Statistical Comparison of Train Accident Rates: Methodology and Decision Support Tool

Xiang Liu Ph.D., Corresponding Author
Assistant Professor
Department of Civil and Environmental Engineering
Rutgers, The State University of New Jersey
96 Frelinghuysen Road, Piscataway, NJ 08854-8018
Tel: (848)445-2868; Email: xiang.liu@rutgers.edu

Daniel F. Rodriguez
Research Assistant
Department of Civil and Environmental Engineering
Rutgers, The State University of New Jersey
96 Frelinghuysen Road, Piscataway, NJ 08854-8018
Tel: (848)218-8294; Email: dfr49@scarletmail.rutgers.edu

5,229 Words + 5 tables/figures × 250 (each) = 6,229 Words

August 1, 2016
ABSTRACT

Freight railroads transport 43 percent of intercity ton-miles of cargo in the United States. Accidents can disrupt rail services and potentially cause significant consequences with respect to human health, property, and the environment. Therefore, it is important to evaluate and manage railroad operational safety based on proper safety metrics. One common metric is train accident rate, which is defined as the number of accidents normalized by traffic exposure. Although empirical (observed) comparisons using this rate have been extensively studied for comparing railroad safety performance, it does not account for random fluctuations in accident occurrences. When these empirical rates differ, it is not well understood what proportion of this difference is caused by stochastic variation and what proportion is reflective of actual safety change. This paper aims to narrow this knowledge gap theoretically and practically. First, a generalized statistical methodology is provided for comparing train accident rates. Based on empirical accident rates, this methodology can determine whether one rate is statistically higher or lower than the other, across different railroads or time periods. This methodology is implemented into a computer-aided decision support tool that allows analysts to perform an automated statistical comparison of train accident rates. The methodology and decision support tool developed in this research can assist the railroad industry in monitoring, evaluating, and comparing these rates. In the future, this research can also support the industry’s ongoing initiative in developing a data-driven, safety management system based on a quantitative understanding of past accident data.

Keywords: Train accident rates, statistical comparison, decision support tool
1 INTRODUCTION

The American economy hinges on a safe and reliable freight rail transportation system, which annually transports over 600 billion dollars’ worth freight that accounts for around 43 percent of ton-miles of intercity cargo (1, 2). Disruptions to this system can affect its operational safety and efficiency, resulting in damage costs, casualties, or environmental impact. For example, in 2015, train accidents had a reported damage cost of about 360 million dollars and were responsible for 859 casualties (3).

Railway safety management is a high priority for the Federal Railroad Administration (FRA) of US Department of Transportation (USDOT) and the railroad industry. It is also an active research area in academia. The first step in railway safety risk management is to develop a proper metric for safety evaluation. A train accident rate is one of the most commonly used metrics employed by government agencies, the railroad industry, and academics. By definition, an accident rate is the number of accidents normalized by traffic exposure (e.g., train-miles, car-miles or gross ton-mile) (4-7). U.S. train accident data is available through the FRA Rail Equipment Accident database (8). Traffic exposure data can be obtained through two databases. Train-mile data is available through the FRA Operational Safety Database (9) and Surface Transportation Board (STB) Class I railroad annual reports (Form R1) (10). Car-mile and ton-mile data are only available from the STB R1 form.

Empirical accident rates have been widely used in numerous railroad safety and risk studies (4, 5, 7, 11, 12). One application of these rates is to compare safety using accident rates. According to the current practice, a lower empirical rate may indicate a better safety performance. For example, suppose that two railroad companies are compared. The first company has an accident rate of 1.8 accidents per million car-miles in 2007 while the second railway has an accident rate of 1.9 accidents per million car-miles in the same year. Purely based on the empirical rates, it may be stated that the first railroad has a better safety performance than the second in that year. Similar logic has been used in many previous studies in which comparisons are solely based on empirical accident rates.

