UMTRI-2010-39 SEPTEMBER 2010

A NEW MODEL OF CRASH SEVERITIES REPORTABLE TO THE MCMIS CRASH FILE

PAUL E. GREEN
DANIEL BLOWER

UMTRI-2010-39

A New Model of Crash Severities Reportable to the MCMIS Crash File

Paul E. Green
Daniel Blower

The University of Michigan
Transportation Research Institute
Ann Arbor, MI 48109-2150
U.S.A.

September 2010

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UMTRI-2010-39		
4. Title and Subtitle A New Model of Crash Severities Reportable to the MCMIS Crash File		5. Report Date September 2010 6. Performing Organization Code
7. Author(s) Green, Paul E., and Blower, Daniel		8. Performing Organization Report No. UMTRI-2010-39
9. Performing Organization Name and Address The University of Michigan Transportation Research Institute 2901 Baxter Road Ann Arbor, Michigan 48109-2150 U.S.A.		10. Work Unit no. (TRAIS) 052702 11. Contract or Grant No. DTMC75-06-H-00003
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Motor Carrier Safety Administration 400 Seventh Street, SW Washington, D.C. 20590		13. Type of Report and Period Covered Special report 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

The Motor Carrier Management Information System (MCMIS) Crash file has been developed by the Federal Motor Carrier Safety Administration (FMCSA) to serve as a census file of trucks and buses involved in traffic crashes meeting a specific crash severity threshold. Each state is responsible for identifying cases that meet the MCMIS Crash file criteria and reporting the required data through the SafetyNet system. The present report is an addition to three previous reports describing models to predict the number of crash involvements a state should be reporting. The model has been updated and changed over time as more data becomes available from additional states.

In each state, the number of fatal involvements is well-known, so all states will start with a known quantity, the number of fatal truck and bus crash involvements. The new model also incorporates a rural/urban (RU) factor that accounts for the relative proportion of rural to urban truck travel in a state.

In the new model, data from 16 states that provide all the information necessary to identify MCMIS-reportable cases were used. A log-linear model is fit to MCMIS data for the states that have information recorded for both fatal and nonfatal crashes. The model is then used to predict the number of nonfatal crashes for a new state in which the number of fatal crashes and the RU factor are known. Ninety percent prediction intervals provide a range of nonfatal crash values to be used for guidance. The new model provides more accurate prediction than previous models and is expected to be updated as data from additional states become available.

17. Key Words		18. Distribution Statement	
MCMIS, crash severity, crash ratios		Unlimited	
19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	26	

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
^oC	Celsius	1.8C+32	Fahrenheit	^oF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

Table of Contents

List of Tables

List of Figures

A New Model of Crash Severities Reportable to the MCMIS Crash File

1 Introduction

This report is part of a series of reports developed by the University of Michigan Transportation Research Institute (UMTRI) to assist the Federal Motor Carrier Safety Administration (FMCSA) in its efforts to improve truck and bus crash data reporting by the States. One part of the effort is to evaluate the completeness and accuracy of reporting by the States to the Motor Carrier Management Information System (MCMIS) Crash file. UMTRI has issued a series of evaluations of individual States.[1] A separate part of the effort is to develop a method of predicting the number of total reportable cases from each State, to help individual States estimate the number of cases they should be reporting, based on the number of fatal involvements in the state, which is known with good reliability. The present work reports on further development of the methodology to predict total reportable crash involvements qualifying for the MCMIS Crash file from the number of fatal involvements. Sixteen States with suitable crash data have been evaluated since the last report on the ratio. Data from those States are incorporated into the methodology, along with a new factor that measures the ratio of rural to urban commercial vehicle travel, and a new prediction equation is developed. This report represents a further development of the previous reports.[1, 2, 3]

The MCMIS Crash file has been created and compiled by the Federal Motor Carrier Safety Administration (FMCSA) to serve as a census file of trucks and buses involved in traffic crashes meeting a specific crash severity threshold. FMCSA maintains the MCMIS file to support its mission to reduce crashes, injuries, and fatalities involving large trucks and buses. Designing effective safety measures requires accurate and complete crash data to understand the dimensions of the crash problem. The data are used to monitor the safety performance of carriers and to identify crash safety trends. The usefulness of the MCMIS Crash file depends upon individual states transmitting a standard set of data items on all trucks and buses involved in traffic crashes that meet the file's crash severity threshold.

The University of Michigan Transportation Research Institute (UMTRI) has prepared a series of reports evaluating the completeness of reporting from selected states. As of this report, UMTRI has completed 37 evaluations, covering 35 states. Reporting rates have ranged from under ten percent of reportable cases to over eighty percent.

Each state is responsible for identifying cases that meet the MCMIS Crash file criteria and reporting the required data through the SafetyNet system. The MCMIS selection criteria are clearly defined and, in theory, easily applied. To comply with the reporting requirements, states have modified and in some cases changed existing crash data collection systems, (which were developed for their own safety and enforcement purposes), to identify and capture the correct cases. Each state has for its own crash file, its own threshold for reportable crashes, its own

[1] These State reports may be found at the MCMIS Evaluation Reports at this website: http://www.umtri.umich.edu/divisionPage.php?pageID=308.

system for classifying vehicles, and its own set of data elements to collect on the crashes. Given the multiple purposes and systems for and by which crash data are collected by states, it is not necessarily an easy matter to identify which crashes should be reported to the MCMIS Crash file and which should not. This is especially true for crashes that do not involve a fatality. Many states do not regularly collect the specific information needed to determine if a nonfatal case is reportable under the MCMIS criteria—e.g., whether an injured person was transported for medical attention or whether a vehicle was towed due to disabling damage.

Because of the sheer complexity of managing crash data systems that may have hundreds of thousands of records, there is often no easy way for the states, or FMCSA, to know if the right number of cases is being reported. Some states may have thought that they were in compliance and fully reporting to the MCMIS Crash file, but the UMTRI evaluation found significant underreporting. Until all states can be directly evaluated, a method of predicting, within a reasonable range, the number of cases that each state should report can serve as a guidepost or benchmark for the states on where they stand. Such a benchmark can alert a state that a problem exists, and motivate a process to identify a solution.

