HSIS

Highway Safety Information System

SUMMARY REPORT

Development of a Speeding-Related Crash Typology

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. Between 1990 and 2006, the frequency of speeding-related (SR) fatal crashes ranged from 11,000 to 13,000 each year, and the percentage of SR total fatal crashes ranged between 30 and 33 percent according to data observed in the Fatal Analysis Reporting System (FARS).[1] Thus, speeding is a significant safety issue warranting attention based on its size and impact on society. While the United States has seen progress in other major safety issues such as occupant restraint use and driving under the influence of alcohol, little if any progress has been made with speeding. In response to this issue, the United States Department of Transportation has instituted the *Speed Management Strategic Initiative*, seeking more effective ways to manage the crash-related effects of speeding.[2] In support of this initiative, this study examined recent crash data through the development of an SR crash typology. Such a typology can help define the crash, vehicle, and driver characteristics that seem to result in a higher probability of SR crashes. Thus, the goal of this study was to determine the "what," "where," "when," and "who" descriptors of SR crashes in order to provide guidance to the future development of new treatments and to better target new and existing treatments to subgroups of drivers and types of roadways (e.g., two-lane rural) or roadway locations (e.g., unsignalized intersections).

The Highway Safety Information System (HSIS) is a multi-State safety database that contains crash, roadway inventory, and traffic volume data for a select group of States. The participating States—California, Illinois, Maine, Michigan, Minnesota, North Carolina, Ohio, Utah, and Washington—were selected based on the quality of their data, the range of data available, and their ability to merge the data from the various files. The HSIS is used by FHWA staff, contractors, university researchers, and others to study current highway safety issues, direct research efforts, and evaluate the effectivness of accident countermeasures.

Literature Review

While numerous research studies have explored the effects of speed on crash frequency and severity and on the effect of treatments aimed at managing speed (e.g., *TRB Special Report 254*), two studies have developed typologies similar to those developed in this effort.[3] Bowie and Walz used data from the 1986 Crash Avoidance Research Data (CARDfile) from six States as well as from the Indiana Tri-Level Study and the 1989 FARS data. (See references 4–6 and 1.) 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 to 16.9 percent of crashes.[6] Key findings from the FARS data indicate that fatal SR crashes are usually single-vehicle and that there is a higher percentage of SR crashes on rural roads, on curves, and at night.[1] In these fatal crashes, males were more likely than females to be speeding, and drivers under the influence of alcohol were more likely to be speeding than those who were not drinking. Occupant restraint usage was lower for SR drivers. In addition, more than 45 percent of all motorcycle drivers involved in fatal crashes were speeding.

Hendricks, et al. examined data from in-depth investigations and driver interviews from a sample of 723 relatively severe crashes involving 1,284 drivers collected in a special study as part of the National Automotive Sampling System Crashworthiness Data System (NASS CDS) program.[7,8] Researchers

U.S. Department of Transportation
Federal Highway Administration

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

determined the specific driver behaviors and unsafe driving acts that led to the crashes as well as the situational-, driver-, and vehicle-based characteristics associated with these behaviors. Where cause could be assessed, excessive vehicle speed was the second leading causal factor (i.e., second to inattention). Excessive speed typically involved drivers exceeding the speed limit, but it sometimes included drivers operating at inappropriate speeds for prevailing weather or roadway conditions (i.e., too fast for conditions). For drivers who were assessed to have partially or fully caused the crash, excessive speed was the sole primary cause of the crash for 6.8 percent of these drivers. Excessive speeding 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 these drivers 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. The 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, at night, on local or collector roadways, and during clear weather. In addition, younger male drivers (less than 35 years old) were more involved, with males younger than 20 years old comprising 46.2 percent of the sample.

Methodology

Databases Used

Since a goal of this study was to explore the "what," "where," "when," and "who" questions as they relate to the United States, the study used 2005 data from the two major national crash databases: the National Automotive Sampling System's General Estimates System (GES) and FARS.[9,1] GES data are derived from a nationally representative sample of police-reported motor vehicle crashes of all types, from minor to fatal. 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 6 million U.S. crashes that occur each year. FARS is a census of all fatal crashes in the United States, with approximately 40,000 crashes occurring in 2005. Because of the definitional issues and lack of data described below that arose with the FARS and GES data, data from two States that are part of the Federal Highway Administration's (FHWA) Highway Safety Information System (HSIS) were also used: 2002–2004 data from North Carolina and 2003–2005 data from Ohio.

Definition of SR

For all databases used in this study, the decision on whether or not a crash is SR is based on an examination of each vehicle in the crash. First, each vehicle is defined as SR or not based on several variables. If any vehicle in the crash is SR, the crash is considered SR. In addition, if all of the vehicles involved in a crash are coded as non-SR, then the crash is defined as not SR. In all other cases, it is unknown whether or not the crash is SR.

The FHWA was interested in examining speeding using a disaggregated definition where "exceeding posted speed limit" (herein referred to as "over speed limit") and "too fast for conditions" would be analyzed separately since the appropriate counter-measures could differ for these two types of SR crashes.

However, while it is theoretically possible to distinguish between these two definitions in FARS and GES, the sample sizes available were not adequate for the more restrictive "over speed limit" definition. Thus, only the broader definition could be used—exceeding the posted speed limit or driving too fast for conditions (referred to as the "combined" definition). In order to better examine the possible effects of using only this combined definition, North Carolina and Ohio data were analyzed since they permit the use of both the over speed limit and the combined definitions, allowing comparisons of results within the same database. Note that the four databases differ; FARS provides a census of all U.S. fatal crashes, GES provides an extrapolated estimate of U.S. crashes of all severities, and the HSIS North Carolina and Ohio databases include crashes on only State system roads (i.e., excluding city streets or county roads that are not State controlled). Thus, differences in results would be expected and were found. Table 1 shows the sample sizes and percent SR for each of the databases.

Overview of Analysis Methods

Two analysis methodologies were used: (1) single variable table analysis and (2) classification and regression tree (CART) analysis. In the first methodology—single variable table analysis—for a series of both crash-based variables (e.g., crash type, weather, etc.) and vehicle- and driver-based variables (e.g., vehicle type, age of driver, etc.), 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, 16- to 19-year-old drivers, etc.). The second method used CART analyses that automatically defined which factors/variables were the most critical with regard to SR crashes or drivers and which combinations of variables/codes were the most important. Similar single variable and CART analyses were also conducted for five high-priority subsets of the data (e.g., pedestrian crashes, intersection crashes, etc.) for a limited number of variables within each subset. Additional detail on each of these methods is presented in the following sections.

Single Variable Table Analyses

As indicated above, single variable tables were produced 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 crashes, particularly on the earlier National Highway Transportation Safety Administration (NHTSA) study.[4] The factors describing the overall nature of the crash (e.g., crash type, crash location, etc.) were examined using a crash-based file where each crash was classified as speeding or not. The vehicle- and driver-based factors were examined in a vehicle-based file where each vehicle was also classified as speeding or not. In general, a category within a variable is defined as being over represented if it is characterized by a high percentage of SR crashes in comparison to percentages for other categories. In a few cases, the high percentage categories were found to have low frequencies of SR crashes (or drivers or vehicles) compared to the total SR frequency. These would not be considered as high-priority categories.

Identification of Critical Factors Using Classification Trees

Although the analyses of single variable tables provide useful information about SR crashes and vehicles and drivers in crashes, they do not automatically indicate which factors/variables and which combinations of variables are the most critical with regard to SR crashes or speeding drivers. The CART methodology provides such information and requires fewer statistical assumptions than do other methods (e.g., logistic regression).[10] 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 and categories within the most important categories of the first level, the third level within the second level, etc. Note that the analyst does not predefine the categories within a variable which should be grouped together based on higher SR percents; instead, CART does that automatically. The outputs of CART are presented as a tree with multiple branches that can be traced down to determine most important combinations (or subsets) of variable categories in terms of predicting SR crashes. A more detailed explanation of the CART methodology is presented in the main report, *Development of a Speeding-Related Crash Typology* (FHWA-HRT-10-024).[11]

Table 1. The number and percent of SR crashes in GES and FARS (2005) and in North Carolina (2002–2004) and Ohio (2003–2005) by definition type.

