SENSITIVITY ANALYSIS OF HUMAN ERROR DURING SECURITY SCREENING OPERATIONS[1]

Isaac Jones, Department of Industrial & Operations Engineering, University of Michigan, 1205 Beal Avenue, Ann Arbor, MI 48109-2117, isaacaj@umich.edu
Jan Irvahn, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352, 509-372-6006, jan.irvahn@pnnl.gov
Robert Brigantic, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352, 509-375-3675, robert.brigantic@pnnl.gov
Lauren Hux, Department of Statistical Sciences & Operations Research, Virginia Commonwealth University, 1015 Floyd Avenue, Richmond, VA 23284, huxlb@vcu.edu
Casey Perkins, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352, 509-372-6651, casey.perkins@pnnl.gov
Bryan Stanfill, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352, 509-375-6403, bryan.stanfill@pnnl.gov

ABSTRACT

Understanding how uncertainty of inputs can affect model performance provides significant information in evaluating a model's sensitivity to parameter variation. Sometime, model inputs fall at two extremes - those that have a substantial impact on model's output and those that do not. Identifying the inputs that do leads to a better understanding of the problem the model is trying to simulate, and, as a result, the creation of simpler, more intuitive models. This paper provides an overview of the statistical methods used in developing the human error module, and the procedures used in assessing how variability of model outputs can be attributed to changes in model inputs. The paper is divided into 3 parts: 1) an explanation of security operations, 2) an overview of our modeling approach, and 3) a summary of the sensitivity analysis and Design of Experiments (DOE) results.

Keywords: Human error probability, Decision making, Human reliability Analysis, Security screening operations, Bayes net

INTRODUCTION

Security screening operations are found in a variety of settings, from screening fans at the entrances to sporting events, screening passengers at airport checkpoints, and screening containers and conveyances at international border crossings. These operations rely on security officers to decide whether they should allow a person, container, or conveyance to pass beyond the screening point or whether they should detain the item. Officers frequently use technology, such as metal detectors, to assist in their decision-making process. A simple model for such devices is a binary outcome: either they raise an alarm or they do not. Even with the assistance of technologies, the potential for human error appears throughout the security screening system. Brigantic et al. [3] developed a mathematical framework designed to model and quantify human error potential and decision outcomes that occur during security

[1] This work has been supported by the U.S. Department of Homeland Security, Domestic Nuclear Detection Office, under competitively awarded IAA HSHQDN-16-X-00060. This support does not constitute an express or implied endorsement on the part of the Government.

screening operations. In this paper, we present a security screening model based on Brigantic et al. [3] and conduct a sensitivity analysis, describing how the probability that illicit material penetrates a checkpoint changes based on model inputs.

We are interested in modeling the probability of allowing illicit material through a checkpoint. This probability is fundamentally driven by the sensitivity of available detectors and the reliability of the officers who perform the screening operations. In the following section, we introduce a framework for modeling human errors, a specific concept of operations (CONOPS) for security screening, and a framework for modeling violations of the concept of operations.

MODELING APPROACH

Modeling the probability of human error has been studied in many settings, including nuclear power plant operations, aviation, healthcare, and manufacturing [2] [10] [7] [9]. There are a variety of techniques that are available for quantifying human error; these include the Technique for Human Error Rate Prediction (THERP), Human Error Assessment and Reduction Technique (HEART), Justification of Human Error Data Information (JEHDI), Cognitive Reliability and Error Analysis Method (CREAM), and a human error probability estimation method called INTENT [8] [6] [5]. We use the HEART methodology because of its versatility, ease of implementation, and existing published research that helps to inform parameter values. See Brigantic et al. [3] for a more detailed overview of the HEART methodology but a brief review is provided next.

HEART

The HEART methodology starts by classifying a task into one of eight categories described in Kirwan [8]. Each category is associated with a nominal human error rate. These nominal rates of human error are exacerbated by error producing conditions, (EPCs) so the second step of the HEART methodology is to identify the relevant error producing conditions which are also provided by Kirwan. Each error producing condition corresponds to a maximum effect multiplier that increases the nominal human error rate. For small error rates, this is not problematic; for larger rates, one should consider applying the multipliers to the odds of a human error to obviate the possibility of producing probabilities greater than one. The odds of an error are the ratio of the probability of the error occurring to the probability of no error occurring. Each error producing condition is paired with a proportion of effect (between 0 and 1) to estimate the fraction of the maximum multiplier that will be realized. Each EPC is combined with its proportion of effect to produce an assessed impact value. The assessed impact value is computed with the formula, ((multiplier-1) x proportion of effect) + 1. Finally, the overall probability of human error is the product of the nominal human error rate multiplied by the product of all of the assessed impacts.