However, recent research has found that empirical rates are subject to a statistical error called Regression to the Mean (RTM). The RTM is a statistical phenomenon in which a random variable (empirical accident rate) will fluctuate around its mean value (“true” accident rate). Therefore, a lower rate may be followed by a higher rate in sequent years, even when there are no changes to the actual safety (Figure 1).

![FIGURE 1 Illustration of the RTM phenomenon](image-url)
Because of the RTM, the empirical comparison of accident rates does not reflect the actual safety change. For example, in Figure 2, the two lines reflect the “true” safety level of two railroads, respectively. Railroad A (red solid line) has a lower “true” accident rate than that of railroad B (black dash line). Each dot represents the empirical accident rate for each railroad (circle for railroad A and triangle for railroad B). Although railroad A has a better safety performance (lower “true” accident rate) than railroad B, for several years (e.g. year 2004), its empirical rates appear to be higher, because of the RTM described above.

FIGURE 2 Illustration of the effect RTM on empirical accident rate comparison

While the effect of RTM on safety analysis has been well studied in highway sector, it has not been formally recognized in the rail sector until recently, when researchers started to prove that train accident occurrence follows a random distribution (13) and formalized the statistical way to analyze train accident data. For example, if the difference of empirical accident rates between two railroads is 0.1 accidents per million car-miles in one year, are we confident enough to declare that this difference is statistically significant in that the two “true” rates are not identical? Put another way, based on historical accident data, to what extent can analysts say that the safety has changed?

The quest for the answer to this question motivates developing this research. This paper aims at analyzing empirical accident rates in a normative, statistical way to understand whether there is a statistical difference between entities (e.g., different railroads, different periods). While this paper is scoped at formalizing a normative, statistical methodology for comparing accident rates, it has a wide application to different railroad safety problems in which accident rates are used to make decisions. Considering that decision makers need a practice-ready tool when using the sophisticated statistical methodology, the algorithm is also implemented into a web-based decision support tool that can facilitate the multi-comparison of accident rates within seconds.

This paper is structured as follows. First, relevant literature is reviewed in order to explain how this work is related to, but differs from the prior effort in railroad safety research. Second, the goals of this research are elaborated and the means to achieve each goal are presented. Third, a statistical methodology is presented alongside a rigorous mathematical proof. Fourth, a web-based decision support tool is introduced that implements the statistical
methodology. Fifth, the decision support tool is applied to several real-world examples as to explain the practical use of the method. Lastly, the contributions of this work are clarified with respect to both research and practice, and suggest possible future research directions to better aid the railroad industry in its long-lasting pursuit of improving operational safety.

2 LITERATURE REVIEW

A number of previous studies have focused on accident rate analysis (5, 6, 11, 14, 15) and accident severity analysis (16-18). This paper focuses on the former. Previous accident rate analyses have emphasized the use of empirical accident rates in an effort to evaluate railroad safety change (4, 5, 11). However, there is little research in regards to how to statistically compare accident rates between railroads or time periods. As Liu (13) proves, the occurrence of train accidents follows a random distribution (13). This means that if the “true” accident frequency is two, the empirical number of accidents could be one, three, or four (any number around two). This reflects the RTM phenomenon described above (19).

The stochastic variation present in train accident frequency has an impact on accident rate comparison. The difference between two empirical rates may be due to (or in part due to) random fluctuation rather than actual changes in safety. Therefore, it is of keen interest to understand what proportion of this difference between empirical accident rates is attributable to statistical variation and what proportion is due to actual safety changes. Separating the change due to random fluctuation from the change in actual safety is important for the development of safety analysis and policy making. To this end, researchers have developed two types of statistical analyses, exploratory and explanatory analysis. The former focuses on identifying what the pattern is so that the latter can then be used to explain it. Some previous studies have employed regression analysis to explore the change of train accident rate, and correlated this change with certain influencing or causal factors (6, 13-15, 20). However, this prior effort exclusively focuses on nationwide safety analysis within multiple years. To the authors’ knowledge, there is no published study yet that explicitly compares train accident rates between different railroads or time periods, based on a state of the art statistical theory. The following sections will describe a statistical methodology to address this knowledge gap.

