Safety Research

Predicting Incident Rates

Artificial intelligence as a forecasting tool

By Abdullah Al-Mutairi and Joel M. Haight

PREDICTING THE FUTURE in any discipline is difficult, yet great strides have been made in many areas. Although investing in the stock market may not always deliver the expected profits, people still invest. Although a company may not sell as many of its products as predicted, production run sizes are still based on forecasted demand. Weather forecasts are a common source of complaints, yet people continue to watch the weather reports when planning activities. The safety community, however, does not seem ready to embrace the prediction of incident rates as anything meaningful.

Safety can be an expensive aspect of industrial operations unless efforts are made to enhance and optimize programs to reduce the long-term cost associated with SH&E-related incidents and damage. The objective of an SH&E program is to minimize or prevent loss involving people, the environment, property and profits.

One step toward achieving this objective would be to quantify and analyze intervention activity and incidents for an existing SH&E program. Using neural networks, a form of artificial intelligence, the researchers attempt to determine and identify a relationship between safety interventions and incident rate. Once the relationship has been established, the analyst would be able to use it as a forecasting tool to predict future incident rates given the level of safety intervention activities.

An artificial neural network (ANN) is an information processing prototype that mimics to some extent the way biological nervous systems such as the brain process information. According to the Defense Department's Advanced Research Projects Agency (DARPA) Neural Network Study (1988), "a neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes."

In this study, incidents recorded were comprised of physical injuries to workers in the forestry division of a utility company, as well as spills and equipment failure. While an ANN is a complex concept with respect to how it works, it is not so difficult to use from a practical point of view. Essentially, an ANN's operating objective is to recognize patterns. Any input-output experimental model that uses input data and generates output results can be analyzed using ANN.

Patterns of output data variation in response to input data variation can become too difficult for the naked eye to recognize or to parse out an individual effect. This is especially true when many variables are involved. Through a series of software, networks or pathways, the pattern of each input-output data run is plotted and *remembered.* The resulting output response is noted during training of the ANN. When a similar input pathway comes up, the system looks for a response (incident rate) that is similar to what it saw under similar input conditions. This then becomes the prediction.

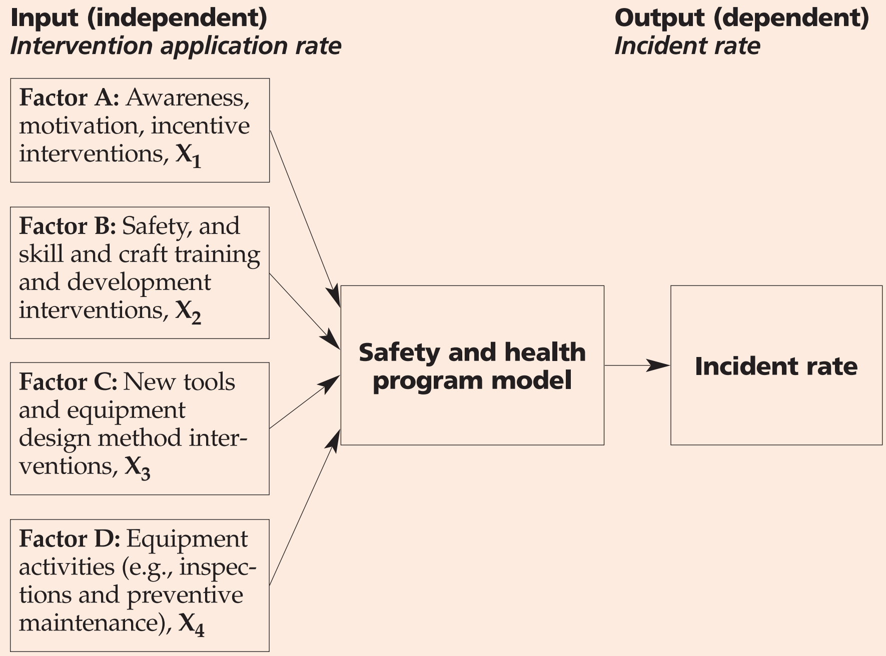
This research is a continuation of work by Haight, Thomas, Smith, et al. (2001a, b) and Iyer, Haight, Del Castillo, et al. (2004; 2005), which focused on quantifying safety intervention activities with the incident rate. It is based on the relationship between four safety intervention factors that are considered inputs; the incident rate is the only output. Figure 1 is a graphical representation of the model established by Haight, et al. (2001) that lays the foundation for quantifying safety intervention activities with the incident rate.

Abdullah Al-Mutairi *is a safety specialist with Chevron Corp., currently working in Thailand. He joined Chevron after graduating from Pennsylvania State University's industrial health and safety program with an M.S. Al-Mutairi holds an undergraduate degree in petroleum and natural gas engineering with minors in political science and economics.*

Joel M. Haight, Ph.D., P.E., *recently became chief of the Mining Injury Prevention Branch at NIOSH's Pittsburgh Research Center. For the nearly 10 years before this, he was an associate professor of energy and mineral engineering at Pennsylvania State University. Haight's master's and doctorate degrees, both from Auburn University, are in Industrial and System Engineering. Before joining Penn State, he was a manager and an environmental and safety engineer for Chevron Corp. He is a professional member of ASSE's Central Pennsylvania Chapter, and also a member of AIHA and the Human Factors and Ergonomics Society.*

Figure 1

Representation of the Safety & Health Program, Mathematical Model



Note. *Adapted from "Intervention Effectiveness Research: Phase 1, Developing a Mathematical Relationship Between Interventions and Incident Rates for the Design of a Loss Prevention System," by J.M. Haight, R.E. Thomas, L.A. Smith, et al., May 2001,* Professional Safety, *46(5), p. 39.*

Research Objectives

The goal of the project is to forecast an incident rate given a set of safety intervention inputs using ANNs. To do so, the following objectives were established:

•Determine and develop the relationship between safety interventions and the incident rate from the forestry division of a power company.

