reliability analyses (HRA) consider human-automation interaction in determining system failure rates. There is no system safety analysis technique that quantifies the impact of human and automation reliability on the risk of hazard exposure.

Method: Classification of the probability and severity of hazard exposure is typically made in terms of linguistic rather than numerical variables. Fuzzy sets are applicable for transforming linguistic classifications to numerical quantities. We focused on using fuzzy sets to define overlapping bands of system risk exposure with reference to the hazard risk categories defined in MIL-STD 882B. Fuzzy sets were also used for human-automated system reliability classification.

Results: Introduction of human and automation reliability assessment in the SHA allows for definition of a system risk-reliability modeling space. The enhanced SHA (E-SHA) technique yields a mishap risk index, which is projected based on a composite assessment of human-automated system reliability at the time of operation. The E-SHA was compared with one of the most advanced HRA techniques.

Conclusion: The E-SHA technique supports broader safety control recommendations and provides comparable, if not more detailed, results than prior systems safety and HRA techniques.

Keywords: systems safety analysis, hazard analysis, risk assessment, human-automaton interaction, system reliability

Address correspondence to David Kaber, ISE Department, North Carolina State University, 400 Daniels Hall, Raleigh, NC 27695-7906, USA; e-mail: dbkaber@ncsu.edu.

HUMAN FACTORS
Vol. 59, No. 5, August 2017, pp. 861–873
DOI: 10.1177/0018720817693357

INTRODUCTION

The present study focused on systems safety analysis. Safety can be considered an outcome of human factors and ergonomics (HF/E) considerations in system design; alternatively, HF/E can also be viewed as an aspect of safety. In either case, safety is a critical technical aspect of what we do in HF/E. Historically, systems safety analysis methods have been developed to account for hazard exposures in automated system operation. However, in many contemporary complex systems, human operators are maintained in control loops in order to account for limitations of automation technologies. With this in mind, it is necessary to develop methods that facilitate assessment of system risk exposure associated with the reliability of humans and automated agents (independently and/or jointly). There is also a need to integrate risk and reliability assessments in order to provide a broader range of hazard control recommendations.

In this research, we demonstrate an advance in system safety analysis and briefly consider a space system in an example analysis. However, there are other domains (e.g., supervisory control in power plant operations, petrochemical process control, air traffic control, etc.) that also require complex human-automation interaction to which the methodology presented in this paper can be applied.

In the following subsections, we describe system safety and human reliability analysis techniques and provide a brief introduction to fuzzy sets as background for our proposed methodology. Fuzzy sets are key to our advanced safety analysis in terms of a method for translating linguistic system risk classifications to numerical values.

System Safety Analysis Methods

There are many formal system safety analysis techniques documented in the literature, and the majority of methods are referred to with acronyms. For concise reporting, we use

these acronyms, and for reader convenience, we provide a list of the acronyms in the Appendix along with definitions. The system safety analysis methods include the preliminary hazard analysis (PHA) and event tree analysis (ETA), which are primarily focused on individual physical component failures (e.g., defects, command faults; Bahr, 1997; Ericson, 2005). Furthermore, very few system safety analysis methods, save the systems hazard analysis (SHA), consider the interaction of components or human and hardware (Clemens, Simmons, & Cincinnati, 1998). There is no system safety analysis method that provides the capability to quantify the impact of human/automation reliability on the level of risk of hazard exposure. (Risk is conceptually defined as the product of the probability of hazard exposure occurring by the severity of outcomes.) In addition, these methods have not made consideration of system aging, cumulative environment exposure, and degradations in human fitness for duty and skills.

Some system safety analysis techniques provide a basis for prioritizing use of engineering resources to control specific types of hazard exposure for equipment and human targets. For example, failure modes, effects, and criticality analysis (FMECA) identifies process failure modes, causes, negative effects, priority of risks, and recommended actions. It uses rating scales for quantifying outcome severity and occurrence of hazard exposure. FMECA also assigns rankings of likelihood of sensor or test technology success (i.e., the potential for revealing failures) and integrates a risk score with the detection ranking to yield a risk priority number (RPN; Ericson, 2005). However, the RPN does not account for system aging and degradations in reliability.

The SHA is the only system safety analysis technique that accounts for hazards due to interactions among components. This method is typically applied after a subsystem hazard analysis has mapped all potential piece-part, subassembly failure modes (Ericson, 2005). The SHA is an "inductive" method that identifies sources of hazards and mechanisms (events) leading to negative outcomes. The method yields baseline system risk assessment scores and revised scores considering recommended controls. In general, system safety analysis techniques are categorized as inductive or "deductive" methods. An example of a deductive method is fault tree analysis (FTA) in which an analyst hypothesizes a negative system outcome and attempts to deduce mechanisms/events and underlying component failure sequences that might have caused the "top" event or fault of the tree. The SHA is rare among formalized inductive system safety analysis techniques in terms of content.

Human Reliability Analysis Methods

Swain (1990) defined human reliability analysis (HRA) as "any method by which human reliability is estimated." Human reliability is likelihood of a task being performed consistently correct by a human operator across repeated trials. Human reliability plays an important role in human-automation interaction in complex systems operation. While reliable human performance leads to accomplishment of missions, human errors can limit the potential for automation to positively impact productivity and safety and result in damage to a system or incomplete missions. The likelihood of erroneous human actions in task performance can be estimated using HRA techniques. Such techniques are typically classified as "first" and "second generation" (Swain, 1990). The technique for human error prediction (THERP) is the most well-known first-generation HRA method. The goal of THERP is to estimate the probability of successful human performance, and results are usually represented in event trees (DeFelice, Petrillo, Carlomusto, & Ramondo, 2012). THERP also classifies human errors as omissions (omitting steps in a process) or commissions (taking incorrect actions or correct actions at the wrong time). Despite the capabilities of THERP, Kim (2001) recommended using second-generation HRA methods for analysis of human-machine systems. The most well-known second-generation HRA methods include: a technique for human error analysis (ATHEANA) and the cognitive reliability and error analysis method (CREAM). ATHEANA was developed to solve some problems with previous HRA methods, such as providing the capability to address dependencies among errors of commission and represent human-system interaction more realistically in error analysis (Cooper, Wreathall, Thompson, Drouin, & Bley, 1996). DeFelice et al. (2012) said that ATHEANA considers a broader

