U.S. Department
of Transportation
**National Highway
Traffic Safety
Administration**

NHTSA
www.nhtsa.gov

DOT HS 811 889

Report to Congress: NHTSA's NASS Data Needs

August 2011

Table of Contents

Executive Summary

The National Highway Traffic Safety Administration was directed to submit a report to the House and Senate Committees on Appropriations that "evaluates the deficiencies of the National Automotive Sampling System's Crashworthiness Data System (NASS CDS) data collection program based on current levels of case investigations and analyzes the improvements in the program that could be achieved through increased levels of case investigation and data collection." This report fulfills that direction.

For the Department of Transportation, safety is the top priority. For NHTSA, this priority means reducing the human and economic cost of motor vehicle traffic crashes and other incidents involving motor vehicles. To accomplish its mission, NHTSA's programs must rely on sound science combined with quality data. Quality data is the backbone of everything NHTSA does.

NHTSA's data systems serve as the preeminent source of traffic safety information in the Nation and provide Federal, State, local, and international entities and the greater safety community with the traffic safety information they need to make informed highway safety decisions. Recognizing the importance of data, NHTSA is pursuing data improvement initiatives through the FY 2012 Budget Request to enhance the quality of the data collected and the effectiveness of its data systems. More robust data will drive more robust traffic safety programs that will prevent or mitigate traffic violations, crashes, fatalities and injuries.

NHTSA continually evaluates its data collection systems and the information contained in these systems for relevancy and accuracy. Input for improvements has been gathered over several years from external safety organizations, such as automobile manufacturers, insurance companies, safety advocacy organizations, researchers, State partners, other Federal agencies, and the public both informally in professional meetings to formally in Congressional testimony or responses to Federal Register Notices. In one of NHTSA's most recent system reviews, interviews were conducted across all offices within NHTSA to determine how data are currently used and how those data needs were likely to change in the future. The message from the previous evaluations and the most recent one has been very clear: NHTSA must modernize and consolidate its data and data systems to allow not only NHTSA to make productive traffic safety programming decisions, but allow States and local communities to do the same.

Some of the key actions that NHTSA needs to take in regards to its data systems and information collected are:

1. Add new data elements to better develop safety countermeasures, including ones related to crash avoidance and behavioral safety issues.
2. Expand the scope of vehicle types and crash types for greater analytical attention, such as motorcycles, medium and heavy trucks, motorcoaches, bicyclists, school buses, and low-speed vehicles.
3. Enhance and improve the analysis of crash data, dissemination of data and the public's access to those data and linkage to other safety information.
4. Obtain a sufficiently large database to identify emerging crash trends and occupant injury trends, including expanding the NASS data collection sites.

Although the actions listed above and NHTSA's need for data modernization and consolidation apply to its data systems in general, this report responds to the Congressional request by providing an in-depth discussion of improvements necessary for the NASS CDS and the National Automotive Sampling System General Estimates System (NASS GES).

NASS CDS is NHTSA's cornerstone for providing detailed crash investigation data on a nationally representative sample of police-reported motor vehicle crashes and related injuries. NASS CDS has been studied by researchers around the world and utilized by NHTSA for implementing and evaluating almost every motor vehicle safety standard that has been created to reduce crash consequences. NASS GES is also a nationally representative sample of police-reported crashes, but is limited to data from police accident reports. NASS GES offers the only source for measuring national level year-to-year changes in crashes from property damage to more severe crashes. As the report explains, the underlying statistical infrastructure for both NASS CDS and NASS GES has not been refreshed for over 20 years and the number of cases and the data collected are insufficient to identify emerging trends. Expanding both NASS databases and the information contained in these databases does not come without a cost. In the agency's FY 2012 Budget Request, NHTSA has proposed a multi-year, phased-in approach to modernize NASS.

As outlined in NHTSA's FY 2012 Budget Request, NHTSA would continue planning the redesign of the NASS sample, determining additional data needed and steps to improve the data collection process. NHTSA would formally engage both internal and external data users that rely every day on NHTSA's data systems to make regulatory, program and policy decisions about traffic safety. A pilot study would be launched to test sample changes and begin phased-in implementation for collecting additional cases in additional sample sites. Parallel to these critical improvements in NASS, NHTSA will also begin to modernize and consolidate the underlying information technology infrastructure to make it more adaptable to current and future trends for collecting, storing, securing and distributing data. This modernization effort provides the necessary foundation for not only NASS, but eventually all of NHTSA's data systems.

1. Introduction

Through education, research, safety standards, and enforcement activity, the National Highway Traffic Safety Administration saves lives, prevents injuries, and reduces economic costs due to motor vehicle crashes.

NHTSA, in partnership with domestic and international safety communities, addresses this mission through primary prevention (e.g., preventing the crash from occurring), motor vehicle crashworthiness (e.g., elimination of injuries and fatalities that result from a crash), and effective post-crash response (e.g., mitigation of crash consequences through a sound system of emergency medical services).

As NHTSA's information center, the National Center for Statistics and Analysis supports the agency's mission by providing the data and the analysis to enable better understanding of the nature, causes, and injury outcomes of motor vehicle traffic crashes, as well as assisting in the identification of strategies and interventions that will reduce crashes and their consequences.

An estimated 16 million traffic crashes occur annually in the United States. Of that total, more than 5.5 million are crashes that are documented by law enforcement officers on police accident reports (PARs). These "police-reported crashes" cover a wide spectrum of crash types – from minor fender-benders to fatal crashes.