•Develop a forecasting tool using ANNs that will be able to predict an incident rate based on the type and amount of safety intervention activities.

The hypothesis is that ANNs are an accurate predictor of incident rates.

As this study involves a new approach to forecasting incident rates, no literature exists to define what *accuracy* means. Therefore, this term is defined as follows in this study:

•absolute average percent error of less than 20% (the average of the percent difference between predicted incident rate and actual incident rate);

•mean absolute deviation (MAD) of less than 1.0 (the mean of the absolute difference in the value of the predicted incident rate and the actual incident rate);

•coefficient of determination (R^2) greater than 0.50 (the measure of strength of the regression; this measures how much variation in the model is due to the model itself versus how much variation is unaccounted for).

Literature Review

The use of ANNs in a multifaceted society is not a new concept, but its use as a means of evaluating SH&E programs is a pioneering application. According to the literature, artificial intelligence has never been used in an attempt to correlate, analyze or forecast safety interventions with the incident rate. In fact, except for Haight, et al. (2001a, b; 2003) and Iyer, et al. (2004; 2005), research dealing with quantification of safety interventions and the incident rate and relationship modeling is minimal.

Guastello (1993) used regression analysis to relate incident rates and intervention programs applied. He evaluated the programs as though the whole program was one intervention within each facility, meaning one input was compared to one output; however, the interactive effects between interventions were lost. Guastello then realized that to determine the optimal level of interventions, one must know all the interventions that affect the incident rate as well as the interactions among and between them.

Haight, et al. (2001a, b) presented an analytical model that established a mathematical relationship between all intervention activities being implemented at the site and the incidents they were designed to prevent. The model provided a tool to develop a quantifiable design and to optimize an SH&E intervention.

Iyer, et al. (2004; 2005) developed a forecasting model and optimization procedure to analyze as well as optimize an SH&E program by minimizing manpower input while concurrently minimizing incidents. This group also produced a forecasting tool that would predict the incident rate given a set of safety intervention inputs. They determined that the carryover effect of an incident rate in a particular week had a statistically significant relationship with the safety intervention activity levels.

They also developed forecasting models based on the results of a study using several statistical techniques such as transfer function modeling and regression analysis. Although Iyer, et al.'s (2005) study suggests that quantifying safety intervention activities with the incident rate is beneficial in terms of cost and fewer losses, further research is needed to establish model reproducibility and its industry-wide applicability.

Methodology

Data Collection

Data were collected on a weekly basis from September 2003 to February 2005. They were then entered in a spreadsheet (Table 1, p. 42). The safety intervention variables are represented by factors A, B, C and D. Factor A represents safety awareness and motivation activities. Factor B represents skill development and training activities. Factor C represents new tools and equipment design methods and activities while Factor D represents equipment-related activities.

The columns in the middle of Table 1 represent the hours spent on each intervention variable by the respective safety center within the forestry division of a power company. The totals column represents the sum of inputs of the various safety intervention

Abstract: *Current research suggests that it may be possible to forecast incident rates. Researchers are using artificial neural networks to learn data patterns associated with safety program and incident rate data. In the study described here, these predicted rates are compared with actual performance data over a 3-year period. Researchers were able to detect indications of promise in predicting incident rates, given current safety and health program configurations.*

Table 1

Data Sheet Used During Data Collection

	Week -	Date from:	5-Jan-04	Date to:	9-Jan-04	
Data Input Representative	No Issue	No Issue	No Issue	No Issue	No Issue	
Cost Center	917/7366	7367	7368	7752	7755	Totals
Safety Activities						
Factor A – Safety Awareness and Motivation Activities						392.00
1. Crew Inspections	2	4			2	43.00
2. Implementing Awards, Incentives, etc. Program						4.00
3. Reviewing and implementing Safety Programs						1.00
4. Implementing Joint Health and Safety Committee Activities and Programs						3.00
5. Developing and delivering safety related communications, bulletins, etc.						22.00
6. Providing Safety Related Feed Back to Employees						15.00
7. Job planning activities		24			24	150.00
8. Tail board conferences		24			37.5	154.00
9. Safety Supervision		78			70	613.00
Factor B - Skill Development and Training Activities:						297.00
1. Safety Training						29.00
2. Technical Training	150					230.00
3. Safety Meetings					17	38.00
4. Drills (emergency, safety, rescue practice and drills, etc.)						0.00
Factor C – New Tools and Equipment Design Methods and Activities:						26.00
1. New tool development activities						4.00
2. New methods and procedure development activities						21.00
3. Audits and Assessments						1.00
Factor D - Equipment Related Activities:						226.00
1. Equipment Inspections					10	111.00
2. Facilities Inspections						12.00
3. Personal Protective Equipment Inspections						27.00
4. Preventive Maintenance Activities						76.00
Total Hours For All Safety Activities/Week:	**152.00**	**130.00**	**0.00**	**0.00**	**160.50**	941.00

Note. *Example only. Adapted from "Intervention Effectiveness Research: Phase 1, Developing a Mathematical Relationship Between Interventions and Incident Rates for the Design of a Loss Prevention System," by J.M. Haight, R.E. Thomas, L.A. Smith, et al., May 2001,* Professional Safety, *46(5), p. 40; and "Intervention Effectiveness Research: Understanding and Optimizing Industrial Safety Programs," by P.S. Iyer, J.M. Haight, T.B. Del Castillo, et al., 2004,* Chemical Health and Safety, *11(2), pp. 9-19.*

Data were entered in a spreadsheet. The safety intervention variables are represented by factors A, B, C and D. Factor A represents safety awareness and motivation activities. Factor B represents skill development and training activities. Factor C represents new tools and equipment design methods and activities while Factor D represents equipment-related activities.

factors. The example presented shows only a fraction of cost centers reporting. It should be noted that *tailboard conferences* is a term used by the company to refer to tailgate safety meetings.

After gathering an adequate amount of data (62 weeks' worth were initially collected; this was separated into training weeks and testing weeks) to proceed with a statistically significant study, the collection phase ended. The next step involved systematically organizing the data, which is explained in the next section, so that they may be entered in ANN.

Data Organization for Use in ANN

In this phase, each of the 20 safety intervention inputs were summed on a weekly basis and placed in a separate spreadsheet. As noted, 62 weeks of data were used due to the availability of that data. Of those 62 weeks, 37 were used to train ANN and 25 weeks were used at the validation phase.

An ANN must first determine what the data pattern is before it can recognize whether that pattern exists in the data it uses to predict an outcome. This is called *training*. During the training phase, a certain number of data points are input in the system where the system is given both independent and dependent variables so that it can recognize the patterns. In this case, 37 weeks' worth of data were used.

In the validation phase, the input data is only the independent variables. The system then predicts what the dependent output will be; 25 weeks' worth of data were used to provide adequate degrees of freedom for statistical significance.

No concrete rules exist regarding the number of weeks required for the two phases. However, guidelines do exist (Masters, 1993). One suggests that the training set be representative of the entire population. Thus, the input data entered during the training phase must encompass the range of incident rates displayed in the 62 weeks of data. Also, no concrete rules define network parameters; the only guidelines are related to how many hidden layers to use and how many times to train the network.

Furthermore, ANN training capacity is partly based on the amount of patterns input. The fewer the data sets, the less capable the system becomes at formulating computational models based on the information given to it and vice versa. Therefore, one needs to determine the correct mix of weeks to input for the training phase and still have enough weeks of data to produce a statistically significant measure of ANN's ability to forecast.

As noted, in this study, that mix involved 25 weeks for validation and 37 weeks for training. ANN was trained with 10, 20, 30 and 40 weeks prior to ending up with 37. There was a tendency for the validation results to improve as the training set size increased. However, since this was not always the case, it is up to the researcher's discretion to determine a suitable mix of training weeks to validation weeks.

The criteria used in this study was the mean square error (MSE) and mean absolute deviation (MAD). Note that for every set of weeks not used for training, the remaining set of weeks were used for validation. This is evident as ANN requires that the training set be representative of the population so that when testing or validating takes place outlier data should be nonexistent. An example would be having a trained ANN with incident rates ranging from 1 to 10, then testing it with time spent on safety interventions that produced an actual incident rate within that range rather than an incident rate of 15.

Throughout the 62 weeks of data gathered, all work centers involved in this study were not able to submit data on their safety intervention activities every week for various reasons. Therefore, a normalized set of data had to be established. Normalization consisted of data representing the entire population of 400 workers or 24 cost centers used in this study.

It is important to note that the researchers attempted to organize and input the data as a percent of available work hours [similar to Iyer, et al. (2004)], but ANN was not able to learn the data nor forecast.

Training & Forecasting

In this context, *training* refers to the act of feeding ANN information and data, then running the program in order to enable it to learn and assimilate that information. In this case, the data are the week-to-week quantified hours of input to the company's SH&E program, the independent variable, and the incident rates, which are the output or dependent variable. These data are input to allow the system to recognize the pattern of incident rates that result from the weekly variations in SH&E activities and the percentage of available workhours applied to their implementation.

Incident rates are sensitive to changes in SH&E program activities and to the extent to which they are implemented (Haight, et al., 2001a, b; 2003; Iyer, et al., 2004; 2005). Therefore, it is reasonable to allow ANN to determine variation patterns.

The term *supervised learning* refers to a process in which the researcher gives ANN specific input patterns with the correct network output, in this case, the incident rate. During the supervised learning phase, the researcher fed the system safety intervention activity inputs with the corresponding output or incident rate. Once ANN was able to fully learn and assimilate the information, the researcher moved to the validation phase, which is the forecasting stage of this research.

During the forecasting stage, the researcher performed validation. The term *validation learning* means that the network is not given any external indication as to what the correct responses should be nor whether the generated responses are correct. It is simply projecting an output or forecasting the incident rate based on the data given to it on a weekly basis. During validation, the system looks back on the various input-output pairs that it learned during training and it learns by the environment, that is, by detecting regularities in the structure of input patterns.

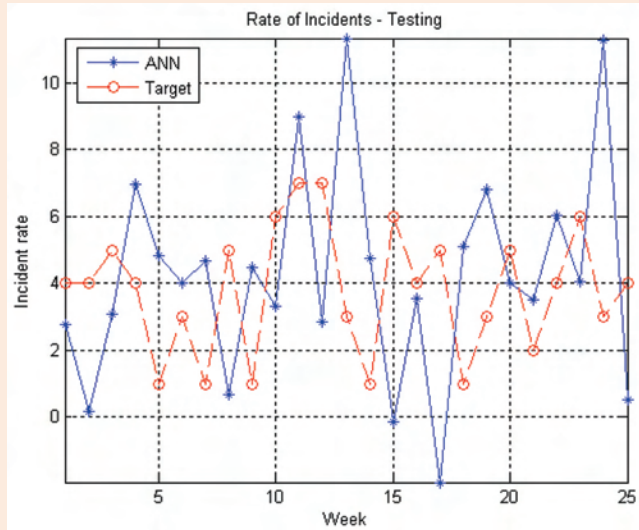
ANN displays the validation results graphically and numerically by comparing the forecasted results to the actual using the MSE formula. The MSE approach was chosen because it lies close to the center of normal distribution; thus, if errors are assumed to be normally distributed, minimizing the MSE corresponds to other preferred optimizations.

Furthermore, the derivative of the MSE can be easily computed relative to other performance measures. This signifies that when the optimization criterion is the MSE, direct methods of optimizing performance can be achieved. To calculate the MSE, sum the squared differences between the predicted output (ANN incident rate) versus the actual incident rate, then dividing by the number of components (in this case weeks) that went into the sum.

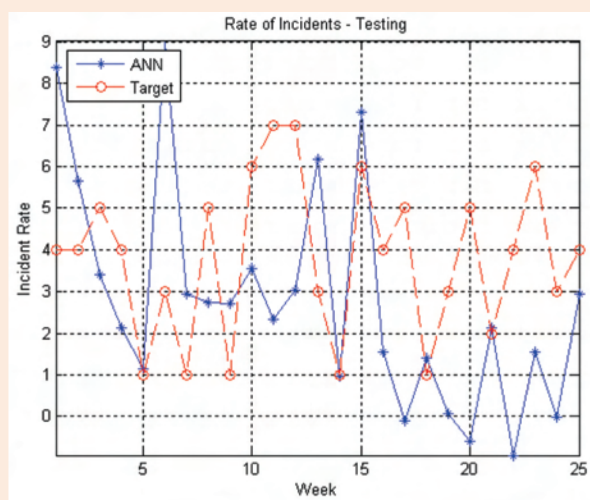
Figure 2

ANN Validation Output

Example of an ANN validation output with its corresponding architecture {40,125,1}, {'logsig','purelin'} MSE = 142.7% R^2 = 0.05. This output shows the actual incident rate as the target and the ANN predicted values—MSE or (mean square error) is relatively high—predictions are not close.



Example of an ANN validation output with its corresponding architecture {40, 125,15,1}, {'logsig','tansig','purelin'} MSE = 76.8% R^2 = 0.00. This output also shows the actual incident rate as the target and the ANN predicted values, this time using different network architectures. No correlation exists between the two with an R = 0.0 and a better MSE than above, but still is relatively high—predictions are not close.



The equation below illustrates how MSE is calculated.

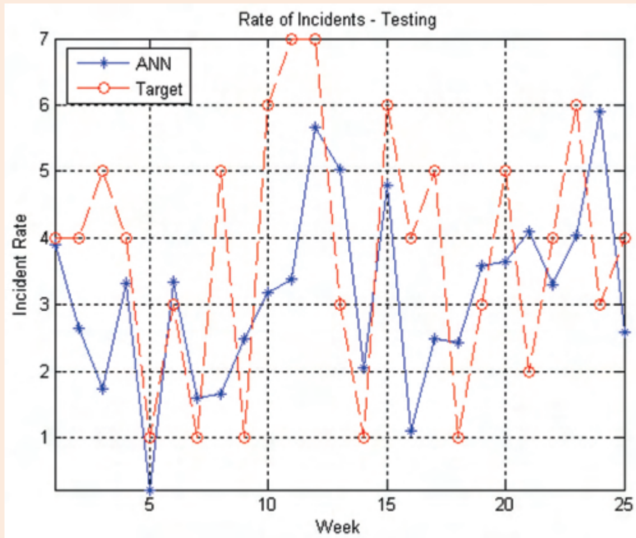
$$MSE = \frac{1}{P}\sum_{p=1}^{p}\sum_{i=1}^{n}(d_{i,p} - a_{i,3})^2 \cdots$$

where: $d_{i,p}$ equals desired output of output unit i for input pattern p
a_i equals observed output of output unit i
P equals total number of patterns in the data set, while n equals the number of output units.

Figure 3

Forecasting Accuracy of Neural Networks

MSE = 55.1% R^2 = 0.13. After a significant amount of training for the ANN, the predicted incident rates versus actual incident rates have improved, but not substantially or statistically significantly.



Results may be improved by altering the architectural structure of the network by changing the number of hidden layers, type of activation functions and the number of neurons used. This is an iterative process. Once enough iterations have taken place that lower the MSE results without overtraining the network, the results are finalized and the forecasting stage of the research is concluded.

Although training the system is not an exact science, reducing the number of hidden neurons helps the system avoid idiosyncrasies. Also, increasing the variety of the training set reduces the probability of overtraining the system. However, one must remember that training usually starts with random initial weights and, thus, there is no exact science of what constitutes adequate learning. Finally, statistical analysis is undertaken to support or refute the hypothesis of whether ANN is an accurate predictor of incident rates.

Moving Average

The moving average part of the research involved all the steps mentioned in the previous sections with one major difference: the inputs of 1 week were compared to the average incident rates for the following 3 weeks (i.e., week 1 inputs were compared to the incident rate for weeks 1, 2 and 3 since it is suggested that the effect from an SH&E program is neither instantaneous nor permanent).

In this study, 40 inputs were entered into ANN per week with a corresponding output. The corresponding output was an average incident rate for 3 weeks, the week in which those 40 inputs originated from and the subsequent 2 weeks. The total number of weeks used in this part of the study is 58 weeks due to data availability. The training phase involved 35 weeks, while the validation phase consisted of 23 weeks. The mix of training weeks to validation weeks was determined using a similar approach to that detailed earlier. A 6-week moving average similar to Iyer, et al. (2004) was not performed due to lack of data, which would have meant loss of degrees of freedom and training strength, as well as statistical significance of the results.

Results, Analysis & Discussion

Prelude to Results

To reach optimal performance of ANN, the network architecture had to be modified. This led to varying results, so the network that produced the best results relative to other network runs was selected. The selection was based on lowest MSE and MAD.