set of performance safety factors (PSFs) in HRA and provides for a more thorough understanding of the role of human performance in accidents than previous first-generation methods. The main goal of ATHEANA is to estimate human error probability (HEP). CREAM is another second-generation HRA method, which was developed by Hollnagel (1998). The general steps as part of the CREAM methodology include: task analysis, context description, initiating events specification, error estimation, and control determination. There are essentially two versions of the method, including: basic and extended. The types of human errors that can be identified by the basic version are actions at the wrong time, in the wrong sequence, and so on. The extended version can identify more specific errors such as incorrect stimulus observation and/or identification. (Chandler et al., 2006).

Another classification of HRA methods is in terms of quantitative versus qualitative methodologies (Embrey, 2004). Qualitative methods attempt to identify most likely errors while quantitative methods provide estimates of HEP. An example quantitative method is THERP. Another quantitative method is standardized plant analysis risk-human reliability analysis (SPAR-H). This method was first developed to estimate HEP at nuclear power plants. Forester et al. (2012) said that one of the drawbacks of SPAR-H is that it relies heavily on an analyst's judgment to decide which PSFs to include in estimating HEP. The only PSF in the method that accounts for human-machine interaction is labeled as *ergonomics and human machine interaction* (Blackman, Gertman, & Boring, 2008). SPAR-H breaks down error probabilities into two categories, including diagnosis and action failures (Gertman, Blackman, Marble, Byers, & Smith, 2005). Qualitative HRA methods include human factors process failure mode and effect analysis (HF-PFMEA) and action error analysis (AEA). HF-PFMEA is considered to be one of the most comprehensive HRA techniques as it identifies human error types as acts of commission or omission and considers the likelihood of human error (*impossible, possible*, and *highly likely*) as well as the consequence of severity in error analysis. It is also the only HRA method that is specifically designed for aerospace applications (Chandler et al., 2006). Although the method decomposes hazard risk exposure based on existing system conditions, HF-PFMEA does not involve determining a revised risk score accounting for any control recommendations or implementation of hazard control methods. Different from the HF-PFMEA method, the AEA method (Suokas, 1982) lists deviations of system states from nominal conditions, including an action occurring *too early, too late*, or for *too long* of a duration. The method also identifies different errors in detail, such as actions applied to incorrect objects, actions not taken (errors of omission), and actions taken in the wrong order (errors of commission). Although we found some studies on advanced AEA (e.g., Bligård & Osvalder, 2014), the method itself is not well defined in the literature.

In summary, the main goal of HRA methods is to identify HEP. None of the existing HRA techniques consider system reliability as a function of system age or degradations in performance capacity, including the age of equipment or declines in human operator perceptual, cognitive, and motor responses. Analysis of human-machine interactions with these methods is basically limited to defining performance safety factors. However, error classifications in HRA methods are more detailed than in SSA methods.

The present study defines a new and expanded approach for overall system reliability estimation and risk exposure quantification. HRA methods can be used for identification of human error types and estimation of HEPs. These outcomes can be combined with automation reliability assessment to produce an overall system reliability estimate. The system reliability estimate depends on the type of automation component and/or human and automation integration, including serial or parallel systems. Once the system reliability is estimated, it can be integrated with other data from the SHA method, as an example, to provide a basis for overall system risk-reliability assessment. The general relationships among the HRA methods, automation reliability analysis, and the proposed enhanced SHA (E-SHA) method are summarized in Figure 1. Developing the E-SHA methodology was the focus of this study.

Introduction to Fuzzy Sets

In this section, we introduce the concept of fuzzy sets as a fundamental basis for enhance-

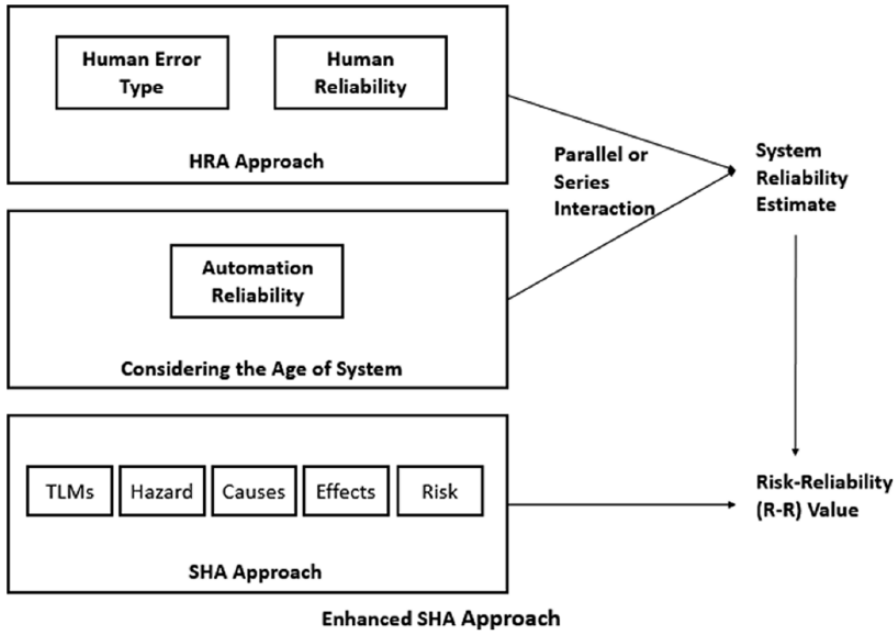

Figure 1. Relation of reliability assessment, human reliability analyses (HRA) methods, and an enhanced system hazard analysis (SHA) method. TLMs = Top level mishaps.

ment of the SHA methodology. In classical set theory, an individual is either a member or not a member of a set, and set boundaries are defined precisely. However, many real-word applications cannot be described using classical set theory. Fuzzy set theory allows for partial membership of multiple sets and represents a more generalized theory. Zadeh (1975) defined linguistic variables as "variables whose values are words or sentences in a natural or artificial language" (p. 199). These variables are best defined using fuzzy numbers. In the system safety literature, classifications of probability (e.g., *frequent, probable, occasional, remote, impossible*), severity (e.g., *catastrophic, marginal, negligible*), and risk (e.g., *unacceptable, undesirable, acceptable with review, acceptable without review*) are common and are based on linguistic variables since their values are linguistic rather than numerical. Such variables can be better defined for systems safety analysis purposes by using fuzzy functions. Ironically, discrete and exclusive risk classifications of hazards have historically been based on subjective ratings or qualitative assessments of system exposure frequency and severity of exposure outcomes (e.g., MIL-STD 882B, 1984).

Most common membership functions in the fuzzy literature include triangular, trapezoidal, and Gaussian functions (Chen & Pham, 2000). A triangular number, which is the simplest fuzzy function, is given as, $A = (a_1, a_2, a_3)$, and an example distribution is shown in Figure 2 (where a_1 = minimum, a_2 = mean, a_3 = maximum). The degree of membership in the fuzzy set, or the grade of membership of the element x, is quantified as $\mu_{(x)}$. In the present study, we used such membership functions to make the classification of hazard risk exposure, as well as human and automation reliability classifications, more realistic. Membership functions allow for continuous and fuzzy classifications of hazard risk as well as agent reliability levels. Such an approach also supports a broader range of safety control measure specification as a result of specific hazard exposures having some degree of membership in multiple classes of risk.

Motivation

As noted previously, the primary objective of HRA methods is to identify HEPs. The error classifications in HRA methods are detailed (as compared to SSA methods) and facilitate accuracy in HEP estimation. Unfortunately, the existing HRA techniques overlook overall

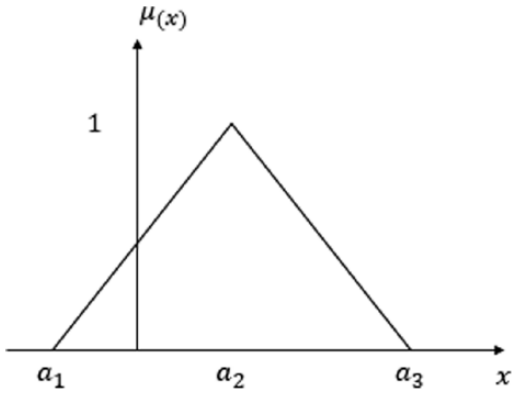

Figure 2. Triangular fuzzy number.

Frequency of Occurrence	Severity Category			
	I Catastrophic	II Critical	III Marginal	IV Negligible
A. Frequent	High (1)	High (3)	Serious (7)	Medium (13)
B. Probable	High (2)	High (5)	Serious (9)	Medium (16)
C. Occasional	High (4)	Serious (6)	Medium (11)	Low (18)
D. Remote	Serious (8)	Medium (10)	Medium (14)	Low (19)
E. Improbable	Medium (12)	Medium (15)	Medium (17)	Low (20)
F. Eliminated	Eliminated (0)			

Hazard-Risk Index	Criterion
1-5	Unacceptable
6-9	Undesirable
10-17	Acceptable with review
18-20	Acceptable without review

Figure 3. Adaptation of risk assessment matrix as part of MIL-STD-882B (1984) along with hazard risk index.

system reliability as a result of system life as well as human interaction with system components.

Beyond the characteristics of current HRA and SSA techniques, risk assessment matrices, such as that included in MIL-STD 882B, are the most common tool by which hazard exposure is currently evaluated through systems safety engineering processes. Unfortunately, little research has been conducted since the MIL-STD development to improve on these matrices for hazard classification and control measure identification. One major limitation of current matrices is a lack of high-resolution scales for ratings of severity and frequency of hazard exposure. Frequency of exposure scales typically include levels of: *improbable* (E), *remote* (D), *occasional* (C), *probable* (B), and *frequent* (A) (see Figure 3 for the MIL-STD 882B scales). Severity scales typically include levels of: *negligible* (IV), *marginal* (III), *critical* (II), and *catastrophic* (I). The overall risk index as part of the MIL-STD is numerical and ranges from a value of 20 (*acceptable without review*) down to a value of 1 (*unacceptable*); that is, a lower risk score indicates a more frequent and potentially severe hazard exposure (also see Figure 3). Risk categories as part of the standard are also defined based on discrete threshold values versus overlapping risk bands (i.e., there is no overlap of the hazard-risk index values in Figure 3). The limited range of scores and number of risk categories does not support hazard association with a broad set of mitigation strategies. Safety control measures are typically limited to general classes of design, safety devices, warnings, training, and personal protective equipment. Furthermore, in applying such methodology, risk scores are typically determined in the detailed system design phase and assume a reliable system operator and automated systems (i.e., fixed risk values at different points in time). Scores do not account for degradations in capacity of the human or automation or their interaction over time.

Toward enhancing the existing SHA technique, the first objective of this study was to use fuzzy sets to define overlapping hazard risk bands and define fuzzy human and automated system reliability classifications. The second objective was to formulate a new integrated system risk-reliability score by defining a higher-dimensional analysis space considering: (1) the likelihood of hazard exposure, (2) severity of potential outcomes, and (3) estimated levels of human-automation reliability.