DEFINITION	FARS	GES	NORTH CAROLINA			OHIO		
	COMBINED	COMBINED	COMBINED	OVER SPEED LIMIT	TOO FAST FOR CONDITIONS	COMBINED	OVER SPEED LIMIT	TOO FAST FOR CONDITIONS
Total crashes	39,189	6,146,907	422,324			461,013		
SR crashes	11,553	1,195,570	62,746	12,802	49,944	51,906	30,677	30,677
Percent SR	29.5	19.5	14.9	3.0	11.9	11.3	6.7	4.6

Just as with the single variable table analyses, CART analyses of factors describing the overall nature of the crash were examined using a crash-based file, and CART analyses of vehicle- and driver-based factors were examined in a vehicle-based file.

Analyses of Data Subsets

In addition to the two sets of analyses described above, additional analyses of subsets of the data were conducted. While the above analyses included both SR and non-SR crashes (and vehicles and drivers), these subset analyses concerned only SR crashes in FARS and GES for five data subsets: (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. In addition, only selected variables were examined within each subset.

Results

Single Variable Table Analyses

While multiple tables were produced in examining the "what," "where," "when," and "who" questions related to SR crashes, figure 1 presents the results of greatest interest. While four different databases and two definitions were used, there were some consistencies across the findings. In addition, many of these are very similar to results found in the earlier two studies related to crash typologies, even though the first of these, the NHTSA study, was based on data from the 1980s, and the second unsafe driving acts study was based on a more limited sample of detailed crash investigations.[3,4] It is also noted that these SR findings point to some of the same descriptors as do current national and State emphasis areas, even though those areas were chosen based on non-SR as well as SR crashes—lane departure crashes, motorcycle crashes, and crashes involving drivers under the influence of alcohol and those not using restraints.

Single Variable Tables for High-Priority Subsets

As noted above, in addition to the analyses of a large series of individual variables in the four databases, additional single variable tables were produced for only SR crashes in FARS and GES for five data subsets. It is pointed out that these results differed from the original single variable results where the goal was to define categories within each critical variable that were more likely to be SR when compared to other categories. Here, the data are restricted to SR crashes, and the goal is to examine the nature of these SR crashes within each subset by examining only a selected list of variables.

Table 2 and table 3 include key results for the intersection and lane departure subsets. Note that the latter subset is predominantly run-off-road crashes (i.e., over 90 percent). Results for the urban and rural subsets

Figure 1. Chart. Single variable results of interest.

What crash characteristics have a higher SR percentage?

- Single-vehicle, run-off-road crashes have higher SR percentage compared to multivehicle crashes.
- SR percentage increases with speed limit in GES and State data, but it is slightly higher at low speed limits in FARS.

Where are crashes with higher SR percentages found?

- Crashes on curves have much higher SR percentages than crashes on tangents (even with the over speed limit definition).
- Functional class findings differ by database, with GES and State (combined definitions) showing higher SR percentages on interstates, while FARS shows higher SR percentages on minor collectors and local roads.
- No consistent pattern of higher SR percentages in work zone crashes.

When do crashes with higher SR percentages occur?

- Nighttime crashes have higher SR percentages.

Who has higher SR percentages in crashes?

- Motorcycle operators (i.e., two to four times SR percentages of car drivers).
- Younger drivers (younger includes 21–25 year olds).
- Males (but only slightly so in GES).
- Non-users of restraints.
- Drivers under the influence of alcohol (i.e., two to four times higher SR percentage than drivers not under the influence of alcohol).
- Drivers with prior speeding convictions (FARS analysis only).
- Drivers with no license or invalid license (FARS analysis only).
- Distracted drivers (GES analysis only).

(where only FARS data are available) and the pedestrian subset (which account for approximately 4 percent of the fatal SR crashes and less than 1 percent of the total SR crashes) are not presented in these tables but can be found in the full report.

CART Analyses

As previously indicated, CART analyses were used to identify the variables and category combinations within those variables that are most critical in describing SR crashes and drivers and the most important combinations of those variables. An example of the most important part of the tree produced when CART was used to examine all crash-level variables in the FARS database is shown in figure 2. The most important SR-predictive variable (the top tree branch) is labeled "First Harmful Event," and the categories with the highest SR percentage include rollovers/overturns, jackknife, and collisions with various fixed objects on the roadside. Then, within that branch, the next 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 limits between 20 and 45 mi/h. In this selected subsample, approximately 60 percent of the crashes are SR. CART did not detect a fourth-level variable within speed limit. In general, this tree indicates that run-off-road

Table 2. Selected characteristics of SR intersection crashes.