Considering that it would cost billions of dollars each year to collect and code into a uniform format the necessary data from each of these millions of police-reported crashes, NHTSA's data program involves an efficient combination of census-based and sample-based data collection. These data collection systems, which were documented in previous a Report to Congress,[1] provide nationally representative traffic crash data on a timely basis and at a small fraction of this cost. One of the systems documented in that report is the National Automotive Sampling System. NASS is composed of two separate data collection programs: the General Estimates System, which provides national estimates of characteristics of traffic crashes based on data coded from PARs; and, the Crashworthiness Data System, which involves detailed investigation of traffic crashes involving light passenger vehicles.

This report has been prepared in response to House and Senate directives. According to the House Report #111-564 of the Departments of Transportation, Housing, and Urban Development and Related Agencies Appropriations Bill of 2011:

> *the committee directs NHTSA to submit a report to the House and Senate committees on Appropriations, by not later than August1, 2011, that evaluates the deficiencies of the NASS CDS data collection program based on current levels of case investigations and analyzes the improvements in the program that could be achieved through increased levels of case investigation and data collection. The report should make recommendations regarding the type of data collection that are needed to improve NHTSA's ability to develop safety countermeasures, the level of NASS CDS case investigations that are needed to obtain a sufficiently robust database to identify emerging crash and occupant injury trends, as well as the types of crashes that should be analyzed and methods that can be used to enhance NASS CDS data collection.*

The Senate Report #111-230 of the Transportation and Housing and Urban Development, and Related Agencies Appropriations Bill of 2011 stated:

[1] Report to Congress NHTSA's Crash Data Collection Programs (DOT HS 811 337), April 2006

> *The Committee directs the FMCSA [NHTSA] to report on the NASS modernization efforts and related expenditures in the President's annual budget submission to Congress. Additionally, within 1 year of the date of enactment of this act, the NHTSA shall provide the House and Senate Committee on Appropriations a report on the results of the data element review and recommendations for revision.*

Given the relationship between NASS CDS and NASS GES, this report considers issues that are inclusive to both data systems as well as issues exclusive to one or the other, though the primary focus is on NASS CDS.

In the succeeding sections, NASS's origin and early development is discussed, followed by an overview of the need and benefits of modernizing NASS. Section 4 provides specific data elements that have been identified as enabling better safety countermeasures. Expanding the scope of vehicle, injury and new technology data collected is the focus of Section 5, while Section 6 considers steps essential to enhancing crash data analysis. Finally, Section 7 addresses what should be done to improve the NASS databases, to include sample redesign and the need for more cases.

2. Development of the National Automotive Sampling System

The design of NASS began in the mid-1970s. The primary objective of the original design was to enhance the understanding of what happens to vehicles and to the people involved in all motor vehicle traffic crashes. At the time NASS was being envisioned, NHTSA's Fatality Analysis Reporting System (FARS) was being implemented to track characteristics of crashes which resulted in the most severe outcome – fatalities – and would be based primarily on the data recorded in PARs with supplemental information from other sources. NASS was envisioned to conduct detailed investigations of all crashes to determine how and why people were injured (or killed) or why they were not. Consequently, NASS focused its data collection efforts on variables that described the vehicle damage, the scene of the crash, and injuries to anyone involved in the crash. The objective of the original NASS was twofold: first, to relate the injuries (or lack thereof) to crash-involved people to potential vehicle or scene characteristics; and, second, to track national characteristics of all traffic crashes. Data collection for NASS began in 1979.

Because it would be impossible for NHTSA to investigate all traffic crashes, NASS needed to be based on a sample of crashes. It also needed to reflect what was happening on the Nation's roads. Consequently, the selection of crashes to be investigated was based on a complex sample design. Because of the expense involved in investigating crashes, teams had to be established in scientifically (statistically) selected geographic areas (primary sampling units or PSUs) across the country. These teams would investigate statistically selected crashes occurring within the PSU. Finally, the statistical selection of crashes had to ensure that a proper distribution of all types of crashes. This initial process was called the Continuous Sampling System.

The original sample design of NASS CSS called for the system to operate 75 primary sampling units. Within those 75 PSUs, a total of 18,000 crashes would be selected and investigated each year. Implementation of this sample design began in 1977 and the first data from the system was produced in 1979. Because of the cost of implementing a system of this size, the addition of PSUs was to be phased in over several years. By 1980, 50 PSUs had been opened and were functioning. However, funding to continue the implementation of the sample design to the other 25 PSUs was never received.

By the mid-1980s, not enough cases were being selected to provide for the two original system objectives: identify injury causation and produce sound national estimates. NHTSA convened a working group to revamp NASS to better achieve agency needs within the limited resources. What resulted was a plan to split NASS CSS into two separate data collection systems. One of those systems would provide national estimates of characteristics of all police-reported motor vehicle traffic crashes. This system would be based on PARs only and would be called the General Estimates System. NASS GES would address the agency's need to track overall motor vehicle traffic safety. The other system would provide detailed information on the causes of injuries to occupants of light passenger vehicles to address NHTSA's mission to improve the crashworthiness of the vehicles in which more than 90 percent of people injured in traffic crashes are riding. This system would be called the Crashworthiness Data System and would be restricted to those crashes in which at least one passenger vehicle was towed from the scene of a crash due to disabling damage. With a more detailed set of data elements, the CDS enabled crashworthiness research. Together, NASS GES and NASS CDS were anticipated to provide NHTSA with data to improve crashworthiness of passenger vehicles as well as to provide national statistical analysis.

Data collection under this revised system began in 1988 at 36 NASS CDS PSUs and at 60 GES PSUs. It was anticipated that NASS CDS would investigate 8,000 cases and 50,000 PARs would be collected for NASS GES. However, due to funding constraints, NASS CDS was reduced to 24 PSUs and 4,500 cases

in 1991. Even with cost saving measures to streamline data collection and processing, NASS CDS struggles to reach 4,500 cases per year.