RESEARCH CONTRIBUTION

Overlapping Risk Bands in SSA

Although the risk categories in MIL-STD 882B (1984) are defined by threshold values, in practice, such certainty in the classification of hazard exposure scores is inaccurate as subjective ratings of exposure frequency and severity are provided by safety engineers as bases for risk score determination. To address this issue, we proposed using triangular fuzzy functions to

TABLE 1: Fuzzy Risk Categories

Criterion	Triangular Fuzzy Number (minimum, mean, maximum)
Unacceptable	C_1 = (1,1,6)
Undesirable	C_2 = (5,7.5,10)
Acceptable with review	C_3 = (9,13.5,18)
Acceptable without review	C_4 = (17,20,20)

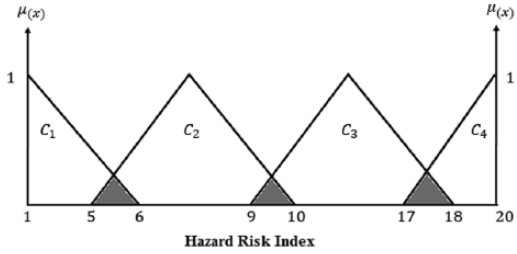

Figure 4. Overlapping fuzzy risk bands.

represent overlapping risk categories (C). Fuzzy membership functions (C_1, C_2, C_3, C_4) were defined based on the hazard risk categories of MIL-STD 882B and are shown in Table 1. The overlapping risk bands are graphically represented in Figure 4 (shaded areas among the triangular distributions). They indicate that a particular risk score can have partial membership in either of the adjacent risk functions. Decisions about threshold risk values depend on an analyst's approach to risk assessment. For example, if an analyst takes a "conservative" approach to risk analysis, any risk value between 5 and 6 could be considered unacceptable (and belonging to C_1). However, in a "risky" approach, the same hazard risk value might be categorized as undesirable (and belonging to C_2). Conservative and risky approaches for all the risk bands are shown in Figures 5 and 6.

Reliability Classification

In order to address our second objective of defining an integrated system risk-reliability index, there is a need to define reliability and identify classes of reliability values from the literature. Lewis (1994) defined reliability as the probability that a system will perform properly for a specified period of time under a given set of operating conditions. We further simplify this

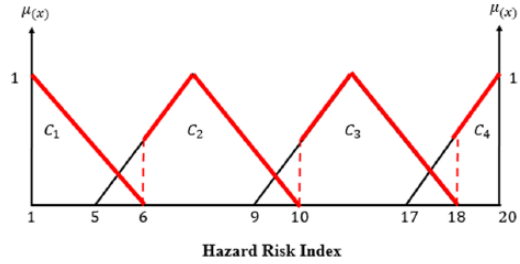

Figure 5. A conservative approach in fuzzy classification of risk scores.

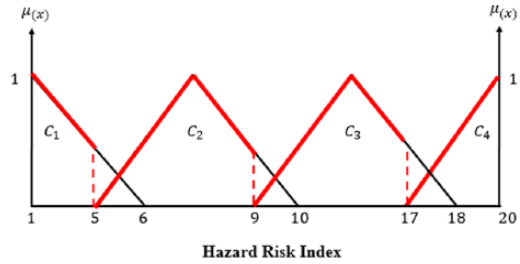

Figure 6. A risky approach to fuzzy classification of risk scores.

definition to an absence of loss, where *loss* refers to damage to personnel, equipment, or an environment during the course of mission operations. For classification of reliability of human-machine systems, Meister (1964) proposed the scheme presented in Table 2, including *high* (H), *medium* (M), and *low* (L) reliabilities. Of course, just like the categories of frequency of hazard exposure and severity of outcomes, Meister's levels of reliability are linguistic variables, and they can be better defined using fuzzy sets in order to account for any subjectivity in interpretation. Related to this, as described by Lewis, continuous random variables have extensive use in reliability analysis for the description of survival times, system loads, or repair rates. Since system reliability is commonly determined using continuous functions, such as exponential and normal (Gaussian), any fuzzy classifications of levels of reliability should also follow this structure. Figure 7a presents system reliability categories using fuzzy membership functions. (Note that this curve is not the classical system reliability/failure rate function based on time; see Figure 7b. The curve in Figure 7a is a reliability state membership function.) The high reliability category identified by Meister was

TABLE 2: Reliability Classifications and Defined Boundaries based on Meister (1964)

Category	Reliability
Low	0.6000 or less
Medium	0.6001 to 0.8000
High	0.8001 or more

defined using a negative exponential function because as reliability decreases, the likelihood of having high reliability (membership of a system in the category) will decrease. The medium reliability category was defined using a normal distribution because as the level of system reliability approaches 0.7, based on Meister's classification, the system is more likely to have medium reliability. Finally, the low reliability category was also defined using an exponential function because as reliability decreases, the likelihood of having low reliability will increase (i.e., membership in this functional state will be greater).

In addition, different phases of the proposed fuzzy classification of system reliability support the classical bathtub-shaped failure curve, as shown in Figure 7b (Lewis, 1994). In the "early system life" region, the initial failure rate caused by design faults or operator mistakes leads to a decrease in human-automated system reliability (decreasing reliability from 1 to 0.8). The constant failure rate region of the curve includes the lowest and most nearly constant failure rate (medium reliability from 0.8 to 0.6). Finally, the last part of bathtub shape, the "wear-out region," is usually caused by system fatigue or component aging, which results in decreasing human-automated system reliability (i.e., decreasing reliability from 0.6 to 0). Addressing this phase of the system lifecycle is of particular concern to the present work, that is, aging of automated and human agents and the impact of degradations in functional capabilities on overall system reliability.

