PRECRASH DATA COLLECTION IN NHTSA’S CRASH DATABASES

Mark Mynatt
Greg Radja
National Highway Traffic Safety Administration
United States of America
Paper Number 13-0371

ABSTRACT

The National Highway Traffic Safety Administration (NHTSA) has been gathering precrash information in its nationally representative crash data collection programs since the early 1980’s. The various precrash elements, which describe the actions of a vehicle and driver leading up to a crash, will be a key source of data as focus on crash avoidance countermeasures and intelligent transportation systems increase in the automotive safety community. The purpose of this paper is to describe the evolution of precrash coding in NHTSA’s crash databases and briefly explain the methodology that provides the basis for these elements. Additionally, the paper will offer an overview of the precrash detail available and differing data collection techniques used in the National Automotive Sampling System (NASS) General Estimates System (NASS-GES), the NASS Crashworthiness Data System (NASS-CDS), and the National Motor Vehicle Crash Causation Survey (NMVCCS). A set of crashes common to all three nationally representative programs will be examined, comparing the precrash element coding and discussing limitations to consider when using data from each of the programs.

INTRODUCTION

NASS began in 1979 with a mission to provide nationally representative data on fatal and nonfatal motor vehicle traffic crashes for use in developing and evaluating federal motor vehicle safety standards and other safety countermeasures. In 1988 the program was re-evaluated and NASS was split into two components to focus more on crash protection performance: NASS-GES and NASS-CDS.

NASS-GES is designed to provide statistical information to monitor large scale trends on the general characteristics of the nation’s police-reported traffic crashes. NASS-GES codes roughly 50,000 police accident reports (PARs) from 60 Primary Sampling Units (PSUs) across the country each year. Approximately 120 data elements are coded, some describing the precrash phase, from the information available on the PAR.

NASS-CDS begins with selection of the PAR. Researchers then conduct detailed investigations on light motor vehicle crashes at 24 PSU’s, collecting data on vehicle damage, injury, injury mechanism, and the precrash phase of the crash. NASS-CDS averages about 4,500 cases annually with an emphasis on fatal and severe injury crashes. Trained researchers inspect the crash scene and vehicles, interview the involved drivers, and obtain occupant medical records. These follow-on investigations are typically initiated within 1-2 weeks of the crash.

The National Motor Vehicle Crash Causation Survey (NMVCCS) was a nationally representative survey of light vehicle crashes conducted by NHTSA from 2005-2007 using Emergency Medical Services (EMS) notification as the primary case initiation criteria. Trained researchers conducted on-scene investigations on nearly 7,000 crashes during the project, focusing on the precrash phase of the crash. Investigating the selected crashes on-scene, in most cases within minutes, allowed the researchers to make better assessments of the events that led up to the crash. The survey collected up to 300 data elements on the driver, vehicle, and environment.

In its early years NASS collected only limited information on the actions of the vehicle and driver leading up to the crash. From 1982-1984 any Avoidance Maneuvers taken and the Last Three Actions Prior to the Avoidance Maneuvers were coded in an initial attempt to add insight into the precrash phase [1]. A key element was added in 1985, Accident Type, which was based on the CALAX collision taxonomy developed by Kenneth Terhune [2, 3]. Accident Type was important because it offered a shorthand method for describing and communicating the essential features of the collision event, allowing analysts to group crashes for countermeasure development [4]. When NASS split into the two separate components in 1988, GES and CDS, Avoidance Maneuver and Accident Type were retained in both systems [5]. GES also included a
new variable, Vehicle Maneuver, which was a precursor to the the current Movement Prior to Critical Event element [6].

Precrash data collection in NASS-GES and CDS as we now know it began in 1992. The elements describe the precrash phase of the crash in further detail based on the concept of a critical crash envelope which was originally outlined by Kenneth Perchonok in the early 1970’s [7]. It’s important to note that when precrash was introduced in NASS, NHTSA elected not to use Perchonok and Terhune’s methodologies in their purest forms, instead implementing adaptations of their ideas updated for use in crashworthiness-based systems.

The idea of a critical crash envelope is initially confusing to many; however, once a few key concepts are understood the precrash elements are relatively straightforward. The most important determination that must be made for each vehicle in the crash is identification of the crash envelope and the critical event. The critical crash envelope begins when the driver recognizes impending danger or when the vehicle is in an imminent path of collision. The envelope ends when the vehicle has an impact or when the driver has made a successful avoidance maneuver, has full steering control, and the vehicle is tracking. There are three types of critical crash envelopes: simple single, complex single, and multiple. The cornerstone of the critical crash envelope is the critical event. The critical event is the action or event that placed the vehicle on a course such that the collision was unavoidable. In other words, the critical event makes the crash inevitable.

A precrash method protocol commonly referred to as the “but for” test, which was borrowed from the Indiana Tri-Level Study [8], is used to assist in determining the critical event. One must ask what action by this vehicle, another vehicle, person, animal, or object was critical to this driver becoming involved in the crash? For example, “but for” the other vehicle going left-of-center, this vehicle would not have been involved in this crash. Or “but for” having entered into the intersection, this vehicle would not have been involved in this crash. Through the years, a set of ten precrash general rules have also been developed to offer guidance for coding consistency [9]. Once the critical crash envelope and critical event are identified, all the remaining precrash elements are coded relative to this selected critical event. It’s essential to not consider culpability as a factor for determining precrash data. Many scenarios suggest fault, but this is considered coincidental rather than by design.

The precrash elements as a whole are designed to identify the following:

1. What was the vehicle doing just prior to the critical event?
2. What made the situation critical?
3. What was the avoidance response, if any, to this critical situation?
4. What was the movement of the vehicle just prior to impact? [10].

In an effort to collect more precrash information in its data collection systems, NHTSA elected to incorporate the seven basic precrash elements into the Fatality Analysis Reporting System (FARS) in 2010 [11]. FARS is a nationwide census providing yearly data regarding fatal injuries suffered in motor vehicle traffic crashes. Similar to NASS-GES, the primary source of information in FARS is the police report.

Figure 1 shows the chronological order of the seven core elements used to describe a single crash envelope in NASS-GES, NASS-CDS, and FARS.

---

**Figure 1**

**Chronological Order of a Single Crash Envelope**

- Driver Distracted
- