U.S. Department of Transportation
Research and Innovative Technology Administration

SR-010

Bureau of Transportation Statistics
Special Report

July 2008

Data-Driven Risk Models Could Help Target Pipeline Safety Inspections

by Rick Kowalewski, Pipeline and Hazardous Materials Safety Administration, and Peg Young, Ph.D., Bureau of Transportation Statistics

Federal safety agencies share a common problem—the need to target resources effectively to reduce risk. One way this targeting is commonly done is with a risk model that uses safety data along with expert judgment to identify and weight risk factors. In a joint effort, the U.S. Department of Transportation's Bureau of Transportation Statistics (BTS) and Pipeline and Hazardous Materials Safety Administration (PHMSA) sought to develop a new statistical approach for modeling risk by *letting the data weight the data*—by using the statistical relationships among the data, not expert opinion, to develop the weights.

Some key findings:

- Weighting data through statistical procedures was superior to judgment-weighting in predicting (targeting) relative risk.
- Statistical modeling can help not only target *which* operators to inspect but also focus *what* to inspect based on a set of risk factors.
- Pipeline infrastructure, operator performance, and incident history appear to be about equally useful in predicting future risk.

Program Background

PHMSA's mission is to protect people and the environment from the risks inherent in the transportation of hazardous materials by pipeline and other modes of transportation. Each year the pipeline safety program inspects several hundred thousand miles of interstate pipelines carrying natural gas and hazardous liquids across the United States. These pipelines are operated by over 1,000 operators who manage systems ranging from a few miles to tens of thousands of miles. While a pipeline might seem to be a very simple system, in fact these systems are very complex, and each system has some unique characteristics.

The general approach for conducting standard inspections until now has been to inspect each major part of each system every 3 years. In 2006, PHMSA initiated a research/pilot project to integrate the various kinds of inspections it conducted, to re-examine the 3-year inspection interval for standard inspections, and to focus the scope of its inspections based on operator risk. Changing inspection intervals from a *periodic*-basis to a *risk*-basis and changing from comprehensive to focused inspections reflect a significant change in approach. Program managers understood from the outset that the new approach would require a better risk model.

The Current Risk Model

For more than a decade, PHMSA has used the Pipeline Inspection Prioritization Program (PIPP) to schedule inspections and allocate resources. PIPP is a data-based model using 10 to 12 data variables (depending on type of pipeline) that are transformed into 9 indexes, which are added together for an overall risk score. The data variables for both hazardous liquid and gas transmission pipelines are listed in table 1.

Beginning with these input variables, each one is transformed into another variable (the individual PIPP scores) ranging from 0 to 9 points, depending on the input variable, and then combined into the final total PIPP score. The variables were selected using expert judgment, and the transformations that determine the weight for each variable also used expert judgment. PIPP results are used with other information to help set scheduling priorities for inspections.

Table 1: Pipeline Inspection Prioritization Program Variables for Hazardous Liquid and Gas Transmission Pipelines

For hazardous liquid pipelines:	For gas transmission pipelines:
1. Number of hazardous liquid pipeline accidents reported per mile of pipeline	1. Miles of coated, but unprotected, pipeline onshore
2. Number of accidents due to corrosion	2. Miles of bare unprotected steel pipeline onshore per mile of bare steel pipeline onshore
3. Number of accidents caused by incorrect pipeline operating practices	3. Miles of bare steel pipeline onshore per mile of steel pipeline onshore
4. Number of accidents due to failed pipe	4. Number of onshore leaks repaired per mile on onshore pipeline
5. Number of accidents due to failure of field welds	5. Known leaks scheduled for repair per mile of onshore pipe
6. Number of accidents due to outside forces	6. Miles of cast iron pipeline
7. Number of failures of longitudinal welds	7. Miles of "other" pipe
8. Total number of barrels of hazardous liquid spilled	8. Number of transmission incident reports
9. Total injuries and deaths not caused by outside force	9. Number of deaths or injuries not caused by third party damage
10. Number of accidents due to corrosion on non-cathodically protected pipe	10. Incidents due to corrosion per mile of onshore pipeline
	11. Incidents due to construction-related defects per mile of onshore pipeline
	12. Number of incidents due to corrosion on onshore pipelines w/o cathodic protection

SOURCE: U.S. Department of Transportation, Pipeline and Hazardous Materials Safety Administration, http://www.phmsa.dot.gov.