This paper presents a refinement of a method to estimate the number of involvements each state should report to the MCMIS Crash file, even if the state's data system cannot readily identify all the crashes that meet the Crash file criteria. It is based on developing a ratio of fatal to nonfatal reportable crashes, using data from states that have sufficiently complete information to identify reportable cases with good confidence. In any given state, the number of fatal involvements of the vehicles that meet the MCMIS reporting criteria is generally well-known, so all states should be able to start with a known quantity, the number of fatal truck and bus crash involvements. It is then hypothesized that a ratio of fatal involvements to nonfatal reportable involvements exists that is independent of any state's data system and that will apply across all the states. If this is true, it is possible, knowing the number of fatal involvements that occurred in a state, to predict the number of involvements of lesser severity, and thus predict the total number of cases that any state should report to the MCMIS Crash file.

The process of evaluating state reporting identified a number of states whose data systems provide the data necessary to apply the MCMIS reporting criteria completely. Virtually all states can identify trucks and buses reasonably well, and all states can identify fatal crashes cleanly. However, most states do not regularly collect the information needed to identify reportable nonfatal crashes: crashes in which an injured person was transported for immediate medical attention or crashes in which at least one vehicle was towed due to disabling damage. States identify persons injured in a crash, and even nominally use the same system to classify injury severity, but not all have taken the next step to capture if the person was transported for medical attention. Similarly, many states record if a vehicle was towed, but not whether the reason for the tow was disabling damage.

UMTRI reviewed all the state evaluations done to date and identified sixteen states whose existing crash data systems can identify all levels of the reporting criteria. In comparison with the prior models of reporting, most of the older reports were dropped, and several new states were added to the observations used in developing the model.

In addition to the new set of observations from the state data, a new factor is used in the model, the ratio of rural to urban commercial vehicle (CMV) travel in a state. This factor was proposed by David Hetzel and Joanne Zhou of the National Institute for Safety Research, Inc. (NISR). The purpose of the adjustment is to account for any variation in the ratio of fatal to nonfatal reportable crash involvements that may be due to variations in the proportion of rural (or urban) CMV travel in a state. Some states with a high proportion of rural truck travel have expressed concerns that the previous model overestimated the number of reportable cases in their states. The inclusion of this factor is meant to address this concern.

The state data, in combination with the rural/urban factor (RU), can be used to model the ratio of reportable crash severities, in this case, the ratio of fatal involvements to nonfatal involvements. [See evaluations in references 4 through 19.]

The General Estimates System (GES) file from the National Highway Traffic Safety Administration (NHTSA) may be considered a possible alternative source of crash data for this purpose. GES is a nationally-representative sample of police-reported crashes. GES cases are sampled from primary sampling units (PSUs) around the country and a standard set of data are coded from the sampled police-reports. The GES data can be used to identify trucks and buses, and crash detail includes whether an injured person was transported for immediate medical attention or a vehicle towed due to disabling damage. Thus, GES nominally can be used to cleanly estimate the number of reportable involvements to the MCMIS Crash file.

However, there are a number of reasons why the GES file is not satisfactory for this application. Though the GES variable recording whether a vehicle was towed has a level for towed due to damage, we know through doing the state reporting evaluations that many state police reports do not include the information to determine the reason for the tow, so this variable likely underestimates the number of vehicles towed due to disabling damage. In addition, the GES file is a sample drawn through a complex stratified, hierarchical sampling system. Truck and bus crashes are a small sample, relative to automobile crashes. The standard errors for small subsets of the file, such as trucks, are relatively large. It is also known that GES underestimates the number of fatal truck and bus involvements. This can be illustrated for truck fatal involvements. For the years 2003 through 2007, GES estimated 2,738 to 5,694 trucks involved in a fatal crash, with an average of about 3,700. The average from UMTRI's Trucks Involved in Fatal Accidents (TIFA) file is over 5,200. The standard deviation of the counts from TIFA is 129, while it is 1,151 for counts from GES. Moreover, the 95 percent confidence interval for an estimate in GES of 3,700 fatal truck involvements is ±1,700, or about 45 percent. The TIFA file is a census file. Since the basis of the ratio is the number of fatal involvements, which it is assumed can be identified reasonably well in state crash data, the GES estimate is too uncertain to be reliable.

2 Problem statement

The purpose of this report is to describe the development of a method that can be used to assist states in determining if they are in compliance with FMCSA's requirement to report all crash involvements reportable to the MCMIS Crash file. This method is not intended to identify the precise number of reportable cases for each state, but to give guidance as to whether a state's reporting is within a reasonable expected range. It is assumed that states can identify fatal involvements with acceptable confidence. Virtually all the states can identify trucks and buses readily. Crashes in which a fatality occurred are equally clearly identifiable. All states identify fatal injuries, and the definition of a fatal traffic accident—death within 30 days of the crash—is standard. Accordingly, one level of the hypothesized ratio, i.e., fatal involvements, should be well established in all states.

It is assumed that the relationship of fatal to nonfatal reportable involvements exists independent of any particular state system. That is, the ratio does not depend on a state's definitions or system of collecting data, so the ratio established in one set of states should hold true for other states. As a counter-example, consider the common system for classifying injury severity. Most states use the KABC0 system, which classifies injuries as fatal, incapacitating, non-incapacitating but evident, complaint of pain, and no injury. Fatal injuries are clear and not subject to much interpretation. But the other injury levels are more difficult to classify consistently and can be subject to more interpretation. As a consequence, states vary widely in the relative proportions of A-, B-, and C-injuries.[2] The differing proportions are related not to some underlying difference in the severity of crashes in different states but to variations in the interpretation and application of standard, widely-accepted definitions.