Figure 2 (p. 43) displays the architecture of the network with its corresponding output. The first number in brackets refers to the number of inputs in the input layer, while the second number refers to the number of neurons associated with the activation function. In this study, the last number will always be one as there is only one output function (incident rate). In Figure 2, the notation {40,125,1} refers to 40 inputs in the input layer, 125 neurons in the hidden layer and one output. The activation function in the hidden layer is logsig. ANN is the forecasting capability and the target or actual is the incident rate for that particular week.

As noted, an effort was made to input the data as a percentage of available workhours but ANN was not able to learn the data nor adequately predict results. It should also be noted that the network not trained as ANN was unable to learn the specific input patterns and correlate it with the output given to it.

Lack of adequate training also leads to poor validation results. This phase of ANN is meaningless without some form of training. The output should attempt to correspond with what actually happened, but in this instance it is insignificant as the system could not be trained. After the attempt to use percent of available workhours failed, data were input as total hours. The total hours represent the sum of hours for each cost center per safety intervention activity.

Forecasting Results & Analysis

After performing several runs and various ANN architectures, the system was trained and 25 weeks of safety intervention data were used for the validation phase. Figure 3 illustrates the finalized results.

The incident rates obtained from ANN were compared to the actual in a pair wise tabulation (Table 2). This comparison produced a residual result of a -0.63, indicating that on average the forecasted results tend to be lower than the actual incident rates. Also, an average percent error of 55% indicates that the forecasted results were not close to the actual incident rates. Furthermore, the standard deviation revealed a relatively low statistical dispersion as the average standard deviation was 1.38.

Table 2

Pair-Wise Comparison Between ANN & Actual Incident Rates

Week	ANN	Actual	Residual (difference)	Absolute percent error
1	3.89	4.00	-0.11	2.70
2	2.64	4.00	-1.36	34.07
3	1.74	5.00	-3.26	65.17
4	3.32	4.00	-0.68	16.90
5	0.22	1.00	-0.78	78.34
6	3.35	3.00	0.35	11.62
7	1.60	1.00	0.60	60.30
8	1.65	5.00	-3.35	67.09
9	2.50	1.00	1.50	149.65
10	3.18	6.00	-2.82	46.94
11	3.39	7.00	-3.61	51.62
12	5.65	7.00	-1.35	19.22
13	5.03	3.00	2.03	67.79
14	2.05	1.00	1.05	104.96
15	4.80	5.00	-0.20	4.05
16	1.10	4.00	-2.90	72.58
17	2.49	5.00	-2.51	50.15
18	2.42	1.00	1.42	142.49
19	3.58	3.00	0.58	19.24
20	3.65	5.00	-1.35	27.07
21	4.09	2.00	2.09	104.48
22	3.31	4.00	-0.69	17.33
23	4.03	6.00	-1.97	32.88
24	5.90	3.00	2.90	96.66
25	2.59	4.00	-1.41	35.31
Average	**3.13**	**3.76**	**-0.63**	**55.14**

To analyze the distribution of the data, a normality test was performed using the Anderson-Darling normality test. Both ANN and actual incident rates followed a normal distribution since both their respective *p*-values were greater than 0.05. The *p* value for ANN and actual were 0.867 and 0.096, respectively. As shown in Table 2, the average of ANN incident rates was 3.13 compared to the 3.76 actual average rate. That amounts to -16.9% error, which indicates closeness among the means but further tests such as a paired *t-test* need to performed to verify this. Note that 55.14% is the average percent error of all 25 weeks, while the 16.9% corresponds with the percent error of the means.

An *F-test* was performed to determine the ratio of two variances. If the two variances are not significantly different, their ratio will be close to 1. The resulting statistic was 0.551 and the associated *p*-value was 0.076. Since *p* was not less than 0.05, it can be concluded that there is no significant difference between the two standard deviations with a 95% confidence interval. This means there is no significant variation between the population means of ANN and the actual incident rates.

After determining a lack of significant difference between the variances, a paired *t-test* was performed. A *p*-value of 0.103 indicates that a statistically significant difference does not exist between the two means.

Also, a box plot of the analysis was performed. Figure 4 (p. 46) illustrates the box plot of ANN and the actual incident rate. The box represents the middle 50% of the differences. The line through the box represents the median difference. The lines extending from the box represent the upper and lower 25% of the differences. The box plots of the data show the closeness in the means of the two data sets.

Finally, to analyze the results, MAD was determined as the measure of accuracy; in addition, a plot graph was created to display ANN forecasted incident rates versus actual incident rates (Figure 5, p. 46). The plot indicates that ANN did not model accurately as the resultant R^2 was 0.13. Furthermore, the points appear to be scattered rather than falling on a straight line. If they were to fall on a straight line, that would indicate that the ANN forecasted incident rates were linearly related. Also a Pearson correlation test was performed to see whether the R^2 value had statistical significance. The test produced a *p*-value of 0.075, which is greater than 0.05, which indicates zero statistical significance in its ability to correlate.

For MAD, the closer the value is to 0 the more accurate one can claim that this prediction is. The equation shown at right displays the manner in which MAD is obtained where the sample size is *N*, the samples have values x_i, the mean is x and fi is an absolute frequency. Furthermore, it shows the average deviation from the actual incident rates.

$$\mathrm{MAD} = \frac{1}{N}\sum_{i=1}^{N} f_i \,|x_i - \bar{x}|,$$

MAD for the 25 weeks of forecasting was 1.63. The result of 1.63 incidents per week means that the predictions made by ANNs were on average within the range of ±1.63 incidents of the actual values. Also, a normal average comparison was done to see whether simply taking the average incident rates of 25 weeks and projecting it on every week produced better results than ANN. Table 3 (p. 47) presents a summary of the results.