NOTE: The difference between accidents (for liquid pipelines) and incidents (for natural gas pipelines) is simply one of terminology. The term "incidents" is often used for both pipeline sectors when a single term is used.

PIPP has been shown to be 3 to 4 times better than random selection in identifying ("predicting") future risk as reflected in the number of pipeline incidents.[1] However, PIPP tends to underestimate risk (substantially) where the actual number of incidents is high, and overestimate risk (somewhat) where the number of incidents is low. This difference is illustrated in the the two PIPP score scatterplots in figure 1 for hazardous liquid pipelines and for natural gas pipelines, respectively.

The New Model

The new model predicts the number of pipeline incidents and the incident rate per mile of pipeline for each pipeline operator. To develop predictions, researchers took several years of historical data to run simulations—using, for example, data from 2002 to 2004 to "predict" 2005. The data were organized conceptually into three sets, each using different data; the results are reflected in the six remaining "risk" scatterplots in figure 1:

1. The *inherent risk* associated with the pipeline—represented by physical and operating characteristics such as age, materials and coatings, diameter, location, and throughput—is estimated using annual reports submitted by each pipeline operator.[2] Inherent risk should be independent of how the pipeline is managed and maintained.

2. The *performance risk* associated with the operator (i.e, the company)—represented by safety deficiencies—is estimated using the results of past safety inspections—particularly those with the broadest scope, known as Integrity Management (or IM) inspections.[3] Performance risk should be independent of the pipeline characteristics.

3. The *historical risk* associated with past incidents is estimated from incident data reported to PHMSA by operators.[4] Historical risk is assumed to reflect the combination of both inherent risk of the pipe and performance risk of the operator.

Each set of data generated separate predictions of future incidents that were also combined into a single prediction for each operator. The diagonal line in each graph in figure 1 represents perfect prediction in which the predicted number of incidents equals the actual number of incidents. The further the data points are from the diagonal line, the poorer the performance of the predictive model. Gas transmission operators were separated from hazardous liquid operators, as they are in PIPP, because they present very different system profiles, different risks, different data, and different numbers of incidents (see table 2). Other breakouts might also make sense (e.g., by product for liquid pipelines, or onshore v. offshore pipeline) but the research has not explored these.

For presentation purposes, small operators (with less than 500 miles of pipeline) were separated from large operators because their operating environment tends to be different and the relatively lower number of incidents makes the results somewhat less reliable. The analysis behind all the models were performed in the statistical software package SAS 9.1.

[1] By scaling PIPP scores to the number of actual incidents, predictive quality was measured by the correct "hits" to determine the percent correct. This was compared to a random selection model where each operator was simply assigned an equal share of points.

[2] See www.phmsa.dot.gov for access to annual reports filed by pipeline operators.

[3] Deficiency data are captured at the point of inspection for Integrity Management (IM) inspections of pipeline operators. Where deficiencies are serious, PHMSA pursues enforcement action. Data on these actions are available at www.phmsa.dot.gov.

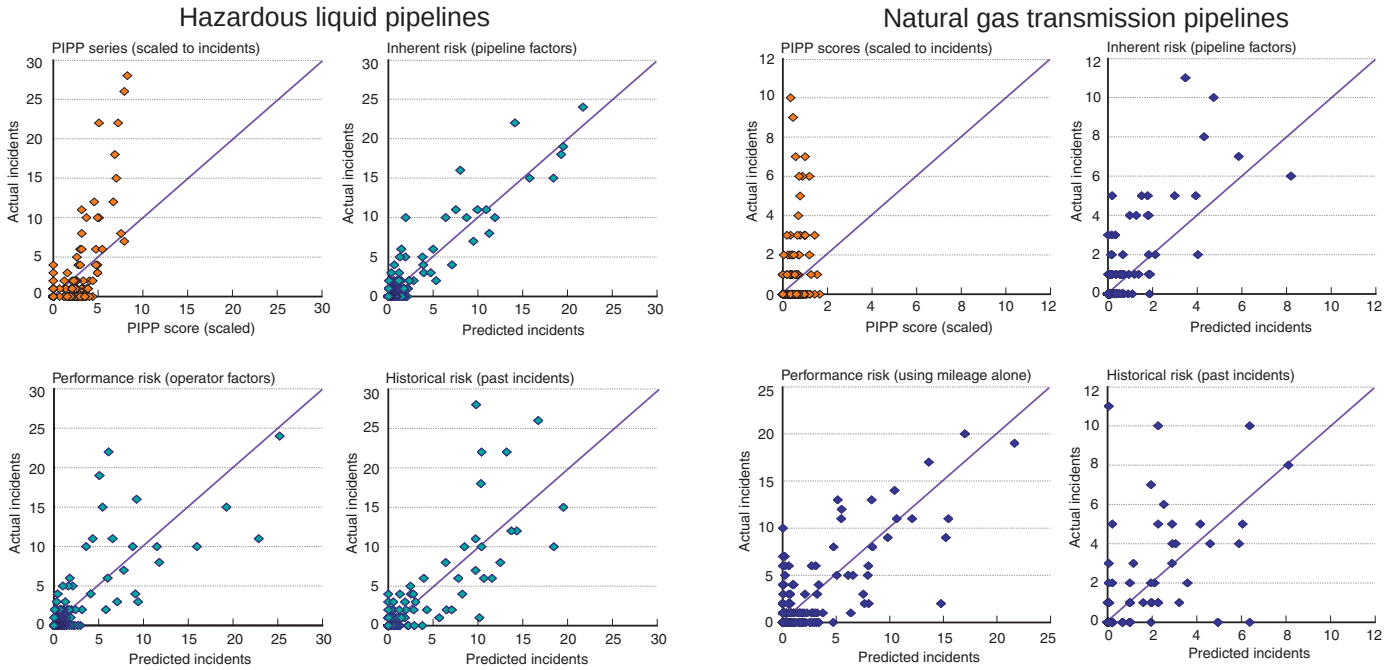
[4] Incident data are available at www.phmsa.dot.gov.

Table 2: Liquid v. Transmission Systems Characteristics

Hazardous liquid pipelines	Gas transmission pipelines
~ 300 operators	~ 900 operators
~ 50 large operators	~ 90 large operators
~ 350 accidents/year	~ 130 incidents/year
More infrastructure data in annual reports	Less infrastructure data in annual reports
Integrity Management inspection data for most operators	Less Integrity Management inspection data for operators
Overall – more data to work with	Overall – less data to work with

SOURCE: U.S. Department of Transportation, Pipeline and Hazardous Materials Safety Administration, http://www.phmsa.dot.gov.

Figure 1: Predicted Incidents v. Actual Incidents for Four Different Models in Each Pipeline Sector



NOTES: See text for explanation of diagonal line in each graph. PIPP = Pipeline Inspection Prioritization Program.

SOURCE: U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, http://www.bts.gov.

Statistical approaches

Three key characteristics of the data influenced the choice of statistical models:

1. Incidents occur infrequently, so the models would have to deal well with small numbers.
2. The number of incidents is a count value, with no fractional or negative values.
3. The number of incidents per operator is highly skewed, with a large number of operators having zero incidents in any given year.

Traditional linear regression, which relies on the assumption of normally distributed data, is inappropriate for count data that are highly skewed towards zero. Two other models—the Poisson distribution and negative binomial regression[5]—can handle such data. Another important quality of these two models is their ability to control for exposure variables, such as miles of pipeline. The negative binomial is the more general model, and this was used to detect and weight risk variables for both inherent risk and performance risk.[6]

[5] In a recent review of the Motor Carrier Safety Status Measurement System, or SAFESTAT, model used by the Federal Motor Carrier Safety Administration, the Government Accountability Office (GAO) recommended a negative binomial regression in place of expert opinion to weight the risk factors used in targeting motor carrier safety inspections. This work by GAO was a strong factor in the risk modeling effort by BTS and PHMSA. See *Motor Carrier Safety: A Statistical Approach Will Better Identify Commercial Carriers That Pose High Crash Risks Than Does the Current Federal Approach,* June 2007 (GAO-07-585).

[6] For a good explanation of the Poisson and negative binomial models and how they are estimated in SAS, see *Logistic Regression Using SAS: Theory and Application*, by Paul D. Allison, 1999 (SAS Institute Inc.).