Safety researchers and policy makers, as well as automobile manufacturers and other users of NASS CDS data, have frequently expressed concern that the current NASS CDS sample size is not large enough to identify trends or to answer detailed questions on specific areas of interest in a timely manner. Three major issues surround the current system:

- The statistical variability of the NASS CDS estimates make analysis difficult.

- A second concern regarding the NASS sample is that the original NASS CDS collection sites were selected in the 1970s. The 1988 redesign of the sample partially updated those sites. But in the last 23 years, the distribution of the U.S. population has shifted. The collection sites for both NASS CDS and NASS GES should be reselected based on the latest Census.

- Additionally, beyond the need for more crashworthiness data, NHTSA has identified the need for the collection of information to address the issue of crash avoidance. Many of the new technologies being researched by the agency are crash avoidance technologies. Some of these technologies are available in a portion of the current vehicle population. Data on the effectiveness of these technologies – and on the types of crashes they prevent – are essential to the agency's decisions on which technologies should be required in all vehicles. These technologies have the potential of vastly reducing the frequency of crashes, so data tailored to enable better implementation of them would play a critical role in saving thousands of lives each year.

The agency plans to update and modernize its data systems, including NASS CDS. Recognizing the importance of data, NHTSA is pursuing data improvement initiatives through the FY 2012 Budget Request to further enhance the quality of the data collected and the effectiveness of its data systems. Included in this request is funding to examine what data need to be collected to meet the agency's and the highway safety community's data needs, funding to update and improve the sample design of NASS CDS and GES and/or its successor, and funding to improve the data processing and data output investments. More robust and accessible data will drive more robust traffic safety programs that will prevent traffic violations, crashes, fatalities and injuries.

Consequently, the requirements of the House and Senate Appropriations Committees to report on the needs of NASS CDS provides an excellent opportunity for NHTSA to address what could be done with its premier data collection system and to report on how it could be modernized.

3. Modernizing NASS

NHTSA continually assesses and reassesses its data needs. An example of this is the Data Integrated Project Team that conducted a comprehensive internal evaluation in 2004. In August 2009, another team began a review of the current uses of NCSA's databases and also conducted an assessment of how the NHTSA's data needs were likely to change in the future. Conclusions drawn from this 2009 assessment helped form the basis for the 2012 Budget Request. The team interviewed 60 NHTSA employees representing every office in the agency and ranging from staff engineers and analysts to senior leadership. Interviewees informed the team about technical challenges they would be facing in the next ten years, what information would be needed to face these challenges, and the changes they recommended in NHTSA's databases to obtain the essential data. The team supplemented the interviews with reviews of the NHTSA Vehicle Safety Rulemaking and Research Priority Plan 2009-2011 and NCSA's own plans for future statistical studies. This compilation and summary of NHTSA's perceptions of its data needs serves as an initial step for the modernization of the NASS. To further define NASS modernization and recognizing that NASS modernization is a long-term process involving multiple stakeholders, NHTSA will seek and consider the opinions of the traffic safety community to further define modernization objectives.

While in the course of the interviews, data needs were identified for other NHTSA databases such as the FARS and the SDS. However, this report will focus on the needs of the NASS, and most specifically on NASS CDS. This report is organized into the following data collection objectives in response to the Congressional directive to "make recommendations regarding the types of data collection that are needed to improve NHTSA's ability to develop safety countermeasures, the level of NASS CDS case investigations that are needed to obtain a sufficiently robust database to identify the emerging crash and occupant injury trends, as well as the types of crashes that should be analyzed and methods that can be used to enhance NASS CDS data collection":

- Adding new data elements necessary to better develop safety countermeasures. These new data elements, identified in Section 4, were subcategorized as follows:
    - Within the scope of CDS as a crashworthiness data system for cars and light truck vehicles (LTVs),
    - Part of a CDS successor system that would also address crash avoidance and behavioral-safety issues for cars and LTVs,
    - For the NASS GES or its successor system, and
    - Candidates for multiple or unspecified NCSA databases.
- Expanding the scope of data collection, to include more vehicle and crash types as well as data concerning injuries and advanced technology (Section 5). The following road users and other data collection objectives were identified as worthy of consideration for greater analytical attention:
    - Adding data on motorcycles, medium and heavy trucks, pedestrians, bicyclists, motorcoaches, school buses, and low speed vehicles,
    - More detailed injury data, and
    - More information about advanced vehicle technologies.

- Enhancing or otherwise improving crash data analysis (Section 6), the following goals and methods were identified:
    - Reduce missing data,
    - Make data more accessible,
    - Improve data linkage, data streamlining, and integration of databases, and
    - Take advantage of new data gathering technologies.
- Obtaining a sufficiently robust database to identify emerging crash and occupant injury trends (Section 7). In order to have a sufficiently robust database to achieve its mission, NHTSA will need to:
    - Update sample design based on 2010 Census data,
    - Examine the need to add more PSUs, and
    - Increase the number of cases selected.

4. Adding New Data Elements to Better Develop Safety Countermeasures

4.1 New data elements for CDS as a crashworthiness data system for cars and light trucks and vans

The following data elements mentioned in the interviews fit within the context of CDS continuing as a crashworthiness data system as it has since it was designed– at least in terms of subject matter, if not necessarily in feasibility of collection by the current data-gathering operations. Some of the elements identified in the interviews already exist in a similar or related form within the existing CDS, but may require modification to best capture future data needs.

