Development of a Speeding-Related Crash Typology

PUBLICATION NO. FHWA-HRT-10-024

APRIL 2010

U.S. Department of Transportation
Federal Highway Administration

Research, Development, and Technology
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, VA 22101-2296

FOREWORD

The overall goal of the Federal Highway Administration's (FHWA) *Speed Management Strategic Initiative* and the FHWA Speed Management Program is to improve the safety of the Nation's highways through the reduction of speeding and speed-related (SR) crashes.

This report summarizes a detailed examination of crash data and the development of an SR crash typology. This typology is intended to help identify the crash, vehicle, and driver characteristics that are associated with SR crashes. The goal was to determine several variables associated with SR crashes—such as what, where, when, and who—in order to aid in the development of new treatments/countermeasures and to more effectively target existing treatments. Two large national databases (the Fatality Analysis Reporting System (FARS) and the National Automotive Sampling System General Estimates System (NASS GES)) and two State databases (North Carolina and Ohio) were analyzed using single-variable table analysis and classification and regression tree (CART) analyses.

This document will be useful to traffic engineers as well as city, State, and local officials who are responsible for highway design and public safety.

Monique Evans
Director, Office of Safety
Research and Development

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA-HRT-10-024	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Development of a Speeding-Related Crash Typology		5. Report Date April 2010
		6. Performing Organization Code: N/A
7. Author(s): Forrest M. Council, Martine Reurings, Raghavan Srinivasan, Scott Masten, and Daniel Carter		8. Performing Organization Report No. N/A
9. Performing Organization Name and Address The University of North Carolina Highway Safety Research Center 730 Martin Luther King, Jr. Blvd. Chapel Hill, NC 27599-3430		10. Work Unit No. N/A
		11. Contract or Grant No. DTFH61-06-C-00013
12. Sponsoring Agency Name and Address Federal Highway Administration Office of Safety Research and Development 6300 Georgetown Pike McLean, VA 22101-2296		13. Type of Report and Period Covered Final Task Report, September 2006–September 2008
		14. Sponsoring Agency Code
15. Supplementary Notes The Contracting Officer's Technical Representative (COTR) was Carol Tan, HRDS-06.		

16. Abstract

Speeding, the driver behavior of exceeding the posted speed limit or driving too fast for conditions, has consistently been estimated to be a contributing factor to a significant percentage of fatal and nonfatal crashes. The U.S. Department of Transportation has instituted the *Speed Management Strategic Initiative* to seek more effective ways to manage the crash-related effects of speeding. In support of this initiative, this study conducted a detailed examination of recent crash data through the development of a speeding-related (SR) crash typology to help define the crash, vehicle, and driver characteristics that appear to result in a higher probability of SR crashes. Thus, the goal is to determine variables associated with SR crashes—such as what, where, when, and who—in order to provide guidance to the future development of new treatments and to more effectively target new and existing treatments. Recent Fatality Analysis Reporting System (FARS) and National Automotive Sampling System General Estimates System (NASS GES) data were used to answer these questions. Because these national databases only allow the use of a definition of SR that combines both exceeding the speed limit and too fast for conditions, two State databases (North Carolina and Ohio) were used to determine if different findings resulted from using the combined definition versus the exceeding the speed limit definition.

Two analysis methodologies were used: (1) single-variable table analysis and (2) classification and regression tree (CART). In the first, for a series of both crash-related and vehicle/driver-related variables (e.g., crash type and age of driver), individual codes within each variable were examined to determine which showed an overrepresentation of SR crashes or SR vehicles/drivers (e.g., rear-end crashes for 16–19-year-old drivers). The second method involved CART analyses which automatically define which factors/variables are the most critical with regard to SR crashes or drivers and which combinations of variables/codes are the most important. Similar single-variable and CART analyses were also conducted for five high-priority subsets of the data (e.g., pedestrian crashes and intersection crashes). As might be expected, the results differed between fatal and total crashes, national and State, and among States. Few differences were seen in the results based on the two definitions. The single-variable table results were consistent with two earlier studies in indicating higher SR percentages in single-vehicle crashes, rural crashes, crashes on curves, nighttime crashes, motorcycle crashes, as well as crashes involving young drivers, male drivers, drivers not using restraints, and drivers under the influence of alcohol. No consistent pattern of speeding was seen in either pedestrian or bicycle crashes or in work zone crashes. The CART results from the different databases were less consistent and more difficult to interpret. The crash-based results consistently identified single-vehicle crashes during adverse weather as a high-priority subgroup. The vehicle-based findings indicated almost no consistency across databases, with young male showing up more than other descriptors.

17. Key Words Speed, Speeding, Crash typology, Traffic safety, HSIS	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161.		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 107	22. Price

From DOT F 1700.7 (8-72) Reproduction of completed page authorized.

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 FIGURES

LIST OF TABLES

ABBREVIATIONS

AADT	Annual average daily traffic
CART	Classification and regression tree
FARS	Fatality Analysis Reporting System
FHWA	Federal Highway Administration
HSIS	Highway Safety Information System
NASS CDS	National Automotive Sampling System Crashworthiness Data System
NASS GES	National Automotive Sampling System General Estimates System
SR	Speeding-related
UDA	Unsafe driving act

INTRODUCTION

Speeding, the driver behavior of exceeding the posted speed limit or driving too fast for conditions, has consistently been shown to be a contributing factor to a significant percentage of fatal and nonfatal crashes. As shown in figure 1, the frequency of fatal crashes and the percentages of total fatal crashes that are speeding-related (SR) have remained fairly constant between 1990 and 2006 according to the Fatality Analysis Reporting System (FARS).[1] However, both are large, with the frequency ranging from 11,000 to 13,000 crashes each year and the percent ranging between 30 and 33 percent.

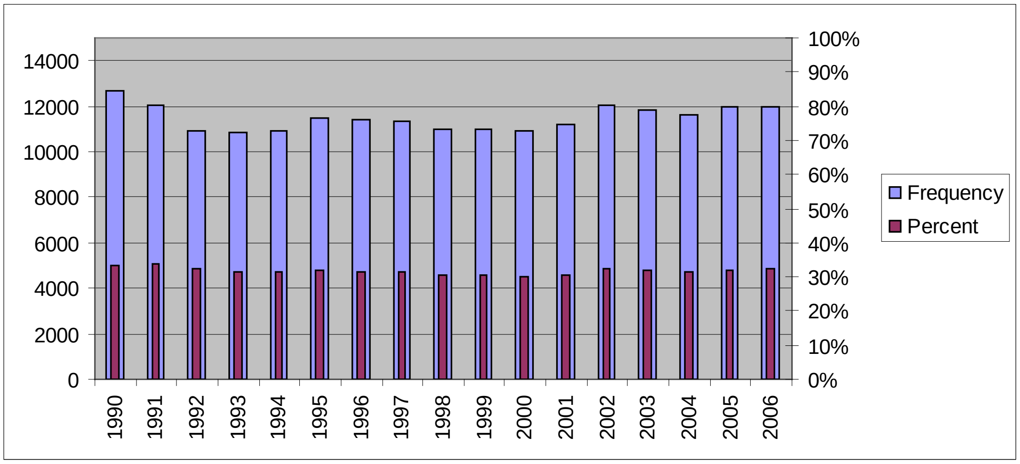

Figure 1. Graph. Frequency of fatal crashes and percentage of total fatal crashes that are SR.

Thus, speeding is a significant safety issue warranting attention. While the United States has seen progress in other major safety issues, such as occupant restraint use and alcohol use and driving, little if any progress has been seen with this issue.

Given the size of this problem and continually increasing speeds on freeways and nonfreeways, the U.S. Department of Transportation has instituted the *Speed Management Strategic Initiative* to seek more effective ways to manage crash-related effects of speeding.[2] One way to search for better speed management techniques is through a detailed examination of recent crash data in order to increase the basic knowledge of the crash-related factors that are associated with speeding. For example, the development of an SR crash typology would help define the crash, vehicle, and driver characteristics that result in a higher probability of SR crashes. Once defined, this information would provide guidance to the development of new treatments for high-risk crash types and better targeting of treatments to the road types, locations, conditions, times, and drivers that disproportionately contribute to crash risk. This report describes the development of such a typology.

The typology requires the identification of relevant crash characteristics most often associated with SR crashes. The typology seeks to answer the following questions:

- What are the characteristics of an SR crash (e.g., type of crash, manner of collision, crash severity, etc.)?

- Where do SR crashes most often occur (e.g., city, intersection proximity, roadway type, roadway surface condition, roadway alignment, etc.)?

- When do SR crashes occur (e.g., time of day, day of week, etc.)?

- Who is most likely to be involved in SR crashes (e.g., age, gender, injury severity, alcohol involvement, vehicle type, safety belt usage, number of vehicle occupants, motorcycle helmet use, etc.)?

PRIOR STUDIES

While numerous research studies have explored the effects of speed on crash frequency and severity as well as on the effect of treatments aimed at managing speed (e.g., *Transportation Research Board Special Report 254*), a previous National Highway Traffic Safety Administration study of particular interest examined an SR crash typology.[3,4] This study used data from the Crash Avoidance Research Data (CARDfile), the Indiana Tri-Level Study, and FARS to examine the scope of the speeding problem and the characteristics of crashes that are SR.[5,6,1] CARDfile includes crash data from six States.[5] In this study, the 1986 CARDfile data were used and included 1.4 million crashes involving about 2.4 million vehicles and drivers.

Based on CARDfile, speed was a factor in about 12 percent of all crashes.[5] Data from the Indiana Tri-Level Study indicated that excessive speed was a factor in 7.1–16.9 percent of crashes.[5,6] However, most of the analyses focused on the 1989 FARS data.[1] The key findings included the following:

- Almost 70 percent of all drivers involved in SR fatal crashes were involved in single-vehicle crashes.
- Approximately 36 percent of all fatal crashes on rural roads were SR, whereas only 30 percent of fatal crashes on urban roads were SR.
- SR crashes accounted for only 27 percent of all fatal crashes on straight roadway sections, but they constituted 54 percent of all fatal crashes occurring on curves.
- While only 27 percent of all daytime fatal crashes were SR, 39 percent of nighttime fatal crashes involved speeding.
- About 25 percent of all male drivers involved in fatal crashes were speeding, while only 16 percent of female drivers involved in fatal crashes were speeding.
- Safety belt usage (lap belt and shoulder belt) for drivers involved in SR fatal crashes was found to be only 19 percent. This compares to the usage rate of 37 percent for drivers involved in non-SR fatal crashes.
- Approximately 41 percent of all drivers under the influence of alcohol who were involved in a fatal crash were speeding.
- More than 45 percent of all motorcycle drivers involved in fatal crashes were speeding.

In a second typology-related study, Hendricks, et al. examined data from a sample of 723 relatively severe crashes involving 1,284 drivers collected as part of the National Automotive Sampling System Crashworthiness Data System (NASS CDS) program.[7,8] Based on in-depth field investigations of the crashes and driver interviews, researchers determined the specific driver behaviors and unsafe driving acts (UDAs) that led to the crashes and the situational driver and vehicle characteristics associated with these behaviors. Each driver's actions were examined to determine whether and how they contributed to the crash, and a ranked list of the most commonly occurring UDAs (e.g., inattention, alcohol impairment, perceptual errors, etc.) was produced. Crash types and driver and behavioral characteristics associated with these higher ranked UDAs were then identified. In situations in which cause could be assessed, excessive vehicle speed was the second leading causal factor, with inattention being the first. Excessive speed constituted exceeding the speed limit; however, it sometimes also included traveling at inappropriate speeds for prevailing weather or roadway conditions (i.e., too fast for conditions). Excessive speed was the primary cause of crashes for 6.8 percent of the drivers who contributed to causation; it was also a primary cause in combination with other causes for an additional 3.8 percent of the contributing drivers and was a contributory cause for an additional 8.1 percent of the drivers. When crash types were examined for the drivers who were excessively speeding, researchers found that speeding was the leading cause of single-driver right- or left-roadside departure with traction loss (i.e., part of run-off-road crashes) and the third leading cause of head-on crashes. Researchers then examined the characteristics of the drivers and roadways in these excessive-speed roadway departure crashes. They found that the more important characteristics of these crashes included that they occurred primarily on curves or on local or collector roadways at night or during clear weather, and the drivers were younger males (less than 35 years old). Males younger than 20 years old comprised 46.2 percent of the sample. Enhanced law enforcement countermeasures were suggested for these crashes.

In summary, two prior studies have developed crash typologies that reference speeding. Both examine data from in-depth crash investigations as part or all of the methodology as well as factors and unsafe driving actions in addition to speeding. This current study is designed to add further information to the findings of those two studies by examining large national databases which include recent crashes as well as by using different methodologies and concentrating solely on SR crashes.

DATA USED

Since a goal of the study was to explore questions as they relate to SR crashes in the United States, data were used from two major national crash databases—the National Automotive Sampling System's General Estimates System (NASS GES, referred to as GES) and FARS.[9,1] While NASS CDS contains more detailed crash investigation data, its sample size of approximately 5,000 cases per year, which is only partially comprised of SR crashes, was considered too small for this analysis.[8] GES data are derived from a nationally representative sample of police-reported motor vehicle crashes of all types from minor to fatal.[9] Approximately 60,000 police accident reports are included each year. Sample weights are assigned to each crash based on a sampling protocol. Using the weight, the sample can be extrapolated to represent the approximately six million U.S. crashes occurring each year. FARS is a census of all fatal crashes in the United States, showing approximately 40,000 fatal crashes in 2005.[1] To ensure that findings were as current as possible, data from 2005, which is the latest year available in each file, were used in these analyses.

The FHWA was interested in examining speeding using a disaggregated definition where "too fast for conditions" and "exceeding posted speed" would be analyzed separately since the appropriate countermeasures could differ for these two types of SR crashes. However, data available in GES and FARS only make it possible to define *SR crashes* with a broader definition, which consists of drivers either exceeding the posted speed limit or driving too fast for conditions. To better examine the possible effects of using only this combined definition, researchers used data from two States in FHWA's Highway Safety Information System (HSIS) that allowed the use of the more restrictive definition for the second part of the study. HSIS is the only multistate database containing annual files of not only crash data but also linkable roadway inventory and traffic volume data. While HSIS is usually used to conduct studies of changes in safety risk due to changes in roadway features (e.g., the effects of an intersection design change on safety), the linkability of the files allows one to enhance the crash variables with additional roadway variables, which was used in this study. Enhanced crash data from North Carolina from 2002–2004 and Ohio from 2003–2005 were used to ensure adequate sample sizes.

DEFINITION OF SR CRASHES

For all databases used in this study, the decision on whether or not a crash is SR (regardless of definition) is based on an examination of each vehicle in the crash. First, each vehicle is defined as SR or not based on the variables, which are described below. Then, if any vehicle in the crash is SR, the crash is considered SR. If all of the vehicles involved in a crash are coded as nonspeeding, then the crash is defined as not SR. In all other cases, it is unknown whether the crash is SR.

GES has three of these variables in its vehicle file: SPEEDREL (SR), VIOLATN (violations charged), and P_CRASH2 (critical event). The first variable is coded as "yes" when excessive speed by the driver is noted as a contributing factor on the police accident report or if a speeding violation has been issued. For the second variable, a speeding violation is charged when a driver travels over the speed limit or too fast for conditions. The third variable is coded "too fast for conditions" when drivers lose control of their vehicles due to driving too fast for conditions. Based on the definitions of these three variables, it was determined that SPEEDREL should be used to indicate SR crashes. Thus, if SPEEDREL is coded as "yes" for one of the vehicles involved in a crash, the crash is considered SR.

For FARS, it is theoretically possible to distinguish between the less restrictive and more restrictive definitions of SR. There are three variables which describe the violations that have been charged, and values of 21–25 and 29 all indicate some type of speeding violation. However, only 1.3 percent of drivers involved in a crash were charged a speeding violation. This is in contrast to data on SR in four related factor driver-level variables. In this instance, for 20 percent of the drivers, speeding was noted as a factor in the crash. Given that the sample of fatal crashes involving SR violations was small, it was decided that the driver-related variables combining "over speed limit" and "too fast for conditions" would be used to identify the SR crashes in FARS.

As indicated previously, FARS and GES only allow the use of the broader definition of SR, and HSIS data were used to allow for a more restrictive definition. It is noted that the issue here is more than just a technical one since the choice of definition can affect how the findings of the analyses are interpreted. Both definitions are based on the investigating officer's judgment made during the crash investigation. There is no exact measure of precrash speed available to the officer. It is perhaps true that fatal crashes would undergo a more detailed crash reconstruction, and thus, the precrash speed estimates would be more accurate; however, the level of accuracy is unknown. In addition, it is difficult to know whether an identified variable shows a true higher association with speed or whether the association shown is partially due to an officer bias in noting the "too fast for conditions" factor. Information from police officers indicates that the "too fast" designator may be systematically used in certain situations—if the condition is present, the officer is more likely to use the descriptor. For example, the officer may be more likely to use this designator in bad weather crashes or crashes on curves. Both are logical uses since drivers should reduce speeds during bad weather and should not drive too fast to maneuver around a sharp curve. However, if this systematic use is present, it may overstate the true effect of these variables, making it difficult to determine whether the overrepresentation of these factors is true or partly due to the effect of an inflated use of this descriptor in certain situations. If the latter is found, treatment programs oriented to these factors may not be as successful as if oriented to other characteristics where such a bias is not expected. The HSIS data were used to attempt to provide further clarity concerning whether such a possible bias matters and whether there are different findings when the two definitions are used for the same variable.

Data formats and data from all nine HSIS States were examined to determine the following: (1) whether available variables allowed the use of both the broader and the more restrictive definitions of SR and (2) whether or not the existing data provided ample sample sizes for use in both definitions. Data from North Carolina and Ohio met both these requirements but in different ways. Crash data in HSIS are divided into three linkable files: (1) crash variables, (2) vehicle variables, and (3) occupant variables. Just as with the GES and FARS data, a decision on whether or not a crash is SR is based on whether one or more vehicles indicated an SR factor.

The vehicle file for North Carolina contains three variables describing contributing factors, and the options for these factors include exceeded authorized speed limit and exceeded safe speed for conditions. The vehicle file for Ohio has only one contributing factor variable with unsafe speed and exceeded speed limit as two of the options. However, when the data were examined, the exceeded speed limit category was only checked for approximately 0.1 percent of the vehicles. Since the officer also recorded an estimate of the precrash speed for each vehicle and the posted speed limit at the crash location, these two variables could be compared to determine additional vehicles that are exceeding the posted limit, providing an adequate sample for analysis. It is noted, however, that because the definitional categories were defined differently in the two States, the resulting percentages for the two definitions differed (see table 1 and table 2).

Finally, note that the North Carolina and Ohio data contained in the HSIS files are for State system roads only and are under the control of the State transportation departments. In North Carolina, where there are no county roads systems, the State system roads include all roads except city streets which are not State-owned highways. In Ohio, the HSIS data do not include rural roadways owned by counties and non-State city streets. Thus, in general, the North Carolina and Ohio data are more rural in nature than the GES and FARS data since they do not include crashes on city streets. As will be noted in certain findings below, the differences in the databases affect some of the outcomes.

Table 1 shows the number of SR crashes in 2005 according to GES and FARS. Comparable data for Ohio and North Carolina are shown in table 2. Note that from this point forward, the more liberal definition combining both over speed limit and too fast for conditions will be referred to as the combined definition. In the tables, the combined definition is referred to as “Total SR.”

Table 1. The number and percentage of SR and non-SR crashes in GES and FARS (2005).

Category	GES		FARS	
	n	**Percent**	***n***	**Percent**
SR	1,195,570	19.5	11,553	29.5
Not SR	4,517,169	73.5	26,034	66.4
Unknown	434,167	7.0	1,602	4.1
Total	**6,146,907**	**100**	**39,189**	**100**

Note: *n* represents the frequency of crashes.

Table 2. The number and percentage of SR and non-SR crashes in North Carolina (2002–2004) and Ohio (2003–2005).

Category	North Carolina Combined		North Carolina Over Speed Limit		Ohio Combined		Ohio Over Speed Limit	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
SR	62,746	14.9	12,802	3.0	51,906	11.3	30,677	6.7
Not SR	343,471	81.3	393,415	93.2	396.135	85.9	417,364	90.5
Unknown	16,107	3.8	16,107	3.8	12,972	2.8	12,972	2.8
Total	**422,324**	**100.0**	**422,324**	**100.0**	**461,013**	**100.0**	**461,013**	**100.0**

Note: *n* represents the frequency of crashes.

As shown, the differences in the databases result in differences in the percentage of crashes that are considered SR even when the combined definition is used—20 percent for GES, 30 percent for FARS, 15 percent for North Carolina, and 11 percent for Ohio. As expected, the restricted over speed limit definition results in much lower percentages of SR for North Carolina and Ohio—3 percent for North Carolina and 7 percent for Ohio. The State percentages probably vary because of the differing definitional procedures used. The conclusion drawn is that all these databases are based on somewhat different crash populations. What will be of interest is whether findings based on categories within individual descriptor variables show any consistency across databases.

The findings also indicate that speeding appears to be a more significant factor in fatal crashes (FARS) than in total crashes (all others). This is a logical finding in that speeding is related to crash energy, which is related to injury. This is further examined in table 3 through table 5, which show the severity distributions of crashes from GES, North Carolina, and Ohio. Note that in all tables from this point forward, only crashes where SR can be defined as yes or no are included, and crashes where SR is unknown are omitted.

Table 3. Frequency and number/percentage of SR crashes regarding crash severity in GES.

Category	Total Crashes	SR Crashes	
	n	*n*	Percent
Fatal	24,712	8,689	35.16
Incapacitating	227,678	51,168	22.47
Non-incapacitating	513,367	117,110	22.81
Possible injury	945,090	218,809	23.15
No injury	3,903,549	774,202	19.83
Injured, unknown severity	29,678	5,145	17.34
Unknown	68,660	20,448	29.78
Total	**5,712,734**	**1,195,571**	**20.93**

Note: *n* represents frequency of crashes.

Table 4. Frequency and number/percentage of SR crashes regarding crash severity in North Carolina.

Crash Severity	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Fatal injury	733	24.0	404	13.2	1,137	37.3	1,913	62.7	3,050	100.0
Disabling injury	835	13.8	1,076	17.7	1,911	31.5	4,156	68.5	6,067	100.0
Evident injury	2,916	7.7	6,427	17.0	9,343	24.7	28,536	75.3	37,879	100.0
Possible injury	3,180	3.2	11,673	11.7	14,853	14.9	84,775	85.1	99,628	100.0
No injury	4,117	1.6	27,776	11.1	31,893	12.8	217,678	87.2	249,571	100.0
Not coded	1,021	10.2	2,588	25.8	3,609	36.0	6,413	64.0	10,022	100.0
Total	**12,802**	**3.2**	**49,944**	**12.3**	**62,746**	**15.4**	**343,471**	**84.6**	**406,217**	**100.0**

Note: *n* represents frequency of crashes.

Table 5. Frequency and number/percentage of SR crashes regarding crash severity in Ohio.

Crash Severity	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Fatal injury	398	20.8	156	8.2	554	29.0	1,355	71.0	1,909	100.0
Incapacitating injury	1,720	15.0	753	6.6	2,473	21.6	8,994	78.4	11,467	100.0
Non-incapacitating injury	5,255	11.0	3,668	7.7	8,923	18.6	38,995	81.4	47,918	100.0
Possible injury	3,956	7.0	2,157	3.8	6,113	10.9	50,076	89.1	56,189	100.0
No injury	19,348	5.9	14,495	4.4	33,843	10.2	296,715	89.8	330,558	100.0
Total	**30,677**	**6.8**	**21,229**	**4.7**	**51,906**	**11.6**	**396,135**	**88.4**	**448,041**	**100.0**

Note: *n* represents frequency of crashes.

In all three databases, regardless of definition, the SR percentage appears to decrease with crash severity. The percentage for fatal crashes is 1.5 to 15 times higher than the percentage for crashes without injuries, depending on the database and definition. The total number of fatal crashes estimated by GES to be SR (8,689) is significantly smaller than the total number of SR crashes in the FARS database (11,553). As has been found by other studies, the total number of fatal crashes estimated by GES is smaller than the true total in FARS, providing further rationale for using both GES and FARS in this study.

OVERVIEW OF ANALYSIS METHODOLOGY

The goal of this study is to determine which crash-, vehicle-, and driver-related factors are more likely to be found in SR crashes. As such, the list of possible variables (e.g., crash type) and combinations of variables (e.g., crash type by urban versus rural roadway class by speed limit) are almost limitless. Finding the most important variables is difficult since it is in some ways determined by the interest of the user or the type of treatment program being considered. For example, roadway-based treatments (e.g., traffic-calming measures) might be better identified or targeted by location-type analyses, while enforcement or educational treatments would be more related to driver variables. It was also difficult to combine vehicle- and driver-related factors in the same analyses as the broader crash factors since the decision of how to code each driver in a crash was complex. For example, while a crash can be classified as SR if one or more vehicles is SR, not all drivers in that crash should be thought of as SR. Indeed, in multivehicle collisions, only one of the drivers would be speeding in many cases, and thus, a comparison of driver age for speeding and nonspeeding drivers must be done on a vehicle basis rather than a crash basis.