The analysis of the historical risk associated with past incidents presented a different set of conditions. The past 3 years of incidents and the next (to-be-predicted) year of incidents most likely are not independent from one another, so the data were transformed to create an "orthogonal" regression model that would allow modeling the 3 years of incidents together to estimate future risk. [7]

Each of these major outputs—inherent risk, performance risk, and historical risk—provide a separate prediction of risk, but they can also be combined to present a single estimate. The approach taken here was to take the average of the three results.[8] Other possibilities not examined here might use another model to weight these three as inputs to an overall risk score, again *letting the data weight the data*, or developing an equation that might relate any one output to the other two. Figure 1 provides a graphical synopsis of the predictive accuracy for estimating the number of accidents per operator based on PIPP scores, inherent risk, operator risk, and historical risk.

The predictive quality of each model tested was compared using a standard statistical measure of error—the mean absolute deviation (MAD)—which averages the absolute difference between the predicted value and the actual value for each operator (see table 3). For example, when the model predicts 7.5 incidents and 5 actually occur, the error is 2.5; when the model predicts 4 incidents and 5 actually occur, the error is 1. MAD provides a sense of "how far off" the model predictions are from the actual values.

[7] "Orthogonal variables" are linearly independent. For details on orthogonal regression, see A. Stuart, J.K. Ord, and S.F. Arnold. 1999. Kendall's Advanced Theory of Statistics, 6th ed. London: Edward Arnold, pp. 764-766.

[8] Although historical risk—using incident data—might reflect the nexus of the inherent risk associated with the pipeline and the performance risk associated with the operator, using equal weights to average provides a simple approximation of overall risk. Other statistical methods might provide a better way to combine these factors.

Testing Inputs to the Model

A key indicator for the effectiveness of any new model was its ability to predict risk better than the existing judgment-weighted model (PIPP ranking). In practice, this should be fairly easy because a statistical model could simply re-weight the 10 input variables in PIPP or the 9 transformed variables for a better prediction using data-weighting. Other obvious inputs to test included:

- the naïve model (which says that what happened last year is likely to happen again next year);
- mileage alone (which suggests that the extent of the system might be the most important indicator of the risk of incidents);
- the input variables into PIPP—reweighted using the new statistical procedures;
- the output variables (L-scores) from PIPP before the PIPP ranking is calculated—reweighted using the new statistical procedures; and
- each of the new indicators of risk—estimating inherent risk associated with the pipeline, performance risk associated with the operator, and historical risk associated with past incidents.

The results demonstrate that PIPP performs the worst in targeting risk, and that reweighting the PIPP variables can improve the predictive quality (reduce the error). Surprisingly, mileage alone and the naïve model both were better (smaller error) than PIPP in predicting future risk, but such simple models offer little guidance in selecting appropriate sites to inspect. The new model performed well (with a MAD of 1.0), although the analysis indicated noticeable differences between gas transmission operators and hazardous liquid operators. Hazardous liquid pipeline incidents are more prevalent and more concentrated (fewer operators), so the data provide a better basis for prediction.

Table 3: Comparison of Model Results Using Mean Absolute Deviation (MAD) for Hazardous Liquid Pipelines

Model	Description	MAD
Naïve model	Previous year's incident count	1.0
Mileage alone	Modeled with negative binomial	1.0
PIPP rank	As presently used by PHMSA	7.6
PIPP total score	Modeled with negative binomial	2.0
PIPP input variables	Modeled with negative binomial	1.1
PIPP output variables (L-scores)	Used to calculate PIPP components and modeled with negative binomial	3.2
Inherent risk – using pipeline data	Infrastructure variables from annual reports and modeled with negative binomial	1.0
Performance risk – using inspection data about operators	Inspection results from IM inspections	1.5
Historical risk – using incident data	Number of past three year's incidents, modeled with orthogonal regression	1.1

SOURCE: U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, http://www.bts.gov.

- Small numbers of incidents each year limit the ability to isolate combinations of factors that might be statistically significant.

Continuing Research

The first line of research, currently underway, is to refine the incident measures to reflect the *consequences* of incidents—to weight incidents by potential severity in terms of harm to people and/or the environment. Using conditional probabilities, we have found so far that three variables help explain whether an incident is likely to be serious: fire/explosion (indicating a violent incident), whether the incident occurred in a high consequence area (indicating proximity to people), and incident cause (e.g., corrosion or excavation damage).

Some general model improvements are planned as well. These would separate out onshore v. offshore systems, interstate v. intrastate operators, and certain commodities that have special risk characteristics. The relationship between inherent risk, performance risk, and historical risk needs to be further explored and modeled. The issue of total number of incidents v. the rate of incidents per mile needs to be addressed; it is not clear which is more important in targeting inspections. And operator relationships—where some operators are part of a larger group of operators that share certain plans and management—need to be addressed because some inspections are targeted at this higher corporate level.