The results indicate that taking the average incident rate over the 25 weeks and comparing it to the incident rate of each of the 25 weeks yields a better MAD of 1.49 as opposed to 1.63 for ANN. On the other hand, the absolute average percent error is far higher when using the normal average incident rate at 75.9% as opposed to ANN at 55.1%. The relative closeness between these results does not strengthen the hypothesis of ANN being an accurate forecasting tool.

A regression analysis was performed to correlate the number of hours of safety intervention activities per week with the actual as well as the ANN forecasted incident rate. The resulting R^2 for ANN was 0.03, which indicates poor correlating power. However, the regression analysis performed for actual incident rates with the number of hours of safety intervention activities per week also produced a poor R^2 value of 0.02. This indicates that the data itself has poor correlation with the incident rate. This may suggest that further studies involving stronger correlation might yield better regression results when using ANN.

Moving Average

As noted, the moving average analysis involved comparing the inputs of 1 week to the average incident rates for the following 3 weeks. Figure 6 displays the optimized network results of the 23 weeks of validation involving a moving average.

To tabulate the results of the forecasting accuracy of ANN for the moving average, a pair-wise comparison was performed. This comparison (Table 4, p. 48) produced a residual result of a -0.122, indicating that on average the forecasted results tend to be lower than the actual incident rates. Also an average absolute percent error of 27.2% indicates that the forecasted results were relatively close to the actual incident rates. Furthermore, the standard was 0.83. To determine the distribution of the data, a normality test was undertaken using the Anderson-Darling normality test. Both ANN and actual incident rates followed a normal distribution since both of their respective *p*-values were greater than 0.05.

The *p*-value for ANN and actual were 0.082 and 0.367, respectively. As shown in Table 4, the average of ANN incident rates was 3.68 compared to the 3.80, which is the average of the actual incident rate. That amounts to a -3.2% error, which suggests closeness among the means but a paired *t-test* is needed to verify this.

Next, an *F-test* was performed to determine the difference of two variances. If the two variances are not significantly different, their ratio will be close to 1. The resulting statistic was 0.590 and the associated *p*-value was 0.112. Since *p* was not less than 0.05, it can be concluded that there is no significant difference between the two standard deviations with a 95% confidence interval.

After determining a lack of significant difference between the variances, a paired *t-test* was performed. A *p*-value of 0.687 indicates that there is not a statistically significant difference between the two means as it is far above the 0.05. A box plot of the analysis was performed to go along with the paired *t-test*. Figure 7 (p. 48) illustrates the box plot for the moving average of ANN and the actual incident rate recorded. Outliers are indicated by an asterisk.

Furthermore, MAD was calculated for the 23 weeks of forecasting; it was 1.01, lower than the 1.63 previously calculated for a non-moving average. The result of 1.01 incidents per week means that the predictions made by ANNs were on average within the range of ±1.01 incidents of the actual values. Also a Pearson cor-

The box represents the middle 50% of the differences. The line through the box represents the median difference. The lines extending from the box represent the upper and lower 25% of the differences.