Given the unlimited number of possible factors of interest, a decision was made to conduct a two-part analysis. In the first part, a series of single-variable tables was produced for key crash, vehicle, and driver variables. Each variable-specific table was examined to determine which categories of that variable had the highest number and SR percentage. Such single-variable tables provide valuable information on SR crashes; however, they do not provide a way of determining which variables are most important in terms of speeding or information on combinations of variables or on the interactions between variables. The second set of analyses attempted to do this through the use of classification trees as produced by the classification and regression tree (CART) software that is available in SAS®.[10]

SINGLE-VARIABLE TABLE ANALYSES

As indicated previously, single-variable tables were created from each dataset/definition for a large number of variables. The choice of variables to be examined was based to some extent on the results of past studies of SR issues, particularly on the earlier study by Bowie and Walz.[4] The factors describing the overall nature of each crash (e.g., crash type, crash location, etc.) were examined using a crash-based file where any involved vehicle was speeding, and the vehicle and driver-based factors were examined in a vehicle-based file where each vehicle was classified as speeding or not. In the results section, three tables are presented for each variable—the first contains GES and FARS results if both are available, and the other two contain results for both definitions for each of the two States. In general, a category is defined as *over-represented* if it is characterized by a high percentage of SR crashes, drivers, or vehicles. Whether this is the most helpful way to characterize these findings if they are to be used in treatment development or targeting is discussed below in the interpretation of results section. A brief discussion describing the consistency of findings across the databases and definitions is included below each table.

IDENTIFICATION OF CRITICAL FACTORS USING CLASSIFICATION TREES

Although the analyses of single-variable tables provide useful information about SR crashes and vehicles/drivers in crashes, they do not automatically indicate which factors/variables are the most critical with regard to SR crashes or speeding drivers. They also do not indicate which combinations of variables are the most important. One way to identify the critical roadway, vehicle, and driver factors associated with an increased likelihood of an SR crash is to estimate a logistic regression model with the roadway, vehicle, and driver factors as independent variables and then to identify the statistically significant factors. *Logistic regression* is a parametric approach that is based on assumptions about error distributions. The CART methodology is nonparametric and does not require any such assumptions. In addition, CART is able to include a relatively large number of independent variables and identify complex interactions between these variables more efficiently compared to logistic regression. For example, CART is able to determine not only the most important variable and categories within that variable in terms of the risk of an SR crash, but also the most important second-level variable within the most important categories of the first-level variable, etc. That is, given the most important variable with respect to the proportion of SR crashes (e.g., manner of collision) and the subgroup of categories within that variable with the highest proportion of SR crashes (e.g., run-off-road crashes), CART is able to determine the next most important variable within these high-risk categories (e.g., road surface condition) and the categories of that variable that are most important (e.g., snow and sleet). It is hoped that these variables and categories are helpful in determining needed treatments. For these reasons, it was decided that classification trees would be used as the second type of analysis in this project.

Thus, the goals of the CART analysis are as follows: (1) to determine which variables available for examination are most important in terms of predicting SR crashes, (2) to determine which categories within that variable predict the highest risk/proportion of SR crashes, (3) to determine the second most important variable and subset of categories in terms of predicting SR crashes within this highest risk subset of categories of the first variable, and (4) to repeat the process to determine the third, fourth, and subsequent variables. This produces a tree with multiple branches that can be traced down to determine the most important combinations (or subsets) of variable categories in terms of predicting SR crashes. In the most simplistic terms, the CART procedure splits the categories of each variable in the database into all possible binary (two-category) combinations (nodes), calculates the SR risk within each part (node) of each pair, and determines which pair (i.e., which two sets of categories) produces the largest difference in SR risk within that variable. By repeating this process for each variable in the database, CART determines the two sets of categories producing the largest difference in risk of SR crash within each variable. This largest difference in risk is then compared across all variables to determine the one variable (and the set of categories) that produces the largest of all differences. This is the top of the tree, and the two categories within that variable are the first two branches of the tree. This process is then repeated within each of the two categories (branches) of the first variable to identify the second, third, and subsequent variables.

For a categorical variable (e.g., manner of collisions, month of crash, etc.), all possible binary combinations of categories are compared (e.g., category 1 versus categories 2–5, category 1 and 2 versus category 3–5, category 1 and 3 versus categories 2, 4, and 5, etc.). For ordinal variables (e.g., speed limit), all cases with the value of that variable smaller than or equal to a certain value go to one node, all other cases go to the other node (e.g., speed limit ≤ 30 mi/h versus speed limit ≥35 mi/h; speed limit ≤ 35 mi/h versus ≥ 40 mi/h, etc.).

CART then outputs a tree showing all branches (i.e., both high and low SR branches). This report shows the section of the tree illustrating up to the first four levels of branches with the highest percentage of SR crashes. Note that CART divides the database being analyzed into a training subset and a validation subset to refine the final output. The results of the training subset are presented in this report, meaning that the total frequency at the top of each tree only shows approximately ⅔ of the total case count shown in the single-variable tables. A description of the results of the CART analysis is provided in the results of CART analyses section.

Further information about CART is available in Breiman et al.[11] For applications of these trees in road safety research, see Stewart and Yan and Radwan.[12,13] Additional statistical details are provided in the appendix of this report.

VARIABLES USED IN THE ANALYSES

To examine the question of what occurs in an SR crash, the following crash characteristics are of interest:

- Manner of collision indicates the orientation of the vehicles in a collision.
- First harmful event indicates the first instance of property damage or injury-producing event in the crash.
- Number of nonmotorists involved captures the presence of pedestrians and cyclists in the crash.

To answer the question concerning where SR crashes mostly occur, the following variables are of interest:

- Location of first harmful event indicates if the crash is an on- or off-roadway crash.
- Relation to junction indicates if the first harmful event is located within a junction or interchange area.
- Roadway functional class indicates the functional class of the road where the crash occurs.
- Speed limit indicates the posted speed limit in miles per hour.
- Number of lanes indicates the number of lanes of travel. In GES and FARS, if the roadway is a divided trafficway, the number of travel lanes counts only lanes in the direction of travel of the first harmful event. If the roadway is an undivided trafficway, the number of travel lanes are all the lanes regardless of their direction of travel. Since this could produce misleading results (e.g., comparing undivided two-lane, two-way roads to two lanes of a divided road), the number of lanes was doubled for physically divided trafficways in all the analyses.

- Annual average daily traffic (AADT) per lane is produced by dividing the average daily flow by the number of lanes, indicating the traffic density at the time of the crash.
- Roadway alignment indicates the horizontal alignment of roadway.
- Roadway profile indicates the vertical alignment of roadway.
- Work zone identifies first harmful events that were related to, but did not necessarily occur in, a construction or work zone.

The variables which might be helpful in deciding when SR crashes mostly occur are as follows:

- The light condition at the time of the crash.
- The condition of roadway surface at the time of the crash.
- The atmospheric condition at the time of the crash.
- The season in which the crash occurred.
- The day of the week when the crash occurred on.

The question concerning who is most likely to be involved in SR crashes is related to the following variables:

- The age of the driver.
- The gender of the driver.
- Police-reported driver use of available vehicle restraints.
- Distraction which may have influenced driver performance and contributed to the cause of the crash.
- Physical impairments for all drivers and nonmotorists which may have contributed to the cause of the crash.
- Police-reported alcohol involvement indicating that the driver had consumed an alcoholic beverage.
- Police-reported drug involvement indicating that the driver had taken drugs.
- Number of previous speeding convictions of the driver (FARS only).
- License type and compliance with license restrictions (FARS only).

Since the “who” question could also involve the vehicle being driven, vehicle characteristics that might be of interest include the following:

- Body type of the vehicle (i.e., vehicle type).
- Hazardous cargo involvement for buses and trucks over 9,909 lb gross vehicle weight rating.

INTERPRETATION OF RESULTS

The overall objective of this research was to obtain information on SR crashes that could be used for basic knowledge, treatment development (i.e., identifying the need for new treatments), and treatment targeting (i.e., targeting existing treatments to populations, times, or locations to maximize the effect). The methodologies used include examining single-variable tables and regression trees. As is the case with all such problem identification efforts, some caution must be observed in drawing conclusions for a number of reasons. As discussed previously, the definition of an SR crash is somewhat problematic in that at least two valid definitions exist, and both are based on the investigating officer's judgment of precrash speeds, which is a difficult judgment to make after the fact.

In addition, given either definition, the method for determining the highest priority categories varies. For treatment-related information, the variable descriptors in the single-variable tables that should be given the highest priority would be those that exhibit not only a high percentage of SR crashes, but also a high number of SR crashes. For example, the single-variable table below illustrates the number and percentage of SR crash categories by the type of location where the crash occurred. According to the percent columns, both FARS and GES indicate that the location category with the highest SR percentage is the interchange area. However, the crash location with the second highest percentage is "Not related to junction," which describes crashes occurring on roadway sections between intersections or interchanges. This category contains 10 times the number of interchange crashes in the GES data and 23 times the number of interchange crashes in the FARS data, implying that while attention should be given to speeding crashes in interchange areas, a higher priority should be given to strategies for reducing SR crashes at between-junction roadway sections.

Table 6. Frequency and number/percentage of SR crashes regarding crash location.

	FARS			GES		
	All Crashes	**SR Crashes**		**All Crashes**	**SR Crashes**	
Crash Location	***n***	***n***	**Percent**	***n***	***n***	**Percent**
Not related to junction	26,966	9,354	34.69	2,491,891	672,697	27.00
Intersection-related	8,078	1,510	18.69	2,340,179	383,900	16.40
Other junctions	1,347	291	21.60	601,811	66,438	11.04
Interchange area	1,007	391	38.83	210,542	64,248	30.52
Unknown	189	7	3.70	68,316	8,286	12.13
Total	**37,587**	**11,553**	**30.74**	**5,712,739**	**1,195,569**	**20.93**

Note: *n* represents frequency of crashes.

Thus, in determining high-priority categories for treatment consideration, both the SR percentage and the SR frequency should be considered.

The CART results must also be viewed carefully. While CART defines those subsets of crashes (i.e., subsets based on categories within multiple variables) with high proportions of SR crashes, even if one only goes to the third- or fourth-level branch, the total numbers of SR crashes and the proportion of total SR crashes is often less than 5 percent.

This report does not offer treatment-related conclusions. The goal of this report was to produce information that could be used by others in developing such conclusions. That information is noted appropriately throughout the report.

RESULTS FOR SINGLE-VARIABLE ANALYSES

The following sections describe the results of the single-variable analyses for the set of variables selected in the GES, FARS, North Carolina, and Ohio databases. For each variable, tables containing data from these four databases (and two definitions for the State data) are presented first, followed by a brief description of findings. It is noted that even though GES and FARS data are presented in the same table (with the State data in separate tables), the percentages of SR crashes within a given category should not be compared across databases. For example, FARS category percentages are likely to be higher than the same GES category percentages since, in general, the FARS data have a higher percentage of SR crashes (30 percent versus 20 percent).

TYPES OF CRASHES MOST LIKELY DUE TO SPEEDING

In this section, tables are presented to examine the types of crashes that are most likely a result of speeding as well as the basic nature of SR crashes. The tables show the percentage of SR crashes by type of collision, first harmful event, and nonmotorist involvement.

Type of Collision

Table 7. Frequency and number/percentage of SR crashes regarding type of collision.

		FARS			GES		
		All Crashes	SR Crashes		All Crashes	SR Crashes	
Type of Collision		***n***	***n***	**Percent**	***n***	***n***	**Percent**
Collision with motor vehicle	Rear-end	2,253	927	41.15	1,730,195	603,805	34.90
	Head-on	3,959	655	16.54	116,689	13,456	11.53
	Angle				1,681,286	97,530	
	Front-to-side, same direction	484	131	27.07			
	Front-to-side, opposite direction	2,243	611	27.24			
	Front-to-side, right angle	5,024	828	16.48			
	Front-to-side, unknown direction	289	33	11.42			
	Sideswipe, same direction	473	141	29.81	350,816	23,243	6.63
	Sideswipe, opposite direction	469	85	18.12	58,714	9,278	15.80
	Rear-to-side	88	45	51.14			
	Rear-to-rear	6	3	50.00	3,788	686	18.11
	Other	75	18	24.00			
	Total	15,363	3,477	22.63	3,941,488	747,998	18.98
No collision with motor vehicle		22,048	8,063	36.57	1,762,624	447,073	25.36
Unknown		176	13	7.39	8,626	499	5.78
Total		**37,587**	**11,553**	**30.74**	**5,712,738**	**1,195,570**	**20.93**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Table 8. Frequency and number/percentage of SR crashes regarding type of collision in North Carolina.

Type of Collision		Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
		n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
Collision with motor vehicle	Rear-end	1,236	1.03	3,394	2.83	4,630	3.86	115,457	96.14	120,087	100.00
	Head-on	205	4.58	733	16.37	938	20.95	3,539	79.05	4,477	100.00
	Backing up	10	0.20	14	0.29	24	0.49	4,855	99.51	4,879	100.00
	Angle	626	1.34	1,820	3.89	2,446	5.23	44,308	94.77	46,754	100.00
	Sideswipe, same direction	453	1.93	1,210	5.16	1,663	7.09	21,781	92.91	23,444	100.00
	Sideswipe, opposite direction	208	3.13	892	13.43	1,100	16.56	5,543	83.44	6,643	100.00
	Left turn	492	1.26	288	0.73	780	1.99	38,411	98.01	39,191	100.00
	Right turn	67	0.92	215	2.96	282	3.89	6,972	96.11	7,254	100.00
	Other	87	2.94	390	13.16	477	16.09	2,487	83.91	2,964	100.00
	Total	3,384	1.32	8,956	3.50	12,340	4.83	243,353	95.17	255,693	100.00
No collision with motor vehicle		9,410	6.27	40,972	27.30	50,382	33.57	99,719	66.43	150,101	100.00
Unknown		8	1.89	16	3.78	24	5.67	399	94.33	423	100.00
Total		**12,802**	**3.15**	**49,944**	**12.29**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 9. Frequency and number/percentage of SR crashes regarding type of collision in Ohio.

Type of Collision		Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
		n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
Collision with motor vehicle	Rear-end	1,236	0.90	4,530	3.29	5,766	4.19	131,817	95.81	137,583	100.00
	Head-on	205	2.88	714	10.03	919	12.91	6,202	87.09	7,121	100.00
	Backing up	10	0.11	70	0.80	80	0.91	8,668	99.09	8,748	100.00
	Angle	626	0.66	6,725	7.11	7,351	7.77	87,299	92.23	94,650	100.00
	Sideswipe, same direction	453	1.43	3,573	11.27	4,026	12.70	27,675	87.30	31,701	100.00
	Sideswipe, opposite direction	208	3.83	344	6.33	552	10.16	4,882	89.84	5,434	100.00
	Left turn										
	Right turn										
	Other										
	Total	2,738	0.96	15,956	5.59	18,694	6.55	266,543	93.45	285,237	100.00
No collision with motor vehicle		9,410	6.45	21,885	14.99	31,295	21.44	114,654	78.56	145,949	100.00
Unknown		8	0.05	1,909	11.33	1,917	11.37	14,938	88.63	16,855	100.00
Total		**12,156**	**2.71**	**39,750**	**8.87**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database

As can be seen, regardless of the database or the definition used, SR crashes are primarily single-vehicle crashes. The only exception is in the over speed limit definition used in Ohio where sideswipe same direction crashes had a higher SR percentage but a much lower frequency. This finding would infer multilane highways. The databases differed with respect to which vehicle-to-vehicle crashes were most likely to be SR. GES and FARS data indicated rear-end crashes, while the State data indicated both types of sideswipe and head-on crashes.

First Harmful Event

Table 10. Frequency and number/percentage of SR crashes regarding first harmful events.

First Harmful Event	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
Rollover/overturn	4,243	1,772	41.76	134,466	58,419	43.45
Jackknife	10	4	40.00	6,317	1,996	31.60
Other noncollision	529	117	22.12	51,431	3,091	6.01
Collision with nonmotorist	4,375	376	8.59	104,323	2,569	2.46
Collision with motor vehicle	15,458	3,490	22.58	4,170,075	776,635	18.62
Collision with animal	172	19	11.05	270,053	1,461	0.54
Collision with other nonfixed objects	363	66	18.18	44,868	5,259	11.72
Collision with ground, ditch, culvert, embankment or boulder	4,114	1,893	46.01	190,667	70,848	37.16
Collision with building, fence, wall, etc.	658	324	49.24	121,563	42,527	34.98
Collision with bush or tree	3,210	1,559	48.57	142,500	63,633	44.65
Collision with other fixed objects	4,399	1,932	43.92	470,975	166,728	35.40
Unknown	56	1	1.79	5,502	2,402	43.66
Total	**37,587**	**11,553**	**30.74**	**5,712,740**	**1,195,568**	**20.93**

Note: *n* represents frequency of crashes.

Table 11. Frequency and number/percentage of SR crashes regarding first harmful events in North Carolina.

First Harmful Event	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Rollover/ overturn	1,055	11.36	3,933	42.37	4,988	53.73	4,295	46.27	9,283	100.00
Jackknife	12	2.86	150	35.71	162	38.57	258	61.43	420	100.00
Other noncollision	87	2.10	343	8.29	430	10.39	3,709	89.61	4,139	100.00
Collision with nonmotorist	30	1.13	35	1.31	65	2.44	2,598	97.56	2,663	100.00
Collision with motor vehicle	3,466	1.34	9,572	3.70	13,038	5.04	245,603	94.96	258,641	100.00
Collision with animal	44	0.13	23	0.07	67	0.19	34,905	99.81	34,972	100.00
Collision with other nonfixed objects	69	1.25	259	4.71	328	5.96	5,172	94.04	5,500	100.00
Ran off road	1,901	8.84	7,254	33.73	9,155	42.57	12,350	57.43	21,505	100.00
Collision with other fixed objects	6,130	8.93	28,359	41.30	34,489	50.22	34,182	49.78	68,671	100.00
Not coded	8	1.89	16	3.78	24	5.67	399	94.33	423	100.00
Total	**12,802**	**3.15**	**49,944**	**12.29**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 12. Frequency and number/percentage of SR crashes regarding first harmful events in Ohio.

First Harmful Event	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Rollover/ overturn	127	10.03	257	20.30	384	30.33	882	69.67	1,266	100.00
Jackknife										
Other noncollision	1,451	10.99	893	6.77	2,344	17.76	10,855	82.24	13,199	100.00
Collision with nonmotorist	263	4.76	22	0.40	285	5.16	5,237	94.84	5,522	100.00
Collision with motor vehicle	15,642	5.48	3,063	1.07	18,705	6.55	266,834	93.45	285,539	100.00
Collision with animal	2,738	5.49	17	0.03	2,755	5.52	47,151	94.48	49,906	100.00
Collision with other nonfixed objects	562	11.21	129	2.57	691	13.78	4,323	86.22	5,014	100.00
Ran off road										
Collision with other fixed objects	8,301	11.73	16,524	23.36	24,825	35.09	45,915	64.91	70,740	100.00
Not coded	1,593	9.45	324	1.92	1,917	11.37	14,938	88.63	16,855	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Reinforcing the findings from table 7 through table 9, table 10 through table 12 show that regardless of database or definition, there was a consistent pattern indicating higher SR percentages and frequencies for various types of single-vehicle, off-roadway crashes. The SR percentage was much lower for vehicle-to-vehicle crashes and even lower for collisions with nonmotorists, which will be further examined in the next section.

Nonmotorist Involvement

Table 13. Frequency and number/percentage of SR crashes regarding nonmotorist involvement.

Nonmotorist Involvement	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
No pedestrians/bicycles	32,937	11,065	33.59	5,605,824	1,192,053	21.26
1+ pedestrians/bicycles	4,646	486	10.46			
Pedestrian accident				59,483	3,084	5.18
Bicycle accident				47,255	388	0.82
Wheelchair involved				177	45	25.42
Total	**37,583**	**11,551**	**30.73**	**5,712,739**	**1,195,570**	**20.93**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Table 14. Frequency and number/percentage of SR crashes regarding nonmotorist involvement in North Carolina.

Nonmotorist Involvement	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
No pedestrians/ bicycles	12,761	3.17	49,860	12.37	62,621	15.53	340,492	84.47	403,113	100.00
1+ pedestrians/ bicycles	41	1.32	84	2.71	125	4.03	2,979	95.97	3,104	100.00
Total	**12,802**	**3.15**	**49,944**	**12.29**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 15. Frequency and number/percentage of SR crashes regarding nonmotorist involvement in Ohio.

Nonmotorist Involvement	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
No pedestrians/ bicycles	30,549	6.88	21,205	4.78	51,754	11.66	392,027	88.34	443,781	100.00
1+ pedestrians/ bicycles	128	3.00	24	0.56	152	3.57	4,108	96.43	4,260	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes.

All databases and definitions of SR crashes indicated that the pedestrian and bicycle crashes were less likely to be SR than other types of crashes. The no pedestrians/bicycles crashes had two to five times the SR percentage. Note that in table 13, the GES and FARS data were probably more meaningful since the State-based findings were more rural; they did not include urban streets where many of the pedestrian and bicycle crashes would occur.

WHERE SR CRASHES OCCUR

In this section, tables are presented that show the frequencies and percentages of SR crashes for variables related to the location of crashes. They show the percentage of SR crashes by location of the first harmful event, relationship to junction, functional class, speed limit, number of travel lanes, AADT per lane, roadway alignment, roadway profile, and work zone.

Location of First Harmful Event

Table 16. Frequency and number/percentage of SR crashes regarding the location of the first harmful event.

Location of First Harmful Event	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
On roadway	20,936	4,241	20.26	4,628,618	790,870	17.09
Treated roadside	3,922	1,552	39.57	195,702	74,028	37.83
Untreated roadside	12,534	5,751	45.88	881,274	329,272	37.36
Unknown	195	9	4.62	7,145	1,400	19.59
Total	**37,587**	**11,553**	**30.74**	**5,712,739**	**1,195,570**	**20.93**

Note: *n* represents frequency of crashes.

Table 17. Frequency and number/percentage of SR crashes regarding the location of the first harmful event in North Carolina.

Location of First Harmful Event	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
On roadway	5,394	1.66	17,162	5.28	22,556	6.94	302,407	93.06	324,963	100.00
Off roadway										
Treated roadside	4,561	8.54	22,195	41.55	26,756	50.09	26,658	49.91	53,414	100.00
Untreated roadside	2,846	10.24	10,578	38.06	13,424	48.30	14,370	51.70	27,794	100.00
Unknown	1	2.17	9	19.57	10	21.74	36	78.26	46	100.00
Total	**12,802**	**3.15**	**49,944**	**12.29**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Table 18. Frequency and number/percentage of SR crashes regarding the location of the first harmful event in Ohio.

Location of First Harmful Event	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
On roadway	22,646	6.00	4,882	1.29	27,528	7.29	349,884	92.71	377,412	100.00
Off roadway	6,627	11.21	15,106	25.56	21,733	36.77	37,370	63.23	59,103	100.00
Treated roadside										
Untreated roadside										
Unknown	1,404	12.18	1,241	10.77	2,645	22.95	8,881	77.05	11,526	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

These tables again support the trend that off-roadway crashes (which are predominantly single vehicle) are more related to speed than on-roadway crashes.

Relationship to Junction

Table 19. Frequency and number/percentage of SR crashes regarding the relationship to the junction.

Relationship to Junction	FARS			GES		
	All Crashes	**SR Crashes**		**All Crashes**	**SR Crashes**	
	n	***n***	**Percent**	***n***	***n***	**Percent**
Not related to junction	26,966	9,354	34.69	2,491,891	672,697	27.00
Intersection related	8,078	1,510	18.69	2,340,179	383,900	16.40
Other junctions	1,347	291	21.60	601,811	66,438	11.04
Interchange area	1,007	391	38.83	210,542	64,248	30.52
Unknown	189	7	3.70	68,316	8,286	12.13
Total	**37,587**	**11,553**	**30.74**	**5,712,739**	**1,195,569**	**20.93**

Note: *n* represents frequency of crashes.

Table 20. Frequency and number/percentage of SR crashes regarding the relationship to the junction in North Carolina.

Relationship to Junction	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
Not related to junction	9,387	3.96	40,788	17.20	50,175	21.16	186,962	78.84	237,137	100.00
Intersection related	1,580	1.64	3,873	4.03	5,453	5.67	90,709	94.33	96,162	100.00
Interchange area	160	1.69	1,071	11.31	1,231	13.00	8,237	87.00	9,468	100.00
Other	1,206	4.26	2,384	8.41	3,590	12.67	24,750	87.33	28,340	100.00
Unknown	469	1.34	1,828	5.21	2,297	6.54	32,813	93.46	35,110	100.00
Total	**12,802**	**3.15**	**49,944**	**12.29**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 21. Frequency and number/percentage of SR crashes regarding the relationship to the junction in Ohio.

Relationship to Junction	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
Not related to junction	22,028	8.39	19,722	7.51	41,750	15.90	220,871	84.10	262,621	100.00
Intersection related	6,845	5.03	956	0.70	7,801	5.74	128,203	94.26	136,004	100.00
Interchange area	802	2.94	366	1.34	1,168	4.28	26,148	95.72	27,316	100.00
Other	851	4.32	143	0.73	994	5.05	18,705	94.95	19,699	100.00
Unknown	151	6.29	42	1.75	193	8.04	2,208	91.96	2,401	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes.

Both the North Carolina and Ohio data indicated that the "Not related to junction" category had the highest percentage of SR crashes. Since these crashes would be predominately run-off-road crashes, this supports the earlier findings showing high SR percentages and frequencies in off-roadway crashes. Interestingly, both the FARS and GES data indicated that the "Interchange area" category had the highest percentage of SR crashes. As noted previously, the category in both FARS and GES with the second highest percentage was "Not related to junction," which represented crashes occurring on roadway sections between intersections or interchanges and would contain off-roadway crashes. This category contained 10 times the number of interchange crashes in the GES file and 23 times the number of interchange crashes in the FARS data.

Functional Class

Table 22. Frequency and number/percentage of SR crashes regarding functional class.