A Three-Dimensional Risk-Reliability Analysis Space

The consideration of system component reliability values in SSA allows for definition of a multidimensional risk-reliability (R-R) modeling space. Multiple risk planes, defined in terms of frequency and severity of hazard exposure, can

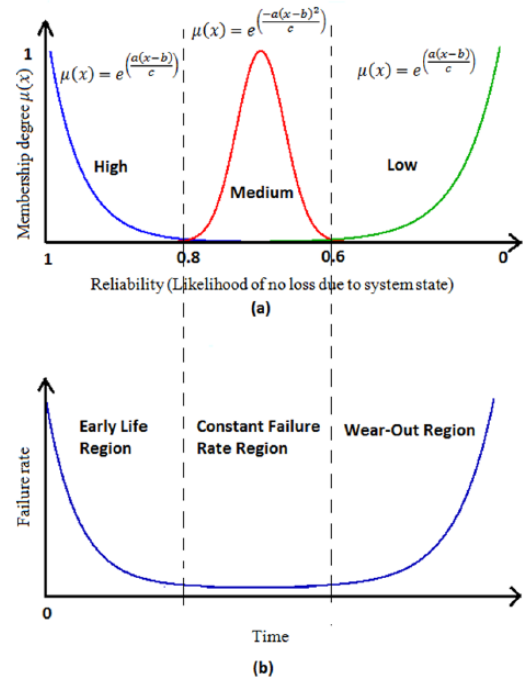

Figure 7. (a) Fuzzy reliability classification. (b) Bathtub-shaped failure curve.

be defined along the dimension of system reliability. Here, we note that system reliability is not simply the complement of risk. Reliability, as defined, is the probability of the absence of loss, whereas risk is the potential for loss given the occurrence of hazard exposure. Any confounding of these constructs is often the result of the misconception that hazard exposure implies loss. In fact, the probability of loss is only conditionally dependent on exposure and is mediated by system resilience and chance success of operation under off nominal conditions (for an accident sequence model, see Ramsey, 1985). Related to this, system failures, or loss of reliability, are not guaranteed with component failures (i.e., not all hazard exposures lead to loss). Figure 8 uses the classic MIL-STD 882B (1984) risk assessment matrix to conceptually represent our new R-R analysis space. Among other factors, reliability can represent system age. Such a multidimensional analysis space can be defined for a human operator or an automated component of a system in the context of complex operations. The range of hazard exposure frequency, severity of potential

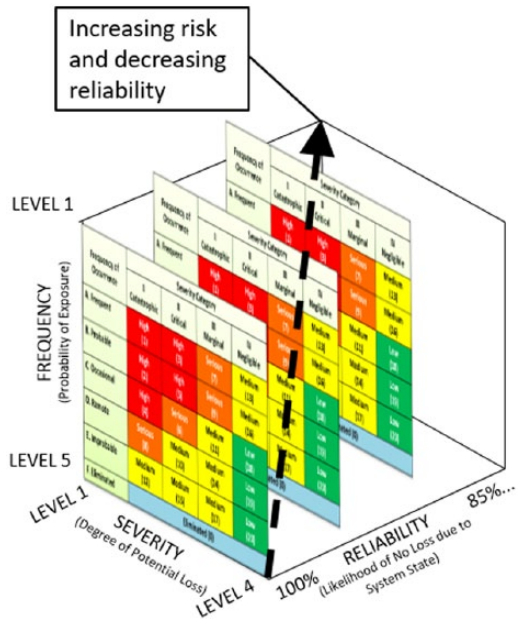

Figure 8. Risk-reliability 3D analysis space.

outcomes, and reliability values may vary among agents (human operator and automation). In addition, the rate of change in any one dimension of the analysis space relative to another may also vary among agents. Following the convention of MIL-STD 882B, lower values for severity and frequency are considered riskier, and increasing risk occurs with degraded system reliability.

Fuzzy Classification of System Hazard Exposure in Risk-Reliability Analysis Space

Since the probability of hazard exposure, severity of outcomes, and agent reliability are all linguistic variables, they should be defined using fuzzy membership functions. Therefore, fuzzy classifications can also be used to represent the multidimensional R-R space. For example, if a hazard probability is estimated as frequent with a membership degree of 1 ($\mu_P = 1$), its severity is identified as catastrophic with a membership degree of 1 ($\mu_S = 1$), and estimation of the composite human-automated system reliability is identified as low with a membership degree of 1 ($\mu_R = 1$), then the R-R value for this hazard would be discretely classified as 1AL (since $\mu_P \times \mu_S \times \mu_R \times = 1$) (see Point 1 in Figure 9). However, any

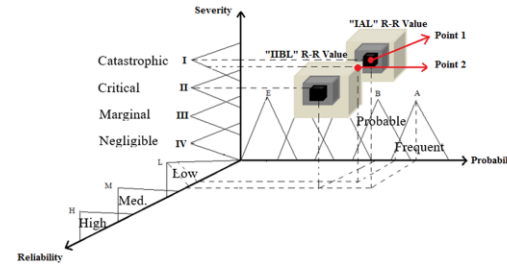

Figure 9. Fuzzy classification applied to 3D risk-reliability (R-R) analysis space.

uncertainty in any of the R-R dimensions would lead to a composite membership degree of less than 1, which would indicate that the R-R value could be categorized as, for example, either IIBL or IAL, depending on analyst's approach to the analysis (conservative vs. risky; see Point 2 in Figure 9).

Consequently, such fuzzy classification of R-R values could lead to recommendation of broader sets of controls for each hazard. That is, mitigation strategies considered appropriate for multiple classifications of risk exposure could be identified as appropriate for a particular hazard. It is important to note that R-R classification can be unidimensional, bidimensional, or three-dimensional depending on crisp or fuzzy classification of the probability of hazard exposure, severity of outcomes, and agent reliability. Figure 9 shows a scenario in which all dimensions are fuzzy. In addition, for simplicity of the graph, we assumed triangular fuzzy functions for all dimensions. However, as mentioned earlier, system reliability can be described using exponential and normal membership functions as well.

One of the shortcomings of "crisp" classifications of the risk of system hazard exposure as well as overall system reliability is that an analyst can be "precisely wrong" in recommending a control measure. Negative outcomes associated with inaccurate system R-R classification, include: (1) recommendation of a control measure that is not adequate for addressing the risk exposure or (2) recommendation of a control measure that is unnecessarily expensive for the system. However, the main advantage of fuzzy classification is that it provides a basis for a broad set of "roughly right" recommendations

of control measures. With this approach, the safety analyst can reduce the possibility of insufficient control measures as well as the probability of excessive cost for the system.