In this work, we have identified six error producing conditions that can potentially increase the overall human error probability (HEP). These six factors include environmental conditions, workload, personnel aptitude, fatigue, stress, and morale. The environmental conditions encompass error producing conditions such as extreme weather and storms. Screening for illicit objects while outside in the middle of a storm is more error prone than screening for illicit material on a mild, sunny day. Workload indicates how busy the security officer is during the screening operation. When the security officer is overworked, errors are more likely to appear. Personnel aptitude is a categorical variable that is informed by two inputs, years of experience and years of training required to be considered fully proficient. Officers with insufficient experience are more likely to make mistakes. Fatigue also plays a role; an officer at the end of a night shift is more likely to slip up than an officer at the beginning of a day shift. When the working environment is stressful, perhaps because of publicized threat, an officer is more likely to make a mistake. If morale is low among the security officers, perhaps because their job security is low, they are more likely to slip up. In our sensitivity analysis, we explore how different levels of these six error producing conditions contribute to the probability of allowing illicit material through a screening checkpoint.

CONOPS

Our CONOPS for security screening requires that each person or object arriving at a venue must first pass through a primary screening station. This station has a detector that (ideally) alarms for illicit materials; some illicit materials can pass through a detector without causing an alarm. When an item alarms, a security officer diverts the item to secondary screening, unless an error occurs, allowing the item to pass. At secondary screening, the item is scanned again and released if no alarm occurs, it is otherwise detained. Within each stage, a security officer has multiple opportunities to make an error. For primary screening, a security officer can fail to acknowledge the alarm, the officer can choose not to send the item to secondary screening, or the officer can send the wrong item to secondary screening. For secondary screening, the officer can fail to acknowledge the alarm or choose not to detain an alarming item.

Bayes Net

While HEART is good for modeling slips and mistakes, for actual human errors, it is not appropriate for modeling violations. One of the failure modes of our CONOPS procedure occurs when a security officer knowingly fails to follow procedure and lets through an alarming item. In this case we are not implying malicious intent, rather the officer may be overworked or possibly think that the object is not a threat. In Figure 1, we model these violations with a Bayesian network (Bayes net) that produces the probability that a security officer knowingly fails to divert an alarming item. Bayes nets are directed acyclic graphs where the nodes represent variables and the edges represent dependencies. The Bayes net contains multiple sets of conditional probabilities that describe the distribution of values associated with each node, conditional on the values of each node’s parents.

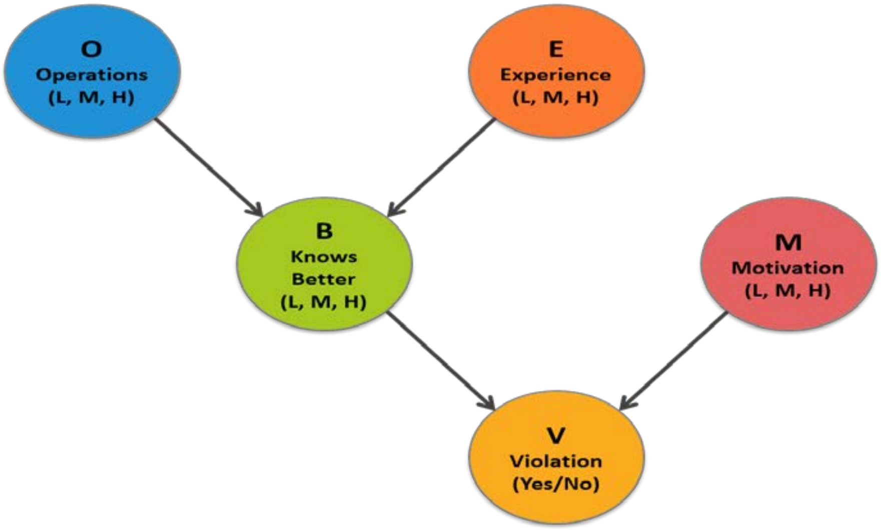

Figure 1. Bayes Network modeled

Our Bayes net has five nodes, the probability of violation (V) is found at the bottom and is informed by the values of its two parent nodes -- motivation (M) and knows better (B). The value of the knows better node is informed by its two parent nodes, experience (E) and operational load (O). The value of the experience node can take one of three values, low, medium, or high depending on the amount of experience the security officer has. The value of the operational load node can take one of three values, low, medium, or high depending on how busy the security screening stage is. The value of the knows better node can take one of two values, knows better or does not know better. Table 1 below shows our assumed probability of knows better based on the intersection between operational load and experience.