Functional Class	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
Interstate				476,336	168,336	35.34
Noninterstate				5,234,875	1,027,500	19.63
Rural interstate	2,600	803	30.88			
Rural principal arterial other	3,866	901	23.31			
Rural minor arterial	3,736	1,052	28.16			
Rural major collector	5,008	1,767	35.28			
Rural minor collector	1,375	549	39.93			
Rural local road	3,754	1,490	39.69			
Unknown rural	240	117	48.75			
Urban interstate	2,231	711	31.87			
Urban freeway/ expressway	1,446	485	33.54			
Urban other principal arterial	4,400	1,003	22.80			
Urban minor arterial	3,206	915	28.54			
Urban collector	1,193	454	38.06			
Urban local street	2,884	994	34.47			
Unknown urban	72	30	41.67			
Unknown	1,576	282	17.89			
Total	**37,587**	**11,553**	**30.74**	**5,711,211**	**1,195,836**	**20.94**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Table 23. Frequency and number/percentage of SR crashes regarding functional class in North Carolina.

Functional Class	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
Rural principal arterial interstate	500	4.02	4	0.03	4,224	33.96	8,214	66.04	12,438	100.00
Rural principal arterial other	931	4.08	3,661	16.04	4,592	20.12	18,235	79.88	22,827	100.00
Rural minor arterial	703	3.13	2,454	10.93	3,157	14.06	19,296	85.94	22,453	100.00
Rural major collector	1,826	4.14	6,520	14.78	8,346	18.92	35,774	81.08	44,120	100.00
Rural minor collector	1,429	6.01	5,156	21.69	6,585	27.70	17,191	72.30	23,776	100.00
Rural local road or street	2,458	6.63	9,156	24.70	11,614	31.33	25,454	68.67	37,068	100.00
Urban principal arterial interstate	796	2.40	5,770	17.36	6,566	19.76	26,662	80.24	33,228	100.00
Urban principal arterial other freeway or expressway	235	2.05	1,338	11.70	1,573	13.75	9,863	86.25	11,436	100.00
Urban principal arterial other	862	1.02	2,770	3.29	3,632	4.32	80,504	95.68	84,136	100.00
Urban minor arterial	1,200	1.76	3,560	5.21	4,760	6.96	63,591	93.04	68,351	100.00
Urban collector	646	3.68	1,826	10.40	2,472	14.08	15,080	85.92	17,552	100.00
Urban local road or street	1,149	4.39	3,741	14.30	4,890	18.69	21,270	81.31	26,160	100.00
Unknown	67	2.51	268	10.03	335	12.54	2,337	87.46	2,672	100.00
Total	**12,802**	**3.15**	**46,224**	**11.38**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 24. Frequency and number/percentage of SR crashes regarding functional class in Ohio.

Functional Class	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Rural principal arterial interstate	3,591	16.65	3,606	16.72	7,197	33.37	14,369	66.63	21,566	100.00
Rural principal arterial other	2,280	7.97	2,262	7.91	4,542	15.89	24,051	84.11	28,593	100.00
Rural minor arterial	1,692	5.36	2,431	7.70	4,123	13.06	27,440	86.94	31,563	100.00
Rural major collector	3,165	5.62	6,299	11.19	9,464	16.82	46,815	83.18	56,279	100.00
Rural minor collector	188	5.15	582	15.94	770	21.09	2,881	78.91	3,651	100.00
Rural local road or street	4	7.55	9	16.98	13	24.53	40	75.47	53	100.00
Urban principal arterial interstate	7,525	10.48	3,153	4.39	10,678	14.87	61,142	85.13	71,820	100.00
Urban principal arterial other freeway or expressway	1,500	8.09	925	4.99	2,425	13.08	16,110	86.92	18,535	100.00
Urban principal arterial other	6,930	4.83	1,059	0.74	7,989	5.57	135,464	94.43	143,453	100.00
Urban minor arterial	3,358	5.22	662	1.03	4,020	6.2%	60,289	93.75	64,309	100.00
Urban collector	269	4.98	216	4.00	485	8.97	4,921	91.03	5,406	100.00
Urban local road or street	11	7.53	5	3.42	16	10.96	130	89.04	146	100.00
Unknown	164	6.15	20	0.75	184	6.90	2,483	93.10	2,667	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes.

There was some inconsistency in the findings in table 22 through table 24. GES did not have a functional class variable, but when all interstates were compared to all noninterstates, the interstates had a higher percentage of SR crashes than noninterstates (35 percent versus 20 percent). This differed from the FARS data in which the detailed functional classes were presented and in which higher SR crash percentages were found in the minor collector and local roads classes. In general, higher SR percentages were seen for rural roads versus urban roads when comparing similar rural and urban classes in both State databases and with both definitions. A much smaller difference was seen in the FARS data when all rural classes were combined and compared to the total SR for all urban classes combined (30 percent versus 28 percent). This could be because the State data did not include the local urban streets and thus included more rural roads in nature. For the State systems in table 23 and table 24, the combined definition showed that rural interstates had the highest percentage of SR crashes. However, the over speed limit definition of SR illustrated a difference between North Carolina and Ohio, which may reflect true differences in speeding. North Carolina data indicated that speeding over the posted limit was a bigger problem on local and minor roads, while in Ohio, speeding was a bigger problem on interstates and arterials. One possible explanation is that Ohio has lower posted speed limits than North Carolina (i.e., 65 mi/h versus 70 mi/h on rural roads and 55 mi/h versus 60 mi/h on urban roads), which might lead to more speeding over the posted limit.

Speed Limit

Table 25. Frequency and number/percentage of SR crashes regarding to speed limit.

Speed Limit (mi/h)	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	***n***	**Percent**	***n***	***n***	**Percent**
≤ 25	1,684	630	37.41	627,227	104,256	16.62
30–35	5,878	2,075	35.30	1,454,310	269,951	18.56
40–45	7,932	2,492	31.42	1,308,034	294,008	22.48
50–60	14,514	4,197	28.92	1,041,497	241,486	23.19
≥ 65	6,503	1,885	28.99	421,216	140,836	33.44
No limit	118	39	33.05	34,984	2,298	6.57
Unknown	953	231	24.24	823,860	142,735	17.33
Total	**37,582**	**11,549**	**30.73**	**5,711,128**	**1,195,570**	**20.93**

Note: *n* represents frequency of crashes.

Table 26. Frequency and number/percentage of SR crashes regarding speed limit in North Carolina.

Speed Limit (mi/h)	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
≤ 25	114	2.41	195	4.12	309	6.53	4,423	93.47	4,732	100.00
30–35	1,723	1.95	4,289	4.86	6,012	6.82	82,156	93.18	88,168	100.00
40–45	1,522	1.84	4,849	5.85	6,371	7.69	76,449	92.31	82,820	100.00
50–60	8,425	4.25	32,984	16.64	41,409	20.88	156,870	79.12	198,279	100.00
≥ 65	1,018	3.16	7,627	23.67	8,645	26.83	23,573	73.17	32,218	100.00
Unknown	0	0	0	0	0	0	0	0	0	0.00
Total	**12,802**	**3.15**	**49,944**	**12.29**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 27. Frequency and number/percentage of SR crashes regarding speed limit in Ohio.

Speed Limit (mi/h)	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
≤ 25	2928	7.75	106	0.28	3034	8.03	34,752	91.97	37786	100.00
30–35	5488	4.76	743	0.64	6231	5.41	108,970	94.59	115201	100.00
40–45	1951	3.81	1,319	2.58	3270	6.39	47,931	93.61	51201	100.00
50–60	11156	6.80	11,476	7.00	22632	13.80	141,373	86.20	164005	100.00
≥ 65	8986	11.64	7,565	9.80	16551	21.45	60,617	78.55	77168	100.00
Unknown	168	6.27	20	0.75	188	7.01	2,492	92.99	2,680	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes.

There is some inconsistency in the findings found in table 25 through table 27. In general, there was agreement between GES, North Carolina, and Ohio when the combined definition was used—the percentage of SR crashes increased as speed limit increased. The fatal crashes were different with FARS, showing a less clear trend and a slightly higher percentage of SR crashes for the lower limits (i.e., 35 mi/h and below). The over speed limit definition produced different results in the two States, with North Carolina showing a higher percentage of SR crashes for the 50–60 mi/h speed limits and Ohio showing a higher percentage of SR crashes for the ≥ 65 mi/h speed limits.

Number of Travel Lanes

Table 28. Frequency and number/percentage of SR crashes regarding number of travel lanes.

Number of Travel Lanes	FARS			GES		
	All Crashes	**SR Crashes**		**All Crashes**	**SR Crashes**	
	n	***n***	**Percent**	***n***	***n***	**Percent**
One lane	455	201	44.18	140,060	36,464	26.03
Two lanes	21,922	7,286	33.24	1,905,333	430,387	22.59
Three lanes	510	136	26.67	253,311	48,508	19.15
Four lanes	8,840	2,352	26.61	949,315	216,818	22.84
Five lanes	196	65	33.16	403,429	100,090	24.81
Six lanes	2,367	626	26.45	453,888	101,775	22.42
Seven or more lanes	2,806	777	27.69	345,821	92,127	26.64
Unknown lanes	491	110	22.40	1,261,582	169,399	13.43
Total	**37,587**	**11,553**	**30.74**	**5,712,739**	**1,195,568**	**20.93**

Note: *n* represents frequency of crashes.

Table 29. Frequency and number/percentage of SR crashes regarding number of travel lanes in North Carolina.

Number of Travel Lanes	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
≤ Two lanes	9,354	4.20	4	0.00	42,085	18.90	180,563	81.10	222,648	100.00
Three lanes	45	1.21	178	4.80	223	6.01	3,487	93.99	3,710	100.00
Four lanes	2,798	1.95	12,824	8.95	15,622	10.91	127,610	89.09	143,232	100.00
Five lanes	14	0.87	74	4.60	88	5.48	1,519	94.52	1,607	100.00
Six lanes	419	1.53	2,512	9.16	2,931	10.68	24,506	89.32	27,437	100.00
Seven lanes	20	3.11	185	28.77	205	31.88	438	68.12	643	100.00
≥ Eight lanes	152	2.19	1,440	20.75	1,592	22.94	5,348	77.06	6,940	100.00
Unknown	0	0	0	0	0	0	0	0	0	0.00
Total	**12,802**	**3.15**	**17,217**	**4.24**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 30. Frequency and number/percentage of SR crashes regarding number of travel lanes in Ohio.

Number of Travel Lanes	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
≤ Two lanes	9,206	5.30	11,129	6.40	20,335	11.70	153,431	88.30	173,766	100.00
Three lanes	194	5.19	61	1.63	255	6.82	3,484	93.18	3,739	100.00
Four lanes	14,002	7.10	7,249	3.67	21,251	10.77	176,006	89.23	197,257	100.00
Five lanes	334	6.49	118	2.29	452	8.78	4,694	91.22	5,146	100.00
Six lanes	5,325	10.84	2,418	4.92	7,743	15.76	41,385	84.24	49,128	100.00
Seven lanes	190	7.78	30	1.23	220	9.01	2,221	90.99	2,441	100.00
≥ Eight lanes	1,258	9.06	204	1.47	1,462	10.53	12,422	89.47	13,884	100.00
Unknown	168	6.27	20	0.75	188	7.01	2,492	92.99	2,680	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes.

The findings in table 28 through table 30 were not consistent across databases. FARS and GES data showed little difference between lane counts. FARS had a slightly higher percentage of SR crashes for two-lane roads, and GES had a slightly higher percentage of SR crashes for seven or more lanes (see table 28). The State data paralleled what was seen in functional class. In North Carolina, using the combined definition of SR, the multilane roads had a higher percentage of SR crashes. However, when using the over speed limit definition of SR, two-lane roads had a higher percentage of SR crashes. In Ohio, the highest percentage of SR crashes was on multilane roads.

AADT Per Lane

Table 31. Frequency and number/percentage of SR crashes regarding AADT per lane in North Carolina.

AADT Per Lane	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
0–2,000	6,693	5.72	4	0.00	30,392	25.99	86,535	74.01	116,927	100.00
2,001–4,000	2,201	2.88	7,674	10.05	9,875	12.93	66,480	87.07	76,355	100.00
4,001–5,500	1,206	2.09	4,197	7.28	5,403	9.37	52,249	90.63	57,652	100.00
5,501–9,000	1,359	1.60	6,097	7.17	7,456	8.77	77,597	91.23	85,053	100.00
9,001–13,000	670	1.86	4,310	11.98	4,980	13.84	30,990	86.16	35,970	100.00
13,001–20,000	458	1.93	2,886	12.15	3,344	14.08	20,411	85.92	23,755	100.00
> 20,000	183	1.83	990	9.89	1,173	11.72	8,834	88.28	10,007	100.00
Unknown	32	6.43	91	18.27	123	24.70	375	75.30	498	100.00
Total	**12,802**	**3.15**	**26,249**	**6.46**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 32. Frequency and number/percentage of SR crashes regarding AADT per lane in Ohio.

AADT per Lane	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
0 - 2,000	3,861	5.77	7,445	11.12	11,306	16.89	55,616	83.11	66,922	100.00
2,001–4,000	6,597	6.37	4,272	4.13	10,869	10.50	92,677	89.50	103,546	100.00
4,001–5,500	4,365	5.68	1,964	2.56	6,329	8.24	70,506	91.76	76,835	100.00
5,501–9,000	6,129	6.35	3,531	3.66	9,660	10.01	86,846	89.99	96,506	100.00
9,001–13,000	3,851	8.86	2,356	5.42	6,207	14.29	37,241	85.71	43,448	100.00
13,001–20,000	4,221	9.94	1,412	3.32	5,633	13.26	36,842	86.74	42,475	100.00
> 20,000	1,485	9.50	229	1.47	1,714	10.97	13,915	89.03	15,629	100.00
Unknown	168	6.27	20	0.75	188	7.01	2,492	92.99	2,680	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes.

Since FARS and GES do not include AADT data, this traffic density-related variable can only be examined with State data. Table 31 and table 32 show that the findings from the two States differed. In North Carolina, while using both definitions of SR, the results showed that the lower the AADT, the higher the percentage of SR crashes. This may well reflect road type because lower volume, rural roads experience higher SR percentages. In Ohio, using the combined definition of SR, the same finding as North Carolina was found; however, the over speed limit definition showed the opposite finding—a higher percentage of SR crashes was found for roads with high AADTs. This may reflect the earlier Ohio finding concerning a higher percentage of SR crashes on interstates.

Roadway Alignment

Table 33. Frequency and number/percentage of SR crashes regarding roadway alignment.

Roadway Alignment	FARS All Crashes	FARS SR Crashes		GES All Crashes	GES SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
Straight	27,286	6,877	25.20	4,585,418	886,794	19.34
Curve	9,930	4,610	46.42	644,542	215,821	33.48
Unknown	371	66	17.79	482,779	92,955	19.25
Total	**37,587**	**11,553**	**30.74**	**5,712,739**	**1,195,570**	**20.93**

Note: *n* represents frequency of crashes.

Table 34. Frequency and number/percentage of SR crashes regarding roadway alignment in North Carolina.

Category	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Straight	6,989	2.18	4	0.00	35,415	11.04	285,322	88.96	320,737	100.00
Curve	5,705	9.73	21,081	35.95	26,786	45.68	31,852	54.32	58,638	100.00
Unknown	108	0.40	437	1.63	545	2.03	26,297	97.97	26,842	100.00
Total	**12,802**	**3.15**	**21,522**	**5.30**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 35. Frequency and number/percentage of SR crashes regarding roadway alignment in Ohio.

Category	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Straight	26,270	6.52	13,717	3.40	39,987	9.93	362,903	90.07	402,890	100.00
Curve	4,279	9.90	7,421	17.16	11,700	27.06	31,538	72.94	43,238	100.00
Unknown	128	6.69	91	4.76	219	11.45	1,694	88.55	1,913	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes.

Regardless of database or definition, curves exhibit SR crash percentages that are three to four times higher than tangents. It is important to recognize that the curve percentage remains higher even using the over speed limit definition of SR.

Roadway Profile

Table 36. Frequency and number/percentage of SR crashes regarding roadway profile.

Roadway Profile	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
Level	26,803	7,734	28.85	3,044,997	494,713	16.25
Grade	9,072	3,313	36.52	799,678	190,204	23.79
Hillcrest	841	278	33.06	77,082	17,499	22.70
Sag	112	49	43.75	8,874	3,858	43.48
Unknown	759	179	23.58	1,782,109	489,297	27.46
Total	**37,587**	**11,553**	**30.74**	**5,712,740**	**1,195,571**	**20.93**

Note: *n* represents frequency of crashes.

Table 37. Frequency and number/percentage of SR crashes regarding roadway profile in North Carolina.

Roadway Profile	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Level	8,421	3.00	4	0.00	39,386	14.03	241,333	85.97	280,719	100.00
Grade	3,455	4.36	5	0.01	18,971	23.93	60,296	76.07	79,267	100.00
Hillcrest	533	3.81	6	0.04	2,353	16.82	11,636	83.18	13,989	100.00
Sag	285	5.28	1,206	22.33	1,491	27.61	3,909	72.39	5,400	100.00
Unknown	108	0.40	437	1.63	545	2.03	26,297	97.97	26,842	100.00
Total	**12,802**	**3.15**	**1,658**	**0.41**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 38. Frequency and number/percentage of SR crashes regarding roadway profile in Ohio.

Roadway Profile	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Level	23,046	6.67	12,008	3.47	35,054	10.14	310,656	89.86	345,710	100.00
Grade	7,503	7.47	9,130	9.09	16,633	16.56	83,785	83.44	100,418	100.00
Hillcrest										
Sag										
Unknown	128	6.69	91	4.76	219	11.45	1,694	88.55	1,913	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Regardless of database or definition used, there was a higher percentage of SR crashes on grades than on level roadway sections as seen in table 36 through table 38. In addition, in the three databases (FARS, GES, and North Carolina) that included categories for "Sags" and "Hillcrest" (i.e., the two ways that two grades can join), sags showed an even higher percentage of SR crashes than grades. This is logical because sags are often located at the bottom of a downgrade where speeds might be highest.

Work Zone

Table 39. Frequency and number/percentage of SR crashes regarding work zone.

	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
Work Zone	***n***	***n***	**Percent**	***n***	***n***	**Percent**
Construction	721	214	29.68	56,367	14,297	25.36
Maintenance	86	23	26.74			
Utility	16	3	18.75			
Unknown work zone	80	14	17.50			
Unknown				40,582	7,242	17.85
Yes, first harmful event related to but not in work/ construction zone				5,407	1,493	27.61
Unknown if first harmful event in or related to work/construction zone				30,297	8,426	27.81
None	36,684	11,299	30.80	5,580,085	1,164,113	20.86
Total	**37,587**	**11,553**	**30.74**	**5,712,738**	**1,195,571**	**20.93**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Table 40. Frequency and number/percentage of SR crashes regarding work zone in North Carolina.

	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
Work Zone	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
Construction/ maintenance										
Construction	211	2.01	930	8.87	1,141	10.88	9,347	89.12	10,488	100.00
maintenance	20	2.06	102	10.52	122	12.58	848	87.42	970	100.00
Utility	2	0.73	11	4.01	13	4.74	261	95.26	274	100.00
Intermittent/ moving	4	1.38	19	6.55	23	7.93	267	92.07	290	100.00
Not work zone	12,565	3.19	48,882	12.40	61,447	15.59	332,748	84.41	394,195	100.00
Total	**12,802**	**3.15**	**49,944**	**12.29**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Table 41. Frequency and number/percentage of SR crashes regarding work zone in Ohio.

	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
Work Zone	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
Construction/ maintenance	1435	11.71	355	2.90	1790	14.61	10,465	85.39	12255	100.00
Construction										
Maintenance										
Utility										
Intermittent/ moving										
Not work zone	29,242	6.71	20,874	4.79	50,116	11.50	385,670	88.50	435,786	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

There was no consistent pattern in the results in table 39 through table 41. FARS showed essentially no difference between work and nonwork zones, GES showed only a slightly higher percentage of SR crashes in the work zones (25 percent versus 21 percent), and the State findings differed from each other. North Carolina showed a lower percentage while Ohio showed a higher percentage of SR crashes for work zones under both definitions of SR.

WHEN SR CRASHES ARE MOST LIKELY TO OCCUR

This section presents tables showing when SR crashes are most likely to occur. It shows the percentage of SR crashes by light condition, road surface, atmospheric condition, season, and day of the week.

Light Condition

Table 42. Frequency and number/percentage of SR crashes regarding light condition.

Light Condition	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
Daylight	18,973	4,950	26.09	3,978,106	805,809	20.26
Dark	11,215	4,124	36.77	671,326	168,480	25.10
Dark but lighted	5,564	1,963	35.28	813,033	169,334	20.83
Dawn	710	219	30.85	86,112	20,785	24.14
Dusk	848	263	31.01	131,615	27,809	21.13
Unknown	277	34	12.27	32,548	3,353	10.30
Total	**37,587**	**11,553**	**30.74**	**5,712,740**	**1,195,570**	**20.93**

Note: *n* represents frequency of crashes.

Table 43. Frequency and number/percentage of SR crashes regarding light condition in North Carolina.

	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
Category	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Daylight	6,416	2.34	30,270	11.04	36,686	13.38	237,423	86.62	274,109	100.00
Dark	4,956	5.83	15,348	18.06	20,304	23.89	64,691	76.11	84,995	100.00
Dark but lighted	878	2.95	1,866	6.27	2,744	9.22	27,026	90.78	29,770	100.00
Dawn	254	3.61	1,237	17.57	1,491	21.17	5,551	78.83	7,042	100.00
Dusk	267	2.77	1,115	11.58	1,382	14.35	8,249	85.65	9,631	100.00
Unknown	31	4.63	108	16.12	139	20.75	531	79.25	670	100.00
Total	**12,802**	**3.15**	**49,944**	**12.29**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 44. Frequency and number/percentage of SR crashes regarding light condition in Ohio.

	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
Category	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Daylight	16,918	5.78	11,895	4.07	28,813	9.85	263,681	90.15	292,494	100.00
Dark	6,007	8.33	6,875	9.53	12,882	17.86	59,259	82.14	72,141	100.00
Dark but lighted	6,031	10.62	1,207	2.12	7,238	12.74	49,567	87.26	56,805	100.00
Dawn	648	6.79	627	6.57	1,275	13.35	8,273	86.65	9,548	100.00
Dusk	700	6.41	443	4.06	1,143	10.46	9,781	89.54	10,924	100.00
Unknown	373	6.09	182	2.97	555	9.06	5,574	90.94	6,129	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes.

With minor exceptions, all databases and definitions in table 42 through table 44 indicated that crashes at night had the highest percentage of SR crashes. The findings in the “Dark but lighted” category varied between databases.

Surface Condition

Table 45. Frequency and number/percentage of SR crashes regarding surface condition.

Surface Condition	FARS All Crashes	FARS SR Crashes		GES All Crashes	GES SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
Dry	31,241	9,196	29.44	4,401,379	774,151	17.59
Wet	4,768	1,627	34.12	866,361	234,023	27.01
Snow or slush	569	299	52.55	209,910	103,822	49.46
Ice	547	317	57.95	153,079	69,146	45.17
Sand, dirt, or oil	48	28	58.33	7,336	2,114	28.82
Other				9,231	3,738	40.49
Unknown				65,444	8,575	13.10
Total	**37,587**	**11,553**	**30.74**	**5,712,740**	**1,195,569**	**20.93**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Table 46. Frequency and number/percentage of SR crashes regarding surface condition in North Carolina.

Surface Condition	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Dry	10,791	3.50	15,558	5.04	26,349	8.54	282,322	91.46	308,671	100.00
Wet	1,868	2.38	21,025	26.75	22,893	29.13	55,703	70.87	78,596	100.00
Snow or slush	24	0.30	5,845	73.30	5,869	73.60	2,105	26.40	7,974	100.00
Ice	73	0.73	7,304	73.34	7,377	74.07	2,582	25.93	9,959	100.00
Sand, dirt, oil, or other	20	4.36	170	37.04	190	41.39	269	58.61	459	100.00
Unknown	26	4.66	42	7.53	68	12.19	490	87.81	558	100.00
Total	**12,802**	**3.15**	**49,944**	**12.29**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 47. Frequency and number/percentage of SR crashes regarding surface condition in Ohio.

Surface Condition	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
Dry	22,418	7.39	4,231	1.40	26,649	8.79	276,565	91.21	303,214	100.00
Wet	6,768	6.66	6,388	6.28	13,156	12.94	88,517	87.06	101,673	100.00
Snow or slush	753	2.91	6,398	24.72	7,151	27.63	18,734	72.37	25,885	100.00
Ice	359	3.22	3,259	29.26	3,618	32.48	7,521	67.52	11,139	100.00
Sand, dirt, oil, or other	211	6.26	900	26.72	1,111	32.99	2,257	67.01	3,368	100.00
Unknown	168	6.08	53	1.92	221	8.00	2,541	92.00	2,762	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes.

The surface condition variable was the first of two variables (the second being weather) in which the findings from the combined versus over speed limit definitions consistently differed (see table 45 through table 47). The combined definition for all databases indicated higher SR crash percentages for snow and ice conditions. In contrast, the over speed limit definition for both States showed that the dry condition had the highest percentage of SR crashes. This is not unexpected, as an investigating officer may be likely to conclude that most drivers go too fast for conditions even without speeding per se in snowy and icy conditions. One conclusion that might be drawn is that SR crashes in snowy and icy conditions might be better treated by roadway (e.g., snow and ice removal) or vehicles treatments rather than by enforcement or education efforts.

Weather

Table 48. Frequency and number/percentage of SR crashes regarding weather.