3 RESEARCH OBJECTIVES

This research aims at formalizing statistical methods to compare train accident rates, accounting for the stochastic uncertainty in accident occurrence. To advance the state of the art and practice, this research has the following objectives:

1) Develop a statistical methodology that compares train accident rates between different railroads or time periods. This is to understand the change of safety, accounting for the fact that some changes might be due to randomness (white noise) while others are actually related to safety activities.

2) Implement the methodology into a decision support tool that can serve a practical use of the advanced statistical theory.

3) Explain the use of the statistical tool based on Class I railroad train accident data from 2000 to 2014.
It is hoped that through attaining these objectives, the railroad industry will be provided with practical tools that adequately apply statistical sciences in learning about past safety performance and improve that in the future.

### 4 METHODOLOGY FOR EVALUATION OF TRAIN ACCIDENT RATES

#### 4.1 Statistical Description of Train Accident Rate

It is assumed that each time a train enters a track segment there is a probability (p) that this train is involved in an accident. Train accident probability is affected by infrastructure conditions, train characteristics, operational factors, environmental factors and many other variables (4, 5). Holding these conditions constant, it can be assumed that accident probability is constant (assuming homogeneous track characteristics, rolling stock, and operational conditions within the study period). Under these assumptions, each train pass can be viewed as a Bernoulli experiment (the Bernoulli probability is denoted as p). The probability theory tells that the sum of independent, identically distributed (IID) Bernoulli variables constructs a binomial distribution (Equation 1):

\[
P(X = n) = \binom{N}{n} p^n (1-p)^{N-n}
\]

Where:

- \(n\) = number of train accidents
- \(N\) = total number of train passes on a given segment during the study period
- \(p\) = probability that a train is involved in an accident each time it enters a segment

Let \(p = \frac{\lambda}{N}\), given a large number of train passes (\(N\) is large) and relatively low accident probability (\(p\) is sufficiently small), Equation (1) can be re-written as:

\[
\lim_{N \to \infty} P(X = n) = \lim_{N \to \infty} \binom{N}{n} \left( \frac{\lambda}{N} \right)^n \left( 1 - \frac{\lambda}{N} \right)^{N-n} = \frac{\lambda^n \exp(-\lambda)}{n!} = \text{Poisson}(\lambda)
\]

Equation (2) indicates that the number of train accidents within traffic exposure can be approximated by a Poisson distribution. This assumption was adopted in several previous studies (6, 12-15) without explanation. In the Poisson distribution, the Poisson mean (\(\lambda\)) represents the expected train derailment count. Consider two railroads, each of which has their respective accident count and traffic volume. Based on the discussions above, the number of accident given a traffic volume approximately follows a Poisson distribution:

\[
y_1 | M_1 \sim \text{Poisson}(\lambda_1) \quad (3)
\]

\[
y_2 | M_2 \sim \text{Poisson}(\lambda_2) \quad (4)
\]

Where:

- \(y_1\) = accident count for the first entity
- \(y_2\) = accident count for the second entity
\[ M_1 = \text{traffic exposure for the first entity (e.g., train-miles, car-miles or ton-miles)} \]
\[ M_2 = \text{traffic exposure for the second entity} \]
\[ \lambda_1 = \text{Poisson mean for the first entity, given traffic volume } M_1 \]
\[ \lambda_2 = \text{Poisson mean for the second entity, given traffic volume } M_2 \]