Figure 4

Box Plot of ANN & Actual Incident Rates

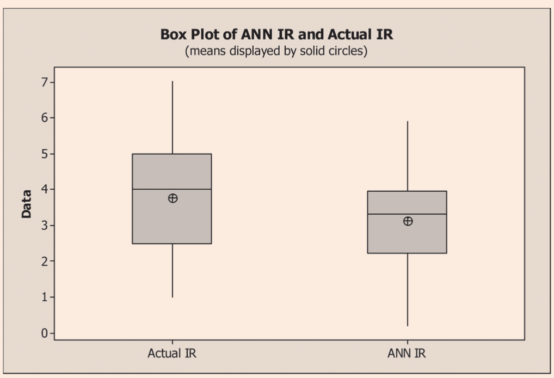


Figure 5

Comparison of ANN & Actual Incident Rates

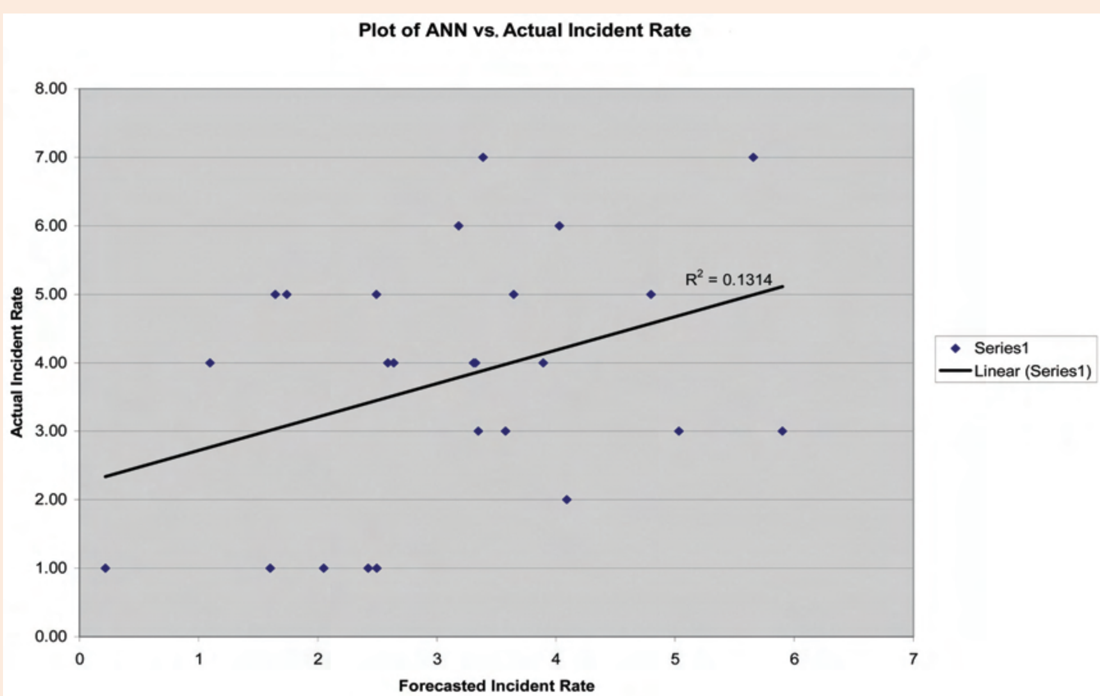


Table 3

Normal Average Comparison

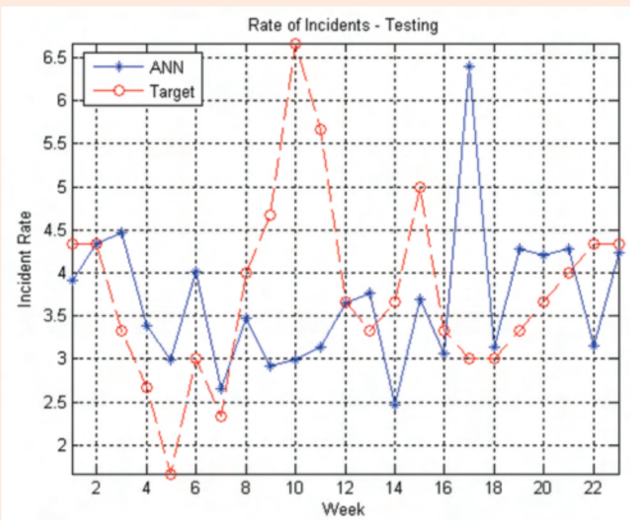
This table summarizes the results of a normal average comparison to determine whether simply taking the average incident rates of 25 weeks and projecting it on every week produced better results than ANN.

	Mean	Average percent error	MAD
ANN	3.13	55.14	1.63
Direct IR Average	3.76	75.86	1.49

Figure 6

Forecasting Accuracy of ANN for Moving Average

MSE = 27.2% R^2 = 0.01. ANN predicted incident rates versus actual incident rates (post-training and validation).



relation test was performed to assess the statistical significance of the R^2 value. This produced a value of 0.639, which is greater than 0.05. This indicates zero statistical significance in ANN's ability to correlate using a moving average.

A regression analysis was performed to correlate the number of hours of safety intervention activities per week with the moving average incident rate. This was done for both actual and ANN forecasted incident rate. The resulting R^2 square for ANN was 0.100, which is a bit higher than the previous regression analysis performed without using a moving average that yielded an R^2 of 0.03. However 0.10 indicates poor correlating power. Also the moving average regression analysis performed for actual incident rates with the number of hours of safety intervention activities per week produced a very poor R^2 value of 0.003.

Finally, a summary of the results comparing the performance of ANN using a moving average as opposed to not using one was tabulated (Table 5, p. 48). The results indicate that a moving average performs better than a direct week-to-week comparison. This is due to a lower absolute average percent error of 27.23, a lower MAD of 1.01 and a higher R^2 of 0.1. However, this does not indicate that ANN is an accurate forecasting tool; it simply performs better with a moving average.

Conclusion & Future Work

After performing the analysis, the hypothesis that an ANN is an accurate predictor of incident rates must be rejected. The low coefficient of determination of 0.10 and a relatively high average percent error indicates low statistical significance in accepting the hypothesis. Furthermore, although this study provided an example in which ANN lacked statistical ability to correlate safety intervention measures with the incident rate, more research is needed to determine whether ANN can be a tool for forecasting incident rates.

The indicator that may show some promise, or at least provide reason to pursue further research, is the relatively small percent error in comparing the overall predicted average incident rate to the actual overall average incident rate. While this is not a strong measure, one could surmise that with additional study, using more data over a longer period, ANN will perform better in predicting incident rates as the frequency distribution of incident rates more closely approximates normal distribution.

It is also important to note that these results are site specific and not applicable industrywide. Some limitations exist with ANNs, such as the exclusion of outlier data and the system's inability to extrapolate the data. ANN effectiveness is as good as the data used to train the system. That said, having an optimized set of input variables can lead to productive results.

This study illustrated the lack of significant statistical difference between the means and the variances as shown by a paired *t-test* and *F-test,* respectively. This does not mean that ANN has the potential to become an accurate predictor of incident rates, but it may prompt further studies and research. More research is needed to gather more data and additional analysis in order to decrease the mean absolute deviation by less than ±1.0 and improving the results of regression analysis. Furthermore different ways of optimizing the data or inputting in the ANN system might produce different results.

One thing is certain, if ANN becomes an accurate predictor by having an MAD less than 1, an absolute percent error of less than 25 and an R^2 value of greater than 0.5, it will unlock doors that will enable companies, firms and businesses to minimize incident rates and safety-related costs by applying the appropriate mix of inputs. If ANNs can show potential for this occurrence, time, injuries and costs can be reduced. ■

(References appear on page 48)

Table 4

Pair-Wise Comparison Between ANN & Actual Incident Rates for a Moving Average

Week	ANN	Actual	Residual (difference)	Absolute percent error
1	3.905	4.333	-0.428	9.882
2	4.329	4.333	-0.004	0.095
3	4.467	3.333	1.134	34.010
4	3.385	2.667	0.718	26.919
5	2.994	1.667	1.327	79.616
6	4.014	3.000	1.014	33.803
7	2.647	2.333	0.313	13.426
8	3.471	4.000	-0.529	13.218
9	2.920	4.667	-1.747	37.426
10	2.992	6.667	-3.675	55.122
11	3.142	5.667	-2.525	44.555
12	3.644	3.667	-0.023	0.626
13	3.758	3.333	0.425	12.740
14	2.456	3.667	-1.210	33.013
15	3.688	5.000	-1.312	26.248
16	3.056	3.333	-0.277	8.314
17	6.392	3.000	3.392	113.067
18	3.139	3.000	0.139	4.637
19	4.273	3.333	0.939	28.184
20	4.200	3.667	0.533	14.537
21	4.279	4.000	0.279	6.983
22	3.148	4.333	-1.185	27.354
23	4.227	4.333	-0.106	2.454
Average	**3.675**	**3.797**	**-0.122**	**27.227**