Table 1. Probability Intersection between O&E

Operational Load	Experience	Probability of Knows Better
Low	Low	0.01
Low	Medium	0.1
Low	High	0.3
Medium	Low	0.1
Medium	Medium	0.3
Medium	High	0.6
High	Low	0.2
High	Medium	0.5
High	High	0.99

The value of the motivation node can take one of three values, low, medium, or high. The values of the probability of violation node are between 0 and 1. Table 2 below shows our assumed probability of violation based on the intersection between knows better and motivation.

Table 2. Probability Intersection between B&M

Knows Better	Motivation	Probability of Violation
No	Low	0.0003
No	Medium	0.0002
No	High	0.0001
Yes	Low	0.0009
Yes	Medium	0.0005
Yes	High	0.0001

The probability of violation in this Bayes net is quite low, reflecting a well-trained, committed officer with good oversight. In addition to the values in the Table 2, we also implement a Bayes net where the probabilities of violation are 1,000 times larger, reflecting an officer who is unlikely to follow the CONOPS, possibly because of a lack of oversight. Finally, we implement a third Bayes net to cover the middle ground with both some large and some small probabilities of violation.

SENSITIVITY ANALYSIS

To determine the extent by which model inputs influenced the probability of an individual or item being inadvertently released, a Design of Experiments (DOE) was conducted. A DOE is a statistical methodology used to quantify effects, learn about interactions among variables, and model relations between independent and/or dependent variables [1]. DOEs are composed of three main parts: 1) determining the independent and uncontrollable variables that affect the response variables, 2) deciding upon an experimental design plan (blocking, factorial, response surface design, etc.) and determining the conditions or levels at which the independent variables will be entered into the experiment, and 3) the statistical analysis of results. This last step commonly consists of, but is not limited to, regression analysis, ANOVA, main-effect plots, and interactions plots. For our DOE, the independent variables or factors and their respective levels are illustrated in Table 3 below.

Table 3. DOE Factors and Respective Levels

Factor	Levels
Stage	primary, secondary
Alarm	gamma, neutron, no alarm
Weather	fair, extreme
Workload	1,1000
Experience	1,10
Training Required	1,5
Stress	low, high
Turnover	low, high
Time	morning, evening, night
Bayes Network(BN)/Oversight	low, mid, high

A cross product of these factor levels resulted in $2^7 \times 3^3$ or 3,456 feature vector combinations. An example feature vector ($\vec{v}$) is illustrated in equation (1). This feature vector corresponds to stage = primary, alarm = gamma, weather = fair, workload = 1, experience = 1, training required = 5, stress = low, turnover = high, Time = morning, and Bayes network = beginner.

$$\vec{v} = \{primary, alarm, fair, 1, 1, 5, low, high, morning, low\} \quad (1)$$

For the two level cases (stage – primary, secondary; weather – fair, extreme; etc.), the levels are the min/max values for the model inputs. However, for factors with three levels (alarm, hour of day, etc.), the entire selection space is contained in the levels shown in Table 3. For each feature vector, an estimated probability of release was determined by computing 10,000 Monte Carlo simulation trials for each feature vector combination. This was done because the Bayes Network and HEART method compares internally computed probabilities to randomly generated uniform distributed numbers between 0 and 1 to determine if the security screening operation will result in release or divert. Essentially, the model’s output is stochastic and not deterministic in nature.

For the DOE, the estimated probability of release was the response variable across the 10,000 trials. This probability represents the likelihood an item or person is released from security screening, even though a banned material is present. Based on this setup, Figure 2 shows a main effects plot which provides the estimated mean probability of release across all 10,000 trials and settings for when each individual factor listed is set at its different defined levels. Also, Figure 3 then shows an interaction plot for the four most influential main effect factors (these were alarm, stress, turnover, and oversight).

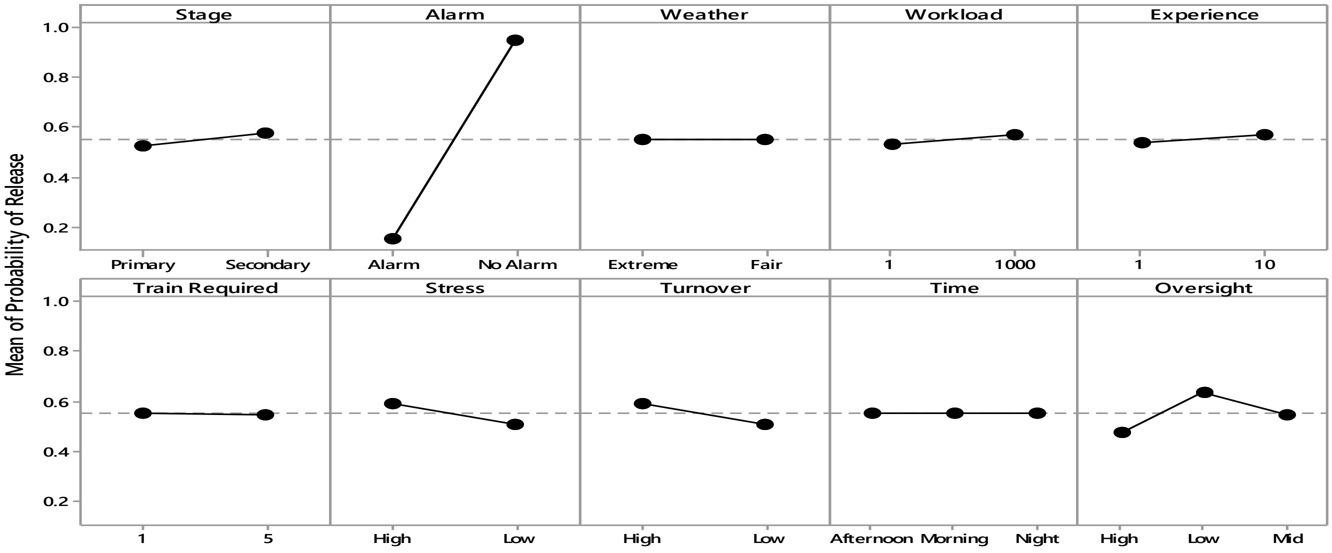

Figure 2. Main Effects Plot

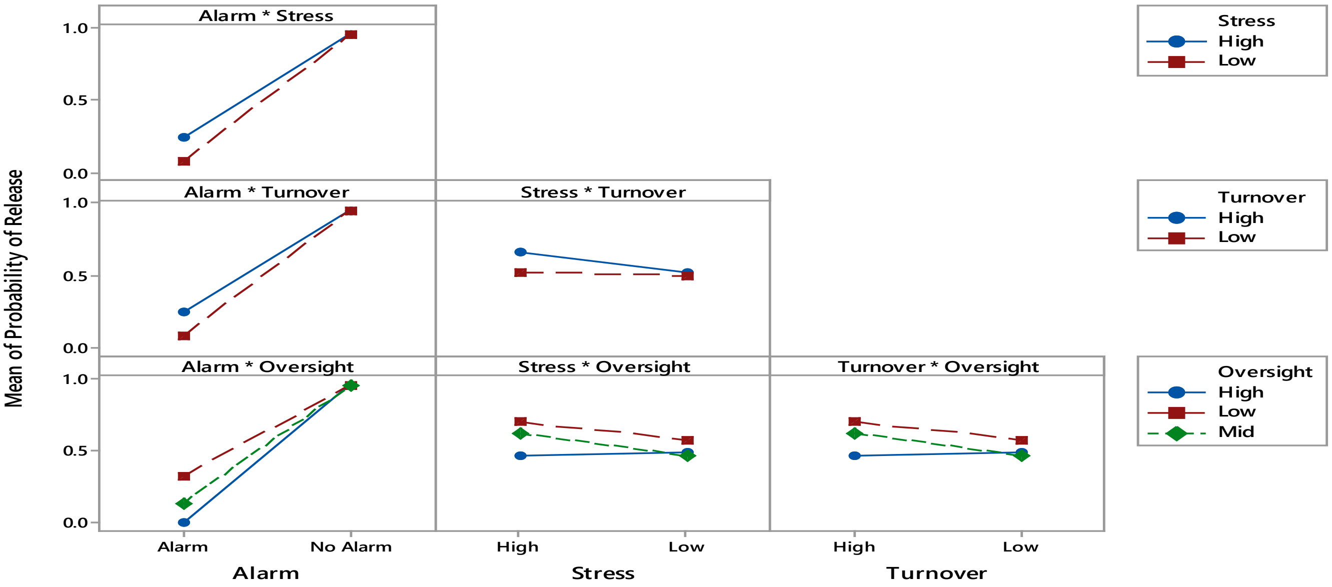

Figure 3. Interaction plot

From the above figures, the factors alarm and Bayes Net (i.e., oversight) had the highest impact on probability of release, followed by stress and turnover. The reason why these two features had the highest impact can be attributed to two reasons. First, when there is no alarm, the probability of an item or person being inadvertently released is high; and, two, the Bayes Nets each modeled a different level of operator experience/susceptibility to security violations. When operators with a high level of oversight are conducting security screening operations, it is reasonable to expect the probability of an unintentional release to be lower than when operators with low oversight are conducting the security screenings.

CONCLUSIONS

In conclusion, independent of the rate at which random screening is used by security operators, the single most important element of a security process is the security alarm. Weather, operator experience, operator training required, etc. have a trivial impact on improving the performance of security operations if an alarm with poor performance is used. Moreover, from the interaction plot, when the alarm does not trigger there is no clear separation between the mean probability of release based on stress, turnover, and oversight for the probability of release. At these instances, the points overlap each other. Therefore, determining the optimal threshold for an alarm to trigger is highly significant for overall security screening success.

In our next steps, we will further investigate additional features that have the potential to be significant to our model's performance. For example, we will consider the notion of "cognitive depletion" as described in [4]. Cognitive depletion helps inform our understanding of how security operator's effectiveness tends to decrease as they become mentally exhausted and could lead us to a better understanding of how and why mistakes happen. In turn, this could enable us to more accurately model the security screening process.

