MID-AMERICA
TRANSPORTATION CENTER

Report # MATC-UI: 479

Final Report
25-1121-0001-479

A Procedure for Matching Truck Crash Records with Hazardous Material Release Incidents and a Comparative Analysis of the Determinants of Truck Crashes with Hazardous Material Releases

Paul Hanley, Ph.D.
Associate Professor
Department of Civil and Environmental Engineering
University of Iowa

Suyun Ma, Ph.D. Candidate
Graduate Research Assistant
Hai Yu, Ph.D. Candidate
Graduate Research Assistant

2012

A Cooperative Research Project sponsored by the
U.S. Department of Transportation Research and
Innovative Technology Administration

A Procedure for Matching Truck Crash Records with Hazardous Material Release Incidents and a Comparative Analysis of the Determinants of Truck Crashes with Hazardous Material Releases

Paul Hanley, Ph.D.
Director of Transportation Policy Research, Public Policy Center
Associate Professor, Department of Civil and Environmental Engineering
University of Iowa

Suyun Ma
Ph.D. Candidate
Department of Civil and Environmental Engineering
University of Iowa

Hai Yu
Ph.D. Candidate
Department of Civil and Environmental Engineering
University of Iowa

A Report on Research Sponsored by

Mid-America Transportation Center

University of Nebraska–Lincoln

June 2012

Technical Report Documentation Page

1. Report No.
25-1121-0001-479

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
A Procedure for Matching Truck Crash Records with Hazardous Material Release Incidents and a Comparative Analysis of the Determinants of Truck Crashes with Hazardous Material Releases

5. Report Date

6. Performing Organization Code

7. Author(s)
Paul F. Hanley, Suyun Ma, and Hai Yu

8. Performing Organization Report No.
25-1121-0001-479

9. Performing Organization Name and Address
Mid-America Transportation Center
2200 Vine St.
PO Box 830851
Lincoln, NE 68583-0851

10. Work Unit No. (TRAIS)

11. Contract or Grant No.

12. Sponsoring Agency Name and Address
Research and Innovative Technology Administration
1200 New Jersey Ave., SE
Washington, D.C. 20590

13. Type of Report and Period Covered
January-June 2012

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

In the current study, we quantified the number and location of hazardous release crashes and identified the events leading to crashes, as well as the type of material released. This study, for the first time, combined two federal databases: the U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration (PHMSA) database, and the Motor Carrier Management Information System crash database (MCMIS). PHMSA and MCMIS data for 1999 through 2009 were obtained and matched using the common attributes of time, day, month, year, county, state, and phase of transportation. Naïve Bayesian, logistic and neural network classification methods were developed and compared. Each method performed well. All possible pairwise combinations of records between the two datasets were identified. Likelihood estimates of a match using these common attributes were calculated, after which a sample of the records was drawn. The sample was manually checked for matches and mismatches, and was used in the calibration of the logistic and neural networks. The matching algorithms were run using all possible pairwise combinations to identify exact matches, as well as the probability of matches. Pairwise comparisons with a probability of a match greater than 0.50 were extracted and used in the statistical analysis of truck crash characteristics. Each of the extracted records was weighed based on the probability of a match, and the weighted total was set to equal the number of MCMIS reported crashes characterized by hazardous material releases. One outcome of this study will be the identification of a probabilistic model that will advance safety regulations of the U.S. trucking industry and fleet.

17. Key Words
truck crashes; hazardous material crashes; material release

18. Distribution Statement

19. Security Classif. (of this report)
Unclassified

20. Security Classif. (of this page)
Unclassified

21. No. of Pages
79

22. Price

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Federal Motor Carrier Safety Administration (FMCSA)
Hazardous Materials Safety Administration Incident Reports (HMSAIR)
Iowa Department of Transportation (DOT)
Mid-America Transportation Center (MATC)
Motor Carrier Management Information System (MCMIS)
Pipeline and Hazardous Materials Safety Administration (PHMSA)

Acknowledgments

We acknowledge The Mid-America Transportation Center for providing the funds to complete this study along with the reviewers of the proposal for their generous giving of their time in strengthening the study. Our thanks is given to the support of the Public Policy Center for housing our research group.

Disclaimer

Abstract

In the current study, we quantified the number and location of hazardous release crashes and identified the events leading to crashes, as well as the type of material released. This study, for the first time, combined two federal databases: the U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration (PHMSA) database, and the Motor Carrier Management Information System crash database (MCMIS). PHMSA and MCMIS data for 1999 through 2009 were obtained and matched using the common attributes of time, day, month, year, county, state, and phase of transportation. Naïve Bayesian, logistic and neural network classification methods were developed and compared. Each method performed well. All possible pairwise combinations of records between the two datasets were identified. Likelihood estimates of a match using these common attributes were calculated, after which a sample of the records was drawn. The sample was manually checked for matches and mismatches, and was used in the calibration of the logistic and neural networks. The matching algorithms were run using all possible pairwise combinations to identify exact matches, as well as the probability of matches. Pairwise comparisons with a probability of a match greater than 0.50 were extracted and used in the statistical analysis of truck crash characteristics. Each of the extracted records was weighed based on the probability of a match, and the weighted total was set to equal the number of MCMIS reported crashes characterized by hazardous material releases. One outcome of this study will be the identification of a probabilistic model that will advance safety regulations of the U.S. trucking industry and fleet.

Executive Summary

Security of the national transportation system is a growing concern, prompting investigations into potential threats. Though the number of truck crashes with hazardous material releases is small in comparison to automobile crashes, the collateral is significant—an FMCSA report (Batelle, Inc. 2001) found the cost related to these events to be in excess of $1 billion per year. In the current study, we quantify the number and location of hazardous release crashes, and identify the events leading to crashes, as well as the type of material released. One outcome of this study will be to identify a probabilistic model that will advance safety regulations of the U.S. trucking industry and fleet.

This study has, for the first time, combined two federal databases: the U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration (PHMSA) database, and the Motor Carrier Management Information System crash database (MCMIS). Naïve Bayesian, logistic and neural network classification methods were developed and applied, and an analysis of their effectiveness in combining the two datasets was performed. Each method performed well. However, the outcome of the logistic and neural network performed equally well in matching known outcomes.

Prior to the current study, we piloted the same strategic concepts for crashes and hazardous material releases within Iowa, matching records by first selecting all cases from PHMSA that were involved in a vehicular accident. We then noted the route, city, state, day, and time of occurrence, searching for records with similar attributes in the Iowa Department of Transportation (DOT) database. The partial preliminary descriptive results showed that of 32 cases, 75% occurred during daylight hours, while the remainder occurred when it was dark. 72% of crashes involved single vehicle truck crashes; of these, almost 40% of vehicles ran off the road. In the year 2008, 35% of all crashes involved single vehicles; of these, 22% were caused by vehicles running off the road. We observed similar values for the years 2006 and 2007. In total, we observed two cases involving fatalities, and two cases where the driver was injured. In one instance, the driver had a heart attack that caused the vehicle crash. In three cases, nearby areas had to be evacuated.

These pilot results suggested that it would be valuable to expand the existing methodology to encompass all states over a longer period of time. Therefore, the PHMSA and MCMIS data for 1999 through 2009 were obtained and matched using the common attributes of time, day, month, year, county, state, and phase of transportation. All possible pairwise combinations of records between the two datasets were identified. Likelihood estimates of a match using these common attributes were calculated, after which a sample of the records was drawn. The sample was manually checked for matches and mismatches, and was used in the calibration of the logistic and neural networks. The matching algorithms were run using all possible pairwise combinations to identify exact matches, as well as the probability of matches. Pairwise comparisons with a probability of a match greater than 0.50 were extracted and used in the statistical analysis of truck crash characteristics. Each of the extracted records was weighed based on the probability of a match, and the weighted total was set to equal the number of MCMIS reported crashes characterized by hazardous material releases.

The current report is divided into two main parts, or, chapters: chapter 1 describes the matching process; chapter 2 presents the results of a statistical analysis of between-truck crashes with hazardous material releases and a comparative analysis of between-truck crashes with and without hazardous material releases

Chapter 1 Procedure for Matching the Federal Motor Carrier Safety Administration (FMCSA) Motor Carrier Management Information System (MCMIS Crash Data) with the Pipeline and Hazardous Materials Safety Administration Incident Reports (HMSAIR)

1.1 Hazardous Material Release by Moving Vehicles from the HMSAIR Dataset

The HMSAIR dataset contains an attribute, “Transportation Phase,” that indicates where in the supply chain a hazardous material release occurred. Table 1.1 shows that a total of 22,582 releases occurred while the hazardous material was in transit, i.e., in a moving vehicle that was in operation on the roadway.

Table 1.1 Occurrence of hazardous material release incidents

Transportation Phase		
	Frequency	Percent
IN TRANSIT	22582	14.2
IN TRANSIT STORAGE	6984	4.4
LOADING	26977	16.9
UNLOADING	102818	64.5
Total	159361	100.0

1.2 Crash Reports of Hazardous Material Releases from the MCMIS Dataset

The MCMIS data set was queried to obtain the frequency of crashes involving hazardous material releases. The results are shown in table 1.2, and reveal that a very small fraction of truck crashes (0.6%) resulted in a release. It was initially tempting to eliminate all crash reports but the 8,987 that involved a hazardous material release. However, this was disregarded since reporting requirements differ between the two data sets, and because the assumed errors and oversights in completing on-site crash reports warn against the initial purging of records.

Table 1.2 Occurrence of hazardous material release listed in crash reports

Hazardous Materials Released	**Frequency**	**Percentage**
No	1511118	99.4
Yes	8987	0.6
Total	1520105	100.0

1.2.1 Comparison of Reported in Transit Incidents (HMSAIR) and Crashes with Hazardous Material Reported (MCMIS)

Over the 11 years of collected data (1999-2009) the number of hazardous material releases that occurred and were reported was 15,9361, with 14.2% occurring while the material was in transit. The reported number of truck crashes that resulted in hazardous material releases was 8,987 over the same 11-year period. The large discrepancy between the reported number of incidents and crashes is troubling, and indicates that the method of limiting analyses to crash records is too restricting; therefore, a method of matching incident records with crash records that also estimates the probability of finding a match between each dataset is required.

1.3 Exploring Possible Attributes for Matching

The HMSAIR data were analyzed for attributes they held in common with the MCMIS data set. Time, day, month, year, county, and state were found to be common factors. These attributes identified the location of an occurrence in time and space (hazardous material releases in the case of HMSAIR, truck crashes in the case of MCMIS). The transportation phase (when a vehicle is operating on the roadway) is directly coded only in the HMSAIR dataset; therefore, its presence was inferred in the MCMIS. This inference was based on the fact that only those truck crashes that resulted in a towed vehicle were recorded in the MCMIS database, thus complying with the definition of the transportation phase.

1.3.1 Descriptive Analysis

1.3.1.1 Time of incident from HMSAIR

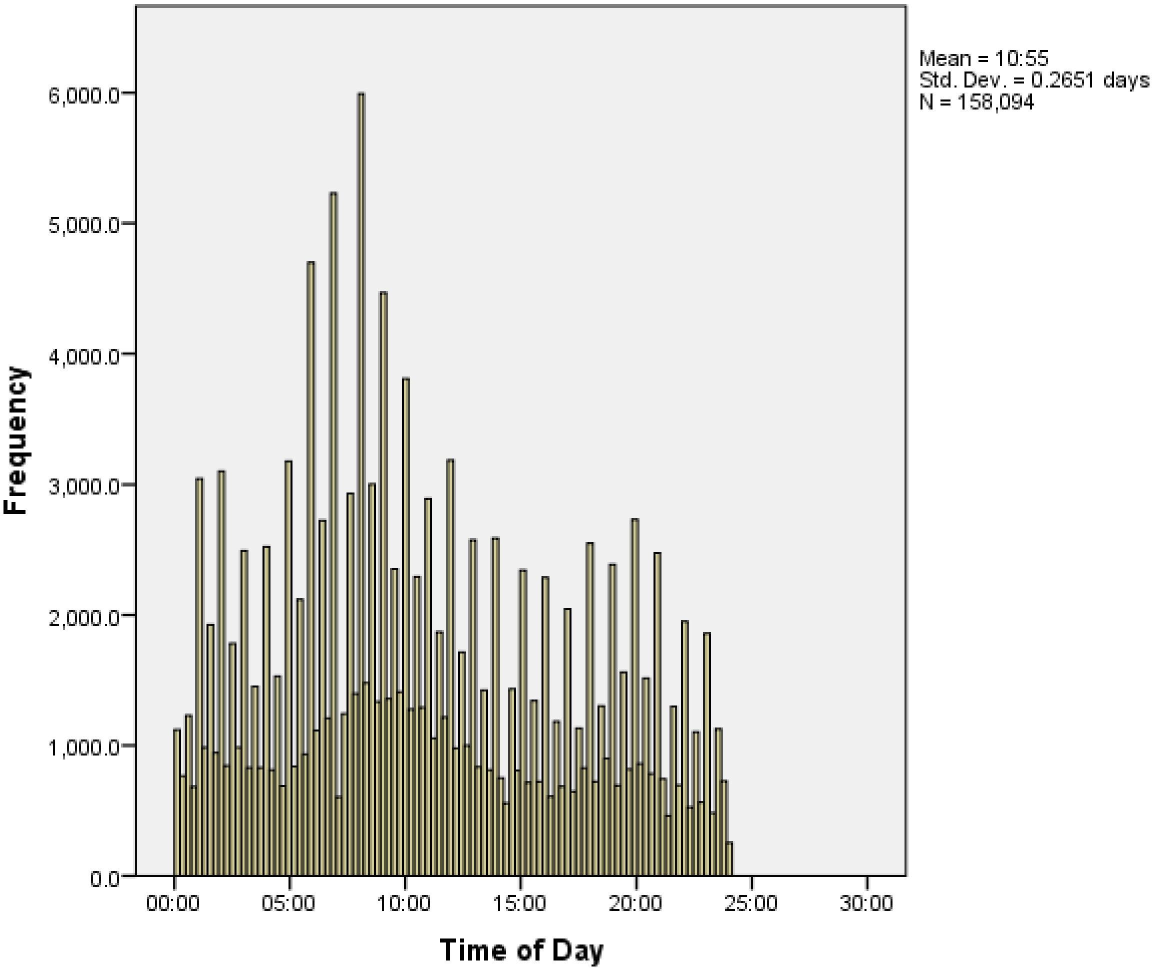

Figure 1.1 Frequency distribution for time of incident report

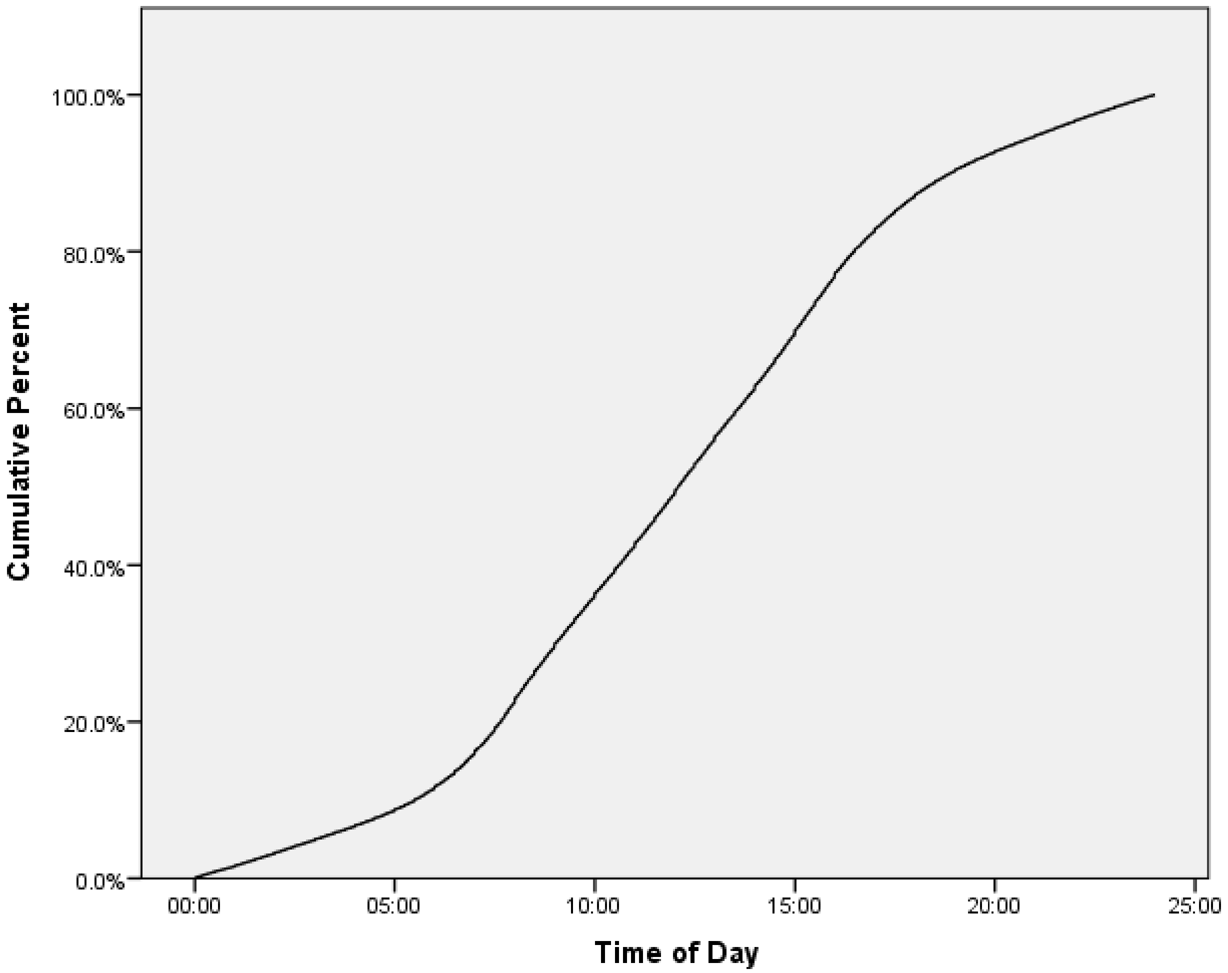

Figure 1.2 Cumulative distribution for time of incident report

Table 1.3 Summary statistics for time of day of incident reports

Statistic	**Value**
Mean	10:55 AM
Median	10:00 AM
Maximum	11:59 PM
Minimum	12:00 AM
Std. dev.	6:21
Observations	158094

1.3.1.2 Time of crash report from MCMIS

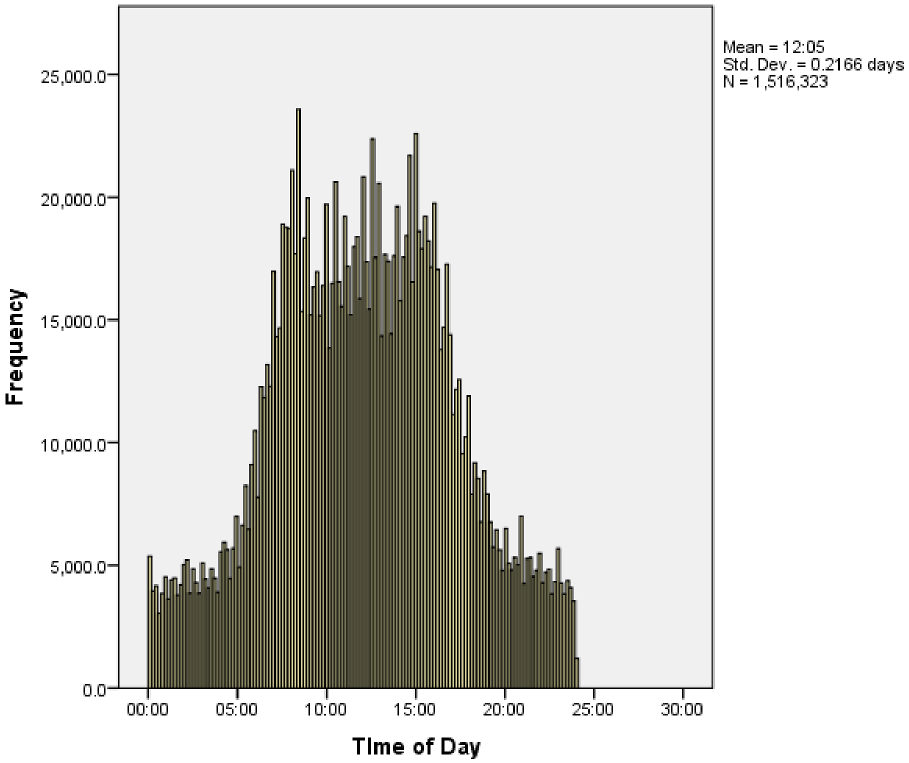

Figure 1.3 Distribution for time of report

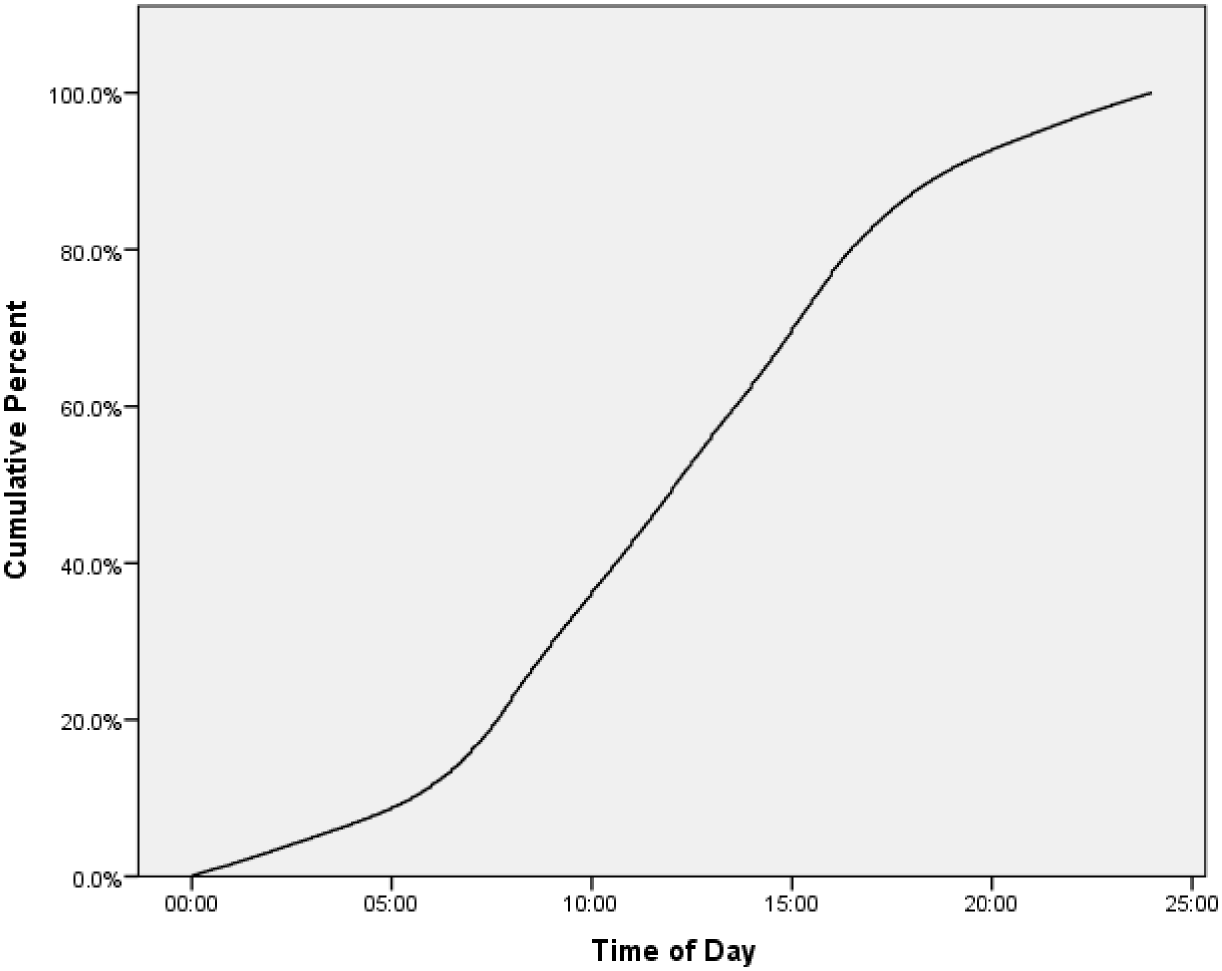

Figure 1.4 Cumulative distribution for time of report

Table 1.4 Summary statistics for time of report for crashes

Statistic	**Value**
Mean	12:05 PM
Median	12:05 AM
Maximum	11:59 PM
Minimum	12:00 AM
Std. Dev.	5:11
Observations	1516323

1.3.1.3 Comparison of time of incident and time of crash report

Tables 1.3 and 1.4, along with figures 1.1-1.4 (above) show the distributions for the time when a hazardous material release occurred and the time a crash report was completed. An interesting commonality between these two distributions is an overrepresentation of times that end on the hour and half-hour, with times ending on a 5 being somewhat overrepresented. This is typical of reported times following an event occurrence. Therefore, there is a strong indication that when using the time attribute as a criteria for matching, one needs to build in a mechanism to compensate for the imprecision of time recordings. As discussed in a proceeding section, matching on time will follow a likelihood procedure.

1.3.1.4 Day of incident from HMSAIR

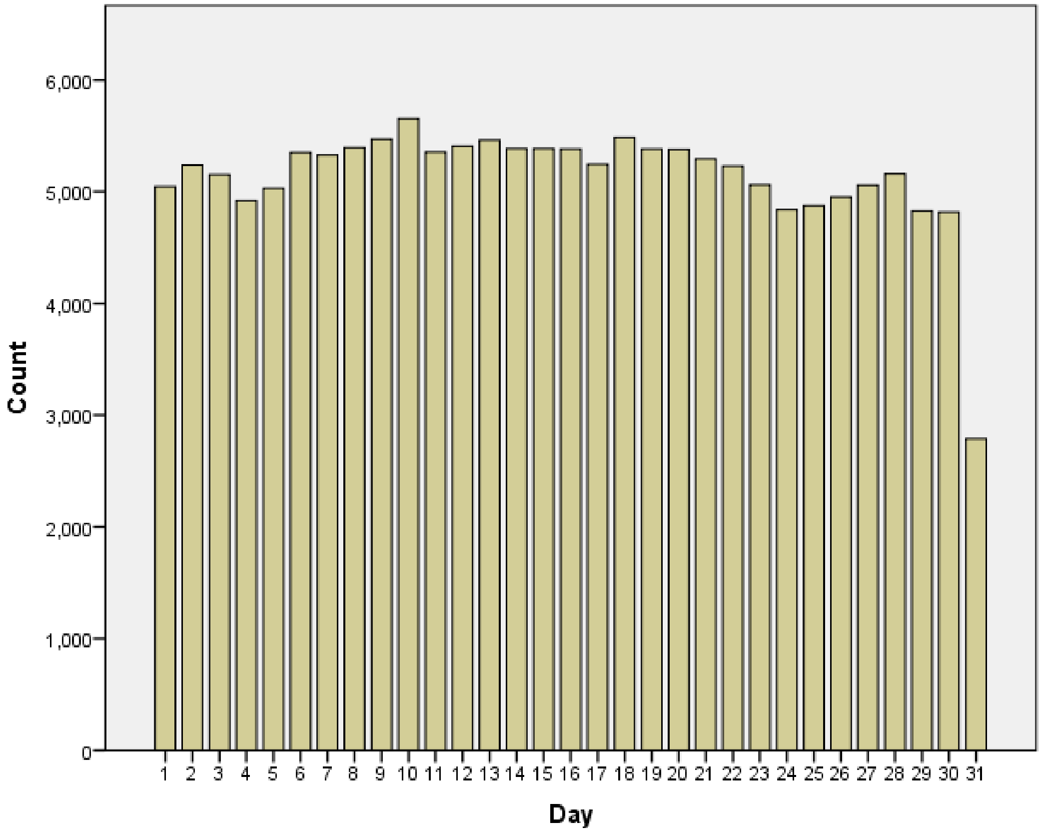

Figure 1.5 Distribution for day of incident

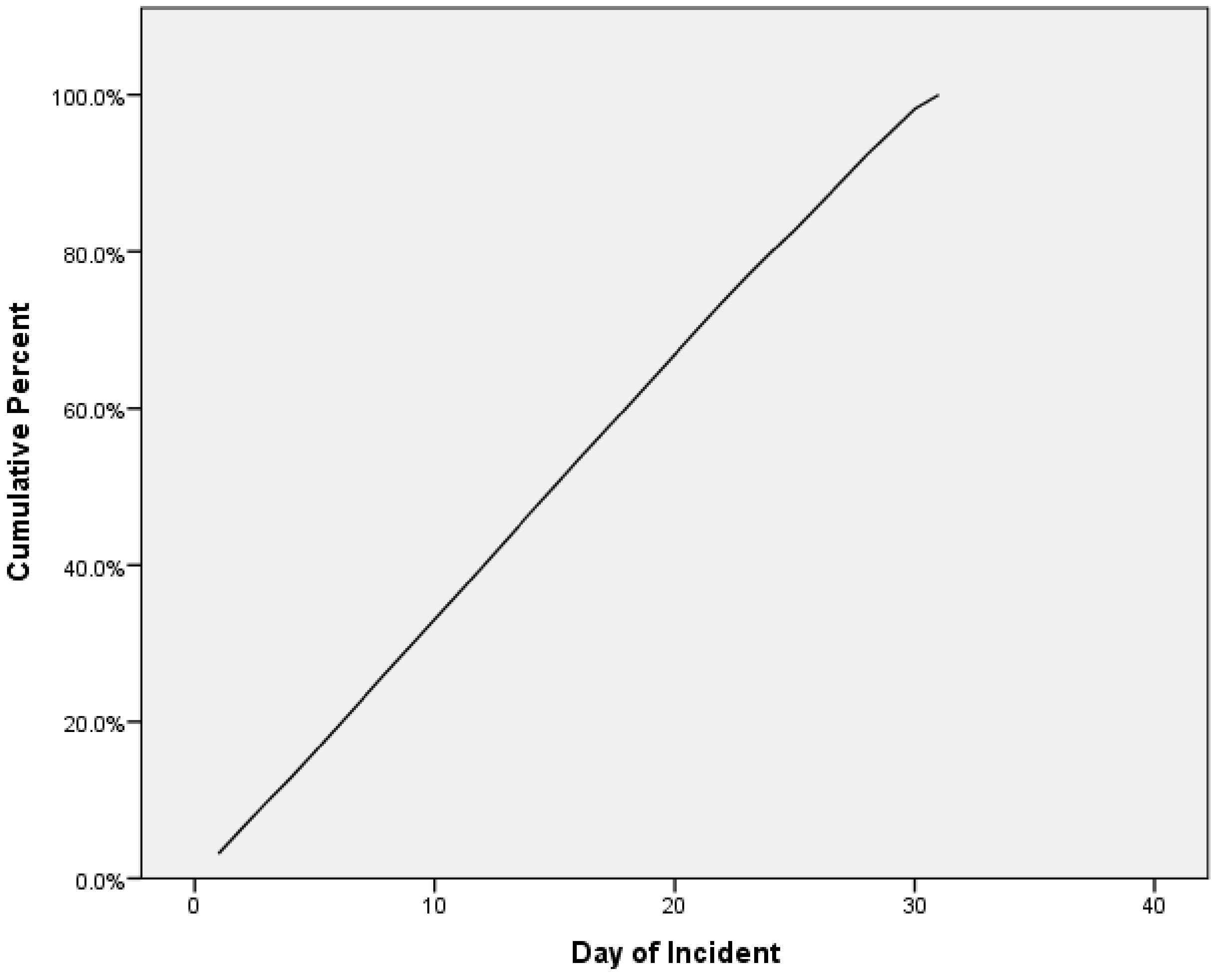

Figure 1.6 Cumulative percent for day of incident

Table 1.5 Summary statistics for day of incident

Statistic	**Value**
Mean	15.63
Median	16.00
Maximum	31
Minimum	1
Std. dev.	8.707
Observations	159361