In contrast, the MCMIS Crash file criteria do not depend on crash severity standards that are known to be applied unevenly, but instead provide a relatively simple definition that should apply in roughly the same way everywhere. Reportable nonfatal involvements include either an injury transported for immediate medical attention or a vehicle towed due to disabling damage. An injury serious enough to be transported for treatment in Maine likely would also be transported if it occurred in California. There may be some variations from state to state, but they are not expected to be large. Similarly, it is not expected that whether a vehicle is disabled enough to be towed will vary much by region. There may be areas where towing following a crash is more common, but less variation, by state, in judging whether a vehicle has suffered *disabling* damage.

In this way, the choice of criteria for the MCMIS Crash file is particularly astute, specifically because the criteria do not depend on how a state may define an injury severity level or train their officers to identify it.

If it is true that there is a fundamental relationship between fatal and nonfatal (injury/transported and towed/disabled) involvements, then the ratio can be discovered by examining state crash files with the information to apply each of the MCMIS reporting criteria. In this process, each state and year of data is one observation, one observation of the underlying ratio. By assembling such observations and fitting a statistical model, it is possible to estimate the true ratio of crash severities that applies across states.

[2] See O'Day, J., *Accident Data Quality*. National Cooperative Highway Research Program Synthesis of Highway Practice, No. 192. Federal Highway Administration, Washington, DC, 1993. O'Day found that the proportion of A injuries varied from 4.9% to 23.8% in a sample of about 20 states. The findings were for 1990-1991 data, but illustrate the point.

This statistical model will allow states to estimate the number of cases that they should be reporting to the MCMIS Crash file, with some margin of error. It is assumed that each state, and FMCSA, will know the correct number of fatal involvements. The model will then predict the number of nonfatal (injury/transported plus towed/disabled) involvements that the number of fatal involvements implies.

The earlier reports presented different methods using data available at the respective times. The first approach was a simple linear regression model that predicted both injury/transported and towed/disabled counts from the number of fatal involvements.[1] That model fit the data well statistically but produced prediction ranges that were large and did not predict well back to the original data. In the second and third reports, log-linear models were developed to predict the nonfatal involvements as a whole, rather than broken down into injured/transported and towed/disabled involvements.[2, 3] The second model used more observations than the first, and the third used more than the second. The current model revisits the selection of states afresh, and includes a total of 16 state observations. The third model weighted the observations by the count of fatal involvements, which gives more weight to the large states, but the current model drops that weighting and instead adds a factor to adjust for the proportion of rural truck travel.

3 Data

States that collect all the detail necessary to reproduce the MCMIS Crash file reporting criteria were selected for modeling the distribution of reportable cases. The essential criteria for reporting are displayed in Table 1. Adequate methods of identifying trucks and buses could be developed for all the states evaluated to date (August 2010), with some qualifications. In some states, light vehicles displaying hazardous materials placards were not identified with high confidence, but the number of such vehicles is so small relative to the number of trucks and buses that it should have only an insignificant impact on the analysis.

Table 1 Vehicle and Crash Severity Threshold for MCMIS Crash File

Vehicle	Truck with GVWR over 10,000 or GCWR over 10,000, or Bus with seating for at least nine, including the driver, or Vehicle displaying a hazardous materials placard.
Accident	Fatality, or Injury transported to a medical facility for immediate medical attention, or Vehicle towed due to disabling damage.

Identifying crashes that meet the reporting criteria is the crux of the problem in estimating reportable cases. Table 2 shows the states that were selected for this problem. In all these states, determining the number of reportable fatal involvements can be done fairly cleanly and with minimal ambiguity. Most of the states directly coded the detail needed to identify the different crash severities. An indicator that an injured person was transported for immediate medical attention was critical for selection. The experience of the UMTRI state evaluations showed that whether an injured person was transported for care does not map cleanly to coded injury severities, so injury severity could not be used as a surrogate. With respect to the towed/disabled criteria, the method some states used to code vehicle damage severity can be used as a substitute for a direct indicator that a vehicle was towed due to disabling damage. The severity scale employed by the states here directly indicates whether a vehicle was disabled. This is not a perfect substitute for towed/disabled, but it is a reasonable surrogate.

Table 2 States Selected for Modeling the Ratio of Crash Severities

State	Data Year	Injured	Transported	Towed	Disabled
Alabama	2005	Yes	Yes	Yes	Yes
Arizona	2005	Yes	Yes	Yes	Yes
Georgia	2006	Yes	Yes	Yes	Yes
Idaho	2006	Yes	Yes	Yes	Yes
Iowa	2004	Yes	Yes	No*	No*
Kentucky	2006	Yes	Yes	Yes	Yes
Louisiana	2005	Yes	Yes	Yes	Yes
Minnesota	2007	Yes	Yes	Yes	Yes
Nebraska	2005	Yes	Yes	Yes	Yes
North Dakota	2008	Yes	Yes	Yes	Yes
Ohio	2005	Yes	Yes	Yes	Yes
Oklahoma	2007	Yes	Yes	Yes	Yes
South Carolina	2006	Yes	Yes	Yes	No*
South Dakota	2005	Yes	Yes	Yes	Yes
Tennessee	2004	Yes	Yes	Yes	Yes
Wisconsin	2006	Yes	No	Yes	Yes

* Vehicle damage severity used as surrogate for towed due to damage

Table 3 shows the data used in modeling the crash severity ratio. Each state and crash year is one observation. For the purposes of the model, it is desirable that the data used in the model cover the range in the number of cases expected to be reported from the fifty states. The observations in the model provide good coverage of the size range of the states. Several relatively small states like North Dakota, South Dakota and Idaho are included, along with one relatively large state, Georgia. The three largest states—California, Florida, and Texas—were not included in the current model. Most of the states available, however, fall into the range between 125 and 200 annual fatalities. There is good coverage of the very small states, but the very largest states are underrepresented in the data available.