VARIABLE	CATEGORY	FARS (PERCENT)	GES (PERCENT)
Daytime		45	73
Predominant speed limit (mi/h)	30–35	30	33
	40–45	30	32
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
Two lanes		54	28
Predominant driver age (years)	16–19	15	21
	20–25	30	20
	36–50	19	22
Alcohol involved		28	6
Total frequency		**1,510**	**383,900**

Table 3. Selected characteristics of SR lane departure crashes.

VARIABLE	CATEGORY	FARS (PERCENT)	GES (PERCENT)
Night		58	46
Predominant speed limit (mi/h)	50–55	39	27
Straight (as opposed to curve/unknown)		51	55
Level (as opposed to grades/hills/unknown)		64	39
Two lanes		71	55
Dry (as opposed to wet, snowy/icy)		80	49
Predominant driver age (years)	16–19	17	24
	20–25	26	25
Alcohol involved		35	13
Total frequency		**7,796**	**415,631**

collisions on curves that are on roads with lower speed limits are more likely to be SR. Note that this final SR subcategory includes 1,258 of the 26,000 fatal crashes included in the sample analyzed—approximately 5 percent.

Similar CART analyses were conducted with the GES data and with both definitions in each of the two State data files (results from the State analyses are not shown in this summary report). Figure 3 shows the top levels of the resulting crash-based CART tree for the GES data. This illustrates the point that the findings from the CART analyses were somewhat less consistent across databases than the earlier single variable tables.

In this figure, like the earlier FARS results, the most critical descriptor is again related to a crash-type variable, but this time, rear-end crashes are included by CART with the single-vehicle, run-off-road crashes (note that *Not MV collision* is any crash in which there is not a collision with another motor-vehicle such as fixed object, pedestrian, etc.). While the next most important combinations involved curves and lower speed limits in FARS, they involved snowy and icy conditions in GES.

CART was also used to examine driver/vehicle characteristics and combinations in all four databases. In general, this analysis of fatal crashes (using the liberal combined SR definition) indicated that the fatal crash subset most likely to include SR vehicles were those involving drivers under the influence of alcohol who were young males (defined as those in the age group 16–35) driving either automobiles or motorcycles. Approximately 52 percent of the crashes in this final subset were SR, and the subset included approximately 2 percent of the total sample and 11 percent of the SR sample.

However, the same pattern does not exist in the GES vehicle/driver tree. There, the most critical factor was driver distraction of some type (e.g., inattentive, sleepy, etc.) in contrast with not distracted. The remaining descriptors for the distracted group were not very informative, grouping nonbelt