1.3.1.5 Day of crash report from MCMIS

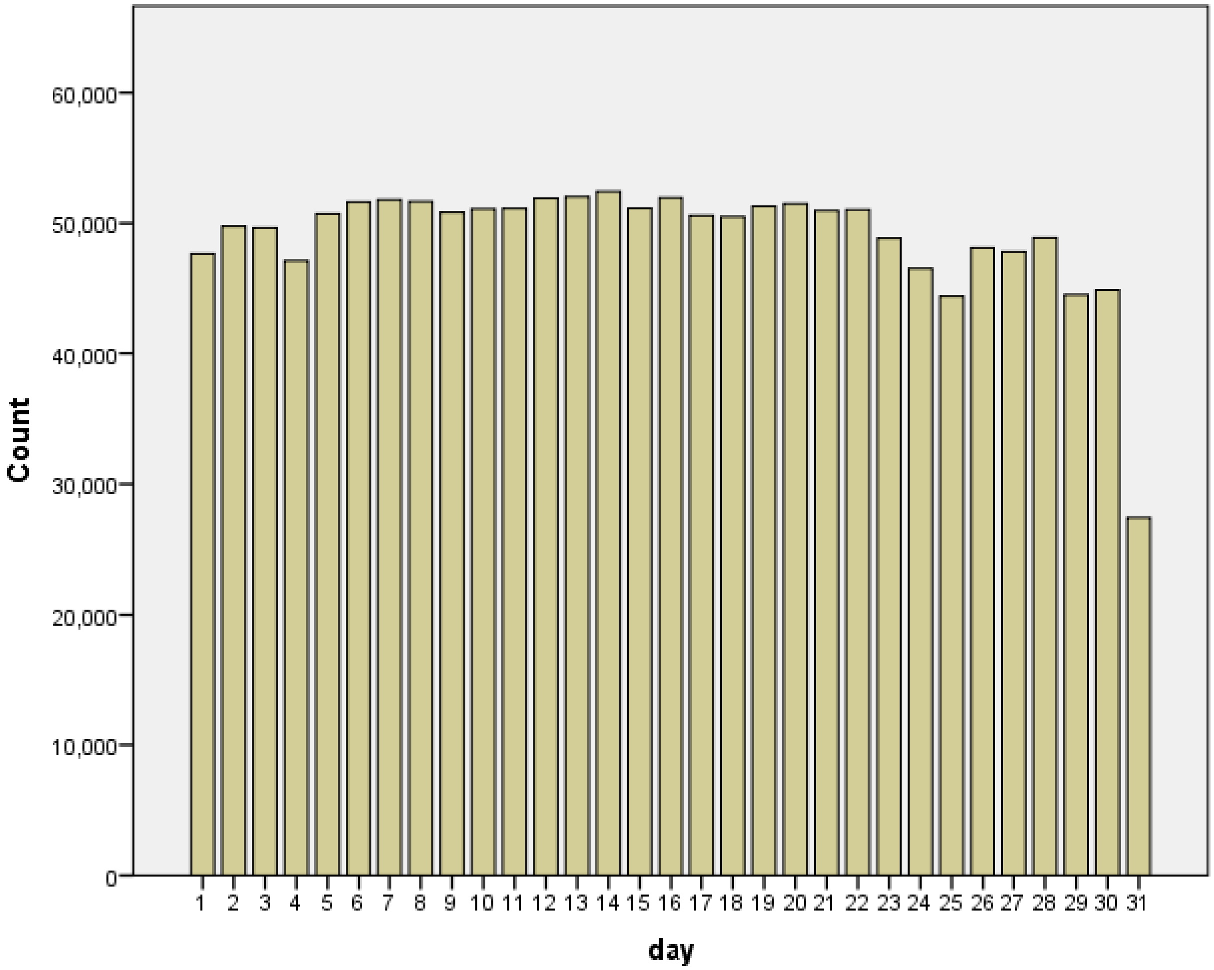

Figure 1.7 Distribution for day of crash report

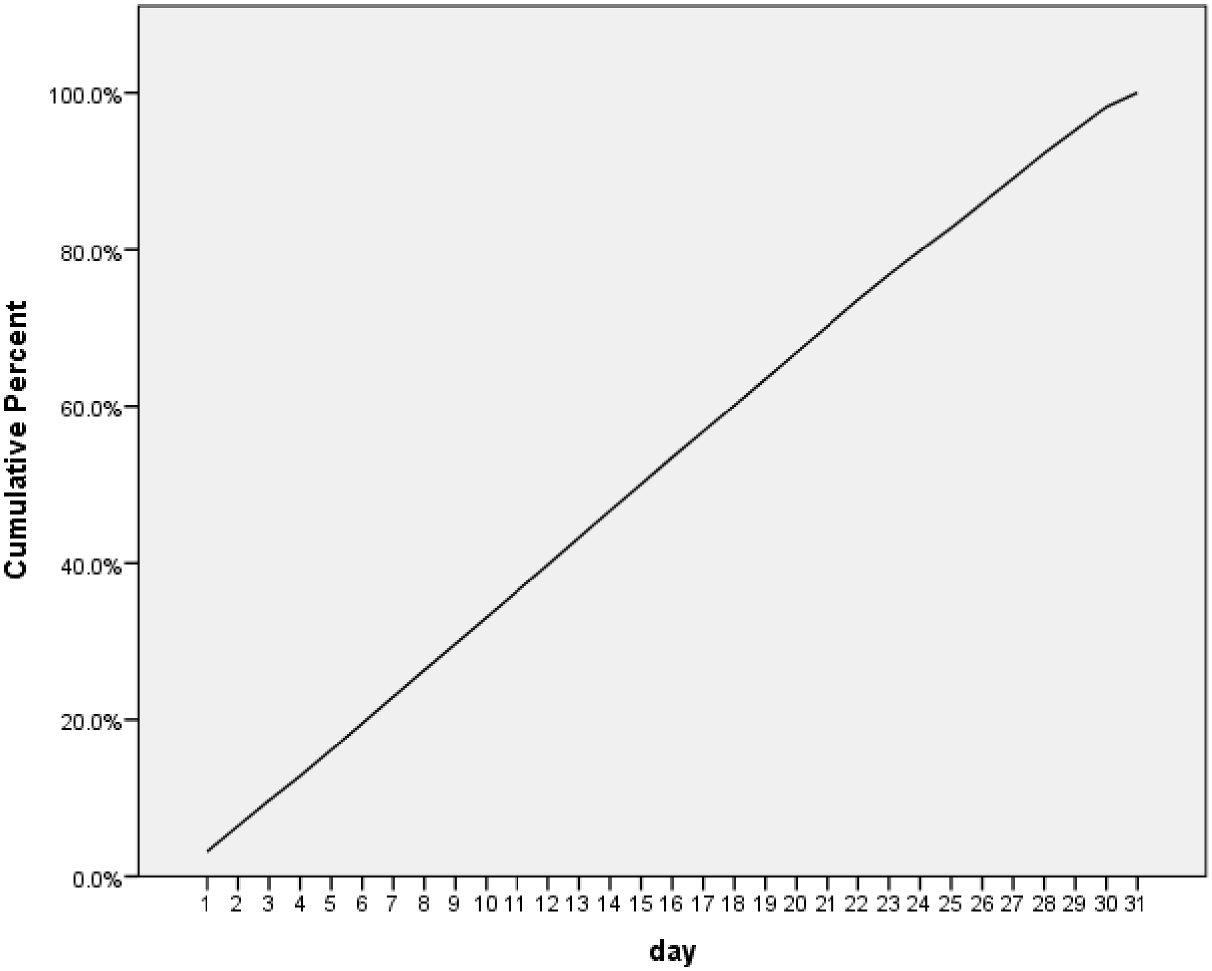

Figure 1.8 Cumulative percent for day of crash report

Table 1.6 Summary statistics for day of crash report

Statistic	**Value**
Mean	15.60
Median	15.00
Maximum	31
Minimum	1
Std. dev.	8.703
Observations	1520105

1.3.1.6 Comparison of day of incident and day of crash report

As shown in figure 1.5 and figure 1.7 records of the day of the crash incident, hazardous material release, and crash report are uniformly distributed. There was a minor decrease in the likelihood of an occurrence falling on the last few days of the month. It was equally likely for an incident or a crash to be reported on *any* day of the month over the 11 year study period. The only exception occurred on the 31st of the month, which was less likely because only seven months contain 31 days. The uniformity of the likelihood across days is further revealed in the cumulative frequency charts in figure 1.6 and figure 1.8. With the exception of the 31st of the month, the plot was a near-perfect straight line at 45 degrees to the horizontal axis.

1.3.1.7 Month of incident from HMSAIR

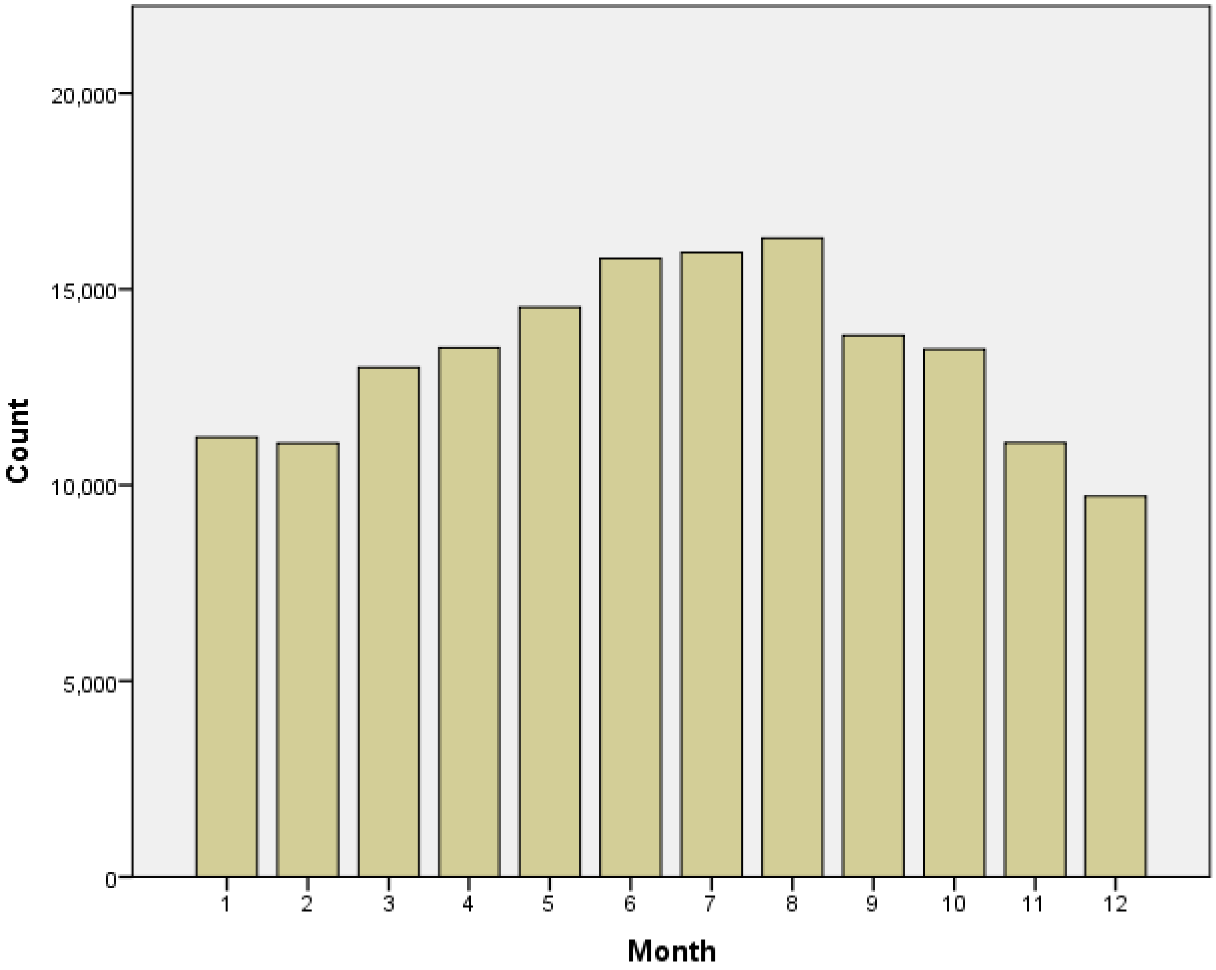

Figure 1.9 Distribution for the month of incident

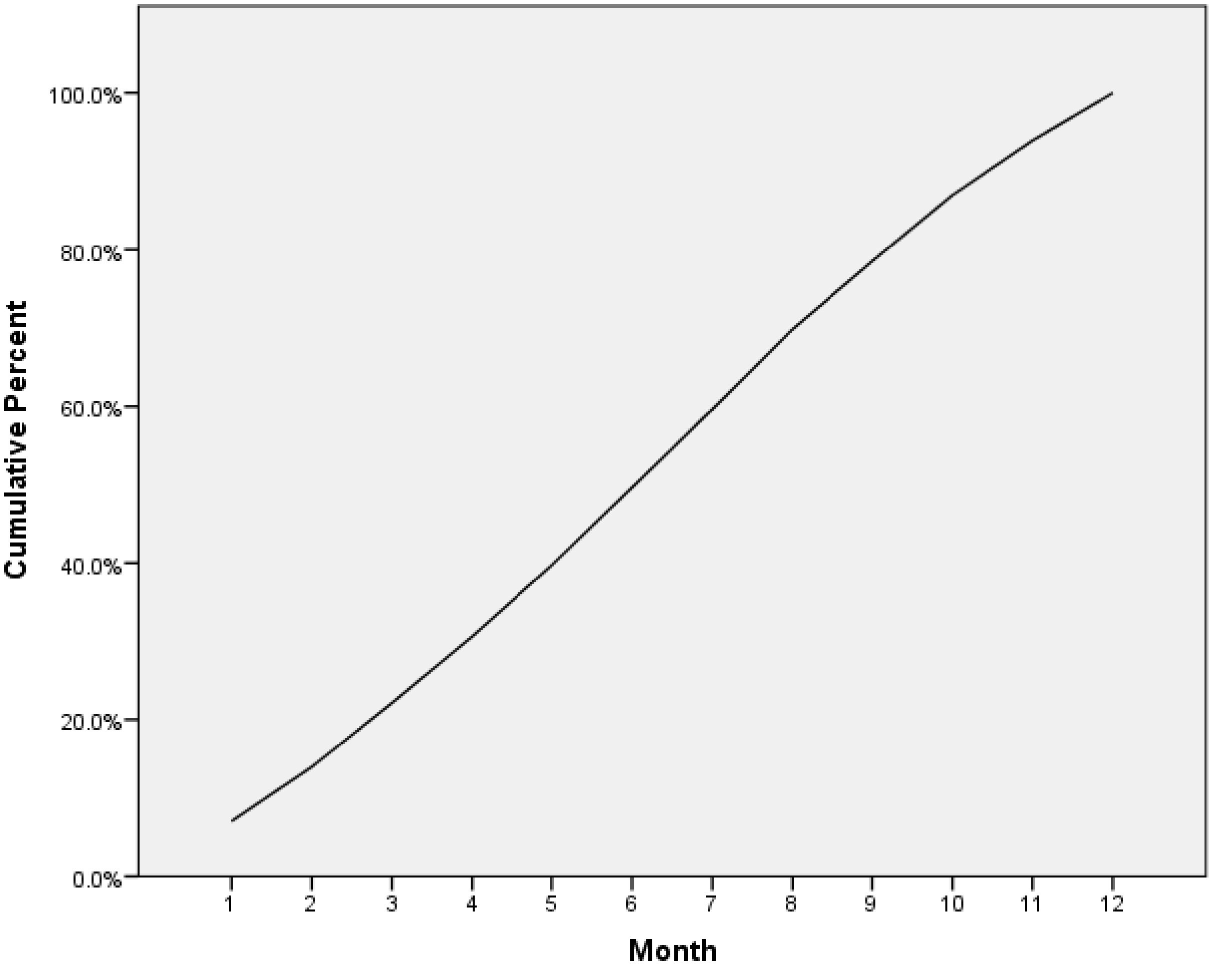

Figure 1.10 Cumulative percent for the month of incident

Table 1.7 Summary statistics for month of incident

Statistic	Value
Mean	6.48
Median	7.00
Maximum	12
Minimum	1
Std dev.	3.221
Observations	159361

1.3.1.8 Month of crash report from MCMIS

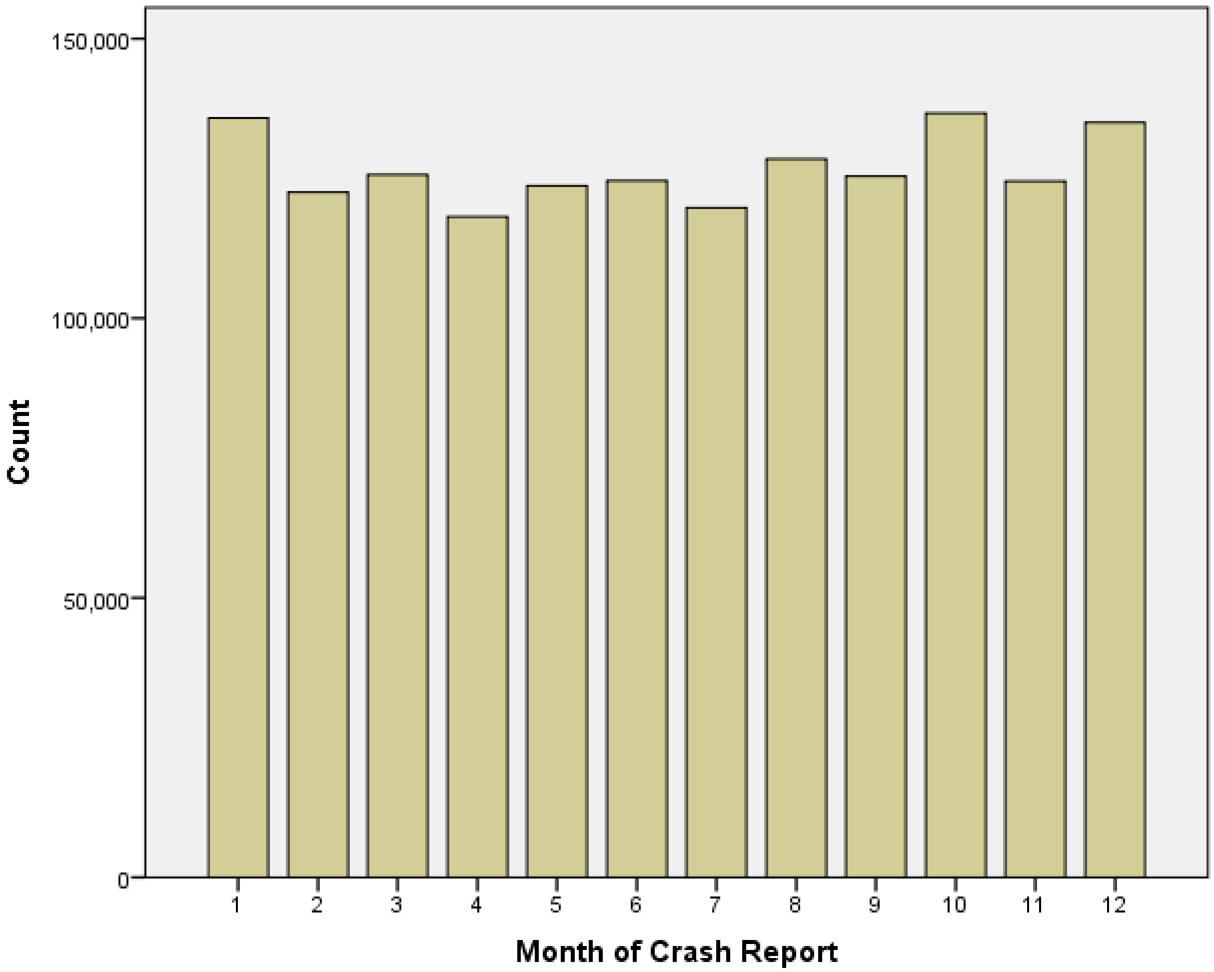

Figure 1.11 Distribution for month of crash report

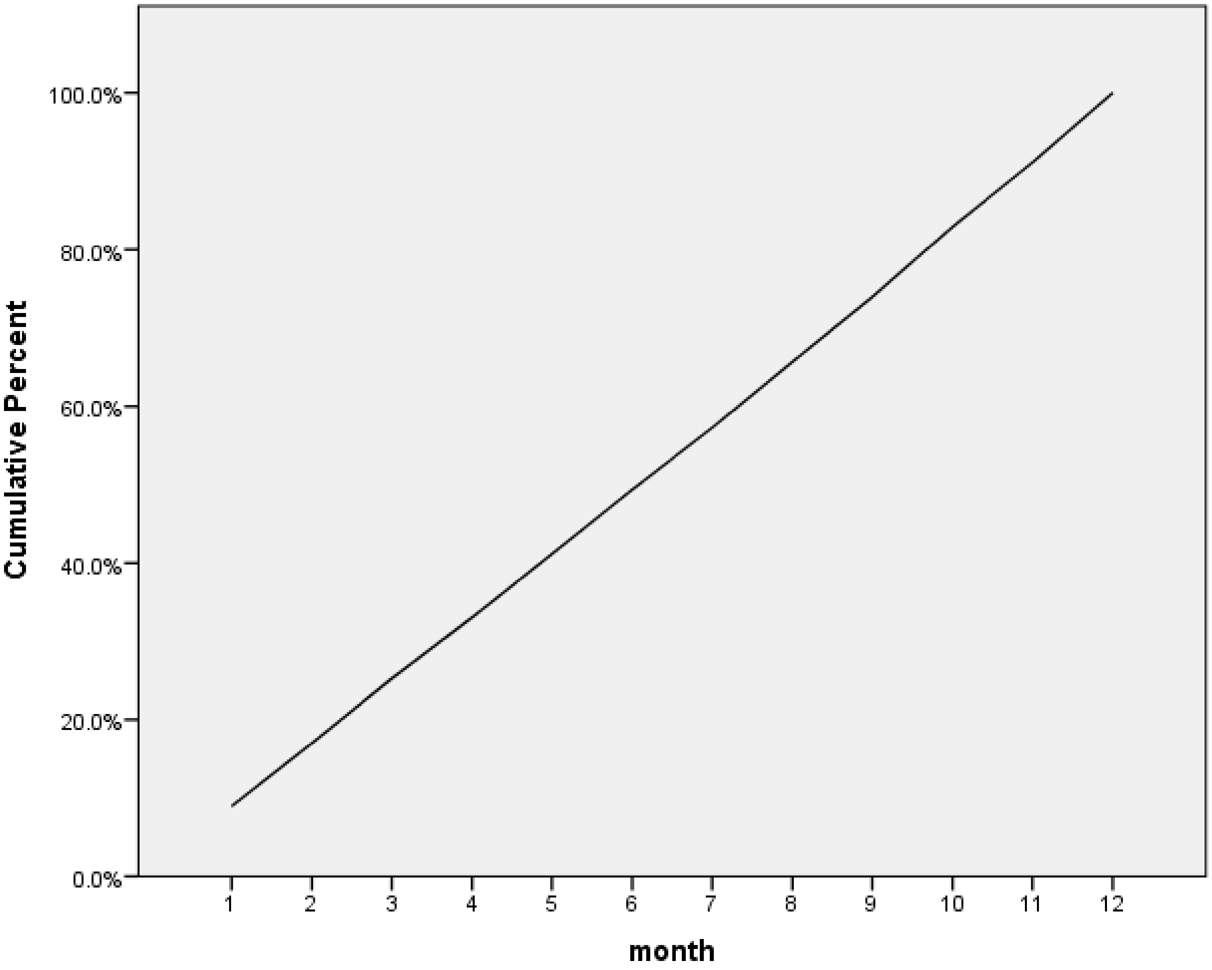

Figure 1.12 Cumulative percent for month of crash report

Table 1.8 Summary statistics for month of crash report

Statistic	**Value**
Mean	6.54
Median	7.00
Maximum	12
Minimum	1
Std. dev.	3.494
Observations	1520105

1.3.1.9 Comparison of month of incident and month of crash report

Figure 1.9 shows a pattern of more hazardous material releases occurring in the months of June, July, and August. The winter months of December and January contained the fewest reported incidents. This pattern differed from that of crash reports: the months with the most crash reports were December and January, typically due to the higher occurrence of poor weather, which increased the likelihood of crashes (see fig. 1.11). As detailed in a subsequent section, this mismatch will be explored by estimating the joint likelihood of crashes and hazardous material releases. Although the pattern differed across the frequency distributions, the individual cumulative percent charts show that each were nearly uniform; that is, it is unlikely that any one month would vary in the likelihood of a reported incident or crash.