According to Hauer (19), “Safety is the number of accidents by kind and severity, expected to occur on the entity during a specified period”. Based on this definition, \( \lambda \) (expected accident count given traffic volume) represents the “true” safety level for a given traffic exposure. Correspondingly, the Poisson rates for the two railroads are \( \lambda_1/M_1 \) and \( \lambda_2/M_2 \), which represent the “safety” per unit exposure of traffic (aka. Poisson rate). Let \( \theta \) represent the ratio of the two Poisson rates:

\[ \theta = \frac{\lambda_1/M_1}{\lambda_2/M_2} \] (5)

If \( \theta = 1 \), it means that the two Poisson rates are equal. If \( \theta > 1 \), the Poisson rate for entity 1 is higher than the second; otherwise, the second entity has a higher Poisson rate. Because \( \lambda_1 \) and \( \lambda_2 \) (expected accident count) are unknown quantities, \( \theta \) is unknown and needs to be estimated based on sample data. Let \( \theta^* \) denote a point estimator of \( \theta \). The following subsection will detail the calculation of \( \theta^* \) and its confidence interval (the interpretation of confidential interval will be provided later).

### 4.2 Statistical Comparison of Train Accident Rates

**Proposition 1**: \( \theta^* = \frac{Y_1 M_2}{Y_2 M_1} \) is a maximum likelihood estimator (MLE) of \( \theta \) (6)

Proof: The joint distribution of independent \( Y_1 \) and \( Y_2 \) is

\[ f(Y_1, Y_2) = \left( \frac{\lambda_1^{y_1} e^{-\lambda_1}}{y_1!} \right) \left( \frac{\lambda_2^{y_2} e^{-\lambda_2}}{y_2!} \right) \] (7)

The log-likelihood function for the joint distribution is

\[ \log(f(Y_1, Y_2)) = y_1 \log \lambda_1 - \lambda_1 - \log y_1! + y_2 \log \lambda_2 - \lambda_2 - \log y_2! \] (8)

Define \( \lambda = \lambda_1/\lambda_2 \), we can replace \( \lambda_1 \) by \( \lambda \lambda_2 \), then

\[ \log(f(Y_1, Y_2)) = y_1 \log \lambda \lambda_2 - \lambda \lambda_2 - \log y_1! + y_2 \log \lambda_2 - \lambda_2 - \log y_2! \] (9)

Take the first-order derivative of log-likelihood function with respect to \( \lambda \), we get

\[ \frac{\partial}{\partial \lambda} \log(f(Y_1, Y_2)) = \frac{y_1 \lambda_2}{\lambda \lambda_2} - \lambda_2, \text{ set this partial derivative be } 0, \text{ we have } \lambda = y_1/\lambda_2 \] (10)

Taking the first-order derivative of log-likelihood function with respect to \( \lambda_2 \), we have
\[
\frac{\partial \log(f(y_1, y_2))}{\partial \lambda_2} = \frac{y_1 \lambda_2}{\lambda_2} - \lambda + \frac{y_2}{\lambda_2} - 1, \text{ set this partial derivative be 0, we have equation (11)}
\]

\[
\frac{y_1}{\lambda_2} - \lambda + \frac{y_2}{\lambda_2} - 1 = 0
\]

Because \(\lambda = y_1/\lambda_2\), based on Equation (11), we have the MLE of \(\lambda_2\) as follows:

\[
\lambda_2^* = y_2 \quad (* \text{ represents a maximum likelihood estimator})
\]

Based on the invariant property of MLE estimators (21), the MLE of \(\lambda\) is:

\[
\lambda^* = \left(\frac{y_1}{\lambda_2}\right)^* = \frac{y_1}{y_2}
\]

By definition, \(\lambda = \lambda_1/\lambda_2\)

Hence, a maximum likelihood estimator of the ratio of two Poisson mean is the ratio of the empirical accident count:

\[
\left(\frac{\lambda_1}{\lambda_2}\right)^* = \frac{y_1}{y_2}
\]

Correspondingly, a maximum likelihood estimator of two Poisson rates is

\[
\left(\frac{\lambda_1 / M_1}{\lambda_2 / M_2}\right)^* = \frac{y_1 M_2}{y_2 M_1}
\]

Therefore, we have the MLE for \(\theta\) is \(\theta^* = \frac{Y_2 M_2}{Y_2 M_1}\). Proof is completed.