- Advanced crash-protection system (e.g., type of advanced frontal or side air bag)
- Roof crush measurement in three dimensions
- Type of alternative power source, such as electric, hybrid, biofuel, etc., if any
- Type of seatbelt reminder system, if any
- Child seat installation errors
- Compatibility of child seat and vehicle seat
- Vehicle structure parameters (bumper height, frame rail height)
- Type of crash-notification system, if any (as well as if system triggered and data was delivered)
- Time of Emergency Medical Services (EMS) notification, EMS arrival, trauma-center type and time of arrival
- For heavy trailer struck by car or light trucks and vans (LTV): model year, style, and type of underride guard on the trailer

The above list does not include requests for more detailed injury descriptions, which will be discussed in a subsequent section, Collect More Detailed Injury Data.

Advanced crash-protection systems, alternative fuels, seatbelt reminders, and crash notification are recent technologies that could affect vehicle performance in crashes and/or occupant survival rates –CDS needs to keep abreast with new technologies and provide data to monitor their performance and possibly evaluate effectiveness. More detailed data on roof crush, child-seat performance and compatibility, and vehicle structural characteristics would aid safety research. The data elements pertaining to EMS times would be needed to evaluate the effectiveness of post-crash care as well as automatic notification systems. Without data about the heavy trailer struck by a car or LTV, it is difficult to explain why the car or LTV performed poorly in the collision or propose remedies.

Many of these data elements would be relatively straightforward to add within the existing CDS framework. EMS times could be obtained by access or linkage to a database such as the National Emergency Medical Services Information System (NEMSIS), but in some cases might be available from police accident reports. Data on heavy trailers is virtually inaccessible under current CDS: within a few hours after most crashes a trucker will resume driving, and the trailer will often be very difficult to locate.

Currently, only one State records trailer VINs on its PARs. Obtaining heavy trailer data would require a long-term effort to extend recording of trailer VINs to other States and, possibly have investigating officers photograph trailers involved in crashes.

4.2 New data elements for a CDS successor addressing crash avoidance and behavioral safety

Without a doubt, the strongest message from the interviews is that the future direction of the agency will increasingly emphasize crash-avoidance technologies, behavioral safety, and technologies that influence or compensate for human behavior – and data systems must adapt to the new emphasis.

The National Motor Vehicle Crash Causation Study (NMVCCS) of 2005-2007 documented the injuries as well as the pre-crash factors and these would need to be captured in a revised or CDS successor system. Here are the data elements pertaining to crash avoidance and human behavior that were explicitly mentioned by interviewees. Some were originally part of NMVCSS; many are not, because they address advanced technologies that were rare or nonexistent in 2005-2007 but that are now available or are expected to come on line during the next 10 years. In addition to these data elements, it may be assumed that most of the other NMVCSS data elements would carry over into the new system. For each of the advanced technologies listed below, interviewees generally wanted to know if the vehicle was equipped with the specific technology, what type of system (if there exist multiple types), whether it was in use at the time of the crash, and how the technology performed in the crash.

- Electronic stability control available as a specification
- Tire pressure monitoring system available as a specification
- Crash involving roll-away, vehicles with keyless ignition, smart pedals available as a specification
- Lane change/merge collision-avoidance system (e.g., lane keeping system or blind spot detection system) available as a specification
- Roadway-departure avoidance system (e.g., lane departure warning system) available as a specification
- Head-on collision avoidance system (e.g., forward looking radar) available as a specification
- Rear-end crash-avoidance/crash mitigation system (e.g., crash imminent braking, brake assist) available as a specification
- Smart cruise control available as a specification
- Drunk-driver sensing system
- Drowsy-driver sensing system available as a specification
- Vehicle-to-vehicle notification system available as a specification
- Night vision system available as a specification

- Advanced headlights available as a specification
- Asymmetric mirrors available as a specification
- Headlamp type and glare parameters
- Adaptive equipment: did driver have professional evaluation of equipment, prescription, and how long has driver been using such equipment?
- Detailed information on lights equipped and whether in use
- Detailed information on signals (and whether they were visible), mirrors, line of vision
- Condition of brakes (malfunction light, etc.)
- Did antilock braking system (ABS) activate?
- Tires: brand, model, size, load range, load index, speed rating, full tire identification number, max psi, actual psi, condition, failure mode
- Roadway parameters that impact rollover
- Tripping mechanism, trajectory, and driver control prior to rollover
- Driver distraction, cell-phone use, texting
- Driver condition prior to crash (e.g., asleep, fatigued)
- Driver reaction before crash (e.g., overcorrection)
- Type of alcohol use (e.g., chronic, binge)

Data elements are needed to capture the use and performance of the emerging and future technologies. After technologies have been around for some years, such as electronic stability control (ESC) or tire pressure monitoring system (TPMS), the data may eventually become sufficient for statistically evaluating crash-reducing effectiveness or at least evaluation of whether systems perform as intended in crashes. Data elements on the condition of brakes, tires, lights, or signals are useful for identifying contributing factors in crashes. Information about environmental factors contributing to rollover can help engineers design vehicles less prone to roll over. Information on behavioral factors such as distraction or alcohol, in combination with the detailed analysis of pre-crash events will help researchers better understand how behavioral factors contribute to crashes and what technologies in the vehicle might compensate for them.

In most cases it would be relatively easy to determine what advanced technologies are installed on vehicles, but more difficult to know if they were in operation at the time of the crash. Data on behavioral factors would largely derive from interviews with drivers or witnesses. It should be noted that if data elements are added for each of the technologies, the duration of each crash investigation would substantially increase.