1.4 Procedure for Estimating Likelihood Values of Matching Record Fields

The two data files were matched based on a likelihood measure using the time, day, month, year, county, state, and phase of transportation. The likelihood for the county, state, and phase of transportation attributes are represented as binary numbers, with a “1” being assigned to an exact match between files and a “0” for all other cases. The likelihood for time, day, month and year were calculated based upon a linear relationship based on the difference between the reported values in each file. The following likelihood definitions were used in the categorization methods described in the next section.

$$L_{time} = 1 - \left(\frac{(time_1 - time_2)}{1440.0}\right) \tag{1.1}$$

$$L_{day} = 1 - \left(\frac{(day_1 - day_2)}{31.0}\right) \tag{1.2}$$

$$L_{month} = 1 - \left(\frac{(month_1 - month_2)}{12.0}\right) \tag{1.3}$$

$$L_{year} = 1 - \left(\frac{(year_1 - year_2)}{11.0}\right) \tag{1.4}$$

The maximum difference in the attribute is the denominator for each likelihood value. For example, the time attribute was recorded in military format at the minute accuracy level; therefore the maximum time difference was 1440. Note that the maximum difference in years was 11, since the dataset ranged from 1999 to 2009.

1.4.1 Joining of HMSAIR and MCMIS Records

The pairings between HMSAIR and MCMIS records proceeded using only those records identified using likelihood values for *state* and *phase* equal to 1.0, as described above. In the process of calculating the likelihood values, pair combinations were recorded if the state and phase equaled 1.0. The generated record contained the individual and the joint likelihood value, along with crash record Id and hazardous material release record Id. The original records were joined based on *ReportNumber* and *Crash_Id* from the HMSAIR and MCMIS, respectively. A one-to-many join was performed, creating the possibility that a hazardous material release record was matched to multiple crash records. This duplication of hazardous material release records was a benefit of using the likelihood matching technique, facilitating the identification of true matches by allowing correction for false positive matches.

1.5 Matching Record Pairs between HMSAIR and MCMIS

The resulting data file contained 2,071,238 pairs of records, from which the pairs most likely to be true matches had to be distinguished. Three methods were tested: Naïve Bayesian

Classification, Binary Logistic Modeling, and a Single-layer Back Propagation Neural Network. The results and a comparison of the classifications are discussed below.

1.5.1 Naïve Bayesian Categorization

The individual likelihoods were assumed to be independent based on the understanding that differences between the two files for the identical crash/incident pair were random errors. Under this assumption, the independence of the differences and the joint likelihood value was calculated for each crash/incident pairing. This procedure resulted in 1,520,105 x 165,464 pairings (1,520,105 crash records and 165,464 hazardous material release incidences). To reduce the number of pairings—and subsequently the data file size—only those pairings that displayed a perfect match on the state and month variables were stored. It was assumed that errors in recording the state and month would be small enough to ignore in the estimation of likelihoods. The joint likelihood value was then calculated as the product of the individual likelihood value.

$$L_{ik} = \prod_{j=1}^{7} L_j \tag{1.5}$$

where,

$L_{ij} =$ joint likelihood value for pair

$L_j =$ individual likelihood value for attribute j

i = crash record

k = hazardous materials release record

$$j \in \begin{cases} time \\ day \\ month \\ year \\ phase \\ county \\ state \end{cases}$$

For example, a pair with a time difference of two hours, and zero difference in day, month, year, phase of transportation, county, and state, would have a likelihood score of 0.9986.

The full distribution of likelihoods for the range falling between 0.0 and 1.00 is shown in figure 1.13, revealing a non-smooth function with interesting spikes. The cumulative likelihood distribution is shown in figure 1.14, which shows a typical sigmoidal shape of an error function. Although the likelihood distribution was not an ideal smooth Gaussian distribution, it was treated as if it possessed all the inferential features to make it so. Under this assumption, pairings that fell within the upper tail of the distribution were examined further for consideration in the analysis of crash and hazardous material releases.

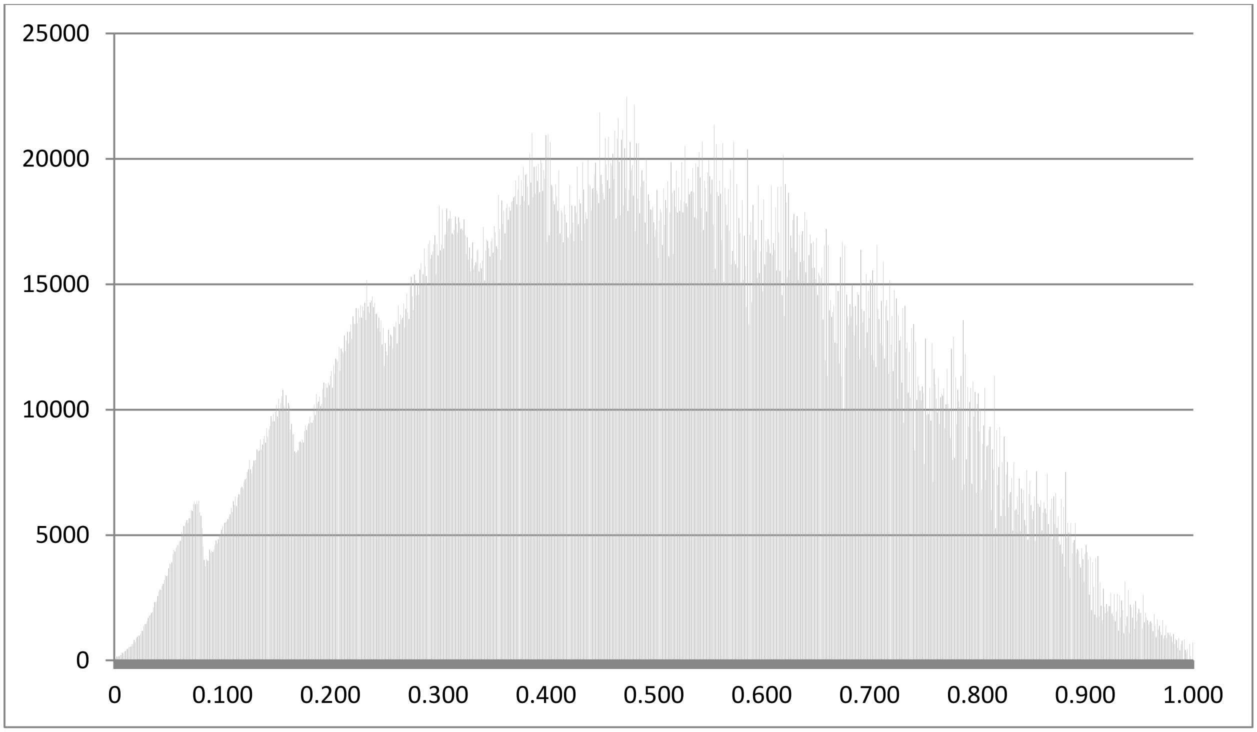

Figure 1.13 Likelihood value distribution

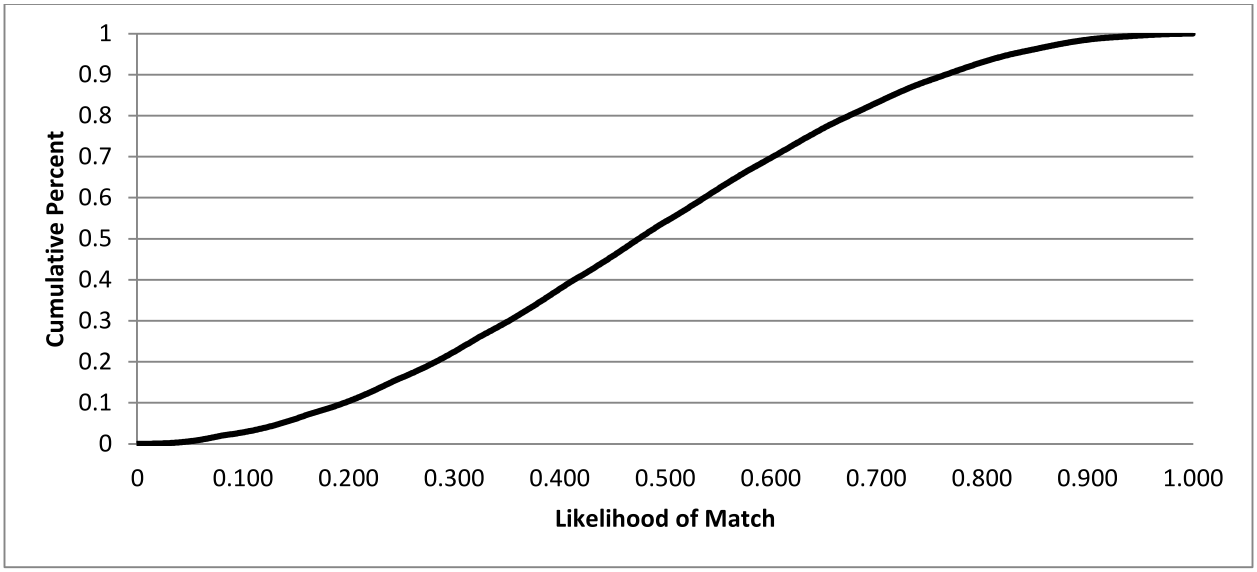

Figure 1.14 Cumulative percent of likelihood values

The likelihood distribution provides a systematic cutoff point for selecting pairings. The cutoff has a likelihood value of 90%. At a 90% likelihood of a match, 1.49% of all possible pairings remained, resulting in 175,454 pairs remaining to be investigated. The likelihood distribution and cumulative distributions are shown in figure 1.15 and figure 1.16, respectively. The non-smooth nature of the likelihood distribution is more evident for the given range of data (0.90 to 1.00). Encouragingly, there were 717 perfect matches between possible pairs.

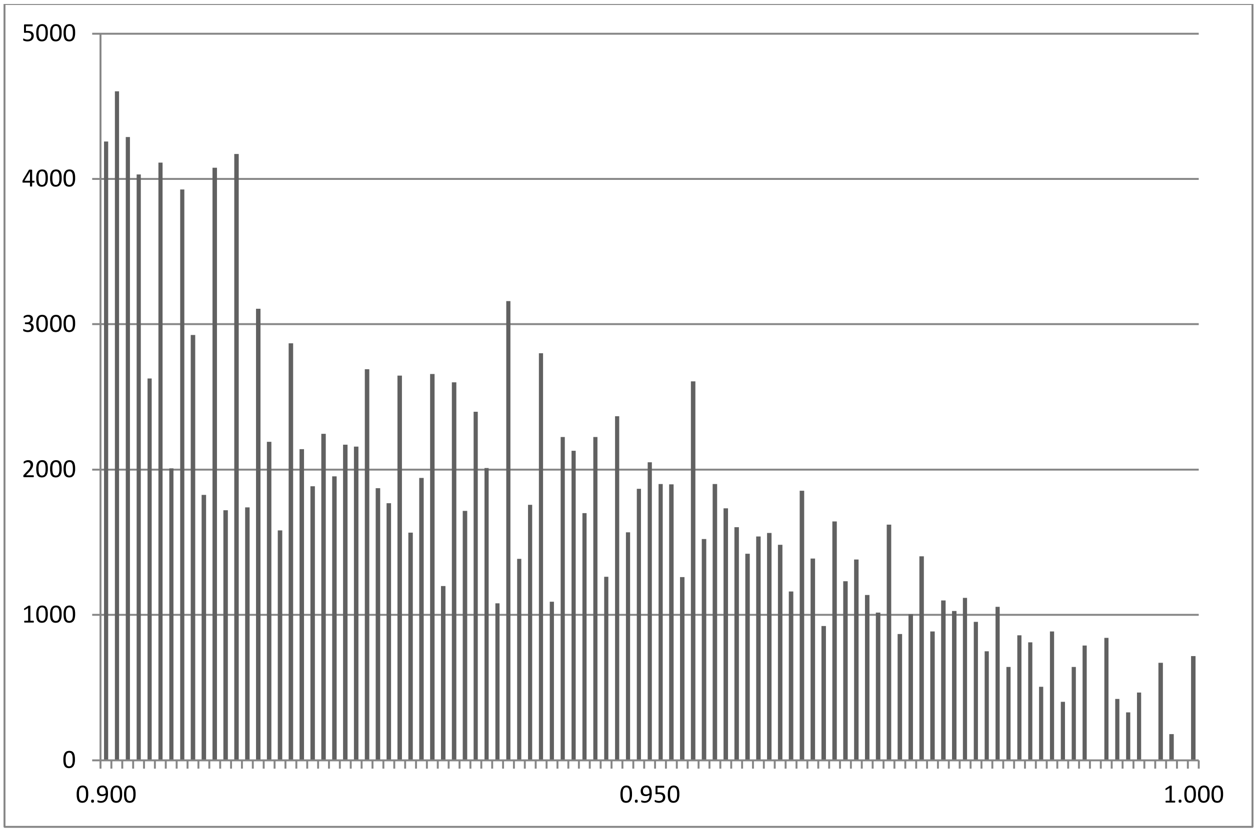

Figure 1.15 Likelihood value of match (range 0.90 to 1.00)

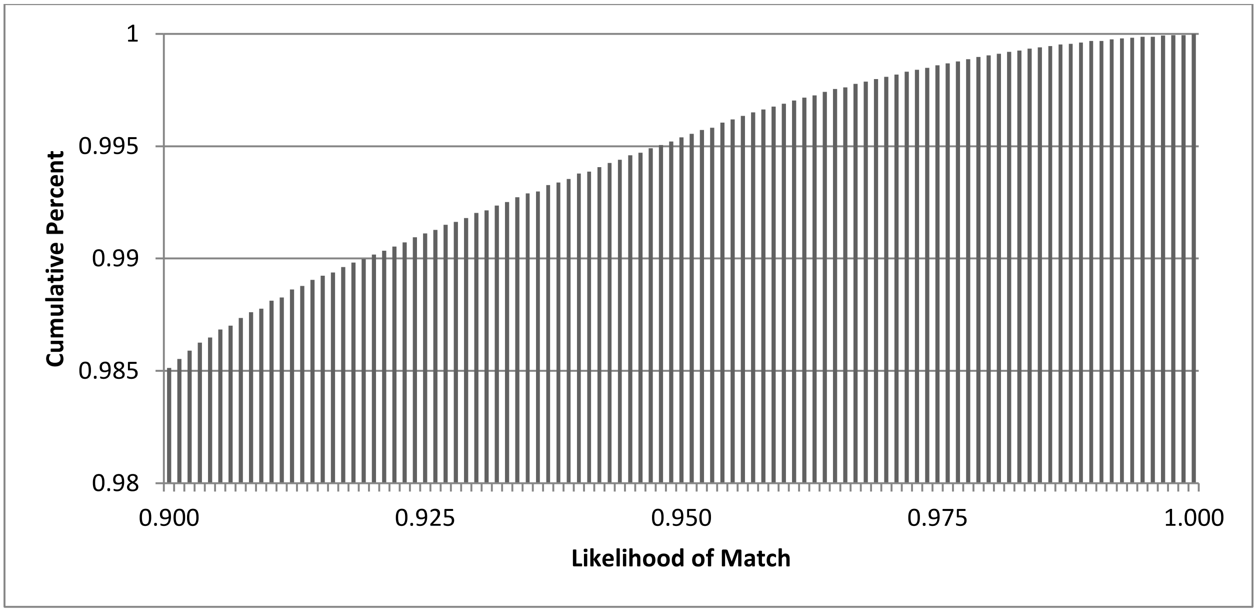

Figure 1.16 Cumulative percent of likelihood value of match (range 0.90 to 1.00)

Although the naïve Bayesian approach determined 717 exact matches, it identified 175,454 possible matches with a cutoff of 0.9. Because the criteria for the naïve Bayesian approach is not well defined, and since a cutoff of 0.5, used as in the binary logistic model, would produce a very large number of identifiable matches ,it was not pursued further.

1.5.2 Binary Logistic Categorization

For the estimation of a binary logistic model, a subset of records was randomly selected from all pairwise combinations; the subset consisted of 825 records, representing approximately 10% of all truck crashes occurring during the study period. Each record in the subset was hand score based to identify records that were true matches and those that were mismatches. The information contained in the previously mentioned record fields was used, as was the location and city of the incident/crash that had been entered into the scoring. The location and city required visual inspection due to the numerous nonstandard methods used to enter data. Of 825 records, 89 (11%) were scored as exact matches; the remaining 736 (89%) were mismatches.