Table 3 also shows the values of the *RU* factor used in the modeling. RU is the ratio of rural to urban CMV travel within a state. The data to calculate the RU factor are taken from Table PS-1 from the Federal Highway Administration's annual *Highway Statistics* publication. Table PS-1 is titled "Selected Measures for Identifying Peer States, and includes a number of general and transportation measures.[3] The table provides estimates of urban and rural travel and the

[3] David Hetzel and Joanne Zhou of NISR, Inc., pointed us to this resource.

proportion of each accounted for by commercial motor vehicles (CMVs).[4] The proportion of CMVs is based on estimates from sampled segments in the Highway Performance Monitoring System data. To determine the RU factor for each state, as shown in Table 3, we first calculate the rural and urban VMT for CMVs and then take the ratio. Values above 1.0 indicate that the state has more rural CMV travel than urban.

Table 3 Data Used in Modeling Crash Severity Ratio

Year	State	Fatal	Nonfatal	Total	RU factor
2005	Alabama	128	4,383	4,511	1.78
2005	Arizona	128	4,283	4,411	0.80
2006	Georgia	260	8,804	9,064	0.79
2006	Idaho	26	870	896	1.84
2004	Iowa	68	1,974	2,042	2.73
2006	Kentucky	119	4,190	4,309	2.08
2005	Louisiana	147	4250	4397	1.11
2007	Minnesota	98	2,878	2,976	1.36
2005	Nebraska	48	1,173	1,221	5.81
2008	North Dakota	15	421	436	5.93
2005	Ohio	205	8,840	9,045	1.10
2007	Oklahoma	110	3,364	3,474	3.24
2006	South Carolina	102	3,260	3,362	1.59
2005	South Dakota	19	437	456	6.33
2004	Tennessee	154	5,892	6,046	1.44
2006	Wisconsin	95	3,773	3,868	1.66

How well do the observations cover the range of States? UMTRI's Trucks Involved in Fatal Accidents (TIFA) and Buses Involved in Fatal Accidents (BIFA) can be used to provide an accurate distribution of the annual expected reportable truck and bus fatal involvements for each state. The two files include all truck and bus fatal involvements that are reportable to the MCMIS Crash file, since the definitions of a reportable truck or bus are compatible with the MCMIS definitions. Only light vehicles transporting hazardous materials that require a placard are not included. Since there are only a small number of such vehicles each year, they would not affect the overall distribution.

Figure 1 shows the distribution of states by the annual average number of truck and bus fatal involvements. The bars show the number of states with counts of fatal truck and bus involvements in different bins. The boxed numbers above some of the bars show the number of states used in the current model in each bin. There are about ten states each year with 25 or fewer MCMIS-reportable fatal involvements, eight with between 26 and 50, and six with between 51 and 75. Almost 60 percent of the states have 100 or fewer fatal involvements annually. The available states cover the range reasonably well, with good coverage of the smaller states and

[4] Up to 2005, the estimate included both trucks and buses. In 2006, the estimate was changed to include only trucks.

excellent coverage of the middle-sized states. The middle two-thirds of states range from 51 to 200 fatal involvements and 11 of the observations in the model are from those 32 states. The top 20 percent of states is somewhat underrepresented, with no observations from the three largest states However, the range between 100 and 225, which includes 17 states, is covered well with nine observations.

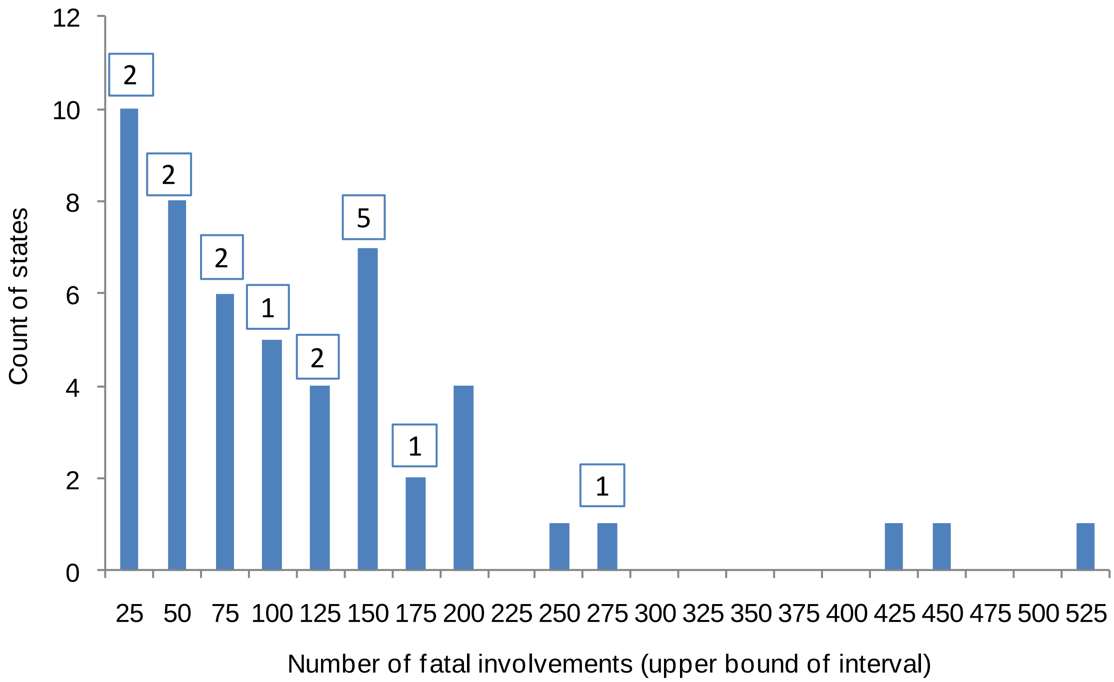

Figure 1 Counts of States by Average Annual Number of Fatal Truck and Bus Involvements TIFA, BIFA 2003-2007

Note also in Figure 1 that there are three states which annually average over 400 fatal involvements, and none between 300 and 400.

The Appendix includes a figure that illustrates for the RU factor how well the states used in the model cover the range of RU in the whole number of states. It is not shown here in the interest of space. The states in the model cover the range of RU in the states quite well, except for the eight states with an RU less than 0.5 and the one state (Montana) with an RU over nine.