REFERENCES

[1] Anderson, Mark J., and Patrick J. Whitcomb. Design of experiments. John Wiley & Sons, Inc., 2000.

[2] Bell, J., and Holroyd, J., 2009, "Review of Human Reliability Assessment Methods," Health & Safety Laboratory, Health and Safety Executive.

[3] Brigantic, R., Perkins, C., Stanfill, B., and Oster, M., 2016, "Quantifying Human Behavior and Decision Errors in Security Screening Operations," Proceedings of the 2016 Industrial and Systems Engineering Research Conference.

[4] Franklin, Lyndsey, Kristina Lerman, and Nathan Hodas. "Will Break for Productivity: Generalized Symptoms of Cognitive Depletion." *arXiv preprint arXiv: 1706.01521* (2017).

[5] Gertman, D., Blackman, H., Haney, L., Seidler, K., and Hahn, H., 1992, "INTENT: A Method for Estimating Human Error Probabilities for Decision Based Errors," Reliability Engineering & System Safety, 35(2), 127–136.

[6] Hollnagel, E., 1998, "Cognitive Reliability and Error Analysis Method (CREAM)," Elsevier.

[7] Ikuma, L., Koffskey, C., and Harvey, C., 2014, "A Human Factors-Based Assessment Framework for Evaluating Performance in Control Room Interface Design," IIE Transactions on Occupational Ergonomics and Human Factors, 2(3-4), 194–206.

[8] Kirwan, B., 1996, "The Validation of Three Human Reliability Quantification Techniques - THERP, HEART and JHEDI: Part 1 - Technique Descriptions and Validation Issues," Applied Ergonomics 27 (6), 359–73.

[9] Murata, A., Nakamura, T., and Karwowski, W., 2015, "Influence of Cognitive Biases in Distorting Decision Making and Leading to Critical Unfavorable Incidents," Safety, 1(1), 44–58.

[10] Nieuwenhuys, A., Savelsbergh, G., and Oudejans, R, 2015, "Persistence of Threat-Induced

Errors in Police Officers’ Shooting Decisions,” Applied Ergonomics, 48, 263–272.