Once the subset was scored, it was randomly split into two groups. The first group of 744 records was used to estimate a binary logistic model. The second group, the hold out group, was categorized using the estimated logistic model, and the forecast was compared to the hand score. The estimated model demonstrated an acceptable fit, as shown by the pseudo-R^2 contained in table 1.9. The coefficients of the binary logistic model are presented in table 1.10. The *year* field was not entered into the logistic model because it was not statistically significant in a previous fitted model. Overall, the estimated logistic model had the ability to correctly categorize 96% of the records used in achieving model fit it. More specifically, the model had a 98% true positive rate for identifying mismatches (2% false positive) and an 85% true positive rate for categorizing matches.

Table 1.9 Binary logistic categorization mode

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	252.522[a]	.298	.597
a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.			

Table 1.10 Coefficients of binary logistic categorization model

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	probTime	10.651	2.834	14.125	1	.000	42245.358
	probDay	15.080	3.978	14.371	1	.000	3540450.887
	probMonth	14.341	3.088	21.575	1	.000	1691833.615
	probCounty	3.511	.613	32.827	1	.000	33.477
	Constant	-42.654	4.518	89.132	1	.000	.000
a. Variable(s) entered on step 1: probTime, probDay, probMonth, probCounty.							

Table 1.11 Classification table for binary logistic categorization model

	Observed		Predicted		
			Match		Percentage Correct
			.00	1.00	
	Match	.00	646	16	97.6
		1.00	12	70	85.4
	Overall Percentage				96.2
a. The cut value is .500					

The estimated model was used to categorize the hold out subset of records. The results of this test are provided in table 1.12. The overall percentage of correctly categorized records increased to 98.8%, with 100% positive in the mismatch category and 85.7% positive matching records.

Table 1.12 Classification table for binary logistic categorization model on holdout data

	Observed		Predicted		
			Match		Percentage Correct
			.00	1.00	
	Match	.00	74	0	100.0
		1.00	1	6	85.7
	Overall Percentage				98.8

a. The cut value is .500

The calibrated binary logistic model was applied to the data file containing 2,071,238 pairs of records (described above). The number of record pairs identified as matches was 894, which was comparable to the result achieved with the naïve Bayesian categorization method.

1.5.2.1 Finalizing matched records with the binary logistic method

Records that were identified by the binary logistic model as matching were extracted from all possible pair combinations, creating a dataset of 894 records. 41 duplicate records (i.e., where the *crash id* from the MCMIS dataset repeated) were removed, resulting in a total of 853 unique matches. For each record, the estimated probability of matching as determined by the binary logistic model was retained and appended to the joined crash-release dataset (described above).

In comparison to the number of truck crash records contained in the MCMIS dataset that were recorded as a hazardous materials release, the numbers for the matched records from the binary logistic method were fewer. The MCMIS records contained 8,987 truck crashes involving a hazardous material release (hazmat_release field coded as 1); therefore, 9.9% of the MCMIS hazardous materials crashes could be linked to the HMSAIR dataset. Perhaps this percentage would be higher if the assumption that both the *state* and *phase of transportation* fields in the MCMIS and HMSAIR were always correctly entered was relaxed—this could be another step in the process following the conclusion of the current study.

1.5.3 Multilayer Neural Network Categorization

The same hand-scored subset of records that was used to estimate the binary logistic model was again used to train and evaluate a multilayer neural network (5 inputs, 1 hidden layer with 4 nodes, input sigmoid activation function, and output softmax activation function). The subset consisted of 825 records—89 (11%) were scored as exact matches, with the remaining 736 (89%) being mismatches. The subset was split in two, consisting of a training set (70%) and a testing set (30%). The result of the training and testing are shown in table 1.13.

Table 1.13 Multilayer neural network description

Input Layer	Covariates	1	probTime
		2	probDay
		3	probMonth
		4	probYear
		5	probCounty
	Number of Units[a]		5
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		4
	Activation Function		Sigmoid
Output Layer	Dependent Variables	1	Match
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit

Table 1.14 Evaluation of the trained neural network

Classification				
Sample	Observed	Predicted		
		.00	1.00	Percent Correct
Training	.00	485	9	98.2%
	1.00	10	54	84.4%
	Overall Percent	88.7%	11.3%	96.6%
Testing	.00	240	2	99.2%
	1.00	2	23	92.0%
	Overall Percent	90.6%	9.4%	98.5%
Dependent Variable: Match				

The trained neural network had an overall accuracy rate of 98.5% when categorizing matching and mismatching records, with a false positive rate of 8%. In this application, a false positive is defined as predicting a match when it was scored as a mismatch.

Table 1.14 contains the results of the training and testing steps. After the training, the neural network was applied to the dataset of all possible record combinations (see above for the description of the dataset). The application resulted in 991 records predicted to be matches. Similar to the binary logistic and naïve Bayesian categorization methods, the neural network approach for this application generated comparable results. To further explore the comparison between the binary logistic and neural network, a cross table of the results from each method was created (see table 1.15).

The comparison between the binary logistic and neural network shows that 788 records were mutually predicted as being matches. The predictions that were discrepant between the methods are of interest: the binary logistic approach categorized 106 records as matches that the neural network predicted as mismatches, leading to a 13.5% rate of false negatives. In comparison to the neural network, the false negative rate of the binary logistic method was 25.8%. Therefore, there is evidence that choosing between the two methods may produce varying final results in the analysis of truck crashes. It should be pointed out that the number of total records commonly predicted to be matches (788) was 71 greater than the naïve Bayesian approach had identified as perfect matches.

Table 1.15 Comparison of the binary logistic and neural network categorization results

		Neural Network Predictions	
		No match	Match
		Count	Count
Binary Logistic Predictions	No match	2018993	203
	Match	106	788

1.5.3.1 Finalizing matched records with neural network

The records that were identified as matching by the neural network were extracted from all possible pair combinations, creating a dataset of 991 records. Those 991 records were scanned for duplicates, of which 50 such records were identified. The duplicate records were then removed, resulting in a final 941 matching records. For each record, the estimated probability of matching as determined by the neural network was retained and appended to the joined crash-release dataset (described above).

In comparison to the truck crash records contained in the MCMIS dataset, the neural network method matched 10.5% of the MCMIS hazardous materials crashes to the HMSAIR dataset. As stated previously, the percentage could perhaps be higher if the assumption that both the *state* and *phase of transportation* fields in the MCMIS and HMSAIR were always correctly entered was relaxed.

Chapter 2 Estimation of the Determinants of Truck Crashes with Hazardous Material Release

2.1 Use of Estimated Probabilities of Matching

An advantage of using either the binary logistic or neural network categorization approach is the probability estimation of a record being a match. These estimated probabilities are used as a measure of certainty, and should be used to influence the inferential analysis of truck crashes. In the analysis that follows, the estimated probabilities enter into the regression analysis as weights. The higher the probability of a match, the more weight ought to be given to that observation—not unlike the weights used to adjust under- and over-sampled populations. The weights were scaled so the total equaled the number of recorded truck crashes involving hazardous material releases.

2.1.1 Results Using the Records of the Binary Logistic Categorization Method

The determinants for the release of hazardous material by material class were identified with the initial analysis. The determinants—independent category variables—used in the analysis were: road surface condition, weather conditions, light condition, vehicle configuration, how the release occurred, and the cause of the release. The coding scheme is presented in table 2.1. The final analysis aggregated *vehicle configuration*, *how the release occurred*, and *the cause of the release*.

Table 2.1 Aggregated categories for vehicle configuration

Combined Category	Original Code and Classification
Bus, other, unknown	1 Passenger Car (Only if Vehicle displays HM Placard)
	2 Light Truck (Only if Vehicle displays HM Placard)
	3 Bus (Seats for 9-15 People, Including Driver)
	4 Bus (Seats for > 15 People, Including Driver)
	99 Unknown Heavy Truck > 10000 lbs., Cannot Classify
Single-Unit	5 Single-Unit Truck (2-Axle, 6-Tire)
	6 Single-Unit Truck (3 or More Axles)
Tractor/Semitrailer	7 Truck/Trailer
	8 Truck Tractor (Bobtail)
	9 Tractor/Semitrailer
Long Combination	10 Tractor/Double
	11 Tractor/Triple

Table 2.2 Aggregated categories for how the release occurred

Combined Category and Code	Original Code and Category
303 Burst or Ruptured	303 Burst or Ruptured
304 Cracked or Crushed	304 Cracked 305 Crushed
308 Leaked	308 Leaked
309 Punctured	309 Punctured
310 Ripped or Torn, Torn off or Damaged	310 Ripped or Torn 312 Torn off or Damaged
99 Other	301 Abraded 302 Bent 306 Failed to Operate 307 Gouged or Cut 311 Structural 313 Vented

Table 2.3 Aggregated categories for the cause of the release

Combined Category and Code	Original Code and Category
501 Other	501 Abrasion 502 Broken Component or Device 506 Corrosion - Exterior 507 Corrosion - Interior 508 Defective Component or Device 510 Deterioration or Aging 532 Stub Sill Separation from Tank (Tank Cars) 533 Threads Worn or Cross Threaded 535 Valve Open 514 Freezing 516 Impact with Sharp or Protruding Object (e.g., nails)
518 Human	505 Conveyer or Material Handling Other Mishap 511 Dropped 513 Forklift Accident 515 Human Error 517 Improper Preparation for Transportation 518 Inadequate Accident Damage Protection 519 Inadequate Blocking and Bracing 520 Inadequate Maintenance 521 Inadequate Preparation for Transportation 522 Inadequate Procedures 523 Inadequate Training 524 Incompatible Product 525 Incorrectly Sized Component or Device 526 Loose Closure, Component, or Device 527 Misaligned Material, Component, or Device 528 Missing Component or Device 534 Too Much Weight on Package 529 Overfilled 530 Over-pressurized
531 Rollover	531 Rollover Accident
537 Crash	509 Derailment 512 Fire, Temperature, or Heat 537 Vehicular Crash or Accident Damage
999 Other	503 Commodity Self-ignition 504 Commodity Polymerization 536 Vandalism 538 Water Damage

The crash records were grouped based on the US DOT hazardous material class codes (see table 2.5). The grouping resulted in six categories: Explosives, Toxic and Radioactive materials, Compressed Gas, Flammable Liquids, Oxidizers, Corrosive Materials, and Miscellaneous. The grouping of Class 1, Class 6, and Class 7 was prompted by having too few observations in the individual classes, which prevented the estimation of the multinomial logistic regression.

Table 2.4 Determinants of hazardous material releases in truck crashes

Category Variable	**Coding**	
Road Surface Condition	Road Surface Wet	2 – Wet 3 – Water 4 – Snow 5 – Slush 6 – Ice 7 – Sand, Mud, Dirt, Oil or Grease 8 – Other 9 – Unknown
	Road Surface Dry	1 – Dry
Weather Condition	Adverse Weather	2 – Rain 3 – Sleet, Hail 4 – Snow 5 – Fog 6 – Blowing Sand, Soil, Dirt, or Snow 7 – Severe Crosswinds 8 – Other 9 – Unknown
	Good Weather	1 – No Adverse Condition

Table 2.4 Determinants of hazardous material releases in truck crashes (cont'd.)

Light Condition	Dark	2 – Dark-Not Lighted 3 – Dark-Lighted 4 – Dark-Unknown 5 – Dawn 6 – Dusk 8 – Other 9 – Unknown or blank
	Daylight	1 – Daylight
Vehicle Configuration	Bus or Unknown	
	Single-Unit Truck	
	Long Combination	
	Tractor/Semitrailer	
How Release Occurred	Burst or Ruptured	
	Cracked or Crushed	
	Leaked	
	Punctured	
	Ripped or Torn Off	
	Other	
Cause of Release	Other	
	Equipment	
	Human Factors	
	Rollover Accident	
	Crash (Non-rollover)	

The multinomial logistic regression revealed a statistically significant set of results that distinguished between six hazardous material classes (see 2.5 for definitions). The six classifications were aggregated using the number of observations and similarity of material from the existing nine classifications defined by the US DOT. The Flammable Liquids class was used as the comparison group in the analysis, because it was the most frequent hazardous material release recorded in the data set.

Table 2.5 Hazardous material reclassification

Material Class	Original DOT Material Classification
Hazard Class 1: Explosives	1.1 mass explosion hazard 1.2 projectile hazard 1.3 minor blast/projectile/fire 1.4 minor blast 1.5 insensitive explosives 1.6 very insensitive explosives
Hazard Class 2: Compressed Gases	2.1 flammable gases 2.2 nonflammable compressed 2.3 poisonous
Hazard Class 3: Flammable Liquids	Flammable (flash point below 141°) Combustible (flash point 141°-200°
Hazard Class 4: Flammable Solids	4.1 flammable solids 4.2 spontaneously combustible 4.3 dangerous when wet
Hazard Class 5: Oxidizers and Organic Peroxides	5.1 Oxidizer 5.2 Organic Peroxide
Hazard Class 6: Toxic Materials	6.1 Material that is poisonous 6.2 Infectious Agents
Hazard Class 7: Radioactive Material	Radioactive I Radioactive II Radioactive III
Hazard Class 8: Corrosive Material	Destruction of the human skin Corrode steel at a rate of 0.25 inches per year
Hazard Class 9: Miscellaneous	A material that presents a hazard during shipment but does not meet the definition of the other classes

Highlights of table 2.6 (below) are as follows:

- All classes of hazardous material were less likely than Flammable Liquids to be released in a truck crash.
- Road surface and weather were not significant factors in distinguishing different types of material releases.
- Time of day was a significant factor in the release of Oxidizers and Compressed Gas. Reference category was Flammable Liquids.

- The type of vehicle was significant in distinguishing which hazardous material release occurred. Reference category was tractor/semi-trailer configuration:
    - Explosive, Toxic, or Radioactive material were more likely to be released from unrecorded vehicles.
    - Unrecorded and single-unit trucks were more highly associated with Compressed Gas release.
    - Oxidizers and Corrosive Materials were more likely to be released from all but tractor/semitrailer vehicle configurations.
    - Single-unit trucks were more likely to leak miscellaneous material.
- The cause of failure was a significant determinant of material release:
    - Burst or rupture failure significantly distinguished flammable material release from all other releases.
    - Cracked containers were a significant causes of releases in all categories except for Oxidizers and Flammable Liquids.
    - Leaking containers were significantly likely to be the best predictor (in comparison to other failures) of failure-related release for compressed gas and corrosive materials.
    - Puncture failures were significantly related to leakage for all but Corrosive Materials.
    - Explosives, Toxic or Radioactive, Corrosive, and Miscellaneous material release was significantly associated with rip or tear failure.
- The reason failure occurred was also a distinguishing factor influencing the release of different material classes in rollover and non-rollover crashes:

- Equipment failure was more significant in the release of Corrosive Materials.
- Rollover crashes were a greater factor in Flammable Liquid and Oxidizer release.
- Failures due to human oversight were significant in Oxidizers and Corrosive Material releases.

Table 2.6 Case processing summary

		N	Marginal Percentage
Hazardous Material Code	Explosives, Toxic, Radioactive	324	4.0%
	Compressed Gas	1259	15.4%
	Flammable Liquids	4717	57.7%
	Oxidizers	359	4.4%
	Corrosive materials	1068	13.1%
	Miscellaneous	449	5.5%
Road Surface	Wet	1779	21.8%
	Dry	6397	78.2%
Weather	Adverse	1996	24.4%
	Non-adverse	6180	75.6%
Light Condition	Dark	3164	38.7%
	Day Light	5012	61.3%

Table 2.6 Case processing summary (cont'd.)

Bus or Unknown	unknown & buses	315	3.9%
	Single-Unit Truck	1214	14.8%
	Long Combination	518	6.3%
	Tractor/Semitrailer	6129	75.0%
How Failure Occurred	Burst or Ruptured	1190	14.6%
	Cracked	1140	13.9%
	Leaked	520	6.4%
	Punctured	1571	19.2%
	Ripped or Torn	903	11.0%
	Other	2852	34.9%
Cause of Failure	Other	767	9.4%
	Other	495	6.1%
	Human	706	8.6%
	Rollover	4651	56.9%
	Crash	1557	19.0%
Valid		8176	100.0%
Missing		0	
Total		8176	
Subpopulation		247[a]	

a. The dependent variable has only one value observed in 156 (63.2%) subpopulations.

Table 2.7 Model fitting information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	9934.869			
Final	8328.848	1606.021	75	.000

Table 2.8 Pseudo R-square

Cox and Snell	.178
Nagelkerke	.193
McFadden	.076

Table 2.9 Likelihood ratio tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	8328.848[a]	.000	0	.
Road Surface	8340.658	11.810	5	.037
Weather Condition	8342.251	13.403	5	.020
Light Condition	8421.197	92.349	5	.000
Bus or Unknown	8885.617	556.769	15	.000
How Failure Occurred	8896.936	568.088	25	.000
Cause of Failure	8556.477	227.629	20	.000

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Table 2.10 Results for multinomial logistic regression for determinants of hazardous material class

Parameter Estimates

Hazardous Material Code		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
Explosives, Toxic, Radioactive	Intercept	-1.728	.159	117.494	1	.000			
	Road Surface Wet	-.034	.189	.032	1	.858	.967	.668	1.400
	Road Surface Dry	0^b	.	.	0	.	.	.	.
	Adverse Weather	-.108	.182	.355	1	.551	.897	.629	1.281
	Good Weather	0^b	.	.	0	.	.	.	.
	Dark	.217	.125	3.025	1	.082	1.242	.973	1.586
	Daylight	0^b	.	.	0	.	.	.	.
	Bus or Unknown	1.194	.225	28.107	1	.000	3.299	2.122	5.129
	Single-Unit Truck	.274	.184	2.227	1	.136	1.316	.918	1.886
	Long Combination	.200	.259	.598	1	.439	1.222	.735	2.030
	Tractor/Semitrailer	0^b	.	.	0	.	.	.	.
	Burst or Ruptured	-1.082	.187	33.474	1	.000	.339	.235	.489
	Cracked or Crushed	-.682	.190	12.855	1	.000	.505	.348	.734
	Leaked	.178	.196	.821	1	.365	1.195	.813	1.756
	Punctured	-2.649	.330	64.563	1	.000	.071	.037	.135
	Ripped or Torn Off	-1.555	.251	38.275	1	.000	.211	.129	.346
	Other	0^b	.	.	0	.	.	.	.
	Other	-1.287	.268	23.078	1	.000	.276	.163	.467
	Equipment	-.925	.351	6.954	1	.008	.397	.199	.789
	Human Factors	-.086	.214	.163	1	.687	.917	.604	1.394
	Rollover Accident	-.549	.144	14.473	1	.000	.577	.435	.766
	Crash (Non-rollover)	0^b	.	.	0	.	.	.	.
Compressed Gas	Intercept	-1.300	.105	153.076	1	.000			
	Road Surface Wet	.202	.111	3.309	1	.069	1.223	.984	1.520
	Road Surface Dry	0^b	.	.	0	.	.	.	.
	Adverse Weather	-.277	.111	6.195	1	.013	.758	.609	.943

Table 2.10 Results for multinomial logistic regression for determinants of hazardous material class (cont'd.)