Risk-Reliability Score

In order to quantify the three-dimensional R-R analysis space, there is a need to further define a R-R score. In the systems safety literature, risk is defined as the product of the frequency of hazard exposure and the severity of outcomes, as shown in Equation 1:

$$Risk = Frequency \times Severity. \quad (1)$$

Frequency, by definition, should represent the number of times system hazard exposure is expected to occur (see Equation 2):

$$Frequency = \frac{Hazard\ Exposure}{Time}. \quad (2)$$

However, this term does not capture the probability of loss (L) given exposure (E; P[L|E]), which can be calculated using Bayes theorem, as shown in Equation 3:

$$P(L \mid E) = \frac{P\left(Loss \cap Exposure\right)}{P\left(Exposure\right)} = \frac{P\left(Loss\right)}{P\left(Exposure\right)}. \quad (3)$$

It is important to note that $\frac{P\left(Loss \cap Exposure\right)}{P\left(Exposure\right)}$ can be written as $\frac{P\left(Loss\right)}{P\left(Exposure\right)}$ since loss is a subset of hazard exposure (i.e., not all hazards lead to loss). Based on Equations 2 and 3, the frequency quotient can be rewritten as:

$$Frequency = \frac{Hazard\ Exposure}{Time} \times P\left(L|E\right) = \frac{Hazard\ Exposure}{Time} \times \frac{P\left(Loss\right)}{P\left(Exposure\right)}. \quad (4)$$

Degradation in agent reliability leads to increased likelihood of loss given exposure. Therefore, $\frac{P\left(Loss\right)}{P\left(Exposure\right)}$ can also be written in terms of reliability, as shown in Equation 5:

$$1 - Reliability = \frac{P\left(Loss\right)}{P\left(Exposure\right)}. \quad (5)$$

Based on Equations 4 and 5, frequency can once again be reformulated as Equation 6:

$$Frequency = \frac{Hazard\ Exposure}{Time} \times \left(1 - Reliability\right). \quad (6)$$

In order to apply the MIL-STD 882B (1984) classifications of risk, there is a need to invert the frequency equation. Therefore, Equation 6 can be rewritten as:

$$FREQ = \frac{1}{\frac{Hazard\ Exposure}{Time}} \times \left(1 - (1 - Reliability)\right) = \frac{Time}{Hazard\ Exposure} \times Reliability. \quad (7)$$

On this basis, we formulate the R-R score as shown in Equation 8:

$$R - R = Severity \times FREQ = Severity \times \frac{Time}{Hazard\ Exposure} \times Reliability. \quad (8)$$

We label Equation 8 as the R-R formula because it reflects the conventional definition of risk as well as reliability. For example, if the severity of a hazard exposure is estimated to be catastrophic and the frequency of the hazard occurrence is assumed to be frequent, then the risk of the hazard (based on MIL-STD-882B) is equal to 1 (with membership in the fuzzy risk category C_1). If there is degradation in the reliability of the HAI (e.g., reliability = 0.95 or a high reliability classification), in association with the hazard exposure, then the R-R value would be calculated using Equation 8, which results in a value of 0.95. The R-R value for this hazard would then be classified as 1AH in the three-dimensional R-R space.

Enhanced SHA Method

Considering the new R-R score, we proposed an enhanced SHA approach to include classification of human error types due to degraded

System: H-R Subsystem/Function: Payload ops.		System Hazard Analysis					Analyst: Kaber, D. B. Date: 10/29/14							
No.	TLM	Hazard	Cause(s)	Human Error Type	Effect(s)	IMRI	RISK	HR	AR	System Type	SR	R-R	Controls	FMRI
HAI-1	Human-robot collision	Heavy payload robot contacts human	(1) Soft control (sensor) failure; (2) degraded astronaut situation awareness	Commission (occupying same location in work cell as robot)	Astronaut head trauma and loss of consciousness; Body blow and internal bleeding; Tear in space suit, loss of atmosphere, death due to suffocation; damage to manipulator system; loss of payload	1D	8	0.95	0.95	Parallel	0.998	7.98 (1DH)	Implement redundant sensor array; provide astronaut training on precautions in robot work cell	1E-1D
								At time of material handling operation						
HAI-2	Human-robot object hand-off failure	Gripper releases CRU prior to astronaut grasp	(1) Gripper contact sensor error; (2) servo motor failure; (3) astronaut vision failure; (4) astronaut motor control failure	Commission (missed point of grasp of CRU; failed to gain control of CRU)	Damage to CRU; damage to spacecraft exterior; CRU impact with astronaut and bodily harm	2D	10	0.80	0.80	Serial	0.64	6.4 (2DM)	Implement contact sensors along all axes of translation; replace servo motors according to regular PM schedule; increase task time by 1.5x to reduce operator time stress in perception and control of materials.	2E-2D
								At time of CRU installation operation						
							Includes: hazards, causes, human errors, effects, IMRI, risk score, HRA, ARA, system type, SRA, R-R score, controls, and FMRI							

Figure 10. Enhance system hazard analysis (E-SHA) worksheet. TLM = top level mishaps (taken from a preliminary hazard list); CRU = component replacement unit; IMRI = initial mishap risk index (based on MIL-STD-882-E, 1984); HR = estimated human reliability level at stage of system operation; AR = estimated automation/robot reliability level; SR = the calculated overall system reliability level based on HR, AR, and the system type; R-R = the composite risk-reliability score for the system (in the defined 3D space); FMRI = Final mishap risk index (with a return to reference MIL-STD-882B-E, 1984); PM = preventive maintenance.