4 Model and Methods

The goal is to predict the number of nonfatal crashes given the number of fatal crashes and the RU factor for a state. The RU factor is available for all states as published by the Federal Highway Administration. Ninety-percent prediction intervals for the estimates are also desired. To accomplish this goal, a log-linear model is fit to MCMIS data for sixteen states that have information recorded for both fatal and nonfatal crashes. The model is then used to predict the number of nonfatal crashes for a new state in which only the number of fatal crashes and the RU factor are known. Prediction intervals are presented, instead of confidence intervals, because data from the state to be predicted were not used to estimate the regression line in the modeling process. Since there is more uncertainty in predicting nonfatal crash involvements for a state not used in the modeling process, prediction intervals are wider than confidence intervals. Table 4 shows the data presented in Table 3 for the sixteen states, with the addition of the logs of the fatal and nonfatal counts. Log counts refer to the *natural* logarithm using base *e*.

Table 4 Data Used in the Modeling Process

State	Fatal	Nonfatal	Log fatal	Log nonfatal	RU factor
Alabama	128	4,383	4.852	8.385	1.778
Arizona	128	4,283	4.852	8.362	0.796
Georgia	260	8,804	5.561	9.083	0.794
Idaho	26	870	3.258	6.768	1.839
Iowa	68	1,974	4.220	7.588	2.731
Kentucky	119	4,190	4.779	8.340	2.078
Louisiana	147	4,250	4.990	8.355	1.114
Minnesota	98	2,878	4.585	7.965	1.365
Nebraska	48	1,173	3.871	7.067	5.815
North Dakota	15	421	2.708	6.043	5.931
Ohio	205	8,840	5.323	9.087	1.105
Oklahoma	110	3,364	4.700	8.121	3.241
South Carolina	102	3,260	4.625	8.089	1.594
South Dakota	19	437	2.944	6.080	6.330
Tennessee	154	5,892	5.037	8.681	1.437
Wisconsin	95	3,773	4.554	8.236	1.659

The decision to analyze the data on the log scale is based on a scatterplot of the log of the nonfatal crashes by the log of the fatal crashes. The scatterplot shows a strong linear association between the logs of the two variables, with a correlation greater than 0.98. Crash counts are often in the thousands, and crash data are often analyzed using log-linear models.

The model for analyzing the data shown in Table 4 takes the form

$$\log y_i = \beta_0 + \beta_1 \log\left[x_{1i}\left(1+\beta_2\sqrt{1/x_{2i}}\right)\right] + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2), \qquad i = 1,\ldots,16$$

where y_i is the number of nonfatal crashes for state i, x_{1i} is the number of fatal crashes, x_{2i} is the RU factor, β_0 and β_1 are the intercept and slope parameters, respectively, β_2 is a multiplier parameter for the RU factor, and ε_i are the error terms. The random component of the model is contained in the error terms. These terms are modeled as normal random variables with mean 0 and variance σ^2. Because of the β_2 parameter attached to the RU factor, the model as shown above is a nonlinear model. Based on the fit of a nonlinear model, the estimate for β_2 is 0.35. When β_2 is known, the model reduces to a linear model. We prefer to treat $\beta_2 = 0.35$ as known and then fit a linear model to estimate β_0 and β_1.

After fitting this model, the estimated equation is

$$\log \hat{y}_i = 2.94 + 1.06 \log\left[x_{1i}\left(1 + 0.35\sqrt{1/x_{2i}}\right)\right].$$

Figure 2 shows a scatterplot the fitted regression line, and a 90 percent prediction band for the model using data from the sixteen states. The linear association is very strong with a correlation greater than 0.98. All points fall within the 90 percent prediction band. The model can now be used to predict the number of reportable nonfatal involvements for a new state, given the number of fatal involvements and the RU factor for that particular state.