	Good Weather	0[b]	.	.	0	.	.	.	.
	Dark	-.381	.077	24.790	1	.000	.683	.588	.794
	Daylight	0[b]	.	.	0	.	.	.	.
	Bus or Unknown	1.547	.150	106.001	1	.000	4.700	3.500	6.309
	Single-Unit Truck	1.502	.082	337.163	1	.000	4.492	3.826	5.273
	Long Combination	.070	.164	.182	1	.669	1.073	.777	1.480
	Tractor/Semitrailer	0[b]	.	.	0	.	.	.	.
	Burst or Ruptured	-.878	.125	49.175	1	.000	.415	.325	.531
	Cracked or Crushed	.384	.099	14.959	1	.000	1.469	1.209	1.785
	Leaked	.483	.142	11.647	1	.001	1.621	1.228	2.140
	Punctured	-.470	.102	21.102	1	.000	.625	.511	.764
	Ripped or Torn Off	-.293	.118	6.171	1	.013	.746	.592	.940
	Other	0[b]	.	.	0	.	.	.	.
	Other	-.144	.138	1.078	1	.299	.866	.660	1.136
	Equipment	-.153	.157	.941	1	.332	.858	.631	1.169
	Human Factors	-.295	.152	3.776	1	.052	.744	.553	1.003
	Rollover Accident	-.144	.089	2.642	1	.104	.866	.728	1.030
	Crash (Non-rollover)	0[b]	.	.	0	.	.	.	.
Oxidizers	Intercept	-3.509	.200	307.383	1	.000			
	Road Surface Wet	-.107	.185	.335	1	.563	.898	.625	1.292
	Road Surface Dry	0[b]	.	.	0	.	.	.	.
	Adverse Weather	-.008	.179	.002	1	.965	.992	.699	1.408
	Good Weather	0[b]	.	.	0	.	.	.	.
	Dark	.739	.118	39.100	1	.000	2.094	1.661	2.640
	Daylight	0[b]	.	.	0	.	.	.	.
	Bus or Unknown	1.133	.261	18.763	1	.000	3.104	1.859	5.182
	Single-Unit Truck	1.448	.139	108.977	1	.000	4.256	3.243	5.586
	Long Combination	.718	.210	11.691	1	.001	2.050	1.358	3.094
	Tractor/Semitrailer	0[b]	.	.	0	.	.	.	.
	Burst or Ruptured	-.004	.164	.001	1	.981	.996	.722	1.375
	Cracked or Crushed	-.558	.217	6.619	1	.010	.572	.374	.876
	Leaked	.342	.229	2.237	1	.135	1.408	.899	2.206
	Punctured	-.546	.181	9.060	1	.003	.580	.406	.827
	Ripped or Torn Off	-.030	.185	.027	1	.870	.970	.674	1.395
	Other	0[b]	.	.	0	.	.	.	.
	Other	.518	.249	4.330	1	.037	1.679	1.031	2.735

Table 2.10 Results for multinomial logistic regression for determinants of hazardous material class (cont'd)

	Equipment	-.479	.361	1.761	1	.184	.619	.305	1.257
	Human Factors	1.016	.218	21.617	1	.000	2.761	1.799	4.236
	Rollover Accident	.456	.169	7.293	1	.007	1.578	1.133	2.197
	Crash (Non-rollover)	0[b]	.	.	0	.	.	.	.
Corrosive materials	Intercept	-1.832	.115	253.018	1	.000			
	Road Surface Wet	-.269	.115	5.535	1	.019	.764	.610	.956
	Road Surface Dry	0[b]	.	.	0	.	.	.	.
	Adverse Weather	.215	.106	4.119	1	.042	1.240	1.007	1.526
	Good Weather	0[b]	.	.	0	.	.	.	.
	Dark	.175	.075	5.529	1	.019	1.191	1.030	1.379
	Daylight	0[b]	.	.	0	.	.	.	.
	Bus or Unknown	.649	.183	12.573	1	.000	1.914	1.337	2.740
	Single-Unit Truck	.410	.109	14.237	1	.000	1.507	1.218	1.865
	Long Combination	.733	.122	36.367	1	.000	2.082	1.640	2.642
	Tractor/Semitrailer	0[b]	.	.	0	.	.	.	.
	Burst or Ruptured	-.299	.125	5.766	1	.016	.741	.581	.947
	Cracked or Crushed	.857	.105	66.896	1	.000	2.356	1.919	2.893
	Leaked	.951	.139	46.654	1	.000	2.589	1.970	3.401
	Punctured	-.045	.107	.176	1	.675	.956	.775	1.179
	Ripped or Torn Off	-.522	.149	12.285	1	.000	.594	.443	.795
	Other	0[b]	.	.	0	.	.	.	.
	Other	.340	.144	5.546	1	.019	1.405	1.059	1.865
	Equipment	.569	.147	14.920	1	.000	1.766	1.323	2.357
	Human Factors	.880	.125	49.587	1	.000	2.411	1.887	3.081
	Rollover Accident	-.294	.096	9.300	1	.002	.746	.617	.900
	Crash (Non-rollover)	0[b]	.	.	0	.	.	.	.
Miscellaneous	Intercept	-1.752	.157	123.744	1	.000			
	Road Surface Wet	-.086	.176	.238	1	.626	.918	.650	1.296
	Road Surface Dry	0[b]	.	.	0	.	.	.	.
	Adverse Weather	-.030	.164	.034	1	.854	.970	.703	1.339
	Good Weather	0[b]	.	.	0	.	.	.	.
	Dark	-.260	.109	5.685	1	.017	.771	.622	.955
	Daylight	0[b]	.	.	0	.	.	.	.
	Bus or Unknown	-.211	.338	.388	1	.533	.810	.417	1.572

Table 2.10 Results for multinomial logistic regression for determinants of hazardous material class (cont'd.)

	Single-Unit Truck	-.685	.211	10.503	1	.001	.504	.333	.763
	Long Combination	.580	.185	9.831	1	.002	1.786	1.243	2.567
	Tractor/Semitrailer	0[b]	.	.	0	.	.	.	.
	Burst or Ruptured	-.455	.145	9.911	1	.002	.634	.478	.842
	Cracked or Crushed	-1.604	.253	40.108	1	.000	.201	.122	.330
	Leaked	-.504	.249	4.098	1	.043	.604	.371	.984
	Punctured	-1.255	.166	57.403	1	.000	.285	.206	.394
	Ripped or Torn Off	-.565	.171	10.870	1	.001	.568	.406	.795
	Other	0[b]	.	.	0	.	.	.	.
	Other	.044	.203	.046	1	.830	1.045	.701	1.556
	Equipment	-.224	.272	.681	1	.409	.799	.469	1.362
	Human Factors	.349	.213	2.677	1	.102	1.418	.933	2.154
	Rollover Accident	.028	.144	.037	1	.846	1.028	.775	1.364
	Crash (Non-rollover)	0[b]	.	.	0	.	.	.	.

a. The reference category is: Flammable Liquids.

b. This parameter is set to zero because it is redundant.

The following tables provide a full comparison of the significant factors leading to release of hazardous materials:

Table 2.11 Impact of road surface conditions

Road Surface Condition Odds i /Odds j	**j**					
i	Explosives, Toxic, Radioactive	Compressed Gas	Flammable Liquids	Oxidizers	Corrosive materials	Miscellaneous
Explosives, Toxic, Radioactive	-	ns	ns	ns	ns	ns
Compressed Gas	ns	-	ns	ns	1.602 (0.001)	ns
Flammable Liquids	ns	ns	-	ns	ns	ns
Oxidizers	ns	ns	ns	-	ns	ns
Corrosive materials	ns	0.624 (0.001)	ns	ns	-	ns
Miscellaneous	ns	ns	ns	ns	ns	-

Comparison is wet road surface

Table 2.12 Impact of weather conditions

Weather Condition Odds i /Odds j	**j**					
i	Explosives, Toxic, Radioactive	Compressed Gas	Flammable Liquids	Oxidizers	Corrosive materials	Miscellan-eous
Explosives, Toxic, Radioactive	-	ns	ns	ns	ns	ns
Compressed Gas	ns	-	ns	ns	6.11 (<0.001)	ns
Flammable Liquids	ns	ns	-	ns	ns	ns
Oxidizers	ns	ns	ns	-	ns	ns
Corrosive materials	ns	1.636 (<0.001)	ns	ns	-	ns
Miscellaneous	ns	ns	ns	ns	ns	-

Comparison group good weather conditions

Table 2.13 Impact of light conditions

Light Condition Odds i /Odds j	**j**					
i	Explosives Toxic, Radioactive	Compre-ssed Gas	Flammable Liquids	Oxidizers	Corrosive materials	Miscellan-eous
Explosives, Toxic, Radioactive	-	1.818 (<0.001)	ns	0.593 (0.002)	ns	1.611 (0.003)
Compressed Gas	0.55	-	0.683 (<0.001)	0.326 (<0.001)	0.573 (<0.001)	ns
Flammable Liquids	ns	1.464 (<0.001)	-	0.478 (<0.001)	ns	ns
Oxidizers	1.686 (0.002)	3.065 (<0.001)	2.094 (<0.001)		1.758 (<0.001)	2.717 (<0.001)
Corrosive materials	ns	1.744 (<0.001)	ns	0.569 (<0.001)	-	1.546 (<0.001)
Miscellaneous	0.621 (0.003)	ns	ns	0.368 (<0.001)	0.647 (<0.001)	-

Comparison group daylight

Table 2.14 Impact of vehicle configuration

Vehicle Configuration Odds i /Odds j					**j**							
	Explosives, Toxic, Radioactive				**Compressed Gas**				**Flammable Liquids**			
i	Bus and Unknown	Single-Unit	Long Combination	Tractor /Semitrailer	Bus and Unknown	Single-Unit	Long Combination	Tractor /Semi-trailer	Bus and Unknown	Single-Unit	Long Combination	Tractor /Semi-trailer
Explosives, Toxic, Radioactive	-	-	-	-		0.293 (<0.001)			3.299 (<0.001)			
Compressed Gas		3.414 (<0.001)			-	-	-	-	4.7 (<0.001)	4.492 (<0.001)		
Flammable Liquids	0.303 (<0.001)				0.213 (<0.001)	0.223 (<0.001)			-	-	-	-
Oxidizers		3.235 (<0.001)							3.104 (<0.001)	4.256 (<0.001)	2.05 (0.001)	
Corrosive materials					0.407 (<0.001)	0.336 (<0.001)	1.941 (<0.001)		1.914 (<0.001)	1.507 (<0.001)	2.082 (<0.001)	
Miscellaneous	0.245 (<0.001)	0.383 (<0.001)			0.172 (<0.001)	0.112 (<0.001)				0.504 (0.001)	1.786 (0.002)	

Table 2.14 Impact of vehicle configuration (cont'd.)

Vehicle Configuration **Odds i /Odds j**						j						
	Oxidizers				**Corrosive materials**				**Miscellaneous**			
	Bus and Unknown	Single-Unit	Long Combination	Tractor/Semitrailer	Bus and Unknown	Single-Unit	Long Combination	Tractor/Semitrailer	Bus and Unknown	Single-Unit	Long Combination	Tractor/Semitrailer
Explosives, Toxic, Radioactive		0.309 (<0.001)							4.073 (<0.001)	2.61 (<0.001)		
Compressed Gas					2.456 (<0.001)	2.98 (<0.001)	0.515 (<0.001)		5.803 (<0.001)	8.911 (<0.001)		
Flammable Liquids	0.322 (<0.001)	0.235 (<0.001)	0.488 (0.001)		0.523 (<0.001)	0.664 (<0.001)	0.48 (<0.001)			1.984 (0.001)	0.56 (0.002)	
Oxidizers	-	-	-	-		2.824 (<0.001)			3.833 (0.001)	8.443 (<0.001)		
Corrosive materials		0.354 (<0.001)			-	-	-	-		2.99 (<0.001)		
Miscellaneous	0.261 (<0.001)	0.118 (0.001)				0.334 (<0.001)			-	-	-	-

Table 2.15 How the release occurred

How Failure Occurred **Odds i /Odds j**					**j**							
	Explosives, Toxic, Radioactive						Compressed Gas					
	Burst or Ruptured	Cracked or Crushed	Leaked	Punctured	Ripped or Torn Off	Other	Burst or Ruptured	Cracked or Crushed	Leaked	Punctured	Ripp-ed or Torn Off	Oth-er
Explosives, Toxic, Radioactive							0.344 (<0.001)		0.113 (<0.001)	0.283 (<0.001)		
Compressed Gas		2.906 (<0.001)		8.834 (<0.001)	3.534 (<0.001)							
Flammable Liquids	2.95 (<0.001)	1.979 (<0.001)		14.14 (<0.001)	4.737 (<0.001)		2.407 (<0.001)	0.681 (<0.001)	0.617 (0.001)	1.601 (<0.001)		
Oxidizers	2.939 (<0.001)			8.195 (<0.001)	4.595 (<0.001)			0.39 (<0.001)				
Corrosive materials	2.187 (<0.001)	4.662 (<0.001)	2.167 (<0.001)	13.52 (<0.001)	2.812 (<0.001)		1.785 (<0.001)	1.604 (<0.001)		1.53 (0.002)		
Miscellaneous		0.398 (0.003)		4.032 (<0.001)	2.693 (0.001)			0.137 (<0.001)	0.373 (<0.001)	0.456 (<0.001)		

Table 2.15 How the release occurred (cont'd.)

How Failure Occurred **Odds i /Odds j**					**j**							
	Flammable Liquids						Oxidizers					
	Burst or Ruptured	Cracked or Crushed	Leaked	Punctured	Ripped or Torn Off	Oth -er	Burst or Ruptured	Cracked or Crushed	Leak -ed	Punctured	Ripped or Torn Off	Oth- er
Explosives, Toxic, Radioactive	0.339 (<0.001)	0.505 (<0.001)		0.071 (<0.001)	0.211 (<0.001)		0.34 (<0.001)			0.122 (<0.001)	0.218 (<0.001)	
Compressed Gas	0.415 (<0.001)	1.469 (<0.001)	1.621 (0.001)	0.625 (<0.001)			0.417 (<0.001)	2.566 (<0.001)				
Flammable Liquids										1.726 (0.003)		
Oxidizers				0.58 (<0.003)								
Corrosive materials		2.356 (<0.001)	2.589 (<0.001)		0.594 (<0.001)			4.116 (<0.001)				
Miscellaneous	0.634 (0.002)	0.201 (<0.001)		0.285 (<0.001)	0.568 (0.001)			0.351 (0.001)		0.492 (0.003)		

Table 2.15 How the release occurred (cont’d.)

How Failure Occurred **Odds i /Odds j**					**j**							
	Corrosive materials						Miscellaneous					
	Burst or Ruptured	Cracked or Crushed	Leaked	Punctured	Ripped or Torn Off	Oth-er	Burst or Ruptured	Cracked or Crushed	Leaked	Punctured	Ripped or Torn Off	Oth-er
Explosives, Toxic, Radioactive	0.457 (<0.001)	0.215 (<0.001)	0.462 (<0.001)	0.074 (<0.001)	0.356 (<0.001)			2.514 (0.003)		0.248 (<0.001)	0.371 (0.001)	
Compressed Gas	0.56 (<0.001)	0.623 (<0.001)		0.653 (0.002)				7.304 (<0.001)	2.684 (<0.001)	2.191 (<0.001)		
Flammable Liquids		0.424 (<0.001)	0.386 (<0.001)		1.685 (<0.001)		1.576 (0.002)	4.973 (<0.000)		3.507 (<0.000)	1.759 (0.001)	
Oxidizers		0.243 (<0.001)						2.847 (0.001)		2.033 (0.003)		
Corrosive materials								11.718 (<0.001)	4.286 (<0.001)	3.354 (<0.001)		
Miscellaneous		0.085 (<0.001)	0.233 (<0.001)	0.298 (<0.001)								

Table 2.16 Cause of the release

Cause of Failure **Odds i /Odds j**					**j**					
	Explosives, Toxic, Radioactive					Compressed Gas				
i	Other	Equipment	Human Factor	Rollover	Crash (non-rollover)	Other	Equipment	Human Factor	Rollover	Crash (non-rollover)
Explosives, Toxic, Radioactive	-	-	-	-		0.319 (<0.001)				
Compressed Gas	3.137 (<0.001)					-	-	-	-	
Flammable Liquids	3.622									
Oxidizers	6.08 (<0.001)		3.009 (<0.001)	2.733 (<0.001)				3.709 (<0.001)	1.822 (0.001)	
Corrosive materials	5.088 (<0.001)	4.454 (<0.001)	2.628 (<0.001)				2.058 (<0.001)	3.24 (<0.001)		
Miscellaneous	3.784 (<0.001)			1.781 (<0.001)						

Table 2.16 Cause of the release (cont'd.)

Cause of Failure Odds i /Odds j				**j**						
	Flammable Liquids					Oxidizers				
i	Other	Equipment	Human Factor	Rollover	Crash (non-rollover)	Other	Equipment	Human Factor	Rollover	Crash (non-rollover)
Explosives, Toxic, Radioactive	0.276 (<0.001)			0.577 (<0.001)				0.332 (<0.001)	0.366 (0.001)	
Compressed Gas								0.27 (<0.001)	0.549 (0.001)	
Flammable Liquids	-	-	-	-				0.362 (<0.001)		
Oxidizers			2.761 (<0.001)			-	-	-	-	
Corrosive materials		1.766 (<0.001)	2.411 (<0.001)	0.746 (<0.002)					0.473 (<0.001)	
Miscellaneous										

Table 2.16 Cause of the release (cont'd.)