Table 5

Summary of ANN Results

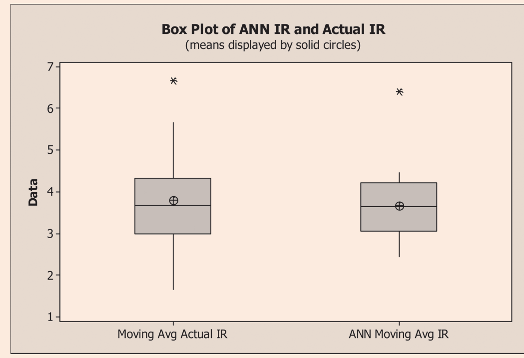
This table summarizes the results comparing the performance of ANN using a moving average as opposed to not using one.

	Residual	Average percent error	MAD	R^2
Direct	-0.63	55.14	1.63	0.03
Moving average	-0.122	27.23	1.01	0.1

On average the forecasted results tend to be lower than the actual incident rates. The average absolute percent error rate indicates that the forecasted results were relatively close to the actual incident rates.

Figure 7

Box Plot of ANN & Actual Incident Rates



References

Bailey, C.W. (1997, Aug.). Managerial factors related to safety program effectiveness: An update on the Minnesota Perception Survey. *Professional Safety, 42*(8), 33-35.

Bailey, C.W. & Petersen, D. (1989, Feb.). Using perception surveys to assess safety system effectiveness. *Professional Safety, 34*(2), 22-26.

Brigham, C.J. (1989, May). Safety program evaluation: One corporation's approach. *Professional Safety, 26*(5), 18-24.

Chandler, B. & Huntebrinker, T.A. (2003, June). Multisite success with systematic BBS: A case study. *Professional Safety, 48*(6), 33-40.

Cleveland, R.J., Cohen, H.H., Smith, M.J., et al. (1979). *Safety program practices in record-holding plants* (NIOSH Publication No. 79-136). Washington, DC: Department of Health and Human Services, CDC, NIOSH.

Defense Advanced Research Projects Agency. (1988). *DARPA neural network study.* Fairfax, VA: AFCEA International Press.

Demuth, H. & Beale, M. (2000). *Neural network toolbox for use with Matlab.* Natick, MA: The Mathworks.

DePasquale, J.P. & Geller, E.S. (1999). Critical success factors for behavior-based safety: A study of 20 industry-wide applications. *Journal of Safety Research, 30*(4), 237-249.

Fausett, L. (1994). *Fundamentals of neural networks.* New York: Prentice Hall.

Geller, E.S., Rudd, J.R., Kalsher, M.J., et al. (1987). Employer-based programs to motivate safety belt use: A review of short-term and long-term effects. *Journal of Safety Research, 18*(1), 1-17.

Guastello, S.J. (1993). Do we really know how well our occupational accident prevention programs work? *Safety Science, 16*(3-4), 445-463.

Haight, J.M & Thomas, R.E. (2003). Intervention effectiveness research: A review of the literature on leading indicators. *Chemical Health and Safety, 10*(2), 21-25.

Haight, J.M., Thomas, R.E., Smith, L.A., et al. (2001a, May). Intervention effectiveness research: Phase 1: Developing a mathematical relationship between interventions and incident rates for the design of a loss prevention system. *Professional Safety, 46*(5), 38-44.

Haight, J.M., Thomas, R.E., Smith, L.A., et al. (2001b, June). Intervention effectiveness research: Phase 2: Design, optimization and verification of the loss prevention system and analysis models. *Professional Safety, 46*(6), 33-37.

Haykin, S. (1994). *Neural networks: A comprehensive foundation.* New York: Macmillan Publishing.

Ingalls, T.S. Jr. (1999, Dec.). Using scorecards to measure safety performance. *Professional Safety, 44*(12), 23-28.

Iyer, P.S., Haight, J.M., Del Castillo, T.B., et al. (2004). Intervention effectiveness research: Understanding and optimizing industrial safety programs. *Chemical Health and Safety, 11*(2), 9-19.

Iyer, P. S., Haight, J. M., Del Castillo, T.B., et al. (2005). A research model: Forecasting incident rates from optimized safety program intervention strategies. *Journal of Safety Research, 36*(4), 341-351.

Masters, T. (1993). *Practical neural network recipes in C++.* San Diego: Academic Press.

O'Sullivan, J.W. (1994, Spring). Neural nets: A practical primer. *AI In Finance.*

Kalsher, M.J., Geller, E.S., Clark, S.W., et al. (1989). Safety belt promotion on a naval base: A comparison of incentives vs. disincentives. *Journal of Safety Research, 20*(3), 103-113.

Petersen, D. (1998, Oct.). What measures should we use and why? Measuring safety system effectiveness. *Professional Safety, 43*(10), 37-40.

Rinefort, F.C. (1992, May). Directions in safety: The economics. *Professional Safety, 37*(5), 42-45.

Rinefort, F.C. (1977, Sept.). A new outlook at occupational safety. *Professional Safety, 22*(9), 36-48.

Smith, M.J, Cohen, H.H., Cohen, A., et al. (1978). Characteristics of successful safety programs. *Journal of Safety Research, 10*(1), 5-15.