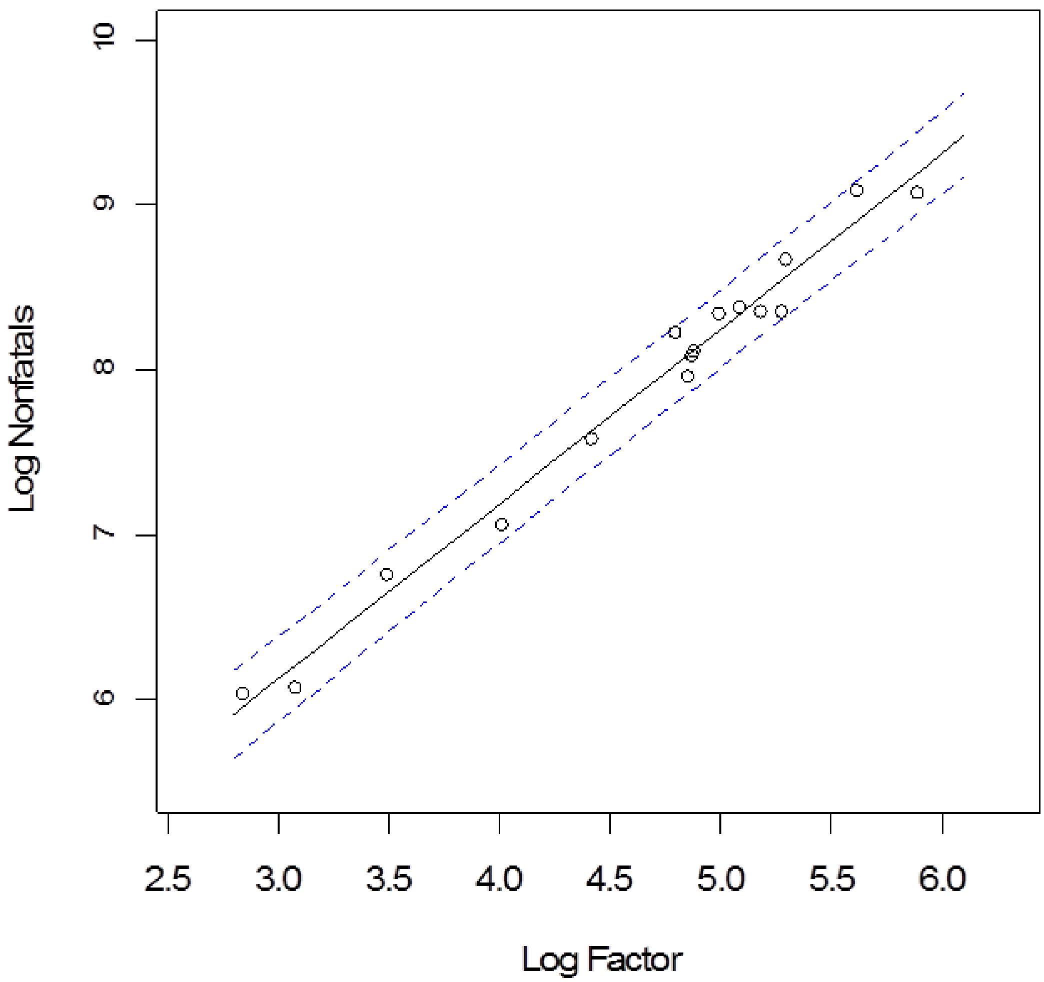

Figure 2 Scatterplot, Fitted Regression Line, and 90 Percent Prediction Intervals for Sixteen States

5 Application

The model can be applied to individual states to estimate reportable cases, although it is generally not valid to predict outside of the range of values covered in the model. However, the model is based on a range that covers most of the states: There are only six states[5] that average fewer than 15 truck or bus fatal involvements annually, which is the smallest observation used in the model. Three states (Florida, California, and Texas) average over 260 (the largest observation used in the model). Thus, the model covers 42 of the 51 (including the District of Columbia) units that supply data to the MCMIS Crash file.

Table 5 compares the predicted and actual values for nonfatal involvements for the states used in generating the model. Generally, the model estimates are reasonably close to the observed values. Seven of the predictions are within 10 percent of the actual, seven are between ten and 20 percent, and two (Ohio and Wisconsin) are within 20.4 percent and 22.4 percent respectively. All the actual numbers are within the 90 percent prediction interval, as shown in Figure 2.

Table 5 Comparison of Actual and Nonfatal and 90 Percent Prediction Intervals

	Nonfatal involvements		90% prediction interval	
State	Actual	Predicted	Lower bound	Upper bound
Alabama	4,383	4,199	3,322	5,308
Arizona	4,283	4,656	3,680	5,891
Georgia	8,804	9,899	7,735	12,668
Idaho	870	770	603	984
Iowa	1,974	2,054	1,627	2,594
Kentucky	4,190	3,822	3,025	4,829
Louisiana	4,250	5,150	4,066	6,523
Minnesota	2,878	3,263	2,585	4,119
Nebraska	1,173	1,337	1,055	1,693
North Dakota	421	388	299	504
Ohio	8,840	7,340	5,767	9,341
Oklahoma	3,364	3,371	2,670	4,256
South Carolina	3,260	3,342	2,647	4,219
South Dakota	437	497	385	641
Tennessee	5,892	5,236	4,133	6,633
Wisconsin	3,773	3,084	2,443	3,892

Figure 3 displays the data in Table 5 graphically. Note that all the observed values for nonfatal involvements are within the 90 percent prediction intervals. The range is relatively large for Georgia. For some states, the predictions are below the actual values while in others the predictions are higher than the actual values, but for most the predictions are very near to the observed, actual values.

[5] The District of Columbia, Rhode Island, Alaska, Hawaii, Vermont, and New Hampshire.