**Proposition 2:** The confidence interval of \(\theta^*\) (the ratio of accident rates) is

\[
LB = \left(\frac{y_1}{y_1 + y_2} + \frac{z^2}{2(y_1 + y_2)} - \frac{z_{\alpha/2}}{2} \sqrt{\frac{y_1 y_2}{y_1 + y_2} + \frac{z^2}{4(y_1 + y_2)} / 4(y_1 + y_2)} / (y_1 + y_2) / (1 + z^2 / (y_1 + y_2)) M_1 \right)^* \left(\frac{y_1}{y_1 + y_2} + \frac{z^2}{2(y_1 + y_2)} - \frac{z_{\alpha/2}}{2} \sqrt{\frac{y_1 y_2}{y_1 + y_2} + \frac{z^2}{4(y_1 + y_2)} / 4(y_1 + y_2)} / (y_1 + y_2) / (1 + z^2 / (y_1 + y_2)) M_2 \right)
\]

\[
UB = \left(\frac{y_1}{y_1 + y_2} + \frac{z^2}{2(y_1 + y_2)} + \frac{z_{\alpha/2}}{2} \sqrt{\frac{y_1 y_2}{y_1 + y_2} + \frac{z^2}{4(y_1 + y_2)} / 4(y_1 + y_2)} / (y_1 + y_2) / (1 + z^2 / (y_1 + y_2)) M_1 \right)^* \left(\frac{y_1}{y_1 + y_2} + \frac{z^2}{2(y_1 + y_2)} + \frac{z_{\alpha/2}}{2} \sqrt{\frac{y_1 y_2}{y_1 + y_2} + \frac{z^2}{4(y_1 + y_2)} / 4(y_1 + y_2)} / (y_1 + y_2) / (1 + z^2 / (y_1 + y_2)) M_2 \right)
\]

\[
(16)
\]
Where:

\[ \text{LB} = \text{lower bound of the confidence interval of the ratio of two accident rates with Type I error } \alpha \text{ (when } \alpha = 0.05, \text{ it represents 95\% confidence interval)} \]

\[ \text{UB} = \text{upper bound of the confidence interval of the two accident rates} \]

\[ y_1, y_2 = \text{accident count on entity 1 and entity 2} \]

\[ M_1, M_2 = \text{traffic volume on entity 1 and entity 2} \]

Proof: Because \( y_1 \) and \( y_2 \) follows a Poisson distribution, let \( n = y_1 + y_2 \). \( y_1 \) given \( n \) (denoted as \( y_1|n \)) follows a binomial distribution (22). Let \( p = \lambda_1/(\lambda_1 + \lambda_2) \) represent the binomial probability \( p \) and an estimator of this probability \( \hat{p} = y_1/n \). Based on these definitions and notations, we have the following relationship based on statistical inference theories (21):

\[
-z_{\alpha/2}\sigma \leq \hat{p} - p \leq z_{\alpha/2}\sigma , \text{ where } \sigma = \sqrt{\frac{p(1-p)}{n}}
\] (17)

To solve \( p \) for inequality \( |\hat{p} - p| \leq z_{\alpha/2}\sqrt{p(1-p)/n} \), taking squared on both sides, we have

\[
(\hat{p} - p)^2 \leq z_{\alpha/2}[p(1-p)/n], \text{ equivalently, we have } (\hat{p}^2 + p^2 - 2\hat{p}p) \leq z_{\alpha/2}[p/n - p^2/n] \] (18)