4.3 New or revised data elements for GES or its successor

One use of GES data is the computation of national estimates. Inclusion of the following elements would enable more national estimates and coverage of more areas of public interest. The national estimates are an essential part of baseline assessment (target population) for a future regulation. GES is also widely used as a standalone file or in combination with FARS to statistically analyze the effect of safety equipment or programs. The requests about vehicle rollaways, EMS, and trailer VINs are for evaluating roll-away prevention technologies, post-crash care, and underride guards, respectively. (The list does not include requests for more detailed injury descriptions than the current KABCO; that will be discussed in the "More Detailed Injury Data" section. Also discussed later are requests for more complete reporting of data elements that are already on GES, such as VINs.)

- BAC test result for drivers who were tested (data not available on the record source)
- Age of child passenger or non-occupant in months, if child is less than 1 (or preferably 2) years old
- If female occupants or non-occupants are pregnant
- Booster seat as a separate category from child seat; differentiate rear- and forward-facing child seats
- More detail on type of heavy truck or bus
- Distinguish transit buses from motorcoaches
- Presence of highway attributes/safety countermeasures (rumble strips, median barriers, and dynamic curve warning systems, etc.)
- Crash involving roll-away
- Time of EMS notification, EMS arrival, trauma-center type and time of arrival
- VIN (or at least make, type, and model year) of heavy trailer

Data elements on GES are primarily derived from PARs. Information such as whether a female is pregnant, or the age of an infant in months, or the type of child seat, or the trailer VIN can only be added to GES if it is supplied on the PAR in the State where the crash occurred. It would be necessary to add or revise data elements in all or at least the majority of States with GES sites: a long-term effort. The BAC test result and the type of bus may already be on PARs in some form (e.g., the bus VIN) in some States and would be easier to add to GES. EMS times might be obtained by linkage to other databases if they are not available from PARs.

4.4 New data elements for multiple or unspecified NCSA databases

The following data element needs are not specific to a single database but could apply to CDS and GES as well as to other NCSA databases:

- Ethnicity.
- Were crash-involved vehicles stolen?
- Were crash-involved vehicles racing?
- Link crashes by ZIP Code to demographic data,

- Link crashes to judicial records or citation files
- Access to PARs, or their narratives and schematics

The interest in ethnicity, stolen vehicles, and drag-racing is primarily for developing national estimates based on GES, FARS, and/or SDS, but these data elements could also be added to CDS. The last three items will be discussed in subsequent sections on "Improving Linkage" and "Making Data More Accessible," but these have been included in the list because they would entail the addition of new data elements.

5. Expanding the Scope of Data Collection

5.1 Extend CDS or its successor to include other types of vehicles and road users

Another key finding from the interviews is that the agency should consider increased concern about the safety of motorcycles, medium/heavy trucks, pedestrians, and other road users. With the substantial reductions in crash fatalities involving only cars and LTVs, the proportion of fatal crashes involving a motorcycle, medium/heavy truck, or pedestrian has increased. In 2000, there were 28,013 fatalities in crashes involving exclusively cars and/or LTVs; that declined to 22,349 by 2008. In 2000, there were 13,932 fatalities in crashes involving at least one motorcycle, medium/heavy truck, pedestrian, and/or other road user; that increased to 14,912 in 2008. In other words, in 2000 only 33 percent of the fatalities occurred in crashes that involved at least one motorcycle, medium/heavy truck, pedestrian, and/or other road user; but that increased to 40 percent in 2008. New regulations or safety programs will require detailed crash data that cannot be obtained from GES, FARS, and SDS alone.

The Continuous Sampling System, the NASS predecessor of CDS, collected full case information on all of these other road users until it was discontinued in 1987. After that, the six-city Pedestrian Crash Data Study (PCDS) of 1994-1998 and the Large-Truck Crash Causation Study (LTCCS) of 2001-2003 in the 24 CDS PSUs temporarily collected detailed data on subgroups of these road users.

Existing CDS operations, in which the vehicles and crash site are usually not inspected until some days after a crash, only collect detailed data on cars and LTVs, not on the other vehicle types and road users. Another system or an expanded CDS is needed. Perhaps the NMVCCS-like system that might provide data on crash avoidance and behavioral safety could also be adapted to collect data on crashes involving motorcycles, medium/heavy trucks, pedestrians, and other road users. The design of NMVCCS successor system is beyond the scope of this report, but here are the types of information that were requested by the interviewees.

- **Motorcycles**
  - Motorcyclist conspicuity
  - Helmet type, detailed helmet information (novelty versus compliant helmet)
  - Motorcyclist injury data
  - Motorcycle ABS-equipped or not
  - Motorcycle rider training

The requests for new data correspond to the agency's threefold response to the increase in motorcyclist fatalities: behavioral safety, such as driver training; crash avoidance, such as motorcycle conspicuity and braking; and crash survivability, such as injury data and helmet information. The data are needed for problem identification, to assist in developing remedies, and eventually for evaluating remedies.

- **Medium/Heavy Trucks**
  - Restraint availability (belts/bags) and use
  - Roll stability control or ESC availability and performance

- Pre-crash braking performance
- Tripping mechanism, trajectory prior to rollover
- Trailers and their underride guards
- Characteristics of loss-of-control crashes
- VIN decoder for detailed identification of medium/heavy vehicles and their crash-avoidance technologies

The information is needed to better understand truck crashes for development of new regulations and to monitor the performance and eventually evaluate the effectiveness of technologies that are already available. Any system to collect data on long-haul trucks and their trailers will need to rely extensively on investigation at the crash scene, because the trucks do not remain in the PSU for long after the crash.

- **Pedestrians**
  - More cases
  - Update the Pedestrian Crash Data Study (PCDS) with crashes involving late-model vehicles
  - Detailed injury data (body region injured versus point of contact)
  - Shape of front end of vehicle hitting pedestrians
  - Pedestrians' disabilities

The agency is participating in developing an international regulation to make vehicles less dangerous when they hit pedestrians. The vehicles in the 1994-1998 PCDS data are no longer representative of today's fleet. NHTSA also receives inquiries about pedestrian crashes from the public and the media that current databases are not always detailed enough to address.