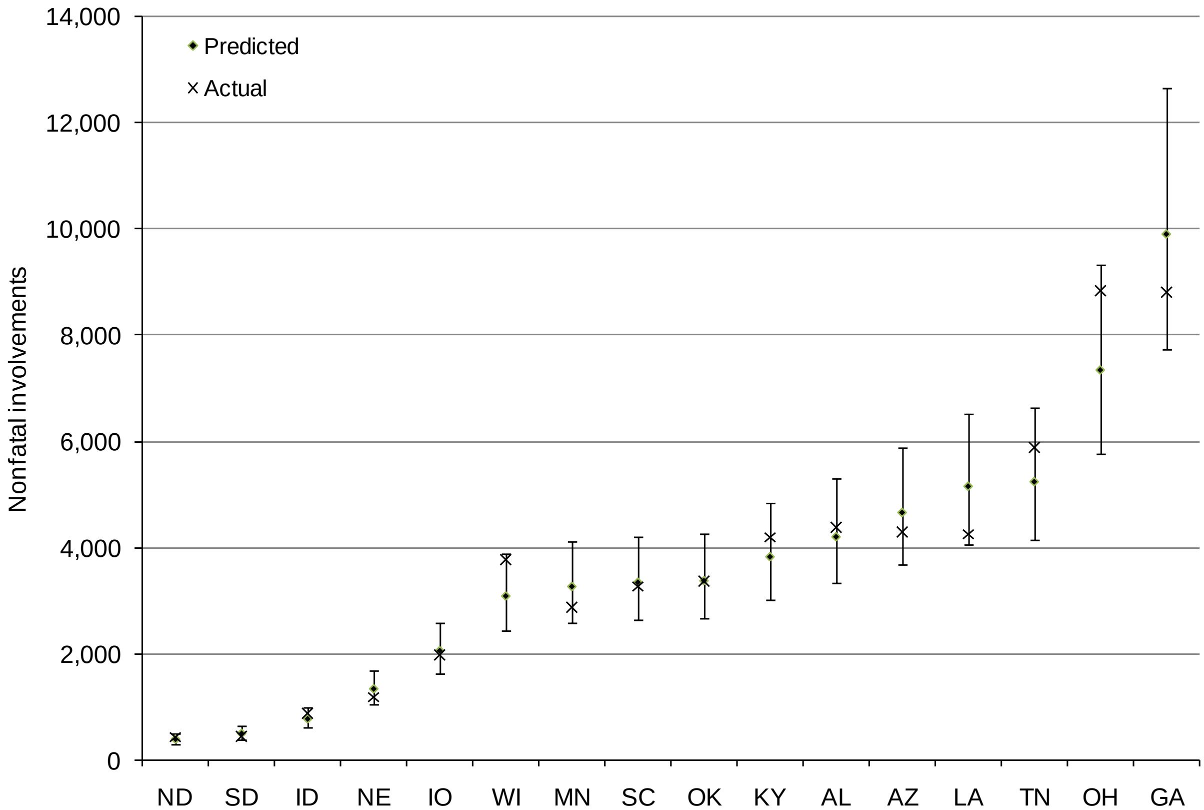

Figure 3 Actual, Predicted and 90 Percent Prediction Interval for Case States

6 Comparison with earlier models

The model presented here continues to improve on the prior models, though it primarily extends that work. In other words, the prior models fit the data well, and the new model presented here fits the data even better. It is clear that the relationship between the number of fatal involvements and nonfatal involvements is very robust. The first effort split the set of nonfatal involvements into injury/transported and towed/disabled components, and tried to work out the relationship between the counts of fatal involvements and the counts of injury/transported and towed disabled crash involvements. The work showed that the relationship is linear and the association between the counts of fatal and injury/transported or fatal and tow/disabled involvements rather good, with R^2 statistics of 0.87 and 0.85 respectively. But the attempt to create statistical models was not entirely satisfactory because they did not predict back to the original data well, and the confidence intervals were too wide to provide useful guidance to the states.[1]

A revised model simplified the approach by modeling nonfatal involvements as a whole, and using the log transform improved the results.[2] We also provided prediction intervals, rather than confidence intervals. In estimating the regression line for predicting the number of nonfatal crashes from the number of fatal crashes, eight observations (six states) were used. These eight observations represent a sample from a larger population of states. The regression model had an R^2 statistic of 0.94, and all the observations fell within the 90% prediction interval. This model was a significant improvement over the previous approach.

The revised model was updated with the addition of five more observations from new state evaluations. The modeling approach was the same as in the previous model, modeling the log transform of the data. The parameter estimates of the revised model were very consistent with those in the prior model, and the fit was comparable, with an R^2 of 0.94.

The present effort extends this model with additional observations, adding data from seven new states, and using a total of 16 states, while the prior model was based on 13 observations from eleven states. The present effort also adds a factor for *RU*, the ratio of rural to urban CMV travel in the state. These seven additional observations are from states evaluated since the earlier work. Some observations used in previous models that are now dated were dropped, as were two states that did not completely satisfy the evaluation criteria. In each of the observations used in the new model, the state data provided apparently reliable information about the number of fatal and nonfatal reportable involvements, and so qualified for inclusion in the model.

Re-estimating the model with these new observations resulted in a new model that was close to the previous model, but fit the data even better and resulted in tighter prediction intervals. The parameters for the new model were quite close to those for the two prior models. The term for the intercept in the new model is 2.944, compared with 3.098 and 3.214 in the prior two. The parameter for the slope is now 1.062, compared with 1.084 and 1.063 in the two prior. Both of the prior models fit the data well, with R^2, a measure of fit, identical at 0.94 for both. The new model fits the data slightly better, with an R^2 of 0.98. Table 6 shows the comparison between the previous and current models of model parameters and measures of fit, and the results for predicting the number of nonfatal involvements for a new state with 100 fatal involvements and an RU factor of 1.31. The predicted number of nonfatal involvements is quite similar for each of the three models, but there is a substantial reduction in the range of the prediction interval because of the additional observations used in the later models. Note that in the new model, the one presented in this report, the range of the prediction interval is substantially reduced from the prior models and is less than half of the range for the first model.

Table 6 Comparison of Previous Models with the New Model

	Revised model	Updated model	New Model
Model parameters			
Intercept	3.0983	3.2143	2.9436
Slope	1.0835	1.0631	1.0617
R^2	0.94	0.94	0.98
Prediction for new state with 100 fatal involvements			
Nonfatal involvements	3,254	3,327	3,349
90% prediction interval, lower bound	1,972	2,345	2,653
90% prediction interval, upper bound	5,371	4,720	4,228
Range of prediction interval	3,399	2,375	1,575

7 Discussion

Since there is uncertainty in a sample, a confidence interval is appropriate for the estimate of the number of nonfatal crashes for a state, given the number of fatal crashes. Often, 90 or 95 percent confidence intervals are calculated for each state. A 90 percent confidence interval is calculated by a procedure such that if this procedure were repeated over and over again, 90 percent of the confidence intervals would *trap* the true number of nonfatal crashes in the population. Thus, we are 90 percent confident that our estimate traps the true number of nonfatal crashes in the population for a particular state. By collecting a sample and calculating our estimates, we only perform this procedure once. Confidence intervals apply to states that were used in the estimation process.

Prediction intervals, on the other hand, apply to out-of-sample states. In other words, prediction intervals are used for new states that were not used to estimate the regression line. Intuitively, prediction intervals are wider than confidence intervals. Thus, the prediction problem begins by first fitting a model to a sample of states. Once the model is fitted, an estimate of the number of nonfatal crashes can be predicted for a new state not used in the estimation process. Since the new state is out-of-sample, a prediction interval should be reported. The interpretation is similar to that of a confidence interval. We are 90 percent certain that the interval traps the population value for the new state.

Prediction intervals provide reasonable guidance to the states for the expected number of nonfatal involvements. The range of the intervals in the new model is about ±20 percent, which is a substantial improvement over the ±30 percent in the previously updated model. A twenty percent range may be regarded as relatively wide, but the extent of variability in the underlying data should be recognized. There are several sources of this variability. One source is in the number of fatal involvements. Whether a person is fatally injured in a particular crash is highly random. When there are many fatal crashes in a state, the randomness tends to wash out, but when there are only a few, the randomness can have a substantial effect on the absolute number of fatal involvements.

There is also no doubt significant measurement error in the counts of nonfatal involvements determined in the state data. Even though the states used in the statistical model coded all the information needed to identify crash involvements that meet the MCMIS Crash file reporting criteria, it is important to remember that the source of the data is ultimately an individual police officer completing a crash report. The UMTRI evaluations have shown that the accuracy of reported cases vary widely. Reporting officers often work in difficult conditions. Protecting life and property, rather than accurate crash data, is the primary mission. Moreover, quality control is difficult and expensive. All these factors contribute to variability in the underlying data.

Adding more states to the model may improve the estimates and narrow the prediction intervals. Of particular interest would be to add states in areas not well-covered by the set of states available for the model at this time. States with fewer than 15 fatal involvements, and representatives of the three largest (Florida, California, and Texas) would help fill gaps in the range of states covered by the model. However, the prediction intervals available in the current model should provide meaningful guidance to the states.

8 References

1 Green, P., and Blower, D., Preliminary Ratio of Crash Severities Reportable to the MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. January 2007. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

2 Green, P., and Blower, D., Revised Ratio of Crash Severities Reportable to the MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. February 2007. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

3 Green, P., and Blower, D., Updated Ratio of Crash Severities Reportable to the MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. October 2008. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

4 Blower, D., and Matteson, A., Evaluation of 2004 Tennessee Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. May 2007. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

5 Blower, D., and Matteson, A., Evaluation of 2005 South Dakota Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. March 2007. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

6 Blower, D., and Matteson, A., Evaluation of Iowa Crash Data Reported to the MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. August 2006. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

7 Green, P.E., and Matteson, A., Evaluation of 2005 Alabama Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. October 2007. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

8 Green, P.E., and Matteson, A., Evaluation of 2005 Arizona Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. June 2007. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

9 Blower, D., and Matteson, A., Evaluation of 2005 Louisiana Crash Data Reported to the MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. December 2006. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

10 Blower, D., and Matteson, A., Evaluation of 2005 Nebraska Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. February 2007. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

11 Green, P.E., and Matteson, A., Evaluation of 2005 Ohio Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. November 2006. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

12 Green, P.E., and Matteson, A., Evaluation of 2006 Georgia Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. November 2007. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

13 Green, P.E., and Matteson, A., Evaluation of 2006 Idaho Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. December 2007. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

14 Blower, D., and Matteson, A., Evaluation of 2006 Kentucky Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. December 2007. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

15 Green, P.E., and Matteson, A., Evaluation of 2006 South Carolina Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. July 2008. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

16 Green, P.E., and Matteson, A., Evaluation of 2006 Wisconsin Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. March 2008. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

17 Blower, D., and Matteson, A., Evaluation of 2007 Minnesota Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. March 2009. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

18 Blower, D., and Matteson, A., Evaluation of 2007 Oklahoma Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann Arbor, Michigan. June 2009. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

19 Blower, D., and Matteson, A., Evaluation of 2008 North Dakota Crash Data Reported to MCMIS Crash File. University of Michigan Transportation Research Institute, Ann

Arbor, Michigan. July 2009. 34 p. Sponsor: Federal Motor Carrier Safety Administration, U.S. D.O.T.

9 Appendix

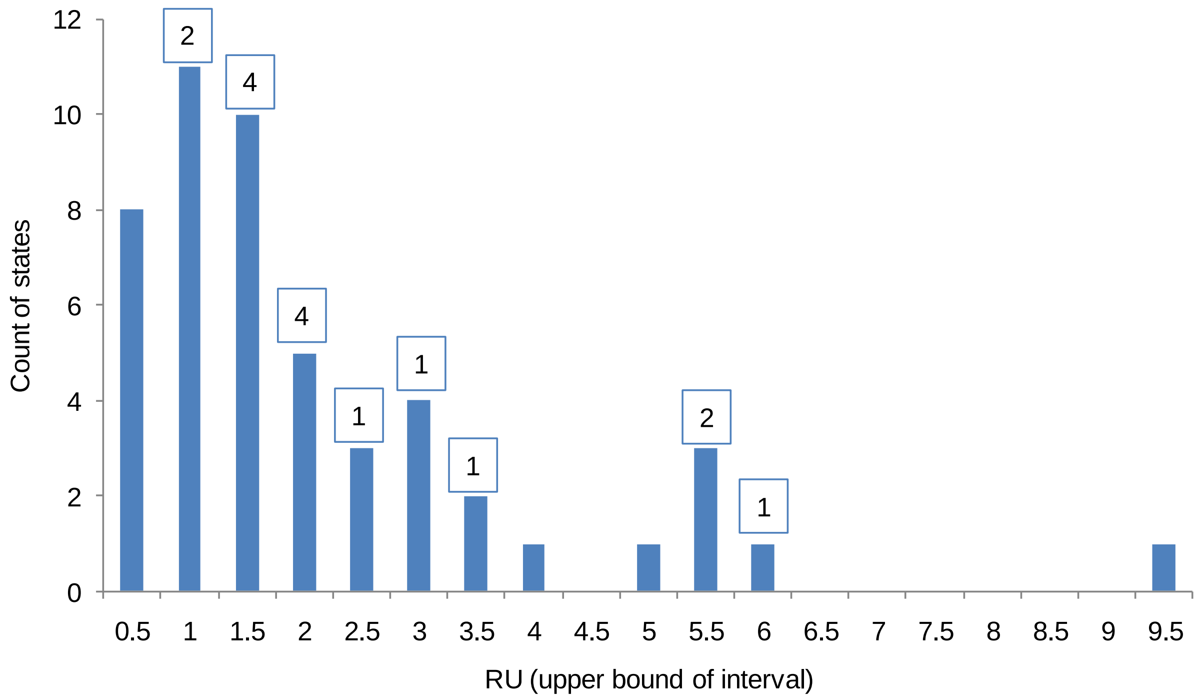

Figure 4 Counts of States by RU and Number of States in the Model for Each Range of RU