Note that it’s quadratic inequality. By quadratic formula, Equation (18) is equivalent to the following equation:

\[
\left(\hat{p} + \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2}\sqrt{\left[\hat{p}(1-\hat{p}) + \frac{z_{\alpha/2}^2}{4n}\right]/n}\right)/(1 + z_{\alpha/2}^2/n) \leq p
\]

\[
\leq (\hat{p} + \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2}\sqrt{\left[\hat{p}(1-\hat{p}) + \frac{z_{\alpha/2}^2}{4n}\right]/n}\right)/(1 + z_{\alpha/2}^2/n)
\] (19)

Therefore, the upper (UB\(_p\)) and lower bound (LB\(_p\)) of the confidence interval for \( p \) are as follows:

\[ \text{LB}_p = \left(\hat{p} + \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2}\sqrt{\left[\hat{p}(1-\hat{p}) + \frac{z_{\alpha/2}^2}{4n}\right]/n}\right)/(1 + z_{\alpha/2}^2/n) \] (20)

\[ \text{UB}_p = \left(\hat{p} + \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2}\sqrt{\left[\hat{p}(1-\hat{p}) + \frac{z_{\alpha/2}^2}{4n}\right]/n}\right)/(1 + z_{\alpha/2}^2/n) \] (21)

By definition,

\[ \text{LB}(\theta) = \text{LB}\left(\frac{\lambda_1 M_2}{\lambda_1 M_1}\right) = \frac{M_2}{M_1} \text{LB}\left(\frac{\lambda_1}{\lambda_2}\right) \text{ (LB is the lower bound)} \] (22)

By definition...
\[ p = \frac{\lambda_1}{\lambda_1 + \lambda_2} \]

Therefore,

\[ LB \left( \frac{\lambda_1}{\lambda_2} \right) = \frac{LB_\rho}{1 - LB_\rho} \] (23)

Similarly,

\[ UB \left( \frac{\lambda_1}{\lambda_2} \right) = \frac{UB_\rho}{1 - UB_\rho} \] (24)

According to Equations (19) to (24), the lower bound and upper bound of the ratio of the two Poisson-based train accident rates can be derived. The proof is completed.

5 DECISION SUPPORT TOOL

As a practical application of the statistical methodology produced within this paper, an online decision support tool based on this methodology has also been developed. Essentially, this tool (Figure 3a) is a calculator that provides an input table for raw data, an output table for the resulting statistical comparison of accident rates (Figure 3b), and a graphic visualization of the resulting data (Figure 3c).

The input data table used for this example (Figure 3a) has four vertical input columns for each year. The variables Y and M are similar to those presented in the methodology with the Y’s representing accident count values and the M’s denoting traffic exposure values. Among these variables, the number 1 denotes values belonging to the first railroad, while the number 2 denotes another railroad. Along the bottom of the table, options are available for creating additional rows of data.
FIGURE 3(a) Decision support tool input data table

Note: This example compares the train accident rate of CSX (denoted as 1) against the average rate of the three other companies (denoted as 2) from 2000-2014.

Imbedded within the tool, a table regarding outputs for the statistical calculations (Figure 3b) is offered. This table provides values for the MLE (point estimate) and its corresponding confidence interval, including both an upper and lower bound. The ratio of the two “true” accident rates (the first railroad rate to the second railroad rate) has a 95% chance of being between the lower and upper bounds of the confidence interval. For example, in the year 2000, the ratio of the two empirical rates is 1.0365, with the 95% confidence interval of 0.7931 to 1.3598. The confidence interval contains 1, indicating that there is no statistical difference of accident rate between the two railroads in that year.
The final part of the decision support tool is a graphic representation (Figure 3c) of the calculated results. The vertical axis denotes the point estimate values whereas the horizontal axis represents years. Confidence intervals are illustrated vertically (red bars). On these bars lie the point estimates (red dots). Lastly, running along the horizontal axis is a reference line (dash horizontal line) signifying where the estimate value of 1 (a benchmark) is. For example, the first railroad has a statistically lower accident rate if the confidence interval is below the reference line. Otherwise, if the confidence interval is above the reference line, the first train accident rate is statistically higher. If the reference of 1 is within the confidence interval, it indicates that the two accident rates are statistically identical.
6 DATA ANALYSIS