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

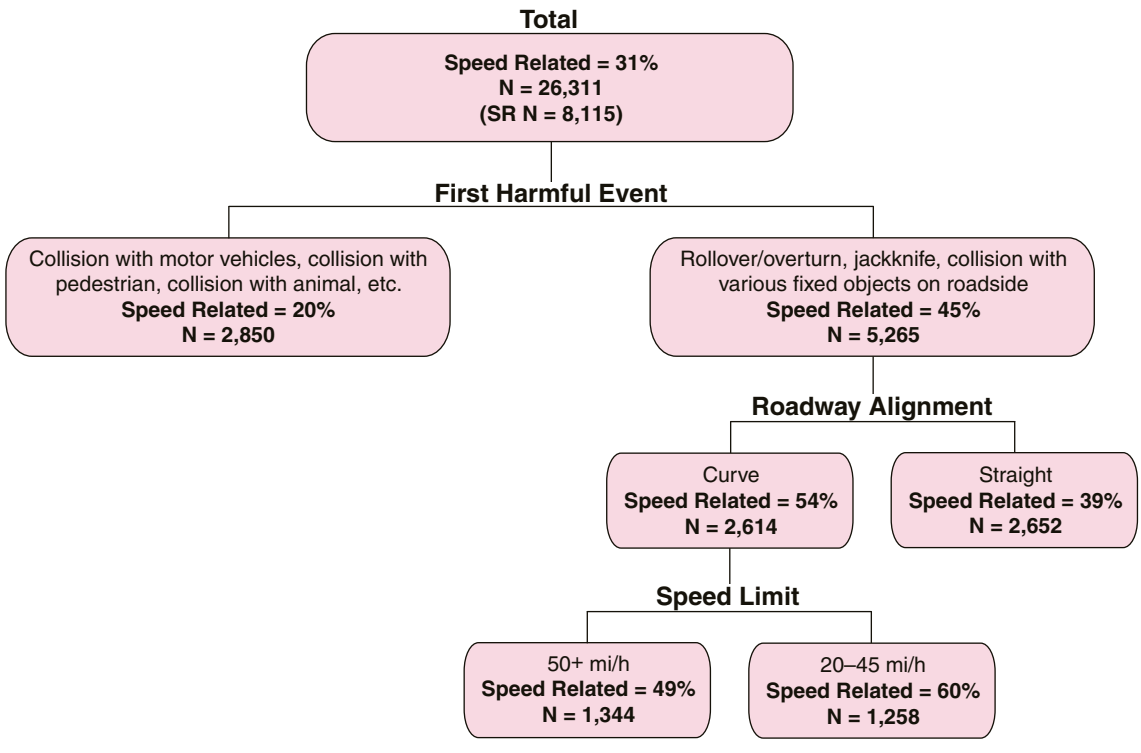
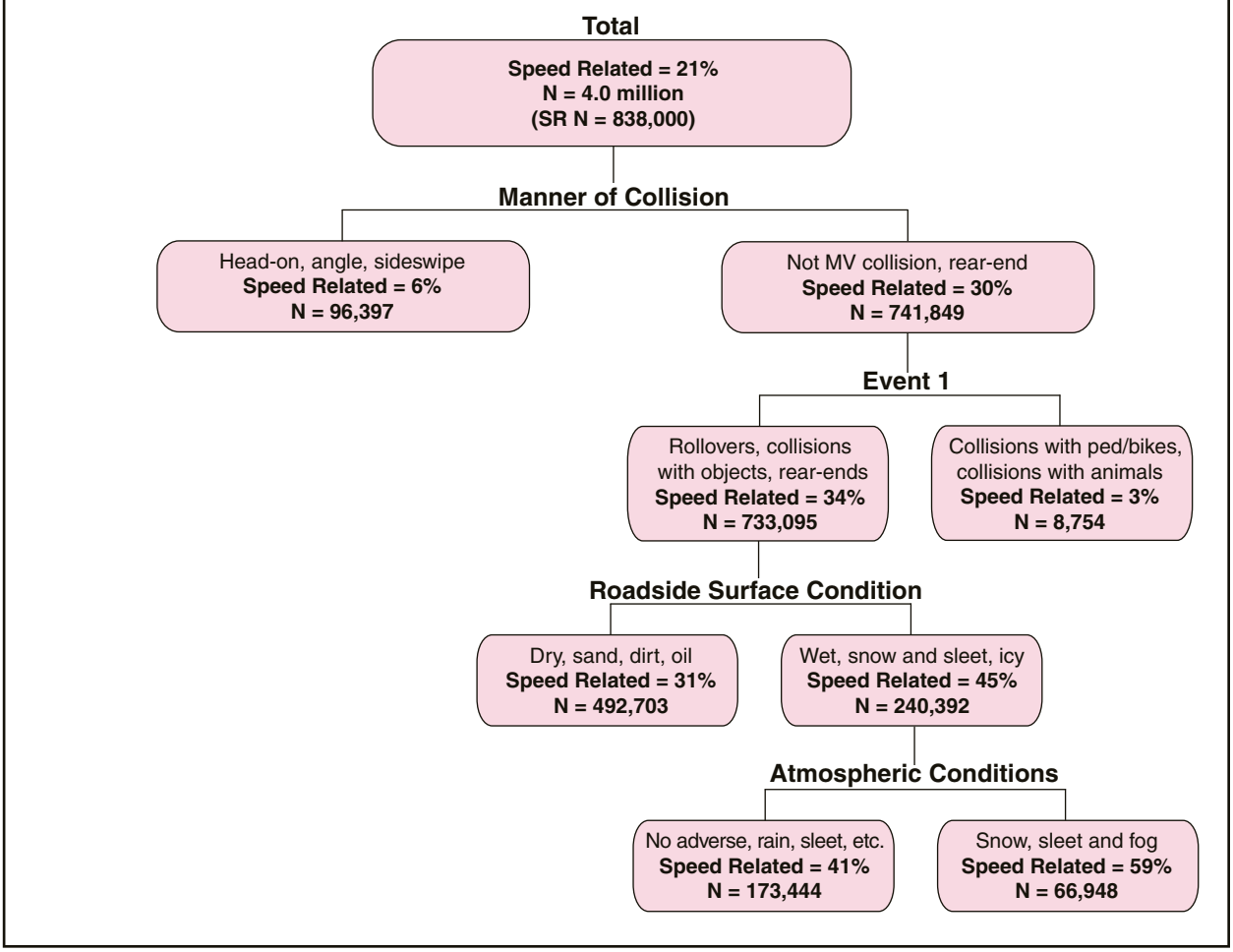