Weather	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
No adverse atmospheric conditions	33,036	9,907	29.99	4,808,369	899,217	18.70
Rain	2,805	992	35.37	547,389	156,006	28.50
Sleet	131	71	54.20	17,339	7,019	40.48
Snow	616	307	49.84	230,049	115,887	50.37
Fog	420	145	34.52	22,127	4,862	21.97
Rain and fog	31	13	41.94	1,819	359	19.74
Sleet and fog	10	7	70.00	503	241	47.91
Other; smog, smoke, blowing sand, snow, dust, crosswind, etc.	184	50	27.17	31,065	6,962	22.41
Unknown	354	61	17.23	54,079	5,017	9.28
Total	**37,587**	**11,553**	**30.74**	**5,712,739**	**1,195,570**	**20.93**

Note: *n* represents frequency of crashes.

Table 49. Frequency and number/percentage of SR crashes regarding weather in North Carolina.

Weather	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
No adverse atmospheric conditions	11,410	3.44	22,922	6.91	34,332	10.34	297,589	89.66	331,921	100.00
Rain	1,164	2.01	17,773	30.67	18,937	32.68	39,003	67.32	57,940	100.00
Sleet	21	0.58	2,528	69.55	2,549	70.12	1,086	29.88	3,635	100.00
Snow	26	0.32	5,636	70.04	5,662	70.36	2,385	29.64	8,047	100.00
Fog	151	4.18	799	22.14	950	26.32	2,659	73.68	3,609	100.00
Rain and fog	11	2.60	170	40.19	181	42.79	242	57.21	423	100.00
Sleet and fog	3	9.68	18	58.06	21	67.74	10	32.26	31	100.00
Other; smoke, dust, or wind	16	2.62	98	16.04	114	18.66	497	81.34	611	100.00
Unknown										
Total	**12,802**	**3.15**	**49,944**	**12.29**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Table 50. Frequency and number/percentage of SR crashes regarding weather in Ohio.

Weather	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
No adverse atmospheric conditions	24,925	7.23	8,027	2.33	32,952	9.56	311,655	90.44	344,607	100.00
Rain	3,691	6.69	3,670	6.65	7,361	13.35	47,794	86.65	55,155	100.00
Sleet										
Snow	1,285	3.73	7,671	22.28	8,956	26.01	25,480	73.99	34,436	100.00
Fog	259	7.11	281	7.71	540	14.81	3,105	85.19	3,645	100.00
Rain and fog										
Sleet and fog										
Other; smoke, dust, or wind	278	4.47	1,489	23.94	1,767	28.41	4,453	71.59	6,220	100.00
Unknown	239	6.01	91	2.29	330	8.30	3,648	91.70	3,978	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

The findings from table 48 through table 50 support those from table 45 through table 47. Using the combined definition of SR (predominantly too fast for conditions), adverse weather conditions had a higher SR crash percentage. Using the over speed limit definition, nonadverse conditions had a higher SR crash percentage.

Season

Table 51. Frequency and number/percentage of SR crashes regarding season.

Season	FARS			GES		
	All Crashes	**SR Crashes**		**All Crashes**	**SR Crashes**	
	n	***n***	**Percent**	***n***	***n***	**Percent**
Winter	8,269	2,540	30.72	1,484,529	355,032	23.92
Spring	9,085	2,813	30.96	1,395,434	278,033	19.92
Summer	10,224	3,153	30.84	1,379,243	275,118	19.95
Fall	10,009	3,047	30.44	1,453,533	287,387	19.77
Total	**37,587**	**11,553**	**30.74**	**5,712,739**	**1,195,570**	**20.93**

Note: *n* represents frequency of crashes.

Table 52. Frequency and number/percentage of SR crashes regarding season in North Carolina.

Season	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Winter	2,684	2.61	19,791	19.24	22,475	21.85	80,405	78.15	102,880	100.00
Spring	3,438	3.57	9,085	9.42	12,523	12.99	83,889	87.01	96,412	100.00
Summer	3,463	3.53	11,077	11.28	14,540	14.80	83,695	85.20	98,235	100.00
Fall	3,217	2.96	9,991	9.19	13,208	12.15	95,482	87.85	108,690	100.00
Total	**12,802**	**3.15**	**49,944**	**12.29**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 53. Frequency and number/percentage of SR crashes regarding season in Ohio.

Season	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Winter	7,049	5.78	11,304	9.27	18,353	15.05	103,560	84.95	121,913	100.00
Spring	7,584	7.31	4,101	3.95	11,685	11.26	92,061	88.74	103,746	100.00
Summer	7,803	7.53	2,545	2.46	10,348	9.99	93,233	90.01	103,581	100.00
Fall	8,241	6.94	3,279	2.76	11,520	9.70	107,281	90.30	118,801	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes.

While GES and FARS data in table 51 showed essentially no differences between seasons for total SR, the State data using the same definition in table 52 and table 53 showed that winter had a higher SR percentage. This could result from the States having more snowy and icy conditions in winter and thus more SR crashes resulting from driving too fast for the condition than the average State represented in FARS and GES. The over speed limit category showed slight differences between seasons.

Day of the Week

Table 54. Frequency and number/percentage of SR crashes regarding day of the week.

Day of the Week	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
Weekend	12,842	4,492	34.98	1,320,236	289,867	21.96
Weekday	24,735	7,059	28.54	4,392,503	905,703	20.62
Unknown	10	2	20.00	0	0	0.00
Total	**37,587**	**11,553**	**30.74**	**5,712,739**	**1,195,570**	**20.93**

Note: *n* represents frequency of crashes.

Table 55. Frequency and number/percentage of SR crashes regarding day of the week in North Carolina.

Day of the Week	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
Weekend	4,583	4.60	14,938	15.00	19,521	19.60	80,097	80.40	99,618	100.00
Weekday	8,219	2.68	35,006	11.42	43,225	14.10	263,374	85.90	306,599	100.00
Total	**12,802**	**3.15**	**49,944**	**12.29**	**62,746**	**15.45**	**343,471**	**84.55**	**406,217**	**100.00**

Note: *n* represents frequency of crashes.

Table 56. Frequency and number/percentage of SR crashes regarding day of the week in Ohio.

Day of the Week	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
Weekend	9,065	8.75	6,728	6.50	15,793	15.25	87,790	84.75	103,583	100.00
Weekday	21,612	6.27	14,501	4.21	36,113	10.48	308,345	89.52	344458	100.00
Total	**30,677**	**6.85**	**21,229**	**4.74**	**51,906**	**11.59**	**396,135**	**88.41**	**448,041**	**100.00**

Note: *n* represents frequency of crashes.

As seen in table 54 through table 56, the percentage of SR crashes was highest on the weekends except for GES in which no difference was seen.

WHO IS MORE LIKELY TO BE INVOLVED IN SR CRASHES

In this section, driver and vehicle variables in SR crashes were examined. Note that while the previously shown tables were based on counts of crashes, these tables focus on counts of drivers or vehicles in crashes to correctly capture situations in which a crash includes a speeding and a nonspeeding driver. The percentage of SR crashes are examined by driver age, driver gender, restraint use, driver distraction, driver physical impairment, driver alcohol use, driver drug use, prior speeding convictions, driver license restriction compliance, vehicle type, and hazardous cargo.

Driver Age

Table 57. Frequency and number/percentage of SR crashes regarding driver age.

Driver Age (Years)	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
16–19	5,523	1,868	33.82	1,267,554	241,674	19.07
20–25	9,272	2,959	31.91	1,729,591	255,737	14.79
26–35	10,781	2,510	23.28	2,055,135	230,252	11.20
36–50	15,321	2,534	16.54	2,735,097	260,753	9.53
51–70	11,264	1,321	11.73	1,782,612	127,775	7.17
71+	4,513	407	9.02	775,128	89,141	11.50
Unknown	46	13	28.26	436	364	83.49
Total	**56,720**	**11,612**	**20.47**	**10,345,553**	**1,205,696**	**11.65**

Note: *n* represents frequency of crashes.

Table 58. Frequency and number/percentage of SR crashes regarding driver age in North Carolina.

Driver Age (Years)	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
16–19	3,122	3.77	9,449	11.40	12,571	15.16	70,338	84.84	82,909	100.00
20–25	3,423	2.83	11,595	9.60	15,018	12.43	105,820	87.57	120,838	100.00
26–35	2,550	1.68	10,997	7.26	13,547	8.94	137,991	91.06	151,538	100.00
36–50	1,867	1.02	10,196	5.59	12,063	6.61	170,297	93.39	182,360	100.00
51–70	565	0.50	4,487	3.93	5,052	4.43	109,030	95.57	114,082	100.00
71+	63	0.22	562	1.98	625	2.21	27,706	97.79	28,331	100.00
Unknown	1,295	5.60	3,055	13.21	4,350	18.80	18,783	81.20	23,133	100.00
Total	**12,885**	**1.83**	**50,341**	**7.16**	**63,226**	**8.99**	**639,965**	**91.01**	**703,191**	**100.00**

Note: *n* represents frequency of crashes.

Table 59. Frequency and number/percentage of SR crashes in relation to driver age in Ohio.

Driver Age (Years)	Over Speed Limit		Too Fast For Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
16–19	5,693	6.51	3,589	4.11	9,282	10.62	78,145	89.38	87,427	100.00
20–25	7,772	6.41	4,671	3.85	12,443	10.26	108,886	89.74	121,329	100.00
26–35	6,528	4.41	4,499	3.04	11,027	7.46	136,877	92.54	147,904	100.00
36–50	6,608	3.07	5,129	2.38	11,737	5.45	203,745	94.55	215,482	100.00
51–70	3,407	2.35	2,791	1.93	6,198	4.28	138,607	95.72	144,805	100.00
71+	662	1.81	371	1.01	1,033	2.82	35,632	97.18	36,665	100.00
Unknown	2,213	7.84	356	1.26	2,569	9.10	25,675	90.90	28,244	100.00
Total	**32,883**	**4.21**	**21,406**	**2.74**	**54,289**	**6.94**	**727,567**	**93.06**	**781,856**	**100.00**

Note: *n* represents frequency of crashes.

Regardless of database or SR definition used, the percentage of SR crashes was highest for the youngest drivers (i.e., 16–19 years old) and decreased with age (see table 57 through table 59). It should be noted, however, that the 20–25-year-old category was also consistently higher than the categories with older drivers.

Driver Gender

Table 60. Frequency and number/percentage of SR crashes regarding drive gender.

Driver Gender	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
Male	41,742	9,405	22.53	5,870,675	736,951	12.55
Female	14,578	2,114	14.50	4,312,121	436,702	10.13
Unknown	400	93	23.25	162,758	32,043	19.69
Total	**56,720**	**11,612**	**20.47**	**10,345,554**	**1,205,696**	**11.65**

Note: *n* represents frequency of crashes.

Table 61. Frequency and number/percentage of SR crashes regarding drive gender in North Carolina.

Driver Gender	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Male	8,709	2.20	30,417	7.67	39,126	9.87	357,240	90.13	396,366	100.00
Female	2,882	1.02	16,884	5.95	19,766	6.97	263,933	93.03	283,699	100.00
Unknown	1,294	5.60	3,040	13.15	4,334	18.74	18,792	81.26	23,126	100.00
Total	**12,885**	**1.83**	**50,341**	**7.16**	**63,226**	**8.99**	**639,965**	**91.01**	**703,191**	**100.00**

Note: *n* represents frequency of crashes.

Table 62. Frequency and number/percentage of SR crashes regarding drive gender in Ohio.

Driver Gender	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Male	20,707	4.71	12,643	2.88	33,350	7.59	405,946	92.41	439,296	100.00
Female	10,080	3.21	8,397	2.67	18,477	5.88	295,516	94.12	313,993	100.00
Unknown	2,096	7.34	366	1.28	2,462	8.62	26,105	91.38	28,567	100.00
Total	**32,883**	**4.21**	**21,406**	**2.74**	**54,289**	**6.94**	**727,567**	**93.06**	**781,856**	**100.00**

Note: *n* represents frequency of crashes.

All databases and definitions seen in table 60 through table 62 showed a higher percentage of SR crashes for male drivers versus female drivers, with GES showing the smallest difference.

Restraint Use

Table 63. Frequency and number/percentage of SR crashes regarding driver restraint.

Driver Restraint	FARS All Crashes	FARS SR Crashes		GES All Crashes	GES SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
None used or not available	17,592	5,592	31.79	300,304	61,274	20.40
Shoulder belt only	256	45	17.58	43,444	4,002	9.21
Lap belt only	496	67	13.51	101,938	7,077	6.94
Lap and shoulder	31,517	3,947	12.52	8,350,726	949,369	11.37
Motorcycle helmet	2,543	990	38.93	57,665	13,041	22.62
Used, type unknown	110	20	18.18			
Belt use improper	56	15	26.79			
Helmet use improper	60	21	35.00			
None available				2,948	185	6.28
Restraint used/specifics unknown or other				536,973	34,911	6.50
Unknown	4,090	915	22.37	951,556	135,836	14.28
Total	**56,720**	**11,612**	**20.47**	**10,345,554**	**1,205,695**	**11.65**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Table 64. Frequency and number/percentage of SR crashes regarding driver restraint in North Carolina.

Driver Restraint	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
None used	1,055	7.93	1,465	11.01	2,520	18.93	10,789	81.07	13,309	100.00
Motorcycle helmet	276	8.55	455	14.10	731	22.65	2,496	77.35	3,227	100.00
Lap and shoulder belt	5,790	1.38	27,377	6.53	33,167	7.92	385,847	92.08	419,014	100.00
Lap belt only	3,056	1.39	15,671	7.12	18,727	8.51	201,295	91.49	220,022	100.00
Shoulder belt only	59	1.38	234	5.49	293	6.87	3,971	93.13	4,264	100.00
Unknown	2,649	6.11	5,139	11.85	7,788	17.96	35,567	82.04	43,355	100.00
Total	**12,885**	**1.83**	**50,341**	**7.16**	**63,226**	**8.99**	**639,965**	**91.01**	**703,191**	**100.00**

Note: *n* represents frequency of crashes.

Table 65. Frequency and number/percentage of SR crashes regarding driver restraint in Ohio.

Driver Restraint	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
None used	2,548	11.38	1,210	5.40	3,758	16.79	18,630	83.21	22,388	100.00
Motorcycle helmet	189	9.49	189	9.49	378	18.99	1,613	81.01	1,991	100.00
Lap and shoulder belt	25,698	3.74	19,216	2.80	44,914	6.53	642,372	93.47	687,286	100.00
Lap belt only	148	3.06	81	1.67	229	4.73	4,609	95.27	4,838	100.00
Shoulder belt only	161	4.82	49	1.47	210	6.28	3,132	93.72	3,342	100.00
Unknown	4,139	6.67	661	1.07	4,800	7.74	57,211	92.26	62,011	100.00
Total	**32,883**	**4.21**	**21,406**	**2.74**	**54,289**	**6.94**	**727,567**	**93.06**	**781,856**	**100.00**

Note: *n* represents frequency of crashes.

All databases in table 63 through table 65 showed that the highest percentage of SR crashes was for those who did not use restraints (at least twice as high) with the over speed limit category for North Carolina and Ohio showing the largest percentage—three to four times higher than other categories. Note that the "Motorcycle helmet" category was an indicator vehicle type, which is examined in a later section.

Driver Distraction

Table 66. Frequency and number/percentage of SR crashes regarding driver distraction.

Driver Distraction	GES		
	All Crashes	**SR Crashes**	
	n	***n***	**Percent**
Not distracted	4,874,677	467,828	9.60
Inattentive	676,562	150,267	22.21
Persons/objects inside or outside of vehicle	125,094	25,600	20.46
The driver is doing something	90,832	20,326	22.38
Sleepy or fell asleep	82,810	16,851	20.35
Unknown	4,495,579	524,825	11.67
Total	**10,345,554**	**1,205,697**	**11.65**

Note: *n* represents frequency of crashes.

Table 67. Frequency and number/percentage of SR crashes regarding driver distraction in North Carolina.

Driver Distraction	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Not distracted	11,365	1.86	45,387	7.42	56,752	9.28	555,026	90.72	611,778	100.00
Inattentive	1,209	1.43	4,411	5.23	5,620	6.67	78,648	93.33	84,268	100.00
Driver distracted	38	1.63	89	3.81	127	5.44	2,207	94.56	2,334	100.00
Sleepy or fell asleep	273	5.67	454	9.44	727	15.11	4,084	84.89	4,811	100.00
Unknown	0	0	0	0	0	0	0	0	0	0.00
Total	**12,885**	**1.83**	**50,341**	**7.16**	**63,226**	**8.99**	**639,965**	**91.01**	**703,191**	**100.00**

Note: *n* represents frequency of crashes.

Table 68. Frequency and number/percentage of SR crashes regarding driver distraction in Ohio.

Driver Distraction	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Not distracted	30,346	4.02	21,337	2.83	51,683	6.84	703,384	93.16	755,067	100.00
Inattentive	668	5.09	0	0.00	668	5.09	12,458	94.91	13,126	100.00
Driver distracted	257	9.26	28	1.01	285	10.27	2,491	89.73	2,776	100.00
Sleepy or fell asleep	1,002	15.98	41	0.65	1,043	16.63	5,227	83.37	6,270	100.00
Unknown	610	13.21	0	0.00	610	13.21	4,007	86.79	4,617	100.00
Total	**32,883**	**4.21**	**21,406**	**2.74**	**54,289**	**6.94**	**727,567**	**93.06**	**781,856**	**100.00**

Note: *n* represents frequency of crashes.

FARS did not include data for this variable. The GES data in table 63 showed consistently higher SR percentages for distracted drivers regardless of the type of distraction. The North Carolina and Ohio databases showed that the type of distraction with the highest percentage of SR crashes was the "Sleepy or feel asleep" category for both the over speed limit and combined (i.e., total SR) definitions. In Ohio, the SR percentage for the "driver distracted" category was also higher than the SR percentage for the "Not distracted" category using both definitions, while in North Carolina, the "Not distracted" category had a higher percentage of SR crashes than any remaining category.

Driver Physical Impairment

Table 69. Frequency and number/percentage of SR crashes regarding driver physical impairment.

Driver Physical Impairment	GES All Crashes *n*	GES SR Crashes *n*	GES SR Crashes Percent
None	9,564,263	1,015,263	10.62
Illness or blackout	30,345	2,944	9.70
Drowsy, sleepy, fell asleep, or fatigued	83,545	18,541	22.19
Deaf	1,418	6	0.42
Physical impairment no details	4,419	1,010	22.86
Other physical impairment	230,676	67,677	29.34
Unknown if physically impaired	430,260	100,254	23.30
Total	**10,344,926**	**1,205,695**	**11.65**

Note: *n* represents frequency of crashes.

Table 70. Frequency and number/percentage of SR crashes regarding driver physical impairment in North Carolina.

Driver Physical Impairment	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
None	10,501	1.58	45,321	6.84	55,822	8.42	606,955	91.58	662,777	100.00
Illness	15	2.96	52	10.28	67	13.24	439	86.76	506	100.00
Sleepy or fatigued	336	5.05	596	8.96	932	14.02	5,718	85.98	6,650	100.00
Medical condition	22	1.43	74	4.81	96	6.23	1,444	93.77	1,540	100.00
Other physical impairment	29	3.05	62	6.51	91	9.56	861	90.44	952	100.00
Unknown	1,982	6.44	4,236	13.77	6,218	20.21	24,548	79.79	30,766	100.00
Total	**12,885**	**1.83**	**50,341**	**7.16**	**63,226**	**8.99**	**639,965**	**91.01**	**703,191**	**100.00**

Note: *n* represents frequency of crashes.

Table 71. Frequency and number/percentage of SR crashes regarding driver physical impairment in Ohio.

Driver Physical Impairment	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
None	27,659	3.83	20,223	2.80	47,882	6.62	674,890	93.38	722,772	100.00
Illness	89	7.81	10	0.88	99	8.69	1,040	91.31	1,139	100.00
Sleepy or fatigued	913	16.21	41	0.73	954	16.94	4,679	83.06	5,633	100.00
Medical condition	127	10.52	13	1.08	140	11.60	1,067	88.40	1,207	100.00
Other physical impairment	339	10.72	38	1.20	377	11.92	2,786	88.08	3,163	100.00
Unknown	3,756	7.83	1,081	2.25	4,837	10.09	43,105	89.91	47,942	100.00
Total	**32,883**	**4.21**	**21,406**	**2.74**	**54,289**	**6.94**	**727,567**	**93.06**	**781,856**	**100.00**

Note: *n* represents frequency of crashes.

FARS did not include data for driver physical impairment. GES data indicated that the "Other physical impairment" category had a percentage of SR crashes approximately three times that of the "None" category (see table 69). Both State databases for total SR showed that the major driver physical impairment with the highest percentage of SR crashes was "Sleepy or fatigued," similar to the distraction variable results seen in table 67 and table 68. In Ohio, the SR crash percentage for the "Medical condition" category was higher than for the "None" category, but this was not true for North Carolina.

Driver Alcohol Use

Table 72. Frequency and number/percentage of SR crashes regarding driver alcohol use.

	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
Alcohol Use	***n***	***n***	**Percent**	***n***	***n***	**Percent**
No alcohol	26,308	3,224	12.25	9,285,808	1,020,367	10.99
Alcohol	8,327	3,486	41.86	293,204	86,007	29.33
Unknown	22,085	4,902	22.20	766,543	99,322	12.96
Total	**56,720**	**11,612**	**20.47**	**10,345,555**	**1,205,696**	**11.65**

Note: *n* represents frequency of crashes.

Table 73. Frequency and number/percentage of SR crashes regarding driver alcohol use in North Carolina.

Alcohol Use	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
No alcohol suspected/ determined	8,781	1.34	43,478	6.61	52,259	7.95	605,419	92.05	657,678	100.00
Alcohol suspected/ determined	2,405	12.11	3,107	15.64	5,512	27.75	14,350	72.25	19,862	100.00
Unknown	1,699	6.62	3,756	14.64	5,455	21.27	20,196	78.73	25,651	100.00
Total	**12,885**	**1.83**	**50,341**	**7.16**	**63,226**	**8.99**	**639,965**	**91.01**	**703,191**	**100.00**

Note: *n* represents frequency of crashes.

Table 74. Frequency and number/percentage of SR crashes regarding driver alcohol use in Ohio.

Alcohol Use	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
No alcohol suspected/ determined	27,977	4.14	20,072	2.97	48,049	7.11	627,304	92.89	675,353	100.00
Alcohol suspected/ determined	1,433	17.38	569	6.90	2,002	24.28	6,244	75.72	8,246	100.00
Unknown	3,473	3.53	765	0.78	4,238	4.31	94,019	95.69	98,257	100.00
Total	**32,883**	**4.21**	**21,406**	**2.74**	**54,289**	**6.94**	**727,567**	**93.06**	**781,856**	**100.00**

Note: *n* represents frequency of crashes.

All databases and definitions in table 72 through table 74 showed correlations between drinking and SR crashes, with drivers exhibiting alcohol use having much higher percentages of SR crashes than those who did not drink and drive (i.e., two to four times as high).

Driver Drug Use

Table 75. Frequency and number/percentage of SR crashes regarding driver drug use.

Drug Use	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	***n***	**Percent**	***n***	***n***	**Percent**
No drugs	15,621	2,468	15.80	9,389,804	1,072,577	11.42
Drugs involved	1,614	491	30.42	44,763	9,562	21.36
Unknown	39,485	8,653	21.91	910,987	123,556	13.56
Total	**56,720**	**11,612**	**20.47**	**10,345,554**	**1,205,695**	**11.65**

Note: *n* represents frequency of crashes.

Table 76. Frequency and number/percentage of SR crashes regarding driver drug use in North Carolina.

Drug Use	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
No drugs suspected/ determined	10,975	1.62	46,389	6.87	57,364	8.49	618,317	91.51	675,681	100.00
Drugs suspected/ determined	211	11.35	196	10.54	407	21.89	1,452	78.11	1,859	100.00
Unknown	1,699	6.62	3,756	14.64	5,455	21.27	20,196	78.73	25,651	100.00
Total	**12,885**	**1.83**	**50,341**	**7.16**	**63,226**	**8.99**	**639,965**	**91.01**	**703,191**	**100.00**

Note: *n* represents frequency of crashes.

Table 77. Frequency and number/percentage of SR crashes regarding driver drug use in Ohio.

Drug Use	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
No drugs suspected/ determined	27,432	4.08	19,823	2.95	47,255	7.03	624,843	92.97	672,098	100.00
Drugs suspected/ determined	1,978	17.20	818	7.11	2,796	24.31	8,705	75.69	11,501	100.00
Unknown	3,473	3.53	765	0.78	4,238	4.31	94,019	95.69	98,257	100.00
Total	**32,883**	**4.21**	**21,406**	**2.74**	**54,289**	**6.94**	**727,567**	**93.06**	**781,856**	**100.00**

Note: *n* represents frequency of crashes.

Like alcohol use, all databases and all definitions in table 75 through table 77 showed correlations between drug use and SR crashes, with drug users having much higher percentages of SR crashes than nonusers (three to six times as high). As expected, the percentage of total drivers using drugs was small.

Prior Speeding Convictions

Table 78. Frequency and number/percentage of SR crashes regarding prior speeding convictions.

Prior Speeding Convictions	FARS		
	All Crashes	SR Crashes	
	n	*n*	Percent
0	44,083	8,347	18.93
1	7,753	1,870	24.12
2	2,172	611	28.13
3+	1,045	339	32.44
Unknown	1,667	445	26.69
Total	**56,720**	**11,612**	**20.47**

Note: *n* represents frequency of crashes.

Only FARS contains data on prior speeding convictions. Based on the data in table 78, the percentage of SR crashes increased as the number of prior speeding convictions increased. For drivers with three or more prior convictions, the percentage of SR crashes was almost twice that than those who had no prior convictions (32.44 percent versus 18.93 percent).

Driver License Restriction Compliance

Table 79. Frequency and number/percentage of SR crashes regarding license type.

License Type	FARS		
	All Crashes	SR Crashes	
	n	*n*	Percent
Not licensed	2,149	765	35.60
Not required	130	44	33.85
Not valid	4,759	1,659	34.86
Valid	48,530	8,875	18.29
Unknown if commercial drivers license	30	1	3.33
Unknown	1,122	268	23.89
Total	**56,720**	**11,612**	**20.47**

Note: *n* represents frequency of crashes.