6.1 Data Sources

In order to implement this statistical method into a series of comparisons, data from four major US Class I freight railroad companies (BNSF Railway, CSX Transportation, Norfolk Southern Railway and Union Pacific Railroad) from 2000-2014 was analyzed. Accident count data was collected through the FRA Rail Equipment Accident (REA) Database. The report (form 6180.54) that gathers this information is required to be completed by a railroad if an accident’s damage cost is above a certain monetary level ($10,500 in 2014) (2, 23). There were three types of traffic exposure data used (train-mile, car-mile, and gross ton-mile data). Regardless of which specific metric is used, traffic exposure essentially measures the volume of traffic that a track could undergo. The monthly train-mile data is collected through Form 6180.55. The other two traffic exposure measures, namely car-mile and gross ton-mile, are reported annually to the Surface Transportation Board (STB) along with their operational data through Form R-1. Recent work has shown that some derailment causes are related to train-miles whereas others regard car-miles or gross ton-miles (16). For example, the car-mile exposure tends to be more relevant in the case of equipment or track failures whereas train-mile tends to be more attributed in the case of human failure. Considering this difference, this research will compare train accident rates using all three traffic metrics.

6.2 Results

The resulting deliverables from the implementation of this paper’s statistical method are a MLE ratio of accident rates and its corresponding confidence interval. This paper focuses on train derailments, which account for about 65 percent of accident frequency on mainlines (24). This paper selects a western railway (BNSF) and an eastern railway (CSX) as an example to illustrate the analytical process using the above-mentioned methodology. The analysis can be adapted to other railroads.

To show how the estimation points and confidence intervals are calculated, the data of BNSF in 2000 is used as an example. The derailment counts of BNSF and the other three
railroads are 144 and 295. The corresponding traffic volumes are 146 million train-miles and 331 million train-miles, respectively. Using Equation (6), the estimated ratio of the BNSF train derailment rate versus the other three railroads combined is 1.1, with a 95% confidence interval (0.91, 1.35). The derailment rate comparison in other years can be conducted in a similar way (Figure 4a, 4b, 4c).

In regards to the charts pertaining to the BNSF rate compared to the other three railroads combined, the confidence intervals contain the benchmark value 1 for most of the years in the study period. This indicates that in general, there is no statistical difference between the BNSF derailment rate and the industry average rate. However, there are a few exceptions. For example, in the year 2003 and 2004, for the train-mile-based rate comparison, the confidence interval is below 1 (Figure 4a). This means that BNSF’s derailment rate is statistically lower than the industry average. The car-mile-based train derailment rate comparison (Figure 4b) has a similar trend to that of the train-mile-based comparison. However, the ton-mile comparison (Figure 4c) is a bit different in that five confidence intervals (years 2002-2005, and year 2013) are below the benchmark of 1. This means that BNSF has a statistically lower train accident rate than the industry average in these years.

FIGURE 4(a): Comparison between BNSF and the other three companies combined based on train derailment rate per train-mile
FIGURE 4(b): Comparison between BNSF and the other three companies combined based on accident rate per car-mile

FIGURE 4(c): Comparison between BNSF and the other three companies combined based on train derailment rate per ton-mile
From the derailment rate comparisons between CSX and the other three railroads combined, CSX has a statistically higher derailment rate in year 2004 based on the train-mile traffic metric (Figure 5a). If car-mile data or ton-mile data is used, the CSX rate is statistically higher in years 2003, 2004, 2006, 2007, 2009 and 2010 (Figure 5b and Figure 5c) because the confidence intervals are all above the benchmark value of 1. In the other years, the CSX safety level is identical with the other three railroads combined, regardless of which traffic metric is used.