- **Bicyclists**
  - More cases
  - Detailed injury data (body region injured versus point of contact)
- **Motorcoaches**
  - Restraint availability (belts/bags) and use
  - Roll Stability Control or ESC availability and performance
  - Pre-crash braking performance
  - Characteristics of loss-of-control and rollover crashes
  - Distinguish transit buses from motorcoaches

  - Detailed injury data

- **School bus** injury data

- **Low speed vehicles**

A database limited to the NASS sites, even if the number of PSUs increased substantially from the current 24, would generate relatively few cases for any one of these specific types of road users. The data could be useful primarily as a small selection of cases to allow researchers and engineers to study how vehicles performed and what happened to the people. FARS and SDS would continue to be the best sources for analyzing statistical data related to these types of road users.

5.2 Collect more detailed injury data

Within NHTSA, Vehicle Safety (NVS) and Traffic Injury Control (NTI) Offices have very different injury data needs. However, both programs need more detailed data on injuries.

NVS requires detailed injury data to shed more light on injury mechanisms to develop more effective countermeasures and regulations for all varieties of crash-related injuries and fatalities. More in-depth information similar to what the NHTSA Crash Injury Research and Engineering Network (CIREN) compiles and maintains has been requested.

NTI needs injury data at the State level to support NHTSA programs. The current structure of NASS provides estimated injuries only at the national level. This does not allow NHTSA researchers and program managers to understand how variations in state or local practices affect injuries. Specific requests for detailed injury data were made for the following.

- State-level injury estimates
- Head impact protection
- Booster seats
- Knee air bags
- Automatic door locks
- Protection for 5th percentile females
- Harm caused by unbelted people to other occupants
- Injury severity
- Injured body region
- Abbreviated injury scale (AIS) severity
- More representative crash scenario for AIS 3+ injuries
- Injuries associated with luggage racks

- Fire-related injuries
- Contact source of injury
- Details on motorcycle helmet type
- Pedestrian Injuries
- Information in CDS as gathered in CIREN

NHTSA needs injury coding at a level more detailed and more consistent than the KABCO injury status scale currently used on most State PARs. Another problem identified in the databases is the under-reporting of less severe injuries.

5.3 Collect more information about advanced vehicle technologies

New vehicle technologies are entering the market at a rapid pace. In order to properly understand their effectiveness, NHTSA needs to be able to identify these new technologies in those vehicles involved in particular crashes. A list of requests about advanced vehicle features follows.

- Derived/auxiliary data elements
- Make VIN data in GES as good as FARS
- VIN-derived make/model
- VIN-derived restraint systems
- Identify trucks
- Crash avoidance technology in light vehicles
- Auxiliary lights (e.g., fog lights)
- Enhanced-viewing or warning for backing
- Brake assist, ESC, ABS
- Tire Pressure Monitoring System
- Distribution of advanced technology in vehicle fleet rear-end crash-avoidance system
- Lane change/merge collision-avoidance system
- Smart cruise control
- Drowsy-driver sensing system
- Night vision
- Motorcycle with/without advanced brakes

- Keyless ignition
- Smart pedal technology
- Bumper program

One idea strongly advocated by NHTSA researchers is to link the VIN of a particular vehicle with its manufacturers "Build Sheet." The Build Sheet is a form created for every vehicle, providing a detailed listing of all options and features installed on that vehicle at the time of manufacture. This would enable the analysis and evaluation groups to correctly classify vehicles by their unique features and to estimate their effect on the crash outcome. Although the car manufacturers maintain records of the vehicles that they produce, other organizations keep similar records. For example, the National Insurance Crime Bureau maintains information about vehicle options. Linkage with these databases would be invaluable to NHTSA's mission of saving lives.

6. Enhancing or Otherwise Improving Crash Data Analysis

6.1 Reduce missing data

The following data elements were mentioned in relation to reducing missing data in the interviews conducted by the NHTSA Data Needs Team.

- Delta-V
- Event Data Recorder downloads
- Complete blood alcohol concentration (BAC) data
- Full 17-digit VIN
- Training for motorcyclists
- Training for young drivers
- Insurance status
- State safety inspection status
- Graduated driver license status
- Better count of underride cases
- Database for tires
- Pre-crash data/first impact
- Vehicle Speed
- Improved drug test results
- Other involved vehicles

The most frequently requested data item was for more complete delta-V data from the CDS. Delta-V is the resulting instantaneous change in speed resulting from the collision between a vehicle and another object. A comprehensive understanding of the factors relating to changes in velocity stemming from vehicle crashes would help to further identify factors that would help enhance vehicle crashworthiness and to potentially reduce injuries.

Event data recorders are another rich source of information, and imaged data from these devices were requested across NHTSA. These instruments could collect a wide array of data including delta-V, enabling analysts to compute correlations between delta-V and many other potential crash-related factors.

BAC is a measurement that understandably receives considerable attention. Currently, up to 60% of crashes in FARS and GES do not have BAC values for drivers and occupants involved. Since 2001, NHTSA has estimated BAC values using the statistical technique of multiple imputation which provides robust estimates of BAC for drivers. However, this estimation is not preferred to having the actual BAC values.

VINs provide a wealth of information both directly and indirectly by linkage with other data sources. But missing VIN data from some crashes is a problem in the NASS GES, especially since these crashes are nationally-representative for all injury-only and property damage crashes.