Table 80. Frequency and number/percentage of SR crashes regarding license restriction compliance.

License Restriction Compliance	FARS		
	All Crashes	SR Crashes	
	n	*n*	Percent
No restriction	38,302	8,520	22.24
Compiled	3,254	539	16.56
Not compiled	520	215	41.35
Compliance unknown	13,443	2,056	15.29
Unknown	1,201	282	23.48
Total	**56,720**	**11,612**	**20.47**

Note: *n* represents frequency of crashes.

Only FARS contained data on licensing status and restriction compliance. Based on FARS data in table 79, drivers in SR crashes with no licenses and invalid licenses had SR crash percentages approximately twice as high as those with valid licenses. Those drivers that did not comply with a license restriction had SR crash percentages approximately twice as high as those with no restrictions (see table 80). Interestingly, those who had restrictions but complied with them had lower SR crash percentages than those with no restrictions.

Vehicle Type

Table 81. Frequency and number/percentage of SR crashes regarding vehicle type.

Vehicle Type	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
Automobiles	24,047	5,486	22.81	5,729,338	699,421	12.21
Automobile derivatives	30	5	16.67	4,317	105	2.43
Utility vehicles	7,869	1,536	19.52	1,520,109	163,178	10.73
Van-based light trucks	3,569	398	11.15	745,788	69,230	9.28
Light conventional trucks	10,567	2,056	19.46	1,553,955	193,896	12.48
Other light trucks	99	8	8.08	201,318	21,882	10.87
Buses	270	12	4.44	49,488	1,510	3.05
Medium/heavy trucks	4,865	352	7.24	362,652	27,955	7.71
Motored cycles	4,768	1,685	35.34	95,310	20,462	21.47
Other vehicles	156	17	10.90	14,115	1,469	10.41
Unknown	480	57	11.88	69,164	6,588	9.53
Total	**56,720**	**11,612**	**20.47**	**10,345,554**	**1,205,696**	**11.65**

Note: *n* represents frequency of crashes.

Table 82. Frequency and number/percentage of SR crashes regarding vehicle type in North Carolina.

Vehicle Type	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**	***n***	**Percent**
Automobiles	8,169	2.07	28,275	7.15	36,444	9.22	358,990	90.78	395,434	100.00
Utility vehicles	1,196	1.34	7,561	8.47	8,757	9.80	80,560	90.20	89,317	100.00
Van-based light trucks	400	0.79	2,106	4.14	2,506	4.93	48,342	95.07	50,848	100.00
Light conventional trucks	1,820	1.59	8,699	7.60	10,519	9.19	103,947	90.81	114,466	100.00
Other light trucks	34	1.60	125	5.89	159	7.49	1,963	92.51	2,122	100.00
Buses	2	0.09	30	1.39	32	1.49	2,122	98.51	2,154	100.00
Medium/heavy trucks	253	0.90	1,548	5.52	1,801	6.43	26,218	93.57	28,019	100.00
Motored cycles	439	8.39	766	14.64	1,205	23.03	4,027	76.97	5,232	100.00
Other vehicles	3	0.31	15	1.54	18	1.84	959	98.16	977	100.00
Unknown	569	3.89	1,216	8.32	1,785	12.21	12,837	87.79	14,622	100.00
Total	**12,885**	**1.83**	**50,341**	**7.16**	**63,226**	**8.99**	**639,965**	**91.01**	**703,191**	**100.00**

Note: *n* represents frequency of crashes.

Table 83. Frequency and number/percentage of SR crashes regarding vehicle type in Ohio.

Vehicle Type	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
Automobiles	20,386	4.56	11,727	2.62	32,113	7.19	414,634	92.81	446,747	100.00
Utility vehicles	3,309	3.61	3,214	3.50	6,523	7.11	85,198	92.89	91,721	100.00
Van-based light trucks	2,187	3.05	1,297	1.81	3,484	4.85	68,292	95.15	71,776	100.00
Light conventional trucks	3,814	3.57	3,703	3.46	7,517	7.03	99,364	92.97	106,881	100.00
Other light trucks	237	7.12	69	2.07	306	9.19	3,022	90.81	3,328	100.00
Buses	37	1.18	14	0.45	51	1.63	3,078	98.37	3,129	100.00
Medium/ heavy trucks	2,150	4.57	994	2.11	3,144	6.69	43,879	93.31	47,023	100.00
Motored cycles	457	9.13	360	7.19	817	16.32	4,189	83.68	5,006	100.00
Other vehicles	62	2.00	28	0.90	90	2.91	3,008	97.09	3,098	100.00
Unknown	244	7.75	0	0.00	244	7.75	2,903	92.25	3,147	100.00
Total	**32,883**	**4.21**	**21,406**	**2.74**	**54,289**	**6.94**	**727,567**	**93.06**	**781,856**	**100.00**

Note: *n* represents frequency of crashes.

As seen in table 81 through table 83, regardless of database or definition, motorcycles had the highest percentage of SR crashes (two to four times as high as cars). Buses had lower SR crash percentages in all categories, and medium/heavy trucks had lower SR crash percentages relative to other types of trucks in most categories.

Hazardous Cargo

Table 84. Frequency and number/percentage of SR crashes regarding hazardous cargo.

Hazardous Cargo	FARS			GES		
	All Crashes	SR Crashes		All Crashes	SR Crashes	
	n	*n*	Percent	*n*	*n*	Percent
No hazmat	56,133	11,566	20.60			
Hazmat present				3,731	402	10.77
Placard	163	20	12.27			
No placard	11	2	18.18			
Unknown placard	14	2	14.29			
Not applicable				10,177,390	1,194,361	11.74
Unknown	399	22	5.51	163,969	10,933	6.67
Total	**56,720**	**11,612**	**20.47**	**10,345,090**	**1,205,696**	**11.65**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Table 85. Frequency and number/percentage of SR crashes regarding hazardous cargo in North Carolina.

Hazardous Cargo	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
No hazmat	12,884	1.83	50,330	7.16	63,214	8.99	639,839	91.01	703,053	100.00
Hazmat present	0	0	7	6.25	7	6.25	105	93.75	112	100.00
Unknown	1	3.85	4	15.38	5	19.23	21	80.77	26	100.00
Total	**12,885**	**1.83**	**50,341**	**7.16**	**63,226**	**8.99**	**639,965**	**91.01**	**703,191**	**100.00**

Note: *n* represents frequency of SR crashes.

Table 86. Frequency and number/percentage of SR crashes regarding hazardous cargo in Ohio.

Hazardous Cargo	Over Speed Limit		Too Fast for Conditions		Total SR		Not SR		Total	
	n	Percent	*n*	Percent	*n*	Percent	*n*	Percent	*n*	Percent
No hazmat	32,399	4.17	21,211	2.73	53,610	6.90	723,459	93.10	777,069	100.00
Hazmat present	22	10.58	17	8.17	39	18.75	169	81.25	208	100.00
Unknown	462	10.09	178	3.89	640	13.98	3,939	86.02	4,579	100.00
Total	**32,883**	**4.21**	**21,406**	**2.74**	**54,289**	**6.94**	**727,567**	**93.06**	**781,856**	**100.00**

Note: *n* represents frequency of crashes.

There were low frequencies of crash-involved vehicles carrying hazardous cargo in all of the databases, making conclusions difficult (see table 84 through table 86). The FARS data showed that vehicles not carrying hazardous materials had a higher percentage of SR crashes. In addition, GES and North Carolina data showed practically no difference. In the Ohio database, the total SR category showed that vehicles containing hazardous cargo had a higher percentage of SR crashes than those that did not carry hazardous materials (see table 86).

ANALYSES OF DATA SUBSETS

Following a review of the above tables, the FHWA requested additional tables on subsets of the data. The request was for specific tables concerning only SR crashes in FARS and GES for five data subsets which included: (1) pedestrian crashes, (2) intersection crashes, (3) lane departure crashes, (4) rural crashes, and (5) urban crashes. *Lane departure crashes* were defined using a standard FHWA definition including single-vehicle, run-off road crashes; multivehicle, head-on crashes; multivehicle, opposite-direction front-to-side crashes; and multivehicle, opposite-direction sideswipe crashes. Within each subset, tables for only selected variables were requested, as follows:

- **Pedestrian crashes:** rural/urban, functional class, day/night (using light condition), speed limit, relationship to junction, number of travel lanes, pedestrian age, and pedestrian alcohol use.

- **Intersection crashes:** rural/urban, functional class, day/night, speed limit, traffic control device, number of travel lanes, alignment, grade, driver age, and driver alcohol use.

- **Lane departure crashes:** rural/urban, day/night, speed limit, number of travel lanes, alignment, grade, surface condition, driver age, and driver alcohol use. Note that tables were run for all lane departure crashes and then only for the single-vehicle, run-off-road crashes.

- **Rural crashes:** lane departure (by category), day/night, speed limit, relationship to junction, alignment, grade, number of travel lanes, surface condition, pedestrian related, driver age, and driver alcohol use.

- **Urban crashes:** lane departure (by category), day/night, speed limit, relationship to junction, alignment, grade, number of travel lanes, surface condition, pedestrian related, driver age, and driver alcohol use.

Note that in all cases, the tables related to pedestrian/driver age and alcohol use were vehicle specific, while the remaining tables were crash specific. In addition, the analyses for rural and urban crashes and the tables for rural/urban and functional class could only be run with FARS data because GES did not include these variables.

The following sections present summary results for each data subset. These results differ from the original single-variable results. For the single-variable results, the goal was to define categories within each critical variable that were more likely to be SR when compared to other categories. However, for these five subsets, the data were restricted to SR crashes, and the goal was to examine the nature of these SR crashes within each subset.

SR Pedestrian Crashes

Table 87. Characteristics of SR pedestrian crashes.

Variable	Category	FARS (Percent)	GES (Percent)
Urban		75	
Urban principal and minor arterials		38	
Nighttime		62	50
Speed limit (mi/h)	≤ 25	17	9
	30–35	30	33
	40–45	22	20
	50–55	13	4
	60+	18	12
Midblock		76	59
Number of lanes	2	38	45
	4	33	17
Predominant pedestrian age (years)	11–20	11	27
	21–30	18	18
	31–40	16	22
	41–50	16	8
Pedestrian alcohol		8	8
Total frequency		**411**	**2,994**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

Table 87 summarizes the key findings from the series of FARS and GES analysis runs related to SR pedestrian crashes. Fatal SR pedestrian crashes were predominantly urban in nature and occurred primarily on arterial roadways. Both fatal and total SR pedestrian crashes occurred mostly during the nighttime on roads with 30–45 mi/h speed limits at midblock (nonintersection) locations and on two-lane roads and streets. Predominant pedestrian age categories varied slightly with the GES database, showing a higher percentage of total SR crashes involving drivers who were 11–20 years old. Alcohol use was found in approximately 8 percent of the pedestrians in both databases.

SR Intersection Crashes

As shown in table 88, fatal SR intersection crashes predominantly occurred in urban areas on principal and minor arterials. While approximately half the fatal SR intersection crashes occurred during the daytime and half at nighttime, a much higher percentage (73 percent) of the total SR intersection crashes occurred in the daytime. Both fatal and total SR crashes were more likely to occur on roads with speed limits of 30–45 mi/h. While the fatal SR intersection crashes were almost equally divided among intersections with stop/yield signs, signals, and no control, over half of the total SR intersection crashes occurred at signalized locations. Both databases indicated that most were on straight and level roads and on two-lane roads with GES having a high percentage of unknown variables in both cases. Predominant driver ages were similar, with slightly more 25–30 year olds involved in the fatal SR intersection crashes. Alcohol was found in a much higher proportion of the fatal SR intersection crashes.

Table 88. Characteristics of SR intersection crashes.

Variable	**Category**	**FARS (Percent)**	**GES (Percent)**
Urban		65	
Urban principal and minor arterials		49	
Daytime		45	73
Speed limit (mi/h)	≤ 25	15	24
	30–35	30	33
	40–45	30	32
	50–55	21	9
	60+	4	1
Traffic control	Stop/yield	33	16
	Signals	31	55
	No control	32	26
Straight (as opposed to curve/unknown)		83	85
Level (as opposed to grades/hills/unknown)		80	43
Number of lanes	2	54	28
	4	28	16
Predominant driver age (years)	16–19	15	21
	20–25	30	20
	26–35	23	17
	36–50	19	22
Alcohol		28	6
Total frequency		**1,510**	**383,900**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

SR Lane Departure Crashes

SR lane departure crashes, which included both single-vehicle, run-off-road crashes and multivehicle, head-on and sideswipe crashes, were mostly single-vehicle, run-off-road crashes. Specifically, over 90 percent of the fatal SR lane departure crashes in FARS and over 95 percent of the total SR lane departure crashes in GES were single-vehicle, run-off-road crashes. Thus, the findings in table 89 are dominated by these single-vehicle crashes.

Table 89. Characteristics of SR lane departure crashes.

Variable	Category	FARS (Percent)	GES (Percent)
Rural		64	
Functional class	Rural major collector	18	
	Rural local road	16	
Night		58	46
Speed limit (mi/h)	≤ 25	6	11
	30–35	18	17
	40–45	21	17
	50–55	39	27
	60+	14	15
Straight (as opposed to curve/unknown)		51	55
Level (as opposed to grades/hills/unknown)		64	39
Two lanes	No control	71	55
Dry (as opposed to wet, snowy/icy)		80	49
Predominant driver age (years)	16–19	17	24
	20–25	26	25
	26–35	21	18
	36–50	22	19
Alcohol		35	13
Total frequency		**7,796**	**415,631**

Note: *n* represents frequency of crashes. Blank cells indicate that the code was not present in this database.

As shown in table 89, fatal SR lane departure crashes predominantly occurred in rural areas on major collectors and local roads. While there were slight differences, both databases indicated that approximately half the crashes occurred in the daytime and half in the nighttime and that a higher percentage occurred on roads with speed limits of 50–55 mi/h. Both showed a practically equal split between curves and straight roadway sections. While the fatal SR lane departure crashes were more likely on level roads, a higher percentage of the total SR crashes were on grades or hills. Both exhibited the majority of such crashes on two-lane roads. While the fatal SR lane departure crashes were more likely to be on dry roadways, approximately half of the GES total crashes were on wet/snowy/icy roadways. Predominant driver ages were similar, with more 16–19 year olds in the total SR lane departure crashes for GES. Alcohol was found in a much higher proportion of the fatal SR lane departure crashes.

SR Crashes in Rural and Urban Areas

Note that since GES does not include a rural/urban variable, the information in table 90 only concerns fatal SR crashes.

Table 90. Characteristics of fatal SR crashes by area type (rural/urban).

Variable	Category	Rural (Percent)	Urban (Percent)
Single-vehicle run-off-road		68	51
Predominant number of lanes	2	77	42
	4	16	28
Dark		48	59
Speed limit (mi/h)	≤ 25	4	14
	30–35	10	28
	40–45	18	26
	50–55	45	16
	60+	23	15
Not related to junction		89	69
Curves		46	30
Grades/hills		37	23
Dry (as opposed to wet, snowy/icy)		77	83
Pedestrian involved		1	7
Predominant driver age (years)	16–19	16	16
	20–25	23	29
	26–35	20	24
	36–50	24	20
Alcohol		31	30
Total frequency		**6,679**	**4,592**

Note: *n* represents frequency of crashes.

As would be expected, there were differences between the characteristics of rural and urban SR fatal crashes. The rural SR fatal crashes were more likely to be single-vehicle, run-off-road crashes as opposed to multivehicle crashes, and they were more likely to be on two-lane roads at night. Almost half of the rural SR fatal crashes were on roads with 50–55 mi/h speed limits, while the urban SR fatal crashes were at lower speed limits, as expected. Both rural and urban crashes were more likely to be nonintersection crashes on straight level roadways that were not characterized by wet or other slippery conditions. Pedestrian involvement was low for both, but it was much higher in urban areas. There was little difference between driver age or alcohol use.

RESULTS OF CART ANALYSES

The following sections present the results of the previously described CART analyses. Since the focus of this effort is national, detailed results will be presented for FARS and GES analyses. Recall that CART analyses on crash-based variables and vehicle-/driver-based variables were separate analyses producing separate trees. In each case, a figure presenting selected branches (the high SR branches) are presented followed by a discussion of the findings.

FATAL FARS SR CRASHES

Figure 2 presents the results of the CART analysis of crash-related variables in the FARS database.

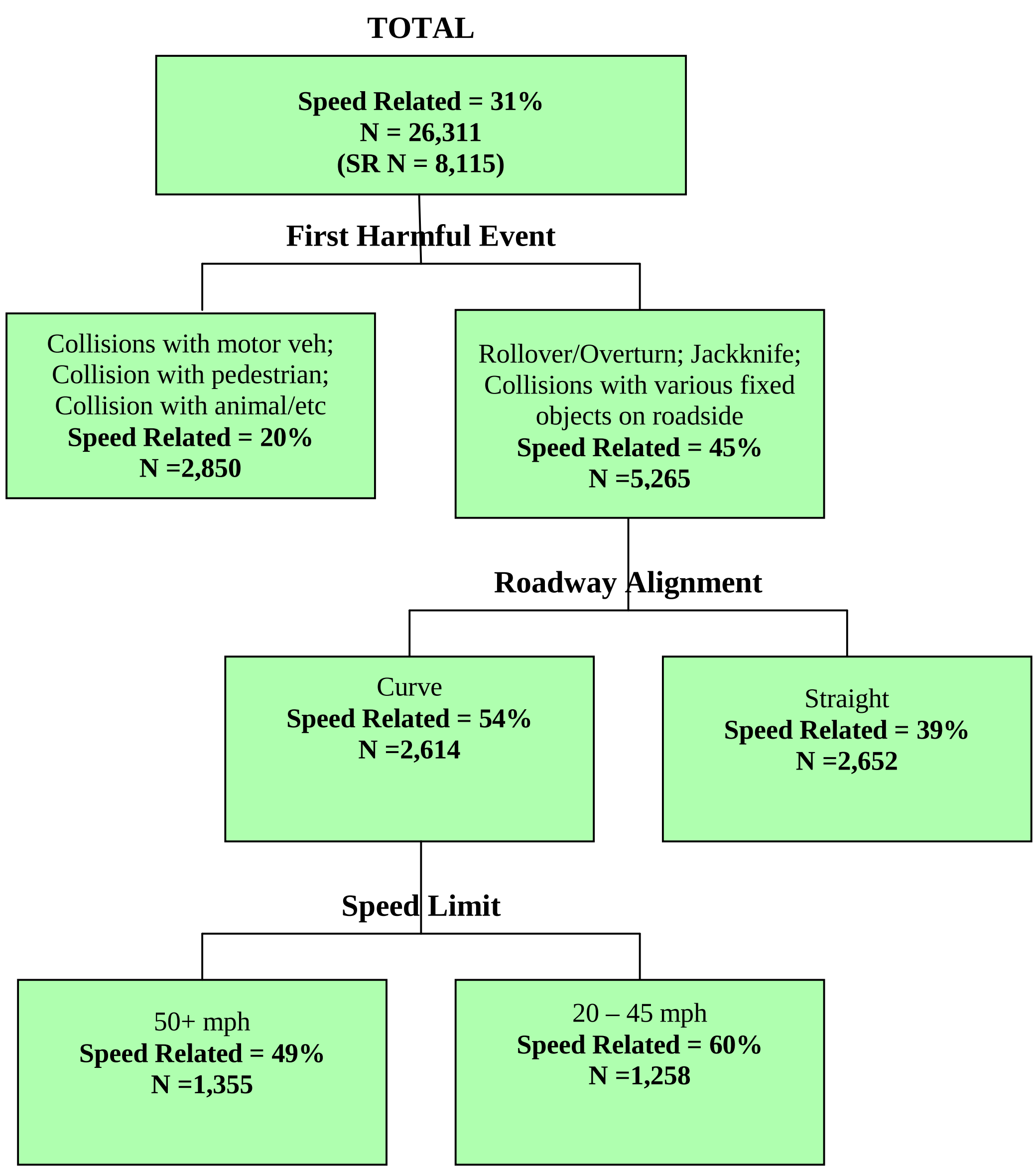

Figure 2. Chart. CART SR output for 2005 FARS crashes.

The most important SR predictive variable shown at the top tree branch labeled is first harmful event. The categories with the highest percentage of SR crashes include rollovers/overturns, jackknife crashes, and collisions with various fixed objects on the roadside. Within that branch, the most important variable is roadway alignment, with the highest SR category being curves. Within that category, the next variable is speed limit, with the highest SR categories being the lower speed limit of 20–45 mi/h. CART did not detect a fourth-level variable branching off of speed limit. In this selected subsample, approximately 60 percent of the crashes are SR. In general, this tree indicates that run-off-road collisions on curves that are found on roads with lower speed limits are more likely to be SR. Note that this final subcategory includes approximately 1,300 of the 26,000 fatal crash analyzed in this training set—approximately 5 percent of the total fatal crashes and 16 percent of the total SR fatal crashes.

TOTAL GES SR CRASHES

Figure 3 presents the results of the CART analysis for crash-related variables in the GES database. The CART analysis involved four levels. The most important SR predictive variable shown at the top tree branch is manner of collision, and the categories with the highest SR percentage are nonmotor-vehicle collisions and rear-end (30 percent). Within that branch, the most important variable is event 1, and the categories with the highest SR percentage include rollovers, collisions with objects, and rear-ends. Within that branch, the most important branch variable is roadway surface condition, and the categories with the highest SR percentage are wet, snow and sleet, and icy. Finally, within that branch, the most important variable is atmospheric conditions, and the categories with the highest SR percentage are snow and sleet and fog. The latter reinforces the adverse weather finding from the above branch, but it adds little additional information. In this final selected subsample, approximately 60 percent of the crashes are SR. In general, these branches indicate that rear-end and off-roadway crashes in bad weather are more likely to be SR. Note that this final SR subcategory includes about 67,000 of the 4 million GES crashes analyzed in the training set—approximately 2 percent of the total crashes and 8 percent of the total SR crashes. If the latter branch is omitted, the third level (including road surface condition) contains approximately 6 percent of the total GES sample and 29 percent of the total SR sample. Approximately 45 percent of this subset is SR.

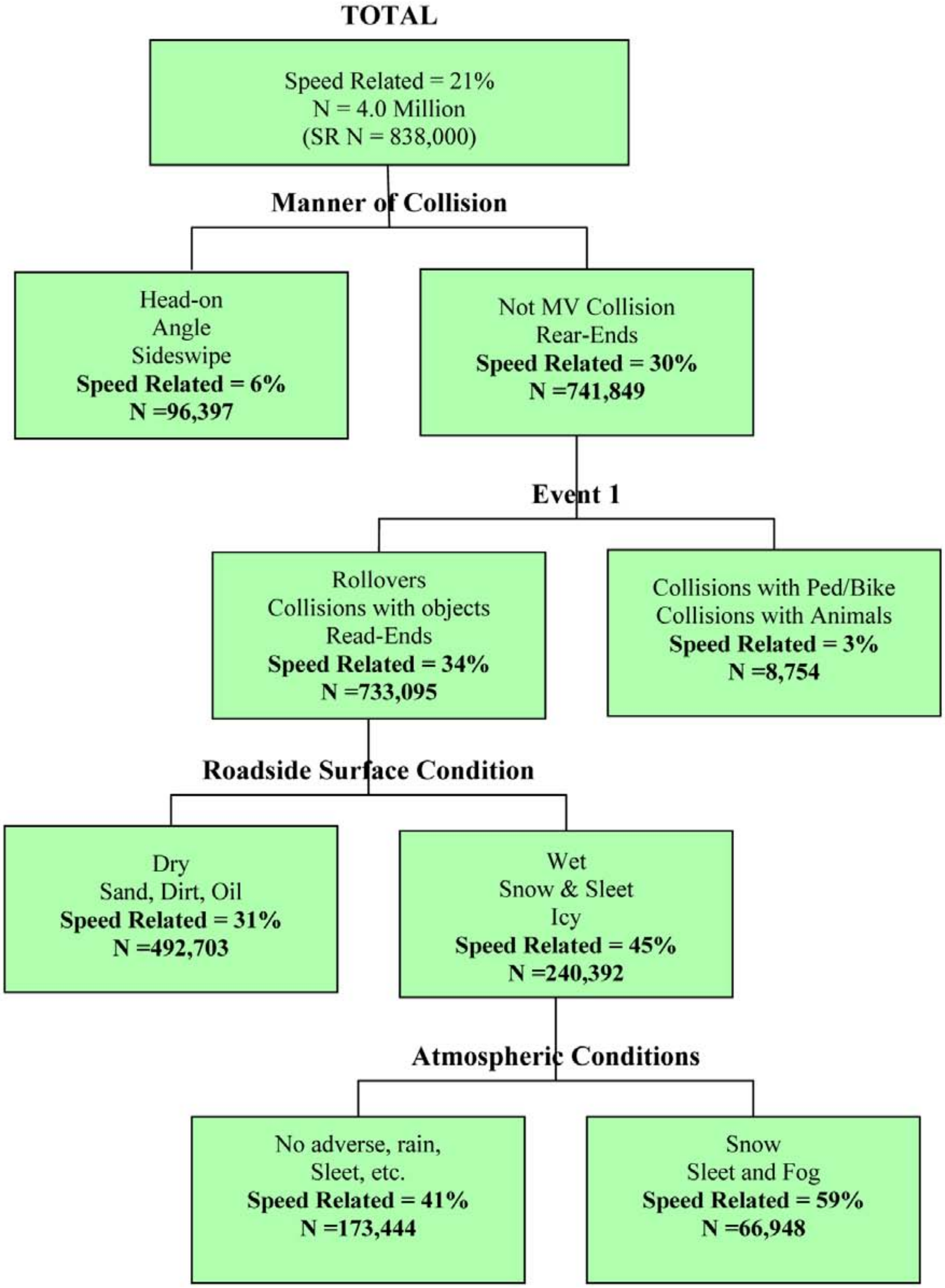

Figure 3. Chart. CART SR output for 2005 GES crashes.