FIGURE 5(a): Comparison between CSX and the other three companies combined based on train derailment rate per train-mile
FIGURE 5(b): Comparison between CSX and the other three companies combined based on train derailment rate per car-mile

FIGURE 5(c): Comparison between CSX and the other three companies combined based on train derailment rate per ton-mile
In summary, the comparisons of railroad-wide derailment rates show two key insights. First, the use of varying traffic exposures can lead to differing comparison results. As such, it is prudent to use different exposure metrics when comparing accident rates across different railroads or time periods. Second, in some years, the studied railroad has a statistically identical derailment rate to the industry average rate, indicating that there is no actual safety variation, even though their empirical derailment rates may differ largely. However, in some other years, one railroad could have a statistically lower or higher derailment rate than other railroads, indicating an actual safety difference that calls for additional analysis.

If a railroad has a statistically higher accident rate than the industry average for several years in a row, it may indicate a systematic safety variation between this railroad and others. If so, safety improvement might be prioritized to this railroad. Similarly, within the same railroad, if a location has a higher rate than the company average, it may lead to an increased emphasis of safety resources on this location. In the absence of a statistical analysis, the comparison solely based on empirical accident rates may lead to a misallocation of safety resources if the seemingly different rates do not reflect an actual safety change.

7 RESEARCH CONTRIBUTIONS
This paper is unique in that it develops a generalized methodology to better analyze past accident rates by identifying the change of actual safety. This is done while accounting for the randomness of accident occurrence. As such, this paper has the potential to provide contributions to both the state of the art and practice:

- The work contained within this paper begins with an advancement of theory. Specifically, a detailed description of Regression to the Mean (RTM) is presented to expand the general understanding of how this particular theory affects the accuracy of empirical accident rate comparisons.
- This paper also advances the current methodology of accident rate comparison by addressing the issue of statistical variation in train accident rates. The sort of statistical formulation used to overcome this problem of variation garners the ability to advance railway safety research from the empirical norm to research that has a stronger base in the statistical modeling of data.
- In order to make this methodology available to practitioners, this research also advances the practice through developing a decision support tool that provides an interactive interface for comparing rates. This tool is intended to automate the process of conducting a sophisticated statistical analysis of railroad accident data.
- This work provides the groundwork for further research pertaining to this topic even for those sectors outside of the realm of freight trains (e.g., passenger trains, roadway safety).

8 CONCLUSIONS
Analysis of accident data is vitally important for managing railway safety. Although empirical (observed) comparisons using this rate have been extensively studied for comparing railroad safety performance, it does not account for random fluctuations in accident occurrences. When these empirical rates differ, it is not well understood what proportion of this difference is caused by stochastic variation and what proportion of this change is reflective of actual safety change. This research develops a novel methodology, accompanied with a practice-ready decision support tool, for statistically comparing train accident rates between railroads and time periods.
Also, this analysis shows that the comparison results may vary with the use of different traffic metrics. This research can be further developed and ultimately aids the railroad industry in intimately understanding past accident data using advanced analytical tools. This level of understanding will support data-driven decisions for railway safety management.

9 FUTURE RESEARCH

- This research focuses on train derailment rate comparisons. The methodology developed in this work can be adapted to other types of train accidents, such as collisions or grade crossing incidents.
- This research primarily pertains to the freight railroad system. The tool can be applied to other transportation sectors such as the commuter railroad system.
- An explanatory or causal analysis can be developed to understand the reasons that account for the accident rate differences between railroads or time periods.

ACKNOWLEDGEMENTS

The authors thank Chang Liu for her help in the statistical analysis in this paper. The authors also thank Yikun Xian and Xiaoyang Xie for developing the Decision Support Tool. The first author was funded by the Department of Civil and Environmental Engineering, and the School of Engineering, both at Rutgers University. The second author was funded by the Aresty Research Center for Undergraduates at Rutgers University. However, the authors are solely responsible for all the analyses and views in this paper.

REFERENCES

10. *Homepage.* Surface Transportation Board (STB).