There are several areas where the measures for inherent risk, performance risk, and historical risk could be enhanced. Improvement would include targeted analyses of certain key variables to better understand why they are or aren't significant risk factors, adding more inspection data, and testing the time-sensitivity of inspection data.

After refinements are made, the model needs to be validated with data from other years, uncertainty should be incorporated into the results, and PHMSA program staff need to be involved in formulating the best presentation of results for the intended use—targeting and focusing inspections.

A parallel effort will extend the concepts from this modeling effort to another safety program—hazardous materials transportation safety—which cuts across four other modes of transportation. The model might be more generally applicable in other federal safety programs as well.

About this Report

This report is the result of joint research by Rick Kowalewski, Senior Advisor of the Pipeline and Hazardous Materials Safety Administration (PHMSA), and Peg Young, Statistician for the Bureau of Transportation Statistics (BTS).

For related BTS data and publications: www.bts.gov

For questions about this or other BTS reports, call 1-800-853-1351, email answers@bts.gov, or visit www.bts.gov.

The three main components of the new model—inherent risk, performance risk, and historical risk—performed about equally well in predicting future incidents.

Findings From the Modeling Research

Modeling inherent risk associated with the pipeline demonstrated that mileage, throughput (barrel-miles per year), date of installation, and pipeline diameter were significant risk factors. Six variables were significant in predicting future incidents for gas transmission systems, and 14 variables were significant for hazardous liquid systems. About half of these variables were negatively correlated with risk, meaning that they had a "protective effect." (Table 4 provides the listing of the significant variables for both models.)

Modeling performance risk associated with the operator demonstrated that a few key inspection areas from Integrity Management[9] inspections were most highly correlated with future risk. One area (*integrity assessment review*) was negatively correlated, suggesting that finding deficiencies in this area helped an operator rapidly improve its safety program. The most significant risk factor was in the area of *continual evaluation and assessment*—which inspection staff have suggested might be a critical indicator of an operator's safety program.

[9] The Integrity Management program was introduced over the last several years, first for hazardous liquid pipelines then later for gas transmission pipelines. This program requires pipeline operators to identify and understand the risks in their systems, identify high consequence geographic areas, establish programs for inspecting and repairing pipelines, and continuously monitoring their systems.

Modeling historical risk associated with past incidents demonstrated that the passage of time rapidly degrades the utility of the data. After 2 years, past incidents do not appear to be useful in predicting future risk. The most recent year is most important, and the model weights this year most heavily.

Significant Data and Modeling Issues

While the model demonstrates the general effectiveness of statistical tools as an alternative to judgment-weighting, several important data limitations and modeling issues remain to be addressed. Some of the more important issues are listed here:

- Data on operators' systems and operator relationships reflect a snapshot in time; changes might not be captured for up to a year, so some data are outdated.
- Deficiency data from inspections are largely limited to one major type of inspection—Integrity Management inspections—representing only a small portion of the inspections conducted.
- The model does not differentiate more serious incidents (the focus of the agency's performance goals) from those with less severe consequences (actual or potential).
- The model introduces an exponential function that can dramatically over-predict incidents when new data are outside the historical range.

Table 4: Significant Variables in the Inherent Risk Models for Gas Transmission and Hazardous Liquid Pipelines
(negative sign indicates a negative correlation with risk)

Gas transmission model	Hazardous liquid model
+ Total Miles	+ Total Miles
+ Miles offshore	+ Miles from 26 to <28" diameter
+ Miles from 4 to <10" diameter	+ Miles pre-1920
+ Miles from 20 to <28" diameter	+ Offshore >20% Specified Minimum Yield Strength
- Miles from 0 to <4" diameter	+ Onshore <20% Specified Minimum Yield Strength
- Miles installed 2000-2009	+ Regulated gathering lines
	+ Total barrel-miles
	+ Highly Volatile Liquid barrel-miles
	- Miles from 8 to <10" diameter
	- Miles from 22 to <24" diameter
	- Miles from 24 to <26" diameter
	- Miles from 30 to <32" diameter
	- Miles >36" diameter
	- Anhydrous ammonia barrel-miles

SOURCE: U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, http://www.bts.gov.