NORTH CAROLINA AND OHIO SR CRASHES

CART analyses for the North Carolina and Ohio crashes were conducted to further examine differences between the combined and over speed limit definitions and to examine the consistency between the State results and between the State and national results. CART figures and their results are included below.

North Carolina Crash-Based CART Results

In the four-level North Carolina analysis using the combined definition shown in figure 4, the most important SR predictive variable is first harmful event. The categories with the highest SR percentage are rollover/overturn, jackknife, ran off road, and collision with fixed objects, which are single-vehicle crashes (49 percent). Within that single-vehicle branch, the most important variable is surface condition, and the categories with the highest SR percentage include wet, slow/slush, and ice. Within that branch, the most important branch variable is again surface condition, with the highest SR subset including only snow and ice. Finally, within this subset, the most important variable is relation to roadway, and the categories with the highest SR percentage include on roadside (treated and untreated). This latter branch reinforces the general finding that the most important crashes are single-vehicle roadside crashes on snowy or icy roads. Approximately 86 percent of the crashes in this final branch are SR, and it includes approximately 5,200 of the 284,000 crashes in the North Carolina training set—approximately 2 percent of the total North Carolina sample and 12 percent of the total North Carolina SR crashes.

In the original North Carolina CART analysis using the over speed limit definition, crash severity is a second-level predictor. However, since it does not appear that dividing crashes into severity categories would give much guidance into crash types of interest, the analysis was redone without including the crash severity variable. The revised CART analysis contains four levels, as shown in figure 5. The most important SR predictive variable is first harmful event. The categories with the highest SR percentage are rollover/overturn, jackknife, ran off road, and collision with fixed objects, which are all single-vehicle crashes (9 percent). Within that single-vehicle branch, the most important variable is surface condition, and the category with the highest SR percentage includes dry. Note that this differs from the combined CART in figure 4 where dry is not the most important category. This clearly shows the difference between using the liberal and conservative definitions of SR. Within the dry roadway branch, the most important branch variable is roadway alignment, with curves being the most important category. Finally, within this branch, the most important variable is light conditions, with the highest categories being dark and dark but lighted. Thus, in general, this analysis using the more conservative over speed limit SR definition indicates that the crash branch most likely to be SR is single-vehicle crashes on dry curves during night. Approximately 20 percent of the crashes in this final branch are SR, and the subset includes approximately 1,600 of the 284,000 crashes in the North Carolina training set—approximately 0.6 percent of the total sample and 18 percent of the SR sample.

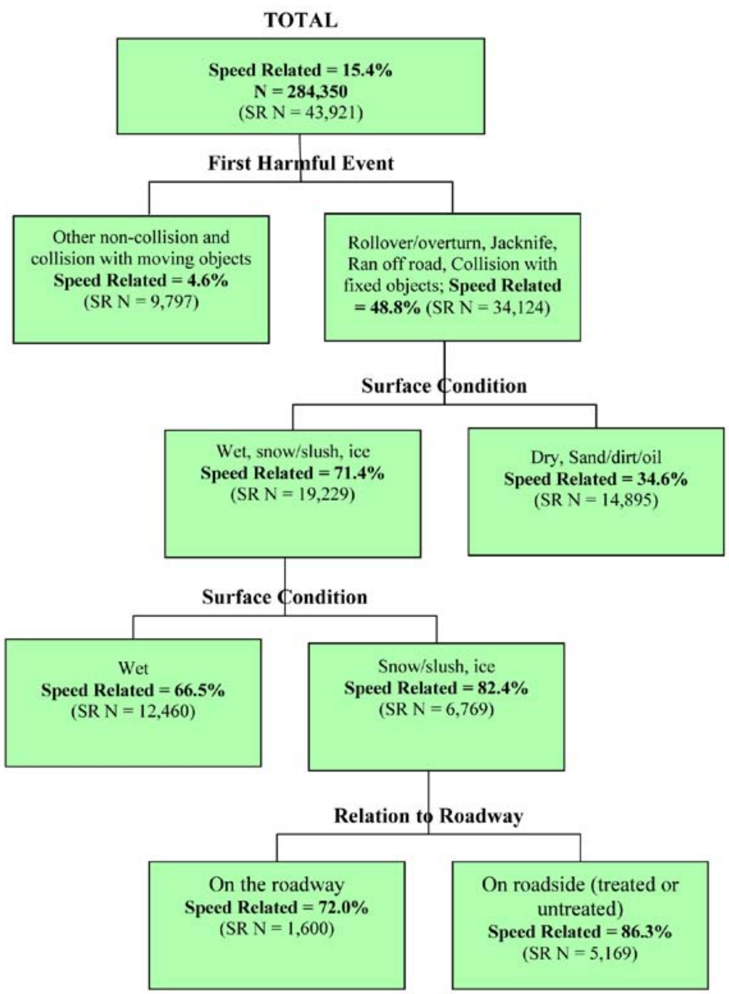

Figure 4. Chart. CART results for North Carolina crash-based variables using the combined SR definition (2002–2004).

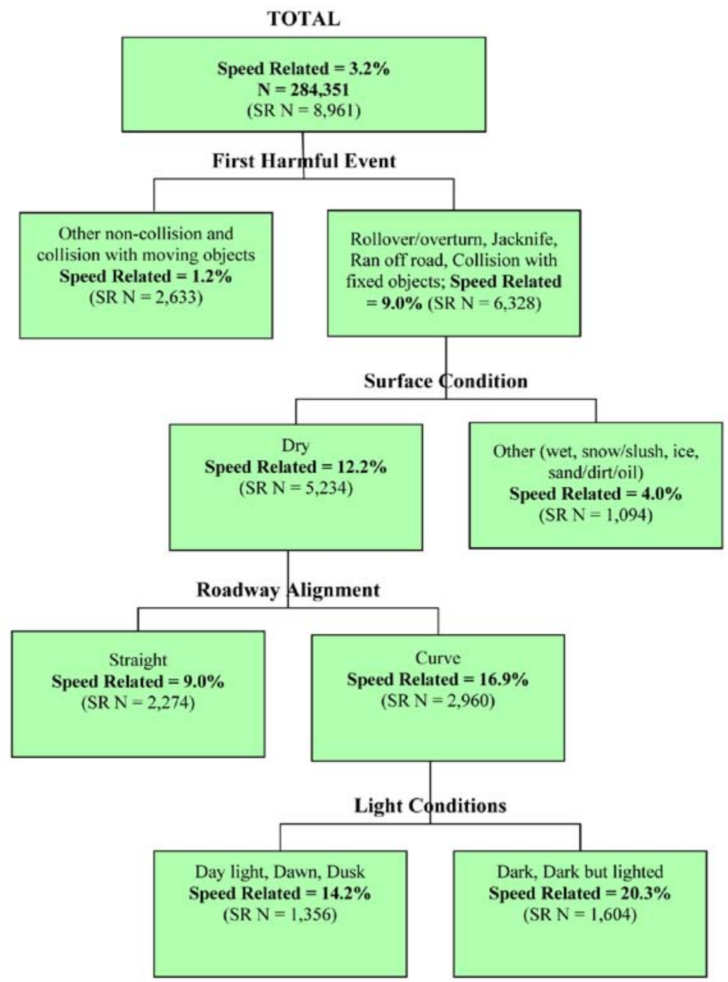

Figure 5. Chart. CART results for North Carolina crash-based variables using the over speed limit SR definition (2002–2004).

Ohio Crash-Based CART Results

In the Ohio CART analysis shown in figure 6, using the combined definition, the most important SR predictive variable is first harmful event. The categories with the highest SR percentage are rollover/overturn and collision with other fixed objects, which are single-vehicle crashes (34 percent). Within that single-vehicle branch, the most important variable is road functional class. The majority are all rural classes (interstate, other principal arterials, major and minor collectors and local roads) except minor arterial and includes the urban local roads. Within this group, the most important branch variable is surface condition, with the important categories including wet, snow/slush, ice, and sand/dirt/oil. Finally, within this branch, the most important variable is relation to roadway, and the category with the highest SR percentage is off roadway. Thus, in general, this analysis using the liberal SR definition indicates that the crash subset most likely to be SR is off-road, single-vehicle crashes within predominantly rural road classes on nondry roads. Approximately 57 percent of the crashes in this final subset are SR by the liberal definition, and the subset includes approximately 6,000 of the 314,000 crashes in the Ohio training set—approximately 2 percent of the total analyzed sample and 17 percent of the SR sample.

In the original CART analyses using the over speed limit definition of SR, crash severity is a fourth-level predictor. As in the North Carolina analyses, because it does not appear that dividing crashes into severity categories would give much guidance into crash types of interest, the analysis was redone without including the crash severity variable. In the revised analysis shown in figure 7, the most important SR predictive variable is first harmful event. The categories with the highest SR percentage are rollover/overturn, other noncollison, and collision with fixed objects, which are all single-vehicle crashes (12 percent). Within that branch, the next most important variable is surface condition, and the categories with the highest SR percentage include dry and wet but not including the snowy/icy conditions. Note that just as with the North Carolina data, this differs from the liberal CART above where dry is not the important category, clearly showing the difference between using the liberal and conservative definitions. Within the dry and wet roadway branch, the most important branch variable is number of lanes and includes roads with three to eight lanes. Finally, within this subset, the most important variable is light conditions, with the highest category being dark but lighted. Interestingly, the dark category is not included here. Thus, in general, this analysis using the over speed limit SR definition indicates that the crash subset most likely to be SR are single-vehicle crashes on either dry or wet dry roads with more than two lanes under a nighttime condition where lighting is present. Approximately 21 percent of the crashes in this final subset are SR, and the subset includes approximately 1.100 of the 314,000 crashes in the Ohio training set—approximately 0.3 percent of the total analyzed sample and 5 percent of the SR sample.

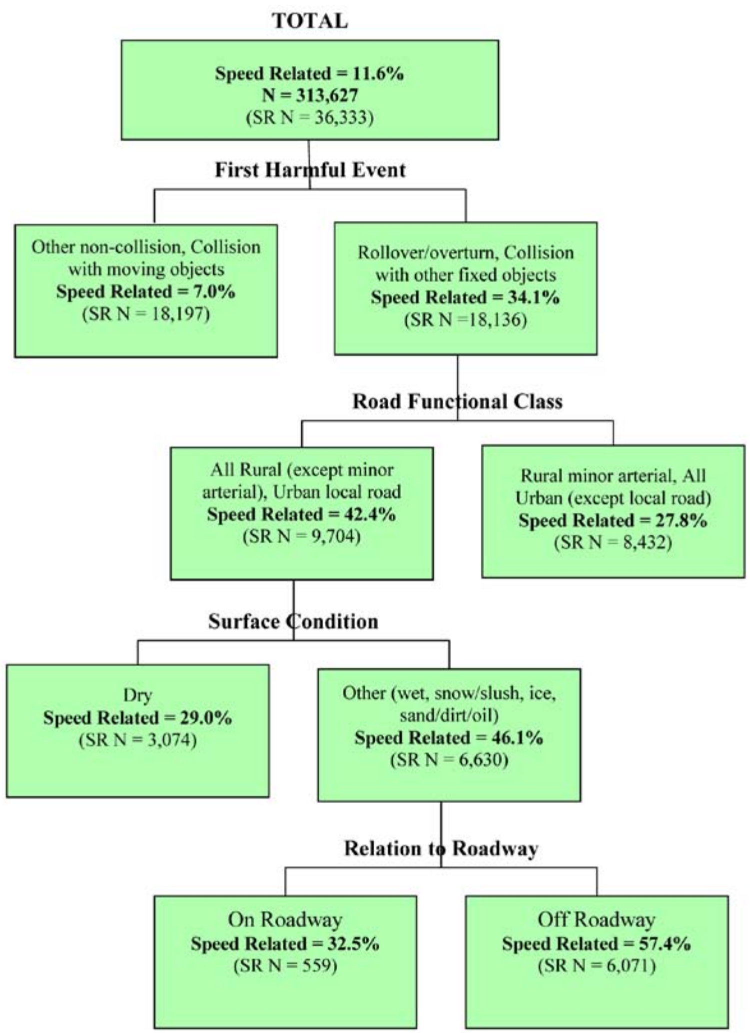

Figure 6. Chart. CART results for Ohio crash-based variables using the combined SR definition (2003–2005).

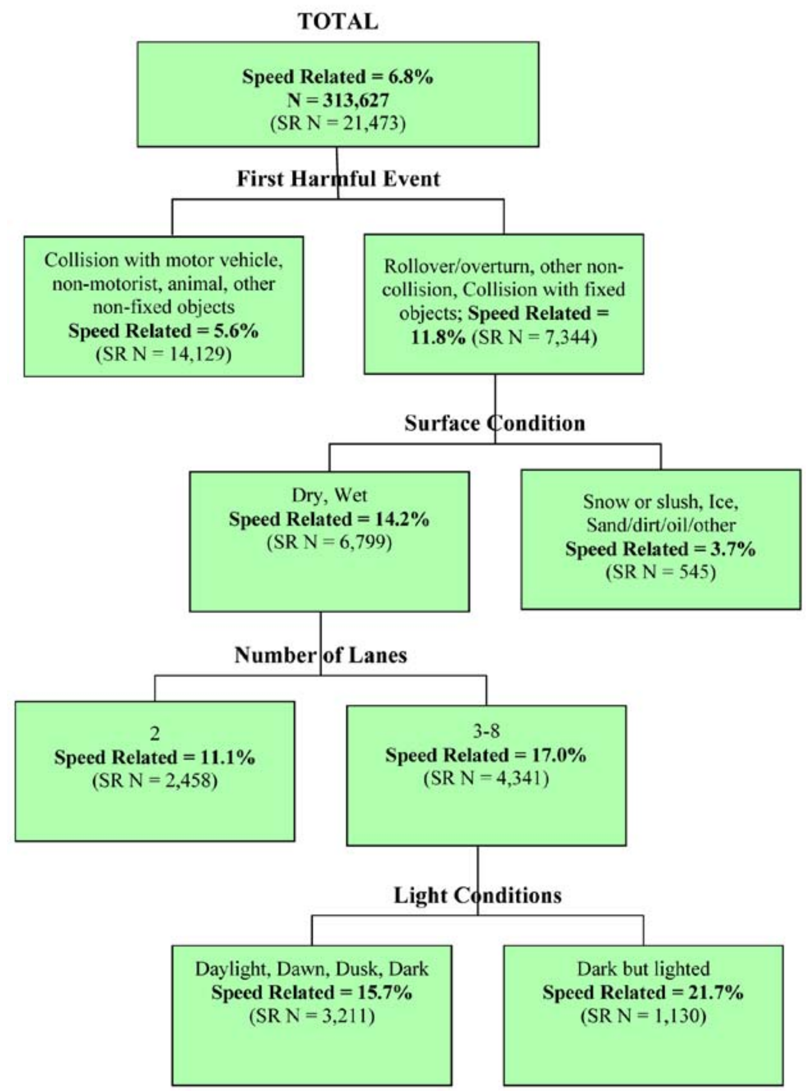

Figure 7. Chart. CART results for Ohio crash-based variables using the over speed limit SR definition (2003–2005).

SUMMARY CONCLUSIONS ON CRASH-RELATED CART ANALYSES

The findings from the CART analysis are not completely consistent; however, there is consistency across all four databases and both SR definitions with respect to the fact that the top-level predictor of SR crashes is first harmful event and that the categories with the highest SR percentage are, in general, single-vehicle, run-off-road crashes. However, it is noted that the top-level GES branch also includes rear-end crashes. The analyses using the liberal combined SR definition indicates that within these run-off-road crashes, the most important predictor is the roadway surface/weather with snowy/icy roads, having the highest SR percentage. Indeed, the GES and North Carolina combined definition findings include no strong predictors other than these two. Third- and fourth-level predictors within FARS and Ohio data differ with FARS data, indicating crashes on curves with lower speed limits and Ohio data, indicating that rural roadway class is an important predictor. The over speed limit findings in North Carolina and Ohio again point to the single-vehicle, run-off-road crashes, but the predictors within this set differ from the above results using the combined definition in that these indicate that dry (or dry and wet) roads predict higher SR percentages than the snowy/icy roads. These over speed limit analyses also indicate that single-vehicle crashes on curves (found in North Carolina), during the nighttime (found in North Carolina and Ohio), and on roads with more than two lanes (Ohio) are important SR predictor categories.

FATAL FARS SR VEHICLES/DRIVERS

Figure 8 below shows the results for the vehicle-based analyses using the FARS data.

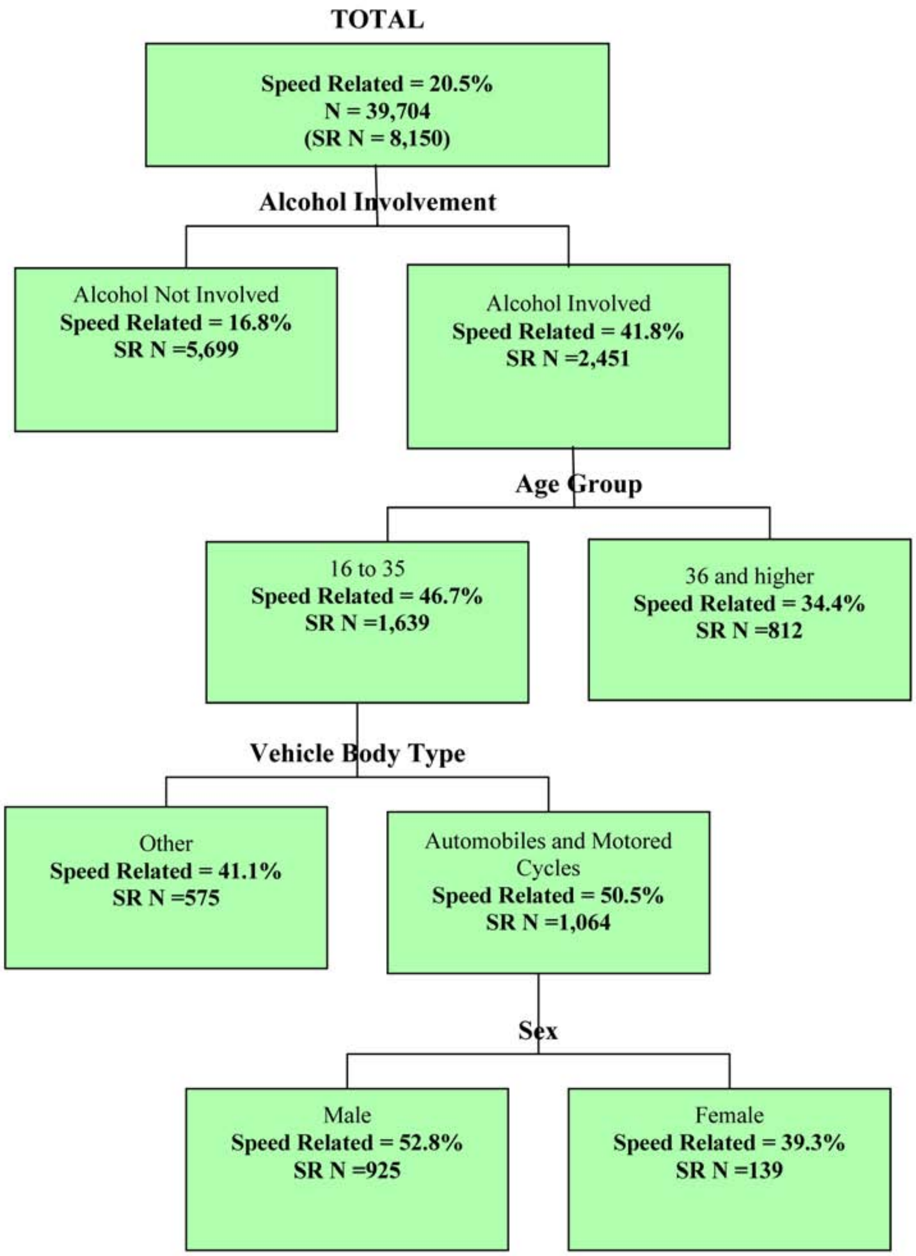

Figure 8. Chart. CART SR output for 2005 FARS vehicles/drivers.

The most important SR predictive variable is alcohol involvement. The category with the highest SR percentage shows that alcohol is involved (42 percent). Within that branch, the most important variable is age group, and the categories with the highest SR percentage include drivers who are 16–35 years old. Within that branch, the most important branch variable is vehicle body type, and the categories with the highest SR percentage include automobiles and motorcycles. Finally, within this branch, the most important variable is sex, and the category with the highest SR percentage is male. Thus, in general, this analysis of fatal crashes (using the liberal combined SR definition by default) indicates that the fatal crash subset most likely to include SR vehicles are those with drivers who are drinking and are young males driving either automobiles or motorcycles. Approximately 52 percent of the crashes in this final subset are SR, and the subset includes approximately 1,800 of the 40,000 fatal crashes in the FARS training set—approximately 2 percent of the total sample and 11 percent of the SR sample.

TOTAL GES SR VEHICLES/DRIVERS

Figure 9 shows the results for the vehicle-based analyses using the GES data. The most important SR predictive variable is driver distraction. The categories with the highest SR percentage are the various types of distraction (doing something, inattentive, sleepy, etc.). Within that branch, the most important variable is restraint system use, and the categories with the highest SR percentage include none, motorcycle helmet, lap/shoulder belt, and shoulder belt. Unfortunately, this grouping of categories is not informative since those who do not use restraint are grouped with those who do. CART does not distinguish between the motorcycle helmet users and nonbelt users versus the users as the single-variable tables have done (probably because the lap and shoulder group is so large, comprising of 80 percent of the data). Within that branch, the most important branch variable is driver visual obstruction, and the categories with the highest SR percentage include no obstruction, weather, and car-related (this does not include exterior obstructions such as hills, curves, trees, or buildings). Finally, within this branch, the most important variable is driver age group, with 16–70-year-old drivers included in category with the highest SR percentage. Thus, this analysis is difficult to interpret given the grouping of categories. In general, the vehicle/driver subset most likely to be SR is distracted drivers using or not using restraints with some visual obstruction and between the ages of 16–70. Approximately 24 percent of the drivers in this final subset are SR, and the subset includes approximately 137,000 of the 7.2 million vehicles in the GES training set—approximately 2 percent of the total sample of vehicles and 16 percent of the SR sample.

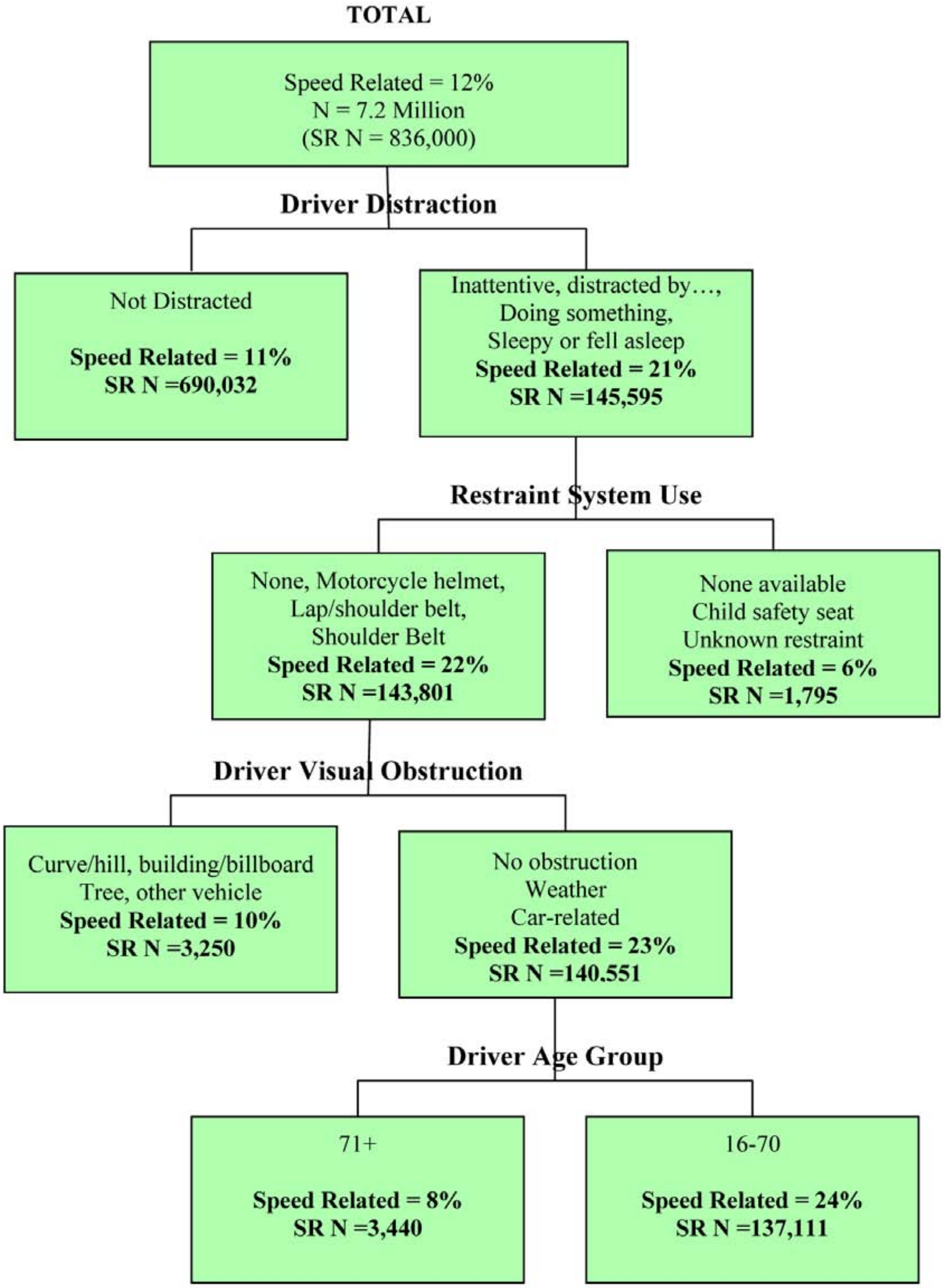

Figure 9. Chart. CART SR output for 2005 GES vehicles/drivers.