Figure 3. Chart. CART SR output for 2005 GES crashes.



users with belt users (versus unknown and no restraint available), certain driver vision obstructions including no obstruction versus others, and grouping all ages between 16 and 70 years of age in the high group. These are neither consistent with the FARS findings nor very informative due to how CART grouped the categories.

This pattern of somewhat inconsistent findings was replicated in the State-based analyses. However, in general, there was consistency with respect to the fact that the top-level predictor of SR crashes was the crash type (first harmful event) and that the categories with the highest SR percentage were, in general, single-vehicle, run-off-road crashes. Note that GES also included rear-end crashes in this set of most critical crash types. Lower-level descriptors included snowy/icy roads using the combined definition and dry roads using the over speed limit definition.

The findings from the vehicle-based analyses differed even more across databases and definitions. Using the combined 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 descriptor. The over speed limit analysis in North Carolina indicated young drivers (up to age 35) under the influence of alcohol not using restraint systems as the most important descriptor, while the Ohio data indicated young males not using restraints as the most important descriptor. A common theme in most (but not all) of the results is perhaps young male.

Discussion

The above sections described some of the specific findings from the multiple analyses conducted in this effort. The more general findings are as follows:

- For some variables, there were differences between FARS and GES findings. Thus, treatments and targets for these treatments (e.g., driver subgroups or roadway location types) may differ if one is focused on fatal crashes versus total crashes.

- There were differences between findings from the two States (and between States and FARS and GES). Thus, it would appear important for each State to conduct its own SR analysis in its Strategic Highway Safety Plan.
- There were few differences in findings between analyses conducted with the combined definition versus the over speed limit definition in the State databases. The exception was findings related to weather or road condition where the over speed limit definition showed "dry" with higher SR percentages and the combined definition showing "nondry" with higher SR percentages. The fact that there were few differences supports the use of the combined definition findings in national databases.

In summary, this study produced a large group of findings, with some findings not being consistent across all four databases and two definitions. This was not completely unexpected because 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 sometimes differ from each other and from a composite national picture. This effort did produce some consistent (and inconsistent) findings that can be used in treatment development and targeting. The findings were very consistent with those from the one similar SR study conducted in 1994, even though the data used there were from the 1980s. 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 both well justified and of critical importance in further reducing the huge cost to society resulting from vehicle crashes in the United States.

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. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984). *Classification and Regression Trees*, Chapman and Hall/CRC, New York, NY.
11. Council, F.M., et al. (2010). *Development of a Speeding-Related Crash Typology*, FHWA-HRT-10-024, Federal Highway Administration, Washington, DC.

FOR MORE INFORMATION

This research was conducted by Forrest M. Council, Raghavan Srinivasan, Scott Masten, and Daniel Carter of the University of North Carolina Highway Safety Research Center and Martine Reurings of the SWOV Institute for Road Safety Research in the Netherlands.

The full final report by the same name can be downloaded from the FHWA Safety Research and Development publications Web site related to "Speed Management" at http://www.tfhrc.gov/safety/speedm.htm.

For more information about HSIS, contact Carol Tan, HSIS program manager, HRDS, (202) 493-3315, carol.tan@dot.gov.

Visit us on the Web at www.tfhrc.gov

FHWA-HRT-10-039
HRDS-05/05–10(1500)E