Cause of Failure Odds i /Odds j			**j**							
	Corrosive materials					Miscellaneous				
i	Other	Equipment	Human Factor	Rollover	Crash (non-rollover)	Other	Equipment	Human Factor	Rollover	Crash (non-rollover)
Explosives, Toxic, Radioactive	0.197 (<0.001)	0.224 (<0.001)	0.38 (<0.001)			0.264 (<0.001)			0.561 (0.003)	
Compressed Gas		0.486 (<0.001)	0.309 (<0.001)			ns	ns	ns	ns	
Flammable Liquids		0.566 (<0.001)	0.415 (<0.001)	1.341 (<0.001)		ns	ns	ns	ns	
Oxidizers				2.116 (<0.001)		ns	ns	ns	ns	
Corrosive materials	-	-	-	-		ns	ns	ns	ns	
Miscellaneous						-	-	-	-	

Tables 2.11 through 2.16 contain odds ratios for each determinant for the different hazardous material classes. As the highlighted sections of table 2.6 illustrate, road surface and weather conditions did not differentially influence the release of any one specific material type over another—that is, all the releases were affected similarly across conditions. The presence or absence of daylight did consistently and significantly correlate with the release of Compressed Gas — Compressed Gas was 1.8, 1.5,3, and 1.7 times more likely to be released in non-daylight than were Explosives, Flammable Liquids, Oxidizers, Corrosive Materials, and Miscellaneous materials, respectively.

Vehicle configuration mapped well with probabilistic expectations based on the physical properties of the hazardous materials. For example, Compressed Gas was 4.5 times more likely to be released from single-unit trucks than from tractor/semitrailers. An insight gleaned from the cause of release is that failure linked to the preparation of the material for transport was significant in the cases of Compressed Gas, Oxidizers, and Corrosive Materials release. A comparatively higher risk of release for these materials was caused by factors such as failing to secure the containers or adequately pack the material. Therefore, an intervention at this stage of transportation could significantly reduce such releases.

2.2 Truck Crashes with and without Hazardous Material Release

The classification of truck crashes with and without hazardous material release provides the opportunity to explore the determinants of these crashes for significant differences. That is, it is possible, for example, to test whether or not weather conditions were a significant factor distinguishing between crashes with and without releases. The following section describes the method used to determine the comparison pool of crashes without releases and the comparison to those with releases.

2.2.1 Estimation Technique

The records of crashes with hazardous material release, matched using the methods described previously, were compared to crashes without releases. The weighted number of records with hazardous material release was 8,176. The number of records for non-release crashes—1,519,252—was several orders of magnitude greater. A Monte-Carlo simulation was applied to the estimation of a multinomial logistic regression with the purpose of distinguishing between types of hazardous material release and non-release crashes.

First, the simulation randomly selected 8,176 non-hazardous material release records, which were subsequently merged with all records reporting a release. The resulting dataset contained 16,352 records, which were used to estimate a multinomial logistic regression. The beta parameter and significance estimates were recorded and stored. The process of randomly sampling from the non-hazardous material release records, merging, estimating the parameters, and storing the estimates, was repeated 500 times. The simulation stopped at 500 after observing that the parameter estimates were stable and the number of draws provided enough for statistical testing. As an example, the distribution for the intercept term on Explosives, Toxic, and Radioactive material releases is summarized and shown in figure 2.1 and table 2.17.

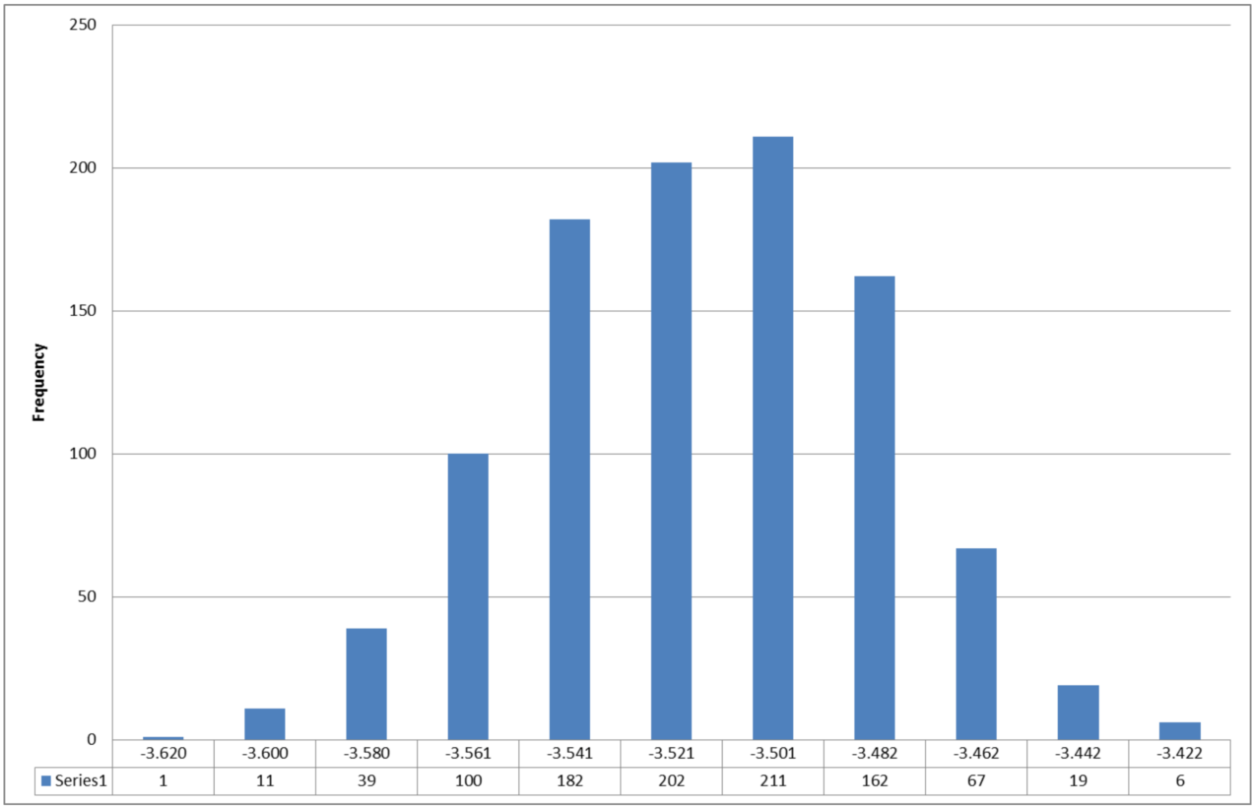

Figure 2.1 Distribution of explosives, toxic, and radioactive material intercept

Table 2.17 Distribution statistics for explosives, toxic, and radioactive intercept

Statistic	**Value**
Minimum	-3.61988
Maximum	-3.42248
Mean	-3.52508
Standard Deviation	0.033487

The dependent variable used in the multinomial logistic regression was the hazardous material class released, as shown in table 2.5, with the addition of a *None* category. The *None* category represented all crashes without a hazardous material release. The determinants for distinguishing between hazardous material classes were: road surface condition, weather condition, light condition, vehicle configuration, access control, and divided travel way. The determinants *how the release occurred* and *the cause of the release* were not used because they did not pertain to all records.

Table 2.18 Simulated significance values for determinants

Hazardous Material Class	**Determinant**	**Minimum**	**Maximum**	**Mean**	**St Dev.**
Explosives, Toxic, Radioactive	Intercept	0	0	0	0
	Road Surface Wet	0.003	0.096	0.02257	0.011121
	Road Surface Dry				
	Adverse Weather	0.12	0.708	0.366132	0.099783
	Good Weather				
	Dark	0	0	0	0
	Daylight				
	Bus or Unknown	0	0.004	0.000463	0.000574
	Single-Unit Truck	0	0	0	0
	Long Combination	0.007	0.316	0.074117	0.044555
	Tractor/Semitrailer				
	Unknown	0.266	1	0.646624	0.150169
	No Access Control	0.028	0.386	0.136902	0.059315
	Partial Access Control	0.114	0.661	0.304848	0.087711
	Full Access Control				
	Unknown	0	0.015	0.002703	0.001854
	Two-way Divided positive median barrier	0.239	0.856	0.503543	0.107092
	Two-way Divided unprotected median	0	0	0	0
	Two-way not divided				
Compressed Gas	Intercept	0	0	0	0
	Road Surface Wet	0.024	0.97	0.261705	0.147484
	Road Surface Dry				
	Adverse Weather	0	0.007	0.000318	0.00073
	Good Weather				
	Dark	0.01	0.837	0.194747	0.120436
	Daylight				

Table 2.18 Simulated significance values for determinants (cont'd.)

	Bus or Unknown	0	0	0	0
	Single-Unit Truck	0	0	0	0
	Long Combination	0	0.073	0.004711	0.006981
	Tractor/Semitrailer				
	Unknown	0.131	0.999	0.727565	0.185264
	No Access Control	0.081	1	0.584676	0.226522
	Partial Access Control	0	0.172	0.012404	0.014725
	Full Access Control				
	Unknown	0	0.201	0.01675	0.020531
	Two-way Divided positive median barrier	0	0	0	0
	Two-way Divided unprotected median	0.004	0.514	0.099154	0.075944
	Two-way not divided				
Flammable Liquids	Intercept	0.011	0.997	0.467217	0.273508
	Road Surface Wet	0	0	0	0
	Road Surface Dry				
	Adverse Weather	0.003	0.989	0.275364	0.229341
	Good Weather				
	Dark	0	0	0	0
	Daylight				
	Bus or Unknown	0	0	0	0
	Single-Unit Truck	0	0	0	0
	Long Combination	0	0.114	0.0019	0.006913
	Tractor/Semitrailer				
	Unknown	0	0.329	0.0164	0.033016
	No Access Control	0	0.95	0.116674	0.148086
	Partial Access Control	0	0.001	0.000001	3.16E-05
	Full Access Control				
	Unknown	0	0.985	0.077185	0.113824
	Two-way Divided positive median barrier	0	0	0	0
	Two-way Divided unprotected median	0	0.163	0.004829	0.012901
	Two-way not divided				
Oxidizers	Intercept	0	0	0	0

Table 2.18 Simulated significance values for determinants (cont’d.)

	Road Surface Wet	0	0.017	0.003415	0.002307
	Road Surface Dry				
	Adverse Weather	0.151	0.87	0.462183	0.125141
	Good Weather				
	Dark	0	0	0	0
	Daylight				
	Bus or Unknown	0	0	0	0
	Single-Unit Truck	0.077	0.738	0.304163	0.106456
	Long Combination	0	0	0	0
	Tractor/Semitrailer				
	Unknown	0	0.01	0.00103	0.000936
	No Access Control	0	0.01	0.002189	0.001526
	Partial Access Control	0	0.001	0.000001	3.16E-05
	Full Access Control				
	Unknown	0.015	0.339	0.097709	0.047289
	Two-way Divided positive median barrier	0	0	0	0
	Two-way Divided unprotected median	0.099	0.817	0.325154	0.111369
	Two-way not divided				
Corrosive materials	Intercept	0	0	0	0
	Road Surface Wet	0	0	0	0
	Road Surface Dry				
	Adverse Weather	0.003	0.412	0.058392	0.04939
	Good Weather				
	Dark	0	0	0	0
	Daylight				
	Bus or Unknown	0	0	0	0
	Single-Unit Truck	0	0	0	0
	Long Combination	0	0	0	0
	Tractor/Semitrailer				
	Unknown	0.001	0.443	0.087084	0.072054
	No Access Control	0.02	0.861	0.258484	0.150773
	Partial Access Control	0	0.121	0.017246	0.016715
	Full Access Control				
	Unknown	0	0	0	0

Table 2.18 Simulated significance values for determinants (cont'd.)

	Two-way Divided positive median barrier	0	0.016	0.001205	0.001777
	Two-way Divided unprotected median	0	0.084	0.009816	0.00956
	Two-way not divided				
Miscellaneous	Intercept	0	0	0	0
	Road Surface Wet	0	0.007	0.000717	0.000802
	Road Surface Dry				
	Adverse Weather	0.166	0.997	0.487352	0.145327
	Good Weather				
	Dark	0.091	0.989	0.483815	0.151752
	Daylight				
	Bus or Unknown	0	0	0	0
	Single-Unit Truck	0	0	0	0
	Long Combination	0	0.003	0.000051	0.00025
	Tractor/Semitrailer				
	Unknown	0.051	0.662	0.283616	0.11092
	No Access Control	0.072	0.866	0.295816	0.120711
	Partial Access Control	0.001	0.041	0.009097	0.006035
	Full Access Control				
	Unknown	0.018	0.384	0.1194	0.055358
	Two-way Divided positive median barrier	0.176	0.995	0.553667	0.166073
	Two-way Divided unprotected median	0.002	0.104	0.019603	0.013378
	Two-way not divided				

2.2.2 Results of the Multinomial Logistic Regression for Crashes with and without Hazardous Material Releases

The summary of the Monte-Carlo simulation technique is presented in table 2.19. The following are the highlights of the results. In all the cases, the comparison material class was *None*—that is, crashes with hazardous material releases were compared to crashes without a release.

Road Surface Conditions

- Road surface conditions were not a factor influencing the likelihood of a release of Explosives, Toxic, or Radioactive materials, Compressed Gas, or Oxidizers. That is, road surface conditions did not increase or decrease the likelihood that these materials would be released in the event of a crash.
- There was a significant reduction in the likelihood of Flammable Liquids, Corrosive Materials, and Miscellaneous Materials being released in a crash when the road surface was wet. That is, in the occurrence of a crash, wet surfaces made it less likely that there would be a release of these materials.

Weather Conditions

- The only release that was affected differently by adverse weather was Compressed Gas. Compressed Gas was 1.5 times less likely to be released in adverse, rather than fair, weather conditions.

Light Conditions

- The release of Explosives, Toxic, or Radioactive materials, Flammable Liquids, Oxidizers, and Corrosive Materials was 1.8, 1.4, 2.8, and 1.5 times more likely to occur in conditions of darkness.

Vehicle Configuration (compared to tractors/semitrailers)

- When comparing vehicle configuration in terms of releases, a wide pattern of results occurred that was linked to the physical properties of the material being transported: Explosives and Toxic or Radioactive Material were less likely to be released from unrecorded or single-unit trucks than they were from tractors or semitrailers.

- Compressed Gas was less likely to be released from unrecorded vehicles and more likely to be released from single-unit trucks in comparison to tractors/semitrailers.
- Flammable Liquids were less likely to be released from unrecorded or single-unit trucks than tractors/semitrailers.
- Oxidizers were less likely to be released from unrecorded trucks and 2.9 times more likely to be released from long combination vehicles than from tractors/semitrailers.
- Compared to tractors/semitrailers, Corrosive Materials were less likely to be released from unrecorded or single-unit trucks and 4.3 times more likely to be released from long combination vehicles.
- In comparison to tractors/semitrailers, the Miscellaneous Material classification was less likely to be released from unrecorded or single-unit trucks, but was 2.1 times more likely to be released from long combination trucks.

Access Control (compared to roads with full access control)

- Releases of Explosives, Toxic, and Radioactive materials, Compressed Gas, Corrosive Materials, and Miscellaneous Material releases were not distinguishable based on the access control of the roadway.
- Flammable Liquids were 1.4 times less likely to be released on roadways with partial access control.
- Oxidizers were 2.2 times more likely to be released on roadways with partial access control.

Divided Travel Way (compared to two-way roads that were not divided)

- Explosives, Toxic, and Radioactive Materials were 2.5 times more likely to be released on two-way, divided, unprotected median travel ways.

- Compressed Gas, Flammable Liquids, and Oxidizers were less likely to be released on two-way, divided, positive median barrier travel ways, in comparison to two-way, non-divided travel ways.
- Corrosive Materials were 3.1 times more likely to be released in crashes on one-way or unknown travel ways.

Table 2.19 Results of the multinomial logistic regression for crashes with and without releases

Hazardous Material Class	**Determinant**	**Betas**				**Sig**	**Exp. (Beta)**
		Minimum	Maximum	Mean	St Dev	Mean	
Explosives, Toxic, Radioactive	Intercept	-3.620	-3.422	-3.525	0.033	0.0000	0.029
	Road Surface Wet	-0.566	-0.323	-0.451	0.035	0.0226	-
	Road Surface Dry	.	.	.	.		
	Adverse Weather	-0.297	-0.072	-0.176	0.036	0.3661	-
	Good Weather	.	.	.	.		
	Dark	0.488	0.670	0.574	0.026	0.0000	1.775
	Daylight	.	.	.	.		
	Bus or Unknown	-0.809	-0.576	-0.698	0.033	0.0005	0.497
	Single-Unit Truck	-0.927	-0.737	-0.833	0.028	0.0000	0.435
	Long Combination	0.256	0.695	0.476	0.072	0.0741	-
	Tractor/Semitrailer	.	.	.	.		
	Unknown	-0.052	0.199	0.083	0.039	0.6466	-
	No Access Control	0.135	0.342	0.238	0.036	0.1369	-
	Partial Access Control	-0.341	-0.095	-0.225	0.040	0.3048	-
	Full Access Control	.	.	.	.		
	Unknown	0.441	0.659	0.554	0.035	0.0027	-

Table 2.19 Results of the multinomial logistic regression for crashes with and without releases (cont'd.)

	Two-way Divided positive median barrier	-0.253	-0.039	-0.146	0.037	0.5035	-
	Two-way Divided unprotected median	0.819	1.029	0.925	0.034	0.0000	2.521
	Two-way not divided	.	.	.	.		
Compressed Gas	Intercept	-1.769	-1.568	-1.669	0.033	0.0000	0.188
	Road Surface Wet	-0.230	-0.004	-0.122	0.035	0.2617	-
	Road Surface Dry	.	.	.	.		
	Adverse Weather	-0.517	-0.292	-0.399	0.036	0.0003	0.671
	Good Weather	.	.	.	.		
	Dark	-0.184	-0.015	-0.099	0.026	0.1947	-
	Daylight	.	.	.	.		
	Bus or Unknown	-0.817	-0.601	-0.701	0.033	0.0000	0.496
	Single-Unit Truck	0.241	0.411	0.327	0.027	0.0000	1.387
	Long Combination	0.291	0.724	0.504	0.070	0.0047	-
	Tractor/Semitrailer	.	.	.	.		
	Unknown	-0.143	0.097	-0.021	0.038	0.7276	-
	No Access Control	-0.061	0.152	0.049	0.034	0.5847	-
	Partial Access Control	-0.439	-0.160	-0.311	0.042	0.0124	-
	Full Access Control	.	.	.	.		
	Unknown	0.109	0.330	0.220	0.034	0.0168	-
	Two-way Divided positive median barrier	-0.718	-0.507	-0.608	0.036	0.0000	0.545

Table 2.19 Results of the multinomial logistic regression for crashes with and without releases (cont'd.)