Finally, development and widespread use of a universal PAR could not only greatly reduce missing data, but also improve the consistency and utility of the crash data. Currently, PARs vary from State to State, and in some States, from jurisdiction to jurisdiction. Lack of uniformity leads to incompatible data, making it more difficult to identify problems and develop appropriate policy.

6.2 Make data more accessible

Not all CDS data elements are available in commonly used statistical software such as SAS, but rather are available in an Oracle database not accessible to all users; making them available in SAS would greatly enhance the utility of this important database. Other data accessibility issues are:

- Restructuring database for ease of use,
- Simplifying the process of repeated inquiries,
- Improve timely processing of NHTSA data,
- Improve query system, and
- Provide SAS training.

6.3 Improve data linkage,[2] data streamlining,[3] and integration[4] of databases

The capabilities of NHTSA data systems could be greatly expanded by linking them with other data, both NHTSA data and with data systems from other government and traffic safety organizations. Some databases that should be considered for linkages with NHTSA databases are:

- Justice/Crime data (through the Bureau of Justice Statistics),
- Public Health data (through National Center for Health Statistics),
- Weather data (through the National Oceanographic and Atmospheric Administration),
- Economic data (through the Bureau of Labor Statistics),
- Injury type,
- Benefit of reduced EMS response and transport times,
- Theft program linkage with FBI, and
- Uniform Tire Quality Grading Standards (UTQGS).

[2] Linkage: forming meaningful connections between databases through data elements common to both.
[3] Streamlining: aligning the sample designs of similar databases that have identical underlying populations so that multiple levels of detail can be obtained about the same or similar crashes.
[4] Integration: transitioning data collected by other organizations into databases maintained by NHTSA; and, consolidating two or more data systems into a single data collection stream.

The issue of "data streamlining" was raised consistently by NHTSA staff. As opposed to the more straightforward approach of connecting two databases by a common data element, streamlining brings two or more data sources together to provide consistency and complementary benefits to each database from the others involved. Streamlining two or more data systems will also allow statistical inferences to be drawn from existing data systems through the combination. An effort underway between CDS and CIREN will attempt to link CIREN crashes to CDS cases and allow broader inferences to be drawn from the streamlining of these two databases.

Currently a particular crash might be recorded in the FARS, GES and CDS databases, but it is very difficult to connect this same crash across these three data systems. Without revealing private information, an indexing system could identify these crashes common to multiple databases, thus expanding our understanding of these crashes.

The CDS and GES databases have the same underlying population. It would be advantageous to align the sample designs so that all CDS cases are also GES cases, for example. This would allow for a much greater understanding of these crashes and might provide some operational efficiency.

The Crash Outcome Data Evaluation System (CODES) has access to a wealth of crash victim health outcome data. Linking CODES with GES and the SDS would provide a greater understanding of the true costs of crashes. For example, NHTSA produces "Lives Saved" estimates that show the positive effects of countermeasures like seat belts and airbags. Similar estimates could be produced to show the corresponding reduction of health care costs with a potential reduction in traffic crash injuries.

The idea of data integration or standardization allows data collection systems to share information technology infrastructure and reduce costs. However, this does require consistency between data systems. Some data elements might have the same name and different definitions; some data elements might have different names with same definition. These issues require standardization. An example of this is the three-year project – currently in the second year – to standardize the complementary FARS and GES data systems. The data from both the NEMSIS and the Strategic Highway Research Program should be considered for integration into NCSA's databases. Both data systems will provide a wealth of information that can be harnessed to address NHTSA's mission of saving lives. Some additional issues raised were:

- Latitude/Longitude on FARS/GES file,
- Linkage between NHTSA's databases and other databases,
- Linkage to impute missing values,
- More precise crash time,
- Injury type,
- Benefit of reduced EMS response and transport times,
- Theft program linkage with FBI, and
- Uniform Tire Quality Grading Standards.

6.4 Take advantage of new data gathering technologies

An EDR is a function or device installed in a motor vehicle to record technical information about the status and operation of vehicle systems for a very brief period of time (i.e., a few seconds immediately before and during a crash), primarily for the purpose of post-crash assessment of vehicle safety system

performance. Part 563 was established on August 2006. It specifies the definition and performance requirements of EDRs including the standardization of pre-crash, crash data elements, data collection and accuracy, survivability, and retrievability of the data. Currently, Part 563 requires vehicle manufacturers who are voluntarily installing EDRs to be in compliance with the regulation by September 1, 2012.

However, retrieving the data from these devices has been problematic. The data retrieval devices are very expensive, making collection difficult. Also, the data formats are not consistent. A standardized format for EDR data and the availability of EDR retrieval devices would maximize the potential of this important source of data.

7. Obtaining a Sufficiently Robust Database

CDS or GES data is used (1) for problem identification and to estimate national totals of different categories of information such as estimating the number of people injured in impacts with trees; (2) to evaluate safety equipment – e.g., by computing the ratio of Abbreviated Injury Scale (AIS) ≥ 3 head-injury rate in vehicles with head-injury protection to the rate in vehicles without such protection; and, (3) to provide a selection of cases of late-model vehicles in crashes of a certain type to allow researchers and engineers to study how the vehicles performed and what happened to the occupants.

For each purpose, more cases and more PSUs improve the quality of the product, but to differing extents. Estimation of national totals is most sensitive to the number of PSUs, because a large number of PSUs improves the national representation. Crashes vary from PSU to PSU not only because of obvious geographical differences (no snow in Florida, few trees in central Oklahoma, and no mountains in Illinois) but also because of wide variations in reporting rates. Statistical tools are available to predict how much the accuracy of an estimate improves with additional PSUs and how much with additional cases per PSU. Effectiveness estimates may be less sensitive to variation between PSUs because ratios of rates tend to cancel out some of the variations in absolute numbers or simple rates. Here, too, it is possible to analyze the relative benefits of adding cases or adding PSUs. The selection of cases appears to be the least sensitive to the number of PSUs and the most easily improved by simply adding cases, especially of late-model vehicles. But even here, there is an implicit (if not quantifiable) assumption that the batch of cases is nationally representative – and the way to achieve that is to have enough PSUs, selected by a good sample design.