capacity (for each specific hazard exposure). The method integrates human reliability values, which should be estimated based on validated cognitive and physical performance measures. The E-SHA also considers automation reliability values, which should be estimated based on prior mission data or manufacturer tests. Designation of system type can be either serial or parallel. A composite human-automated system reliability value can be calculated based on individual agent reliabilities and system type. As noted, the R-R score is calculated using Equation 8. In the original form of the SHA, an initial mishap risk index (IMRI) is determined assuming system reliability. A revised final mishap risk index (FMRI) is then projected based on recommended hazard controls. In the enhanced form of the SHA (see Figure 10), we include the IMRI but revise the FMRI by accounting for the estimated human-automated system reliability level at the time of operation as well as any recommended controls. Here, we also note that degradations in system reliability with time not only contribute to the likelihood of loss, given hazard exposure, but can also lead to new system hazard exposures not identified in a "baseline" analysis during the detailed system design phase. That is, not only does the reliability estimation as part of the E-SHA lead to a more precise assessment of potential loss with exposure, but it also supports a more complete hazard analysis at the various stages of system life. Figure 10 shows an example application of the E-SHA (worksheet) to the domain of human-robot interaction in spacecraft maintenance. Spacecraft systems were selected as an example for our development of the methodology as they are often considered as one of the most complex forms of human-in-the-loop systems. However, our approach can be applied

to other domains (including supervisory control in power plant operations, air traffic control, etc.) in which human, automation, and joint system reliabilities can be determined and are influential in system performance and safety.

DISCUSSION AND CONCLUSIONS

In this study, we enhanced the existing SHA technique by introducing the concept of overlapping hazard risk bands as well as human and automation agent reliability classification using fuzzy sets. We showed how introduction of overlapping risk bands could lead to a broader set of control measures, which can be considered as roughly right recommendations for various risk levels posed by a particular hazard versus crisp classifications of risk and reliability, which could result in a precisely wrong control measure and system or mission failures. In addition, we mathematically formulated the concept of a R-R score, which can be used to quantify the space of system risk exposures across a range of agent reliabilities. We also demonstrated an application of the E-SHA approach through an example of a complex human-in-the-loop system.

Comparison of the E-SHA Approach With HF-PFMEA

Based on our review of literature on HRA methods, we observed that the primary objective of such methods is to calculate HEPs. None of the existing techniques, save HF-PFMEA and AEA, consider the severity of hazard outcomes along with the probability of system exposure as bases for calculating risk scores. Therefore, we identified HF-PFMEA and AEA as the two closest HRA methods to our E-SHA approach. However, as mentioned in the review, the AEA method is not a well-defined technique, whereas the HF-PFMEA has been identified as one of the most comprehensive HRA techniques. Comparing HF-PFMEA with the E-SHA technique, one can identify a number of advantages of the new E-SHA approach, including:

- The E-SHA allows for a safety engineer to account for the type of human-automation interaction, as part of system design, as well as individual agent and overall system reliability assessments.
- The E-SHA integrates the new R-R score to allow for precise assessments of hazard risk exposure throughout a mission and the potential for loss, given exposure.
- The FMRI as part of the E-SHA also supports more accurate hazard control strategy identification and the potential for improved safety measure effectiveness and economics.
- Classification of human error types allows for the E-SHA method to support operator training program design, as a potential safety control measure, in a more specific way than the existing HF-PFEMA.

Summary of Advantages and Limitations of New Risk Assessment and Hazard Analysis Methods

Risk assessment in systems safety analysis has been in need of additional research and development for some time. Prior approaches to hazard risk categorization are subjective, discrete, and limited in categories due to scaling. This situation leads to gross classification of hazard exposures with the potential for inaccurate identification of mitigation strategies. Furthermore, risk thresholds used with assessment matrices can lead to hazards with very similar risk scores resulting in dramatically different classifications. For example, when using MIL-STD 882B (1984), a RISK = 9 is considered undesirable, whereas a RISK = 10 is considered acceptable with review. We developed a fuzzy classification approach to hazard risk assessment that provides for overlapping risk bands and partial membership of various hazards in different risk categories. This approach leads to richer specification of hazard control strategies for any specific exposure.

Prior notions of frequency of system hazard exposure have been "fully loaded" and assume hazard exposure is equivalent to system loss. This notion is simply incorrect in terms of accident sequence models but has unfortunately been implied through many safety publications and educational forums. Related to this, if there exist different outcomes (losses) for a single system hazard exposure, then each outcome should have a separate likelihood estimate. The new R-R score effectively resolves this issue by capturing likelihoods of hazard exposure, degrees of potential

loss, and probability of system loss. The values can be used to predict system vulnerability levels at future points in time at which degradations in human and/or automation reliability might occur. In addition, R-R values can be used to further differentiate hazards with different severities of outcome (e.g., catastrophic vs. critical) and support broader control recommendations. For example, the trend of human and automated system reliability/capacity against frequency of exposure and severity of outcome may be nonlinear; therefore, the trend of R-R values will also be nonlinear.

The E-SHA approach that we presented also has some limitations that may influence practitioner application to various systems. For example, the new method does not provide a structured approach for translating R-R scores to hazard controls for degraded human or automation capacity/reliability states. This issue needs to be addressed in any further revision of the E-SHA methodology. In addition, there is a need to define overlapping ranges of R-R values for fuzzy classification of system hazard exposures, according to MIL-STD 882B-E (1984) categories (e.g., unacceptable, etc.). Fuzzy classification should lead to recommendation of broader sets of controls for each hazard. Furthermore, R-R bands should account for variations in actual human-automated system capacity/reliability levels relative to estimates. Finally, although we provided a mathematical approach to objectively demonstrate the validity of our new R-R analysis, there is a need to further validate the integration of fuzzy sets and the E-SHA approach through analysis of hazard exposures in a real-world human-in-the-loop system.