	Two-way Divided unprotected median	-0.267	-0.061	-0.164	0.034	0.0992	-
	Two-way not divided	.	.	.	.		
Flammable Liquids	Intercept	-0.052	0.141	0.042	0.034	0.4672	-
	Road Surface Wet	-0.689	-0.455	-0.579	0.036	0.0000	0.560
	Road Surface Dry	.	.	.	.		
	Adverse Weather	-0.036	0.187	0.079	0.037	0.2754	-
	Good Weather	.	.	.	.		
	Dark	0.237	0.421	0.321	0.026	0.0000	1.379
	Daylight	.	.	.	.		
	Bus or Unknown	-2.328	-2.117	-2.218	0.033	0.0000	0.109
	Single-Unit Truck	-1.329	-1.158	-1.244	0.028	0.0000	0.288
	Long Combination	0.145	0.575	0.363	0.070	0.0019	-
	Tractor/Semitrailer	.	.	.	.		
	Unknown	-0.299	-0.058	-0.171	0.040	0.0164	-
	No Access Control	-0.214	-0.003	-0.103	0.036	0.1167	-
	Partial Access Control	-0.473	-0.231	-0.355	0.040	0.0000	0.701
	Full Access Control	.	.	.	.		
	Unknown	-0.001	0.229	0.118	0.036	0.0772	-
	Two-way Divided positive median barrier	-0.725	-0.503	-0.617	0.037	0.0000	0.539
	Two-way Divided unprotected median	-0.296	-0.080	-0.191	0.035	0.0048	-

Table 2.19 Results of the multinomial logistic regression for crashes with and without releases (cont'd.)

	Two-way not divided	.	.	.	.		
Oxidizers	Intercept	-3.927	-3.715	-3.824	0.035	0.0000	0.022
	Road Surface Wet	-0.693	-0.443	-0.559	0.037	0.0034	-
	Road Surface Dry	.	.	.	.		
	Adverse Weather	-0.259	-0.030	-0.136	0.038	0.4622	-
	Good Weather	.	.	.	.		
	Dark	0.957	1.130	1.044	0.027	0.0000	2.841
	Daylight	.	.	.	.		
	Bus or Unknown	-1.172	-0.945	-1.060	0.036	0.0000	0.347
	Single-Unit Truck	0.043	0.230	0.137	0.030	0.3042	-
	Long Combination	0.860	1.285	1.074	0.072	0.0000	2.927
	Tractor/Semitrailer	.	.	.	.		
	Unknown	0.484	0.751	0.633	0.040	0.0010	-
	No Access Control	0.456	0.652	0.550	0.036	0.0022	-
	Partial Access Control	0.654	0.932	0.795	0.043	0.0000	2.215
	Full Access Control	.	.	.	.		
	Unknown	0.149	0.376	0.263	0.036	0.0977	-
	Two-way Divided positive median barrier	-1.476	-1.251	-1.365	0.038	0.0000	0.255
	Two-way Divided unprotected median	0.036	0.257	0.157	0.036	0.3252	-
	Two-way not divided	.	.	.	.		
Corrosive materials	Intercept	-2.364	-2.135	-2.252	0.034	0.0000	0.105
	Road Surface Wet	-1.000	-0.757	-0.894	0.037	0.0000	0.409

Table 2.19 Results of the multinomial logistic regression for crashes with and without releases (cont'd.)

	Road Surface Dry	.	.	.	.		
	Adverse Weather	0.089	0.325	0.218	0.038	0.0584	-
	Good Weather	.	.	.	.		
	Dark	0.288	0.476	0.378	0.027	0.0000	1.459
	Daylight	.	.	.	.		
	Bus or Unknown	-1.446	-1.227	-1.337	0.035	0.0000	0.263
	Single-Unit Truck	-0.840	-0.658	-0.748	0.028	0.0000	0.473
	Long Combination	1.222	1.666	1.449	0.074	0.0000	4.260
	Tractor/Semitrailer	.	.	.	.		
	Unknown	-0.321	-0.076	-0.183	0.040	0.0871	-
	No Access Control	-0.228	-0.017	-0.119	0.036	0.2585	-
	Partial Access Control	-0.445	-0.189	-0.306	0.040	0.0172	-
	Full Access Control	.	.	.	.		
	Unknown	1.018	1.250	1.138	0.036	0.0000	3.121
	Two-way Divided positive median barrier	0.276	0.509	0.389	0.037	0.0012	-
	Two-way Divided unprotected median	0.194	0.404	0.305	0.035	0.0098	-
	Two-way not divided	.	.	.	.		
Miscellaneous	Intercept	-2.354	-2.153	-2.256	0.034	0.0000	0.105
	Road Surface Wet	-0.681	-0.455	-0.582	0.036	0.0007	0.559
	Road Surface Dry	.	.	.	.		
	Adverse Weather	0.009	0.222	0.113	0.037	0.4874	-
	Good Weather	.	.	.	.		

Table 2.19 Results of the multinomial logistic regression for crashes with and without releases (cont’d.)

	Dark	-0.001	0.181	0.077	0.027	0.4838	-
	Daylight	.	.	.	.		
	Bus or Unknown	-2.345	-2.126	-2.229	0.034	0.0000	0.108
	Single-Unit Truck	-1.915	-1.752	-1.834	0.028	0.0000	0.160
	Long Combination	0.540	0.958	0.754	0.069	0.0001	2.125
	Tractor/Semitrailer	.	.	.	.		
	Unknown	-0.310	-0.069	-0.176	0.040	0.2836	-
	No Access Control	0.023	0.244	0.147	0.036	0.2958	-
	Partial Access Control	-0.623	-0.383	-0.500	0.040	0.0091	-
	Full Access Control	.	.	.	.		
	Unknown	-0.367	-0.135	-0.248	0.036	0.1194	-
	Two-way Divided positive median barrier	-0.199	0.019	-0.090	0.038	0.5537	-
	Two-way Divided unprotected median	-0.476	-0.245	-0.360	0.036	0.0196	-
	Two-way not divided	.	.	.	.		

2.3 Discussion and Conclusion

2.3.1 Descriptive Analysis

The descriptive analysis using the time, day, and month of the hazardous material release incidents and truck crash reports provided several meaningful insights. An interesting commonality of the distributions of the time of crash and time of hazardous material release report was an overrepresentation of times that ended on the hour and half-hour, with a similar pattern (though not as strong), for times ending with a five. This is typical of reported times after an event occurs. Therefore, there is a strong indication that when using the time attribute as a criterion for matching, one needs to build in mechanism to compensate for the imprecision of time recordings.

Day of report for both hazardous material releases and crash reports were uniform distributions, with a minor decrease in the likelihood of the occurrence falling in the last few days of the month; it was equally likely for an incident or a crash to be reported on any day of the month over the 11-year study period.

Hazardous material releases were more frequent in the months of June, July, and August. The winter months of December and January had the fewest reported incidents. This pattern differed for crash reports—the months with the most crash reports were December and January, typically due to a higher occurrence of poor weather, increasing the likelihood of crashes. Although the pattern differed across frequency distributions, statistically significant variation shows it was unlikely for any one month to vary in its likelihood of exhibiting a reported incident or crash over time.

2.3.2 Outcome of Categorization Methods

The results of the three categorization methods showed similar outcomes. The naïve Bayesian method identified 717, the binary logistic method 894, and the neural network method identified 991 records as matches. Overall, the estimated logistic model had the ability to correctly categorize 96% of the records used in fitting, whereas the neural network correctly categorized 99% of the records in the training set. However well the methods performed on the training sets, only about 10% of the MCMIS hazardous materials crashes could be linked to the HMSAIR dataset.

2.3.3 Distinguishing Factors between Crashes with Hazardous Material Releases

The identified matching records were analyzed using a multinomial logistic regression to distinguish factors between types of crashes that released hazardous material. Highlights of the findings follow:

The impact of road surface and weather conditions did not distinguish between the releases of different materials; all releases were affected similarly across conditions. Compressed Gas was 1.8, 1.5, 3, and 1.7 times more likely to be released in non-daylight than were Explosives, Flammable Liquids, Oxidizers, Corrosive Materials, and miscellaneous materials, respectively. Vehicle configuration mapped well with expectations based on the physical properties of the hazardous materials; for example, in comparison to Flammable Liquids, Compressed Gas was 4.5 times more likely to be released from single-unit trucks than from tractors/semitrailers. One insight gleaned from the cause of release is that failure linked to the preparation of the material for transport was significant in the case of Compressed Gas, Oxidizers, and Corrosive Materials. The comparatively higher risk of release for these materials over the other materials investigated was caused by factors such as failing to secure the containers or adequately pack the material. Therefore, an intervention at this stage of transportation could significantly reduce such releases.

2.3.4 Distinguishing Factors between Crashes with and without Hazardous Material Releases

In the interest of identifying potential interventions for the event of trucks carrying hazardous materials, a comparison of crashes with and without a hazardous material release was conducted. Several factors that were commonly believed to influence truck crashes were used as explanatory variables, and hazardous material classification was used as the dependent variable in a multinomial regression. In all the cases, the comparison material class was *None*—that is, crashes with hazardous material releases were compared to crashes not having a release.

It was found that there was a significant reduction in the likelihood of a crash releasing Flammable Liquids, Corrosive Materials, and miscellaneous materials (1.8, 2.4, and 1.8 times less likely) when the road surface was wet. Weather differentially affected the release of Compressed Gas only; compared non-release crashes, Compressed Gas release was 1.5 times less likely to occur in adverse weather conditions. Compared to non-release crashes, Explosives, Toxic, and Radioactive materials, Flammable Liquids, Oxidizers, and Corrosive Materials were 1.8, 1.4, 2.8, and 1.5 times more likely to occur during darkness. When comparing releases by vehicle configuration, a wide pattern of results occurred that was linked to the physical properties of the material being transported. Two interesting results occurred for long combination trucks—in comparison to non-release crashes, Oxidizers were 2.9 times more likely and Corrosive Materials 4.3 times more likely to be released from long combination vehicles in crashes.

Two explanatory variables used in the comparison of crashes with and without hazardous material release dealt with roadway characteristics: access control, and whether or not the roadway was divided. Flammable Liquids were 1.4 times less likely to be released on roadways with partial access control, whereas Oxidizers were 2.2 times more likely to be released on roadways with partial access control. The type of roadway division had the most varied influence, and could distinguish between types of hazardous material release. It was found that Explosives and Toxic or Radioactive materials were 2.5 times more likely to be released on two-way, divided, unprotected median travel ways. Compressed Gas, Flammable Liquids, and Oxidizers were more likely (1.8, 1.8, and 3.9 times) to be released on two-way, not-divided travel ways than on two-way, divided travel ways with positive median barriers. Finally, Corrosive Materials were 3.1 times more likely to be released in crashes on one-way or unknown travel ways than on two-way, not-divided travel ways.

2.3.5 Conclusions

The study concluded that it is possible to apply competing categorization methods—specifically, naïve Bayesian, binary logistic, and multi-layer neural networks—to solve the matching problem that is created when one needs to combine two publicly accessible datasets that were not designed to be joined. In this study, the categorization methods were successful in matching only 10% of the records, which was disappointing. However, with additional effort, adding several of the text fields (e.g., location of crash/incident) or narratives would increase matching. This would require the calculation of likelihood scores based on the text recognition algorithms currently employed in matching medical records.

The finding that road surface and weather conditions were not useful in distinguishing the types of hazardous materials more likely to be released underscores the impact of driver decision-making; that is, it is possible that, when a driver is responsible for trucking hazardous materials, it is more likely that the trip will not be made on wet or in adverse weather. This conclusion is supported when comparing the truck crashes with releases in this report to those without releases—it was more likely for a truck to crash on wet surfaces when it was not carrying hazardous materials.

Evidence that helps guide interventions to reduce hazardous material releases points toward more conscious effort being placed on the packing of and preparation for transporting hazardous material. The most significant cause of the release of hazardous materials in a crash occurred due to failure to secure containers or adequately pack the material.

Previous studies report that the shift during which the driver is working impacts their safety. In this study, time of day was a significant influence on the likelihood of releasing hazardous materials. It was much more likely for releases to occur after dark than during daylight hours.

Regarding the physical features of vehicles and the roadways traveled, it is seen that, in comparison to tractor/semitrailer configurations, long combination vehicles stood out as being more likely to be involved in the release of Oxidizers and Corrosive Materials. One can conclude that when long combination vehicles carrying one of these materials are involved in a crash, there is a greater chance of a hazardous material release. This finding encourages the creation of more barriers on two-way travel ways. When trucks carrying hazardous material crossed out of their travel way, material release was highly likely. In summary, this study revealed that both human factors and physical features of vehicles and roadways contribute to the safety of our roads.

References

Battelle, Inc. 2001. "Comparative Risks of Hazardous Materials and Non-hazardous Materials Truck Shipment Accidents/Incidents." Prepered for the Federal Motor Carrier Safety Administration, U.S. Department of Transportation. Washington, D.C. http://www.fmcsa.dot.gov/documents/hazmatriskfinalreport.pdf, Final.

Bureau of Economic Analysis. n.d. "Interactive Access to Industrial Economic Data." U.S. Department of Commerce. Accessed July 30, 2012. http://www.bea.gov/iTable/iTable.cfm?ReqID=5&step=1.

Hanley, Paul F., Yu, Hai and Ma, Suyun. Under review. "A procedure for matching truck crash records with hazardous material release incident reports." *Transportation Resesrch Board 92nd Annual Meeting*. Transportation Research Board of the National Academies, Washington, D.C.

Appendix A

A.1 Descriptive Analysis for Truck Crashes

A.1.1 Truck Crashes without Hazardous Material Release

Table A.1 Road surface condition: Truck crashes without release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Wet	451468	29.7	29.7	29.7
	Dry	1067784	70.3	70.3	100.0
	Total	1519252	100.0	100.0	

Table A.2 Weather condition: Truck crashes without release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Adverse	437309	28.8	28.8	28.8
	Non-adverse	1081943	71.2	71.2	100.0
	Total	1519252	100.0	100.0	

Table A.3 Light condition: Truck crashes without release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Dark	482997	31.8	31.8	31.8
	Day Light	1036255	68.2	68.2	100.0
	Total	1519252	100.0	100.0	

Table A.4 Vehicle configuration: Truck crashes without release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	unknown & buses	222166	14.6	14.6	14.6
	Single-Unit Truck	375332	24.7	24.7	39.3
	Long Combination	39262	2.6	2.6	41.9
	Tractor/Semitrailer	882492	58.1	58.1	100.0
	Total	1519252	100.0	100.0	

Table A.5 Access control: Truck crashes without release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unknown	407590	26.8	26.8	26.8
	No Access Control	454575	29.9	29.9	56.7
	Partial Access Control	191886	12.6	12.6	69.4
	Full Access Control	465201	30.6	30.6	100.0
	Total	1519252	100.0	100.0	

Table A.6 Travel way: Truck crashes without release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unknown	412278	27.1	27.1	27.1
	Two-way, divided, positive median barrier	299423	19.7	19.7	46.8
	Two-way, divided, unprotected median	291657	19.2	19.2	66.0
	Two-way, not divided	515894	34.0	34.0	100.0
	Total	1519252	100.0	100.0	

A.1.2 Truck Crashes Hazardous Material Release

The records for truck crashes with hazardous material release are the result of the logistic categorization method described above. The records are weighted based on the likelihood values of a match between the HMSAIR and MCMIS datasets. The weighted crash total was constrained by the total number of crashes reported in MCMIS that had released hazardous materials.

Table A.7 Hazardous material classification: Truck crashes with release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Explosives, Toxic, Radioactive	324	4.0	4.0	4.0
	Compressed Gas	1259	15.4	15.4	19.4
	Flammable Liquids	4717	57.7	57.7	77.1
	Oxidizers	359	4.4	4.4	81.4
	Corrosive materials	1068	13.1	13.1	94.5
	Miscellaneous	449	5.5	5.5	100.0
	Total	8176	100.0	100.0	

Table A.8 Road surface condition: Truck crashes with release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Wet	1779	21.8	21.8	21.8
	Dry	6397	78.2	78.2	100.0
	Total	8176	100.0	100.0	

Table A.9 Weather condition: Truck crashes with release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Adverse	1996	24.4	24.4	24.4
	Non-adverse	6180	75.6	75.6	100.0
	Total	8176	100.0	100.0	

Table A.10 Light condition: Truck crashes with release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Dark	3164	38.7	38.7	38.7
	Day Light	5012	61.3	61.3	100.0
	Total	8176	100.0	100.0	

Table A.11 Vehicle configuration: Truck crashes with release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	unknown & buses	315	3.9	3.9	3.9
	Single-Unit Truck	1214	14.8	14.8	18.7
	Long Combination	518	6.3	6.3	25.0
	Tractor/Semitrailer	6129	75.0	75.0	100.0
	Total	8176	100.0	100.0	

Table A.12 Access control: Truck crashes with release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unknown	2228	27.3	27.3	27.3
	No Access Control	2531	31.0	31.0	58.2
	Partial Access Control	855	10.5	10.5	68.7
	Full Access Control	2562	31.3	31.3	100.0
	Total	8176	100.0	100.0	

Table A.13 Traffic way: Truck crashes with release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unknown	2587	31.6	31.6	31.6
	Two-way, divided, positive median barrier	1233	15.1	15.1	46.7
	Two-way, divided, unprotected median	1623	19.9	19.9	66.6
	Two-way, not divided	2733	33.4	33.4	100.0
	Total	8176	100.0	100.0	

Table A.14 How release occurred: Truck crashes with release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Other	2852	34.9	34.9	34.9
	Burst or Ruptured	1190	14.6	14.6	49.4
	Cracked	1660	20.3	20.3	69.7
	Punctured	2474	30.3	30.3	100.0
	Total	8176	100.0	100.0	

Table A.15 Cause of release: Truck crashes with release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Other	1262	15.4	15.4	15.4
	Human	706	8.6	8.6	24.1
	Rollover	4651	56.9	56.9	81.0
	Crash	1557	19.0	19.0	100.0
	Total	8176	100.0	100.0	