Through the interviews of personnel from NHTSA various offices, the Data Needs Team identified the following examples of how more CDS cases or more GES PSUs could be used:

- Providing detailed baseline injury assessments – a larger sample would provide a more meaningful injury assessment.
- Evaluating head impact protection, booster seats, LTV compatibility improvements, knee air bags, automatic door locks, protection for 5th percentile female occupants, and harm caused by unbelted people to other occupant – uncertainty is increased by the limited number of PSUs and the sample size is too small.
- Studying vehicle performance, specifically side impact protection, roof crush, frontal crash protection, compatibility, and offset. More case data is also needed to update FMVSS Number 213.
- More late-model vehicle cases to conduct statistical effectiveness evaluations of crash-avoidance technologies.
- Developing more representative crash scenarios for AIS 3+ injuries and more accurate target population estimates for advanced technologies.

As stated earlier in this report, safety researchers and users of our data recognize the current NASS sample size is not large enough to identify trends or problems at vehicle make/model level in a timely manner. The original NASS sample design completed in 1977 called for 75 PSUs. Full funding for NASS was never realized. In the early 1980s, funding was limited to support 50 PSUs. Budget constraints and system operating costs have further reduced the number of PSUs funded to 24. In addition, the demographics of the country and data needs have changed significantly since the 1977. As a result, an

updated NASS CDS and GES sample design would be initiated prior to any expansion of the program based on data needs assessment. In FY 2012, NHTSA is requesting funding for NASS to continue redesigning the sample, data and collection process and to initiate a pilot study and phase-in implementation for collecting additional cases to facilitate vehicle safety and behavioral programs performance and decision making.

Significantly increasing the number of collection sites and potentially doubling the number of cases will greatly improve the national representativeness of our data systems and our ability to answer detailed questions on specific areas of interest in a more timely and accurate manner. National representativeness is required to verify how weather, behavior, rural/urban and other factors related to location effect crashes. The increased number of cases provides enough data to study the effects of behavior (such as driver distraction, pedal placement), vehicle-based new technology in crash avoidance (lane departure warning systems, forward looking radar, etc.), defect identification (to include air bag non-deployments and component failures), and crashworthiness (rollovers, side impacts, child safety seats, etc.) in a more timely manner. The increased sample size will enable engineers and scientists to answer detailed research questions with a single year of data instead of many years of data as currently required.

8. Conclusion

This report was prepared in accordance with Congress's directive that NHTSA evaluate the deficiencies of and identify possible improvements in the National Automotive Sampling System's Crashworthiness Data System. NHTSA continually evaluates all of its data collection systems and the information contained in these systems for relevancy and accuracy. Input for improvements has been gathered over several years from external and internal safety organizations both formally and informally. In one of NHTSA's most recent system reviews, interviews were conducted across all offices within NHTSA to determine how data are currently used and how those data needs were likely to change in the future. The message from the previous evaluations and the most recent one has been very clear: NHTSA must modernize and consolidate its data and data systems, not just the NASS CDS, to enable a multitude of data users to make informed decisions to improve traffic safety.

These key actions were identified as the steps that NHTSA needs to take:

1. Add new data elements to better develop safety countermeasures, including ones related to crash avoidance and behavioral safety issues.
2. Expand the scope of vehicle types and crash types for greater analytical attention, such as, motorcycles, medium and heavy trucks, motorcoaches, bicyclists, school buses, and low-speed vehicles.
3. Enhance and improve the analysis of crash data, dissemination of data and the public's access to those data and linkage to other safety information.
4. Obtain a sufficiently large database to identify emerging crash trends and occupant injury trends, including expanding the National Analysis Sampling System data collection sites.

Although the actions listed above and NHTSA's need for data modernization and consolidation apply to its data systems in general, this report, in response to Congress, focused on providing an in-depth discussion of improvements necessary for the National Automotive Sampling System Crashworthiness Data System and the National Automotive Sampling System General Estimates System.

The next steps to implement these key actions were outlined in NHTSA's FY 2012 Budget Request. NHTSA would continue planning the redesign of the NASS sample, further evaluate the additional data needed and assess updates to the data collection process. As part of this process, NHTSA would formally engage the Federal Motor Carrier Safety Administration, the Federal Highway Administration, the Research and Innovative Technology Administration, and external data users that rely on NHTSA's data systems to make regulatory, program and policy decisions about traffic safety. NHTSA would also initiate a pilot study to test the changes and begin phased-in implementation for collecting additional cases in additional sample sites. Parallel to these critical improvements in NASS, NHTSA will also begin an effort to modernize and consolidate the underlying information technology infrastructure to make it far more adaptable to current and future trends for collecting, storing, securing and distributing data. This modernization effort, also described in NHTSA's 2012 Budget Request, provides the necessary foundation for not only NASS, but eventually all of NHTSA's data systems.

NHTSA's data systems serve as the preeminent source of traffic safety information in the Nation. Without the aggressive pursuit of the data improvement initiatives described in this report and the FY 2012 Budget Request, decision makers at every level of government and the private sector will not be able to make sound, data-driven decisions that lead to the reduction of traffic crashes, injuries and fatalities.

DOT HS 811 889
February 2014

U.S. Department
of Transportation
**National Highway
Traffic Safety
Administration**

10146-021414-v2