NORTH CAROLINA AND OHIO SR VEHICLES/DRIVERS

North Carolina Vehicle-Based CART Results

The initial analysis with the North Carolina vehicle file using the combined SR definition includes airbag availability/deployment as the second-level branch variable. Because this variable is difficult to interpret (i.e., it is both an indication of car age and of the severity of the crash), it was removed, and a second CART analysis was conducted. As shown in figure 10, the revised analysis shows that the most important SR predictive variable is driver age. The categories with the highest SR percentage are 16–25-year-old drivers (14 percent). The next most important variable is gender, and the category with the highest SR percentage is male. Within that branch, the most important variable is driver distraction, and the categories with the highest SR percentage include not distracted and sleepy/fell asleep. Finally, the most important fourth-level branch variable is vehicle body type, and the categories with the highest SR percentage are automobiles, utility vehicles, motorcycles, and other vehicles except trucks and buses. This final level provides little additional information. Thus, in general, this analysis using the liberal combined SR definition indicates that the driver/vehicle subset most likely to be involved in SR crashes are young drivers who are males and who are not distracted by anything other than being fatigued/sleepy. Approximately 18 percent of the vehicles in this final subset are SR, and the subset includes approximately 9,200 of the 492,000 vehicles in the North Carolina training set—approximately 2 percent of the total North Carolina vehicle sample and 21 percent of the SR sample.

As shown in figure 11, when the North Carolina vehicle sample is analyzed using the conservative over speed limit definition, the most important SR predictive variable is alcohol involvement, with the highest SR category being yes (12 percent SR). The next variable is restraint system used, with the highest SR categories including none, motorcycle helmet, and shoulder belt only. The motorcycle helmet indication is a vehicle-type indicator. Thus, in general, this analysis using the over speed limit SR definition indicates that the driver/vehicle subset most likely to be involved in SR crashes are drivers who are drinking and who do not use restraints. Approximately 20 percent of the vehicles in this final subset are SR, and the subset includes approximately 500 of the 492,000 vehicles in the North Carolina training set—less than 0.1 percent of the total vehicle sample and approximately 5 percent of the SR sample, which are both small percentages.

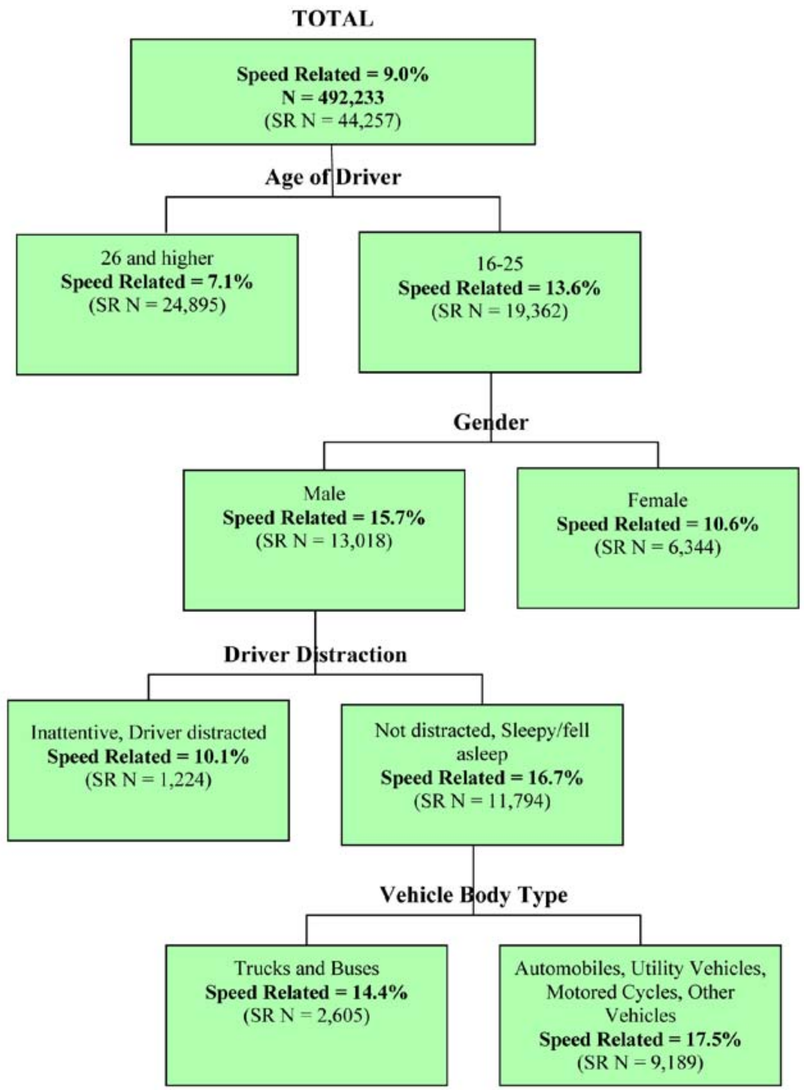

Figure 10. Chart. CART results for North Carolina vehicle-based variables using the combined SR definition (2002–2004).

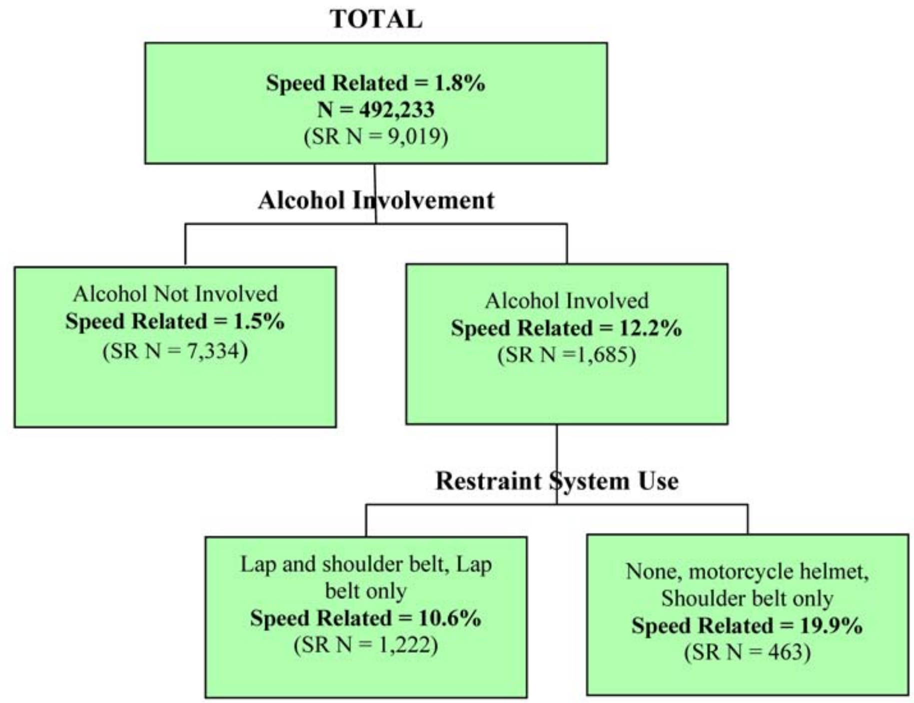

Figure 11. Chart. CART results for North Carolina vehicle-based variables using the over speed limit SR definition (2002–2004).

Ohio Vehicle-Based CART Results

The initial output from the Ohio vehicle file using the combined SR definition includes airbag availability/deployment as the second-level branch variable. Because this variable was difficult to interpret (i.e., it was both an indication of car age and of the severity of the crash) it was removed, and a second CART analysis was conducted. In this analysis reviewing airbag removal in figure 12, the most important SR predictive variable is driver age. The category with the highest SR percentage is 16–25-year-old drivers (10 percent SR). Within that branch, the most important variable is gender, and the category with the highest SR percentage is male. Within that branch, the most important branch variable is vehicle body type, and the categories with the highest SR percentage include automobiles and motorcycles (i.e., no utility vehicles, trucks, etc.). Finally, within this branch, the fourth-level variable is again driver age, but the difference in SR percentage between the 16–19-year-old drivers and the 20–25-year-old drivers is small. Thus, in general, this analysis using the liberal SR definition indicates that the driver/vehicle subset most likely to be involved in SR crashes is younger drivers (up to age 25) who are male and who are driving either automobiles or motorcycles. Approximately 12 percent of the vehicles in this final subset are SR, and the subset includes approximately 6,400 of the 547,000 vehicles in the Ohio training set—approximately 1 percent of the total vehicle sample and 17 percent of the SR sample.

When the Ohio vehicle file was analyzed using the over speed limit definition, the initial run also included the difficult-to-interpret airbag availability/deployment variable. This variable was then removed, and a second CART analysis was conducted. In figure 13, the most important SR predictive variable is driver age. The category with the highest SR percentage is 16–25-year-old drivers (7 percent). The most important variable is gender, and the category with the highest SR percentage is male. Within that branch, the most important branch variable is driver restraint system use, and the categories with the highest SR percentage include none, motorcycle helmet, and lap belt only. The none category dominates this grouping. Finally, within this branch, the fourth-level variable is vehicle body type, and the categories with the highest SR percentage include automobiles, utility vehicles, van-based light trucks, and motorcycles, which is a mix of types that does not provide much useful information. Thus, in general, this analysis using the conservative over speed limit SR definition indicates that the driver/vehicle subset most likely to be involved in SR crashes is younger drivers (up to age 25) who are male, who do not use restraints, and who drive a variety of vehicle types. Approximately 20 percent of the vehicles in this final subset are SR, and the subset includes approximately 615 of the 547,000 vehicles in the Ohio training set—approximately 0.1 percent of the full sample and 3 percent of the SR sample, which are small percentages.

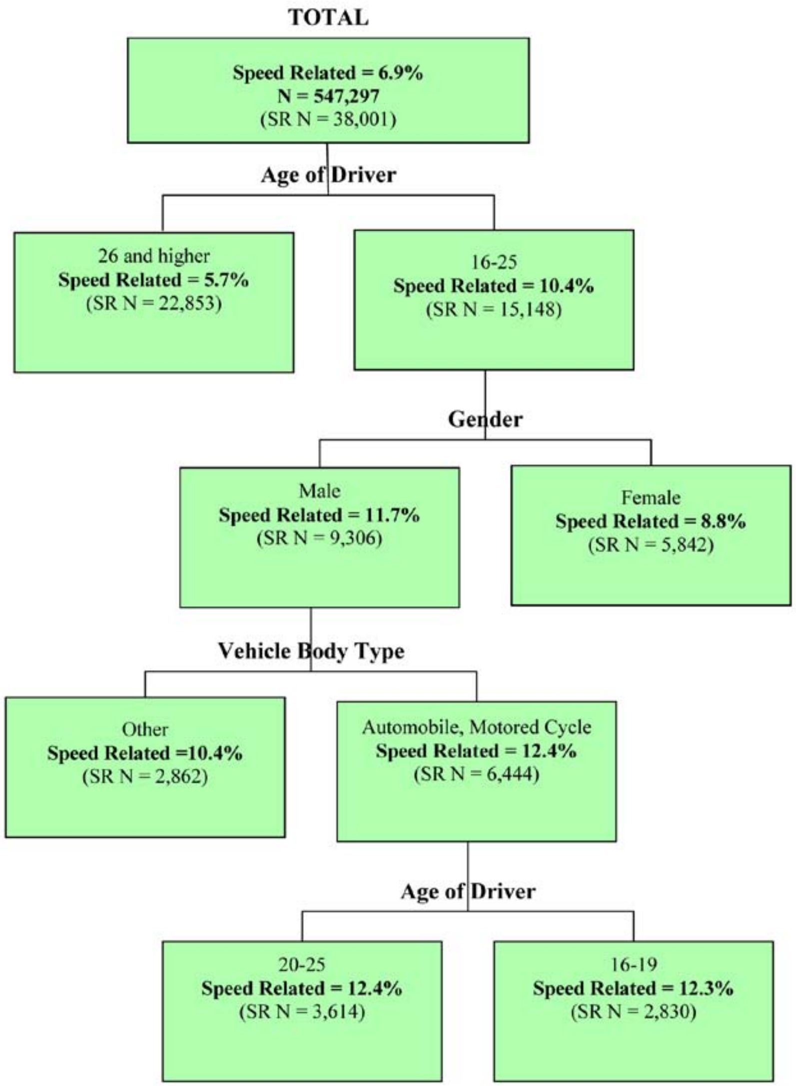

Figure 12. Chart. CART results for Ohio vehicle-based variables using the combined SR definition (2003–2005).

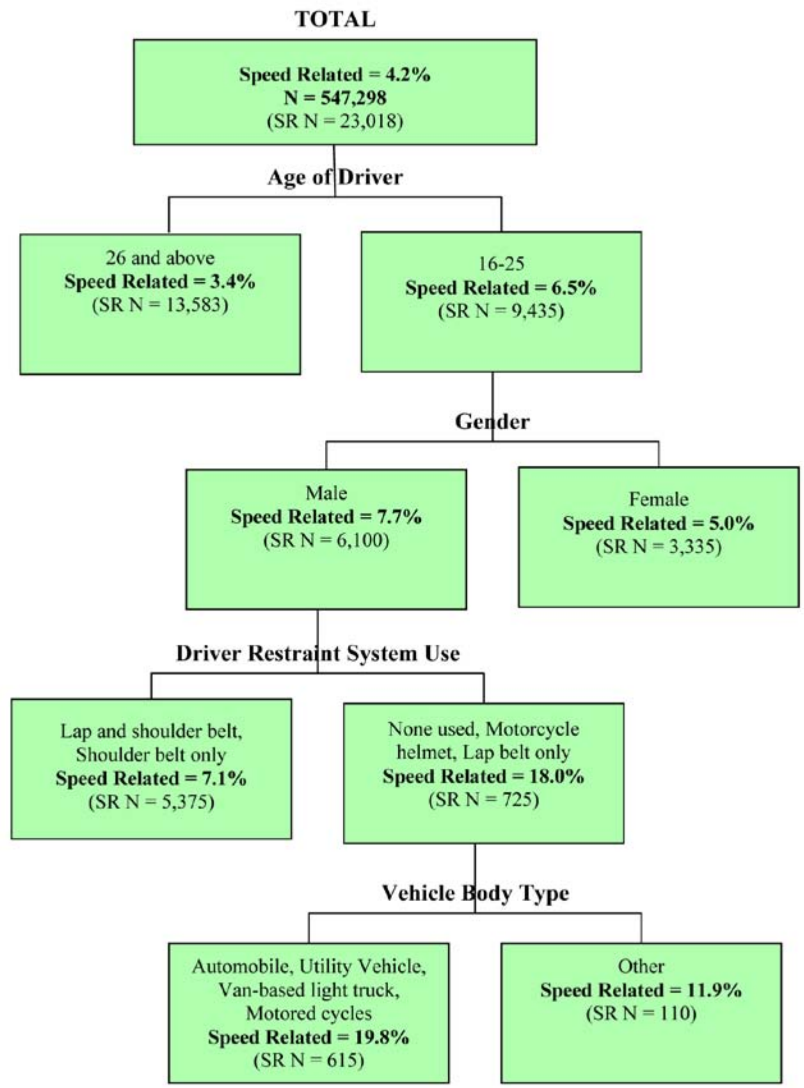

Figure 13. Chart. CART results for Ohio vehicle-based variables using the over speed limit SR definition (2003–2005).

SUMMARY CONCLUSIONS ON VEHICLE/DRIVER-RELATED CART ANALYSES

The findings from the vehicle-based analyses differed across databases and definitions. Using the liberal combined SR definition, FARS noted driver alcohol use as the most important descriptor, GES indicated distracted drivers as the most important descriptor, and North Carolina and Ohio data indicated young drivers as the most important descriptors. The over speed limit analysis in North Carolina indicated young drivers up to age 35 who are drinking while not using restraint systems as the most important descriptors, while the Ohio data indicated young males not using restraints as the most important descriptors. The one theme in most (but not all) of the results is that they all include young males.

CART Results for Data Subset Analyses

As described in the single-variable results, additional analyses were conducted for SR crashes in five subsets of the FARS and GES data: (1) pedestrian SR crashes, (2) intersection-related SR crashes, (3) lane departure SR crashes, (4) rural SR crashes, and (5) urban SR crashes. In additional to the previously described single-variable analyses, CART analyses were conducted for the FARS and GES crash-related variables in the first three of these subsets. Unlike the previous CART analyses in which all crash-related variables were examined by the CART software, only the selected group of variables were examined. These included the following:

- **Pedestrian crashes:** rural/urban, functional class, day/night (using light condition), speed limit, relationship to junction, and number of travel lanes.

- **Intersection crashes:** rural/urban, functional class, day/night, speed limit, traffic control device, number of travel lanes, alignment, and grade.

- **Lane departure crashes:** rural/urban, day/night, speed limit, number of travel lanes, alignment, grade, and surface condition.

Note that the actual subset that CART uses was all pedestrian, intersection, and lane departure crashes. It then examined each of the variables and determined which was the best SR predictor in the initial step. The results of these limited CART analyses are described below. By definition, these are analyses of a restricted set of variables and that CART results for the full set of crash-related variables within each subset can produce different findings.

Pedestrian Crashes

The CART results for the fatal FARS SR crashes using the limited set of crash-related variables only produced one branch/level (see figure 14). The most important SR predictive variable is speed limit, and the categories with the highest SR percentage are speed limits of 65–75 mi/h. Approximately 16 percent of these pedestrian crashes are SR, and this subset accounts for approximately 2 percent of the total fatal pedestrian crashes and approximately 17 percent of the total SR sample, which is a relatively small subset.

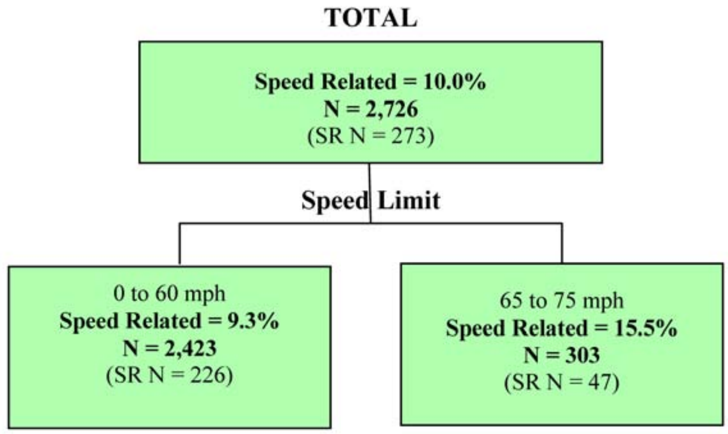

Figure 14. Chart. CART results for FARS pedestrian subset using limited crash-based variables.

While a CART analysis was conducted for the companion SR pedestrian crashes in the GES database, it did not produce any branches. Thus, the analyses of these limited sets of variables for SR pedestrian crashes indicates that fatal crashes are more likely to be roads with the highest speed limits, but there are no variables that predict higher SR involvement for the full crash dataset found in GES.

Intersection-Related Crashes

The CART analyses for the fatal FARS SR intersection crashes using the limited set of variables produced a two-level tree (see figure 15). The most important SR predictive variable is light condition, and the categories with the highest SR percentage include dark, dark but lighted, and dawn (28 percent). Within that branch, the next variable is roadway alignment, with the highest SR category being curves. This final subcategory is 53 percent and accounts for approximately 2 percent of the total fatal intersection-related sample and approximately 10 percent of the total SR intersection sample.

The companion CART analyses of the GES intersection crashes using the limited set of variables produced a three-level tree (see figure 16). The most important SR predictive variable is light condition, and the categories with the highest SR percentage include dark, dark but lighted, and dawn—the same as in the FARS results. Then within that branch, the most important variable is speed limit, with the highest SR category being zero to 35 mi/h. Within that branch, the most important variable is traffic control, with the highest SR categories being no control (as opposed to traffic signals and stop/yield). This final subset is 30 percent SR. It includes approximately 0.8 percent of the total GES intersection crashes in this training set and approximately 5 percent of the total SR intersection crashes.

The FARS and CART results were consistent with respect to the most important predictor variable, which shows that SR crashes are more likely at night. There is less consistency after that, with the fatal intersection results showing that intersections on curves are of interest (which might imply intersection in rural areas where curves are more likely), while the total (GES) results indicate the most important intersections are those with lower speed limits (likely urban) and no stop/yield sign or traffic signal in uncontrolled urban intersections.

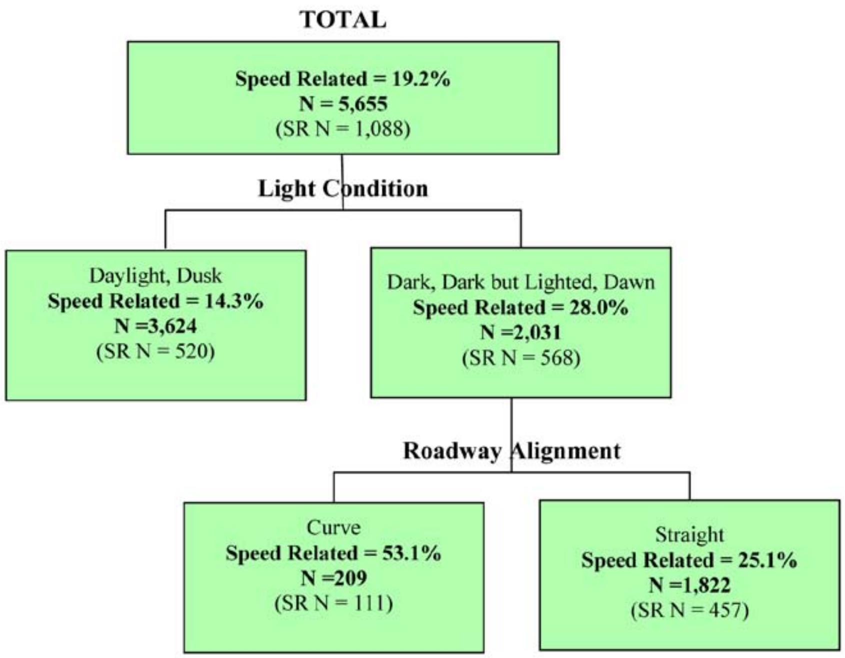

Figure 15. Chart. CART results for FARS intersection subset using limited crash-based variables.

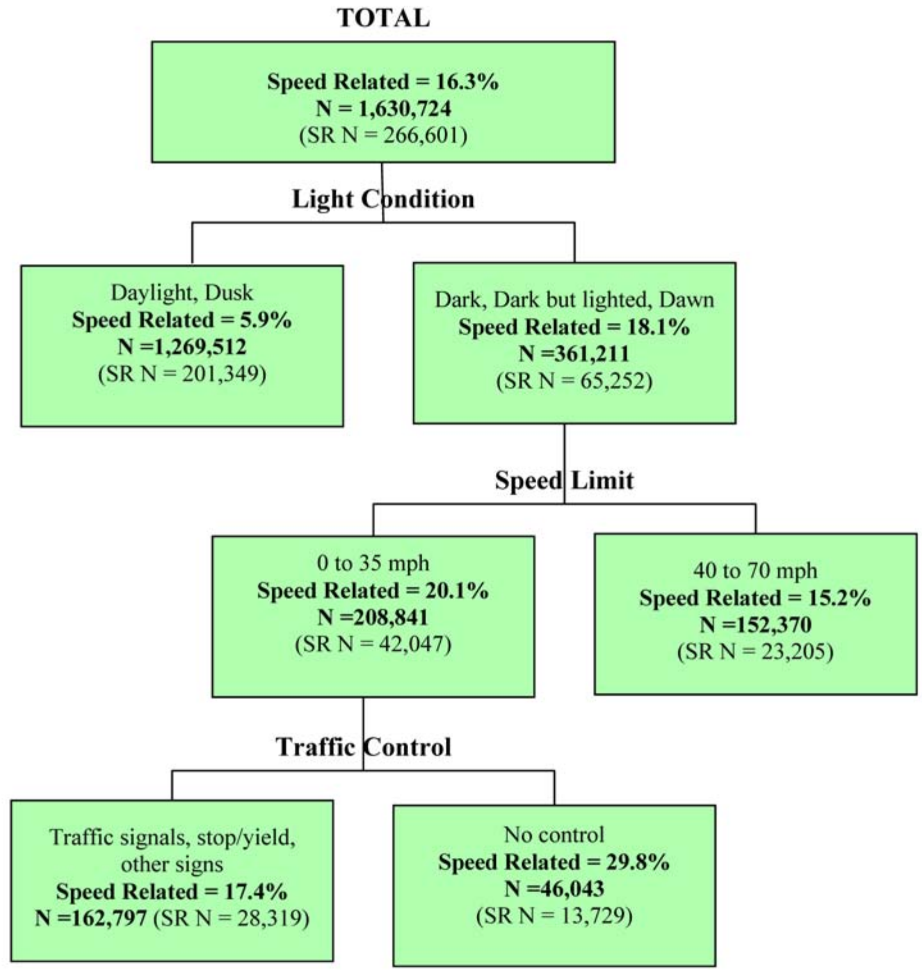

Figure 16. Chart. CART results for GES intersection subset using limited crash-based variables.