APPENDIX A

List of Abbreviations

Notation	Description
HRA	Human reliability analysis
SHA	System hazard analysis
E-SHA	Enhanced system hazard analysis
HF-PFMEA	Human factors process failure modes and effects analysis
PHA	Preliminary hazard analysis
ETA	Event tree analysis
FMECA	Failure modes, effects, and criticality analysis
RPN	Risk priority number
FTA	Fault tree analysis
THERP	The technique for human error prediction
ATHEANA	A technique for human error analysis
CREAM	Cognitive reliability and error analysis method
PSF	Performance safety factor
HEP	Human error probability
SPAR-H	Standardized plant analysis risk-human reliability analysis
AEA	Action error analysis
R-R	Risk-reliability
IMRI	Initial mishap risk index
FMRI	Final mishap risk index

ACKNOWLEDGMENTS

The authors would like to thank Dr. Eduardo Salas and the anonymous reviewers for their thoughtful comments and recommendations on the original manuscript submission. David Kaber's work on this project was supported in part by a grant from the U.S. National Institute for Occupational Safety and Health (NIOSH: No. 2 T42 OH008673-08). The opinions expressed in this report are those of the authors and do not necessarily reflect the views of NIOSH.

KEY POINTS

- Fuzzy sets are applicable in the systems safety analysis as the classification of risk probability and severity are based on linguistic rather than numerical variables.
- An enhanced system hazard analysis (E-SHA) technique was introduced by defining the concept of overlapping hazard risk bands and a reliability classification using fuzzy sets.
- A new risk-reliability (R-R) score was introduced by using a multidimensional analysis space considering the likelihood of hazard exposure, severity of potential outcomes, and levels of human-automation reliability.
- The E-SHA method was compared with human factors process failure modes and effects analysis (HF-PFMEA), which was considered to be the most similar human reliability analysis (HRA) technique, and showed advantages for safety engineer use.

REFERENCES

Bahr, N. J. (1997). *System safety engineering and risk assessment: A practical approach*. Boca Raton, FL: CRC Press.

Blackman, H. S., Gertman, D. I., & Boring, R. L. (2008). Human error quantification using performance shaping factors in the SPAR-H method. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 1733–1737). Santa Monica, CA: Human Factors and Ergonomics Society.

Bligård, L. O., & Osvalder, A. L. (2014). Predictive use error analysis—Development of AEA, SHERPA and PHEA to better predict, identify and present use errors. *International Journal of Industrial Ergonomics*, *44*(1), 153–170.

Chandler, F., Chang, Y., Mosleh, A., Marble, J., Boring, R., & Gertman, D. (2006). *Human reliability analysis methods: Selection guidance for NASA*. Washington, DC: NASA Office of Safety and Mission Assurance.

Chen, G., & Pham, T. T. (2000). *Introduction to fuzzy sets, fuzzy logic, and fuzzy control systems*. Boca Raton, FL: CRC press.

Clemens, P. L., Simmons, R. J., & Cincinnati, O. (1998). *System safety and risk management: A guide for engineering educators* (NIOSH Instruction module). Washington, DC: U.S. Department of Health and Human Services.

Cooper, S. E., Wreathall, J., Thompson, C., Drouin, M., & Bley, D. (1996, October). *Knowledge-base for the new human reliability analysis method: A technique for human error analysis (ATHEANA)*. Presented at the International Topical Meeting on Probabilistic Safety Assessment Moving Toward Risk Based Regulation, Park City, UT.

De Felice, F., Petrillo, A., Carlomusto, A., & Ramondo, A. (2012). Human reliability analysis: A review of the state of the art. *IRACST–International Journal of Research in Management & Technology (IJRMT)*, *2*(1).

Embrey, D. (2004). Qualitative and quantitative evaluation of human error in risk assessment. In C. Sandom & R. S. Harvey (Eds.), *Human factors for engineers* (pp. 151-202). London: IET.

Ericson, C. A. (2005). *Hazard analysis techniques for system safety*. New York, NY: John Wiley & Sons.

Forester, J., Dang, V. N., Bye, A., Boring, R., Liao, H., & Lois, E. (2012, June). *Conclusions on human reliability analysis (HRA) methods from the International HRA Empirical Study*. Presented at the 11th International Probabilistic Safety Assessment and Management Conference (PSAM11) and the Annual European Safety and Reliability Conference (ESREL 2012), Helsinki, Finland.

Gertman, D., Blackman, H., Marble, J., Byers, J., & Smith, C. (2005). *The SPAR-H human reliability analysis method, NUREG/CR-6883*. Washington, DC: US Nuclear Regulatory Commission.

Hollnagel, E. (1998). *Cognitive reliability and error analysis method—CREAM*. Oxford, UK: Elsevier Science.

Kim, I. S. (2001). Human reliability analysis in the man-machine interface design review. *Annals of Nuclear Energy*, *28*(11), 1069–1081.

Lewis, E. E. (1994). *Introduction to reliability engineering* (2nd ed.). New York, NY: Wiley.

Meister, D. (1964). Methods of predicting human reliability in man-machine systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *6*, 621–646.

MIL-STD-882B. System safety program requirements (Technical report). (1984). Washington, DC: Department of Defense.

Ramsey, J. D. (1985). Assessment of warnings based on an ergonomic accident sequence model. *International Journal of Industrial Ergonomics*, *4*, 195–199.

Suokas, J. (1982). Safety analysis of a liquefied gas storage and loading system. *Journal of Occupational Accidents*, *4*(2), 347–354.

Swain, A. D. (1990). Human reliability analysis: Need, status, trends and limitations. *Reliability Engineering & System Safety*, *29*(3), 301–313.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning—I. *Information Sciences*, *8*(3), 199–249.

David Kaber is a distinguished professor of industrial and systems engineering at North Carolina State University. He is also director of research for The Ergonomics Center of North Carolina. Kaber received his PhD from Texas Tech University in 1996. He is a Certified Human Factors Professional and Fellow of HFES and IIE.

Maryam Zahabi is a doctoral candidate in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. She received her MS in industrial and systems engineering in 2013 from Sharif University of Technology (Iran).

Date received: May 15, 2016
Date accepted: January 18, 2017