Lane Departure Crashes

The CART analyses for the fatal FARS SR lane departure crashes using the limited set of variables produced a four-level tree (see figure 17). The most important SR predictive variable is the lane departure variable produced for this analysis, and the category with the highest SR percentage is the single-vehicle, run-off-road category. Within that branch, the most important variable is light condition, and the categories with the highest SR percentage include dark, dark but lighted, and dusk. Within that branch, the next variable is roadway alignment, and the category with the highest SR percentage is curves. Finally, within the curve branch, the final variable is speed limit, and the category with the highest SR percentage includes all speeds of zero to 50 mi/h. This final subcategory is 63 percent of SR crashes and includes approximately 6 percent of the fatal lane departure crashes and approximately 16 percent of the fatal SR lane departure crashes.

The companion CART analyses of the GES lane departure crashes using the limited set of variables produced a two-level tree (see figure 18). The most important SR predictive variable is surface condition, and the categories with the highest SR percentage include wet, snow, slush, sand, dirt, oil, and other. Within that branch, the next variable is speed limit, with the highest SR percentage being for speed limits of 65–75 mi/h—the highest speed limit categories. This final subcategory is 62 percent of SR crashes and includes approximately 3 percent of the total lane departure crashes in this training set and approximately 10 percent of the SR lane departure crashes.

The results from FARS and GES analyses for fatal and total lane departure crashes are not consistent. The most important fatal lane departure crashes in terms of SR are those at night, on curves, and on roads with speed limits less than 55 mi/h. The most important total lane departure crashes are those during bad weather on roads with the highest speed limits (i.e., interstate roads).

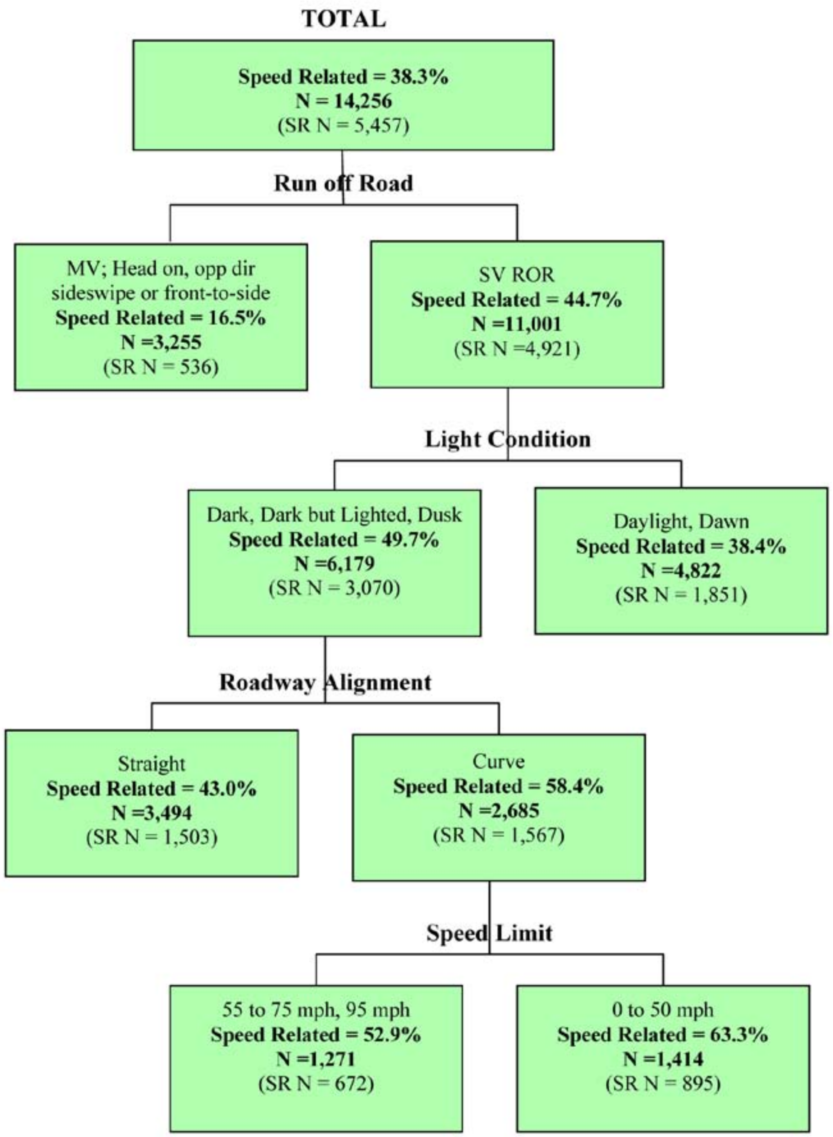

Figure 17. Chart. CART results for FARS lane departure subset using limited crash-based variables.

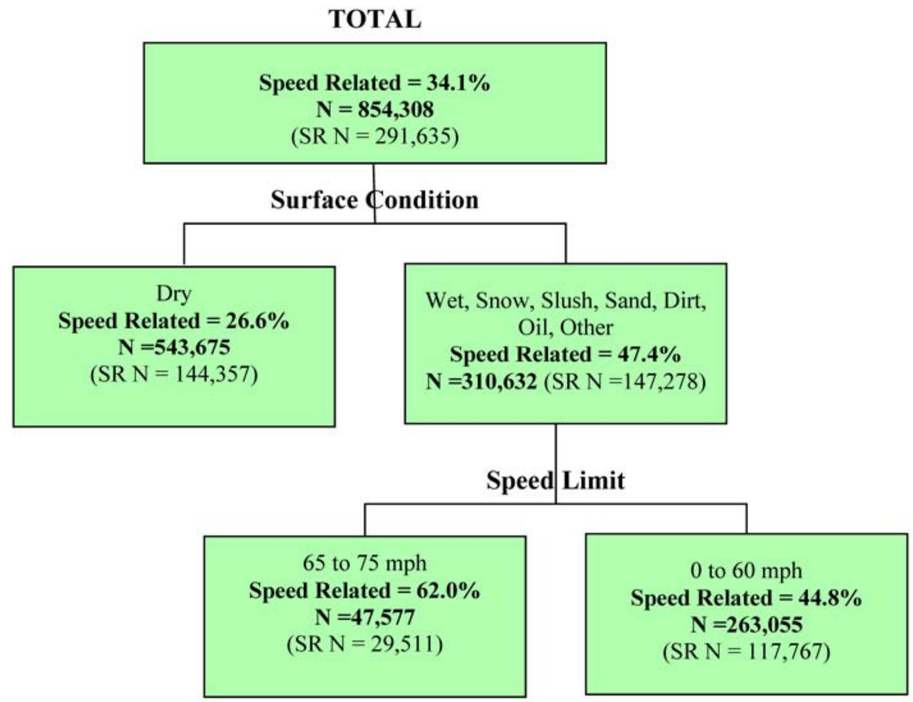

Figure 18. Chart. CART results for GES lane departure subset using limited crash-based variables.

SUMMARY AND CONCLUSIONS

This report described a large set of analyses aimed at developing information for use in better defining new treatments that could reduce SR crashes or crash severity and in better targeting existing treatments. It is difficult to summarize the findings of these analyses due to the number of outputs produced and the fact that the outputs were from two different methodologies—analysis of single-variable tables and regression tree development. It is also difficult because four different databases were used. General findings based on the overall results include the following:

- There were differences between the fatal and total crash data, between the State and national crash data, and between the States. The differences between factors related to fatal and total (i.e., primarily nonfatal) crashes in the FARS and GES data were not unexpected since they were paralleled by similar difference in many other studies—fatal SR crashes differed to some extent from nonfatal crashes. This must be kept in mind during treatment development and may sometimes lead to a choice between whether to attempt to affect fatal or total crashes. The expected differences between the State and national data and the States only indicated that while speed reduction treatments developed for the United States as a whole should clearly be applicable to all or many of the States, the priorities for the use of different treatments and the targeting of the treatments may change from State to State, reinforcing the need for each State to develop a data-driven plan, namely their strategic highway safety plans.

- Few differences were seen between the results based on the two definitions. The use of the two definitions for SR—the combined definition including over the posted speed limit and too fast for conditions and the over speed limit definition—did not reveal many changes in findings when used in the two States. The somewhat expected difference between weather and road condition variables was found (e.g., dry conditions having more SR crashes in the over speed limit definition and the opposing nondry conditions being most important in the combined definition), but most other findings were somewhat consistent under both definitions. This lead to supporting the national finding in which only the combined definition was possible.

- The CART findings were at times more difficult to interpret than expected. The CART regression tree analyses provided information on interactions between variables and on defined subsets of the data that could be targeted for future efforts. However, there were times when allowing CART to choose the combinations of categories that were high SR and low SR within a given variable led to some combinations that did not provide clear information (e.g., the combination of no restraint and lap and shoulder use in the higher SR branch of the GES vehicle-based tree). In addition, even though the full analyses were restricted to only four levels of branches, there were times when the most important subcategory in terms of percentage of SR crashes included only a small percentage of either total crashes or total SR crashes. CART is not a perfect methodology; it requires careful interpretations and usage.

With respect to more specific findings concerning high-priority variables and categories, in general, the findings were consistent with the findings from the study by Bowie and Walz and somewhat consistent with the 2001 study by Hendricks, et al.[5,8] Higher SR percentages in single-vehicle crashes, rural crashes, crashes on curves, nighttime crashes, motorcycle crashes as well as crashes involving male drivers, drivers not using restraints, and drivers using alcohol were found in either one or both of those earlier studies and in these current findings. In addition, key findings (including some which were inconsistent) from the current single-variable analyses include the following:

- Younger drivers were more likely to be involved in SR crashes including drivers who were 21–25 years old. Regardless of database or definition, the percentage of SR crashes was highest for the youngest drivers (16–19 years old) and decreased with age. It should be noted, however, that the 20–25-year-old age group was consistently higher than the older categories.

- In fatal crashes, drivers with prior speeding convictions were more likely to be involved in SR crashes. Based on FARS data, the percentage of SR crashes increased with prior speeding convictions. For drivers with three or more prior convictions, the SR percentage was almost twice that of those with no prior convictions (32.4 percent versus 18.9 percent).

- Speeding did not seem to be as important a factor in crashes involving pedestrians and bicycles. All databases and definitions indicated that the pedestrian and bicycle crashes were less likely to be SR than other type crashes. The nonpedestrian/bicycle crashes had two to five times the SR percentage as the crashes involving pedestrians/bicycles. This could result partly from the fact that, even at low speeds, pedestrians and bicyclists were likely to be injured.

- For SR crashes, high-priority, functional classes differed by database. While findings from GES and the two States showed agreement using the combined definition for SR crashes (i.e., interstates, particularly rural interstates in the State data, had higher SR percentages) the FARS results differed, showing minor collectors and local roads as having higher SR percentages. The use of the over speed limit definition in the State data provided even more complexity with North Carolina data, showing that speeding over the posted limit was a bigger problem on local and minor roads. However, Ohio data showed a bigger problem on interstates and arterials. These same trends in findings were reflected in the analyses of speed limits and number of lanes. GES and the State data using the combined definition indicated that the SR crash percentages increased with speed limit and the number of lanes. FARS data indicated roadways with lower speed limits and two lanes have higher SR crash percentages. The conclusion is that fatal crashes differed from total crashes, and crashes likely differed between States.

- There was no consistent indication that speeding was a critical factor in work zone crashes; no consistent pattern was seen. FARS data showed essentially no difference between work and nonwork zones, GES data showed only a slightly higher SR percentage in the work zones (25 percent versus 21 percent), North Carolina data showed a lower SR percentage for work zones under both definitions, and Ohio data showed a higher SR percentage in work zones for both definitions.

With respect to the CART analyses, there were inconsistencies with the findings from the different databases as follows:

- CART crash-based findings consistently identified single-vehicle crashes during adverse weather as high-priority subgroups. GES included rear-end crashes as a top priority target. While there were no other significant findings for the combined definition in the GES and North Carolina data, FARS added curved roads with lower speed limits to the key descriptors of fatal SR crashes. Analyses of the North Carolina and Ohio data using the over speed limit definition indicated a key difference from the combined definition findings in that dry (or dry and wet) conditions replaced the snowy/icy conditions as key descriptors. This implied that roadway-based treatments such as improved snow and icy removal and speeding-enforcement treatments might both be appropriate for treating SR single-vehicle crashes.

- CART vehicle-based findings indicated that there was almost no consistency across databases, with perhaps young male showing up more than other descriptors. FARS data indicated that in fatal crashes, alcohol use was the primary predictor, with drivers being young and male as further descriptors. In the total crash datasets using the combined definition, GES noted distraction as the only important vehicle-based predictor, while the North Carolina and Ohio databases indicated young males as the primary intervention targets. The North Carolina and Ohio data produced completely different results using the over speed limit definition. North Carolina data indicated that the most important target subgroup was drinking drivers up to age 35 who were not using restraints, and Ohio data indicated 16–25-year-old males who were not using restraints as the most important target subgroup.

This study produced a large group of findings which were not all consistent across the four databases and two definitions. This was not totally unexpected, as prior studies of other crash types not related to speeding have shown that the characteristics of fatal and nonfatal crashes do differ, and States would be expected to differ from each other at times and from a composite national picture. This effort produced some consistent (and inconsistent) findings that can be used in target development and targeting. The findings were consistent with those from the SR study conducted in 1994 by Bowie and Walz.(4) In essence, the problem characteristics have not changed much, and the problem is still a significant one that demands attention. The current focus on the issue is well justified and of critical importance in further reducing the cost to society resulting from motor-vehicle crashes in the United States.

APPENDIX: IDENTIFICATION OF CRITICAL FACTORS USING CLASSIFICATION TREES

Analyses of single-variable tables do not automatically indicate which factors/variables are the most critical with regard to SR crashes or speeding drivers. They also do not indicate which combinations of variables are the most important. One way to identify the critical roadway, vehicle, and driver factors associated with an increased likelihood of an SR crash is to estimate a logistic regression model with the roadway, vehicle, and driver factors as independent variables and then to identify the statistically significant factors. Logistic regression is a parametric approach which is based on assumptions about error distributions. The CART methodology is nonparametric and does not require any such assumptions. In addition, CART is able to include a relatively large number of independent variables and identify complex interactions between these variables more efficiently compared to logistic regression. For example, CART is able to determine not only the most important variable and categories within that variable in terms of the risk of an SR crash but also the most important second-level variable within the most important categories of the first level variable, etc. That is, given the most important variable with respect to the proportion of SR crashes (e.g., manner of collision) and the subgroup of categories within that variable with the highest proportion of SR crashes (e.g., run-off-road crashes), CART is able to determine the next most important variable within these high-risk categories (e.g., road surface condition) and the categories of that variable that are most important (e.g., snow and sleet). It would be hoped that these variables and categories would be helpful in determining needed treatments. For these reasons, it was decided that classification trees would be used as the second type of analysis in this project. A classification tree divides the data set into smaller sets which are more homogeneous in the values of certain variables.

The following provides a limited overview of classification trees. Further information about CART is available a study conducted by Breiman et al.(11) For applications of these trees in road safety research, see Stewart and Yan and Radwan.(12,13)

OVERVIEW OF CLASSIFICATION TREES

The goals of CART are as follows: (1) to determine which of the many possible crash, driver, or vehicle variables available for examination is most important in terms of predicting SR crashes, (2) to determine which categories within that variable predict the highest risk/proportion of SR crashes, (3) to determine the second most important variable and subset of categories within this highest-risk subset of categories of the first variable, and (4) to repeat the process to determine the third, fourth, and subsequent variables. This produces a tree with multiple branches that can be traced down to determine the most important combinations (or subsets) of variable categories in terms of predicting SR crashes. The CART procedure basically splits the categories of each variable in the database into all possible binary (two-category) combinations (nodes), calculates the SR risk within each part (node) of each pair, and determines which pair (i.e., which two sets of categories) produces the largest difference in SR risk within that variable. By repeating this process for each variable in the database, CART determines the two sets of categories producing the largest difference in risk of SR crash within each variable. This largest difference in risk is then compared across all variables to determine the one variable (and the set of categories) producing the largest of all differences. This is the top of the tree, and the two categories within that variable are the first two branches of the tree. This process is then repeated within each of the two categories (branches) of the first variable to identify the second, third, and subsequent most important variables. For a categorical variable, all possible binary combinations of categories are compared (e.g., category 1 versus category 2–5, category 1 and 2 versus category 3–5, category 1 and 3 versus categories 2, 4, 5, etc.). For ordinal variables (e.g., speed limit), all cases with the value of that variable smaller than or equal to a certain value go to one node, all other cases go to the other (e.g., ≤ 30 mi/h versus ≥ 35 mi/h; ≤ 35 mi/h versus ≥ 40 mi/h, etc.). The following describes this process in statistical terms.

In general, assume that there is a dataset D with N cases, in this case, crashes or vehicles/drivers involved in crashes. In order to develop a classification tree, a target variable is needed, which is the variable which will be used by the tree to classify a case in some optimal way. A target variable should always be categorical. The other variables might be ordinal or categorical. The target variable is the variable describing whether a crash was SR or whether a vehicle involved in a crash was speeding. This is indeed a categorical variable because it has only two categories (yes and no). In the following overview of classification trees, there is also a target variable with two categories. Since the objective is to estimate the probability that a crash is SR, the procedure for developing a class probability tree will be discussed. This means a class probability estimator d, should be constructed as follows:

$$d(x) = (d(1 \mid x), d(2 \mid x)), \quad d(1 \mid x), d(2 \mid x) \geq 0, \quad d(1 \mid x) + d(2 \mid x) = 1, \quad \forall x \in D. \tag{1}$$

Building a classification tree starts with dividing case D in two parts. One part, D_1, is the training sample, which is used to build the complete tree. The other part, D_2, is called the validation set and is used to prune the complete tree back to a simpler tree such that it will have a better predictive ability. The number of cases in the training and validation set will be denoted by N_1 and N_2, respectively. In this case, 70 percent of the cases were included in the training sample.

All the cases in the training sample form the root note, t_0, of the tree. Assume that $p(j \mid t_0)$, j = 1, 2, the proportion of this set for which the target variable falls in category j. The impurity $i(t_0)$ of the root node then describes how mixed the node is according to the target variable. This means that the impurity should be equal to zero when all the cases in the root node belong to the same category and that the impurity should have its maximum value when half of the cases fall in one category and half of the crashes fall in the other category. There are several functions possible which satisfy these conditions, but Breiman et al. suggest the use of the Gini index seen in equation 2.[11]

$$i(t_0) = p(1 \mid t_0) p(2 \mid t_0). \tag{2}$$

The classification tree algorithm now the impurity of the nodes computes for all independent variables which would result from splitting the root node in two new nodes. If an independent variable is ordinal, then all splits of the following type are considered: all cases with the value of that variable smaller than or equal to a certain value go to one node and all other cases to the other. For a categorical variable, all splits of the following types are considered: if $J = \{j_1, \ldots, j_k\}$ is the set of possible values of that variable, then all cases for which the variable has a value in $J_{\text{sub}} \subset J$ go to one node, all other observations go to the other. Let S be the set of all possible splits, and assume that a split s in this set S results in two new nodes denoted by $t_{1,s}$ and $t_{2,s}$. Then the impurity of node $t_{i,s}$, i = 1,2 is given by equation 3 as follows:

$$i(t_{i,s}) = p(1 \mid t_{i,s})p(2 \mid t_{i,s}), \tag{3}$$

where p is analogously defined as for the root node. The decrease of impurity due to a certain split s is then equal to the following:

$$\Delta i(s,t_0) = i(t_0) - i(t_{1,s}) - i(t_{2,s}). \tag{4}$$

The split maximizing $\Delta i(s,t_0)$ is chosen. For each of the two new nodes, this splitting procedure is repeated. Splitting stops when the end nodes splitting will not decrease impurity anymore (all observations in that node are then either SR or not SR) or when the end nodes consist of less than a certain number of observations that is set in the beginning.

In most cases, this splitting algorithm results in a very large tree, T_m, where m denotes the number of terminal nodes (also called leaves) which are not split anymore. Each node gives the proportions of cases in that node falling in each of the categories of the target variable. These proportions can be interpreted as the probabilities that a case, which is run through the tree and ends up in a certain node, falls in each of the categories. This tree is helpful in predicting the proportions of the cases in the training sample in each because this set was used to build the tree. When the tree is applied on the other 30 percent of the original data set, it may not do well in predicting the proportions, especially for the nodes resulting from a high number of splits. That is why the classification tree procedure continues with pruning back the resulting tree from the previous step.

Before the actual pruning phase, a sequence of subtrees of T_m is constructed by using the training sample. This sequence consists of n – 1 subtrees of T_m. Subtree T_k, k = 1,…,m – 1, satisfies the following conditions:

- It has exactly k leaves.
- It is a subtree of T_{k+1}.
- It is in some sense the most optimal subtree of T_m with k leaves.

It goes too far to explain this last condition in detail; its optimality has to do with its total Gini index (sum of the Gini indices of all its terminal nodes) and its number of terminal nodes.

In the final step of the classification tree building process, the best subtree is determined using the validation sample. Before explaining what the best subtree is, some notation is introduced. Let $D_{2,j}$ be the part of the validation set D_2 consisting of all cases falling in category j. The number of crashes in $D_{2,j}$ will be denoted by $N_{2,j}$. Further, d is the class probability estimator, which means that if x is a crash in D_2 which ends up in node t while ran through the tree, then the probabilities that x is of category 1 and that x is of category 2 are given by d as follows:

$$d(x) = (p(1 \mid t), p(2 \mid t)), \tag{5}$$

Where, $p(j \mid t)$ is the proportion of cases of the training set D_1 in node t which are in category j. Finally, the map z_i is defined as follows:

$$z_i(x) = \begin{cases} 1, & \text{if } i = j(x); \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Where $j(x)$ is the real category of x.

To determine the best subtree, all the cases in $D_{2,1}$ are first run down all the subtrees, followed by the cases in $D_{2,2}$. For each subtree, the following values are computed as follows:

$$R_j(T_k) = \frac{1}{N_{2,j}} \sum_{\substack{x \in D_{2,j} \\ i \in \{1,2\}}} \left(z_i(x) - d(i \mid x)\right)^2, \qquad j = 1,2. \tag{7}$$

The subtree for which $R_0(T_k)\pi(1) + R_1(T_k)\pi(2)$ (which is an estimate of the mean square error of d and is sometimes called the total leaf impurity) is minimal will be considered the best subtree and will be the final classification tree. In this expression, $\pi(j)$ is the proportion in the entire dataset (so N_1 and N_2) belonging to category j.

Dealing with Missing Values

If the target variable is missing, the crash is not used for growing or pruning the tree. Therefore, these observations are removed from the datasets before developing the trees. Classification trees are able to handle missing values of the independent variables. If the value of a variable for a case on which a node will be split is missing, then a surrogate split is used. This means that based on the value of another variable, it is decided to which of the new nodes the case goes. There are a lot of variables and possible splits to choose from, but the one that gives the most similar results as the real split will be chosen.

In GES and FARS, unknown values of variables are coded as 9, 99, or similar. Because these are not automatically recognized by SAS® Enterprise Miner as missing values, these codes are recoded as blank.[10]

REFERENCES

1. National Highway Traffic Safety Administration. (2009). *Fatality Analysis Reporting System (FARS)*, Washington, DC. Obtained from: ftp://ftp.nhtsa.dot.gov/FARS/.

2. National Highway Traffic Safety Administration. (2005). *Speed Management Strategic Initiative*, DOT-HS-809-924, U.S. Department of Transportation, Washington, DC

3. Transportation Research Board. (1998). *TRB Special Report 254: Managing Speed—Review of Current Practice for Setting and Enforcing Speed Limits*, National Research Council, Washington, DC.

4. Bowie, N.N. and Walz, M. (1994). "Data Analysis of the Speeding-Related Crash Issue," *Auto & Traffic Safety*, 31–38.

5. National Highway Traffic Safety Administration. (1984–1986). *Crash Avoidance Research Data File (CARDfile)*, Washington, DC

6. Treat, J.R., Tumbas, N.S., Mcdonald, S.T., Shinar, D., and Hume, R.D. (1979). *Tri-Level Study of the Causes of Traffic Accidents*, Executive Summary, Report No. DOT-HS-805-099, National Highway Traffic Safety Administration, Washington, DC.

7. Hendricks, D.L., Fell, J.C., and Freedman, M. (2001). *The Relative Frequency of Unsafe Driving Acts in Serious Traffic Crashes—Summary Technical Report*, Report No. DOT-HS-809-205, National Highway Traffic Safety Administration, Washington, DC. Obtained from: http://ntl.bts.gov/lib/26000/26000/26077/DOT-HS-809-205.pdf.

8. National Highway Traffic Safety Administration. (2009). *National Automotive Sampling System Crashworthiness Data System (NASS CDS)*, Washington, DC. Obtained from http://www.nhtsa.dot.gov/people/ncsa/nass_cds.html.

9. National Highway Traffic Safety Administration. (2009). *National Automotive Sampling System/General Estimates System (NASS GES)*, Washington, DC. Obtained from: ftp://ftp.nhtsa.dot.gov/GES/.

10. Enterprise Miner 4.3. SAS® 9.1, SAS® Institute, Cary, NC.

11. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984). *Classification and Regression Trees*, Chapman and Hall/CRC, New York, NY.

12. Stewart, J.R. (1996). "Applications of Classification and Regression Tree Methods in Roadway Safety Studies," *Transportation Research Record 1542*, 1–5, Transportation Research Board, Washington, DC.

13. Yan, X. and Radwan, E. (2006). "Analysis of Rear-End Crashes Based on Classification Tree Models," *Traffic Injury Prevention*, *7*, 276–282.

HRDS-05/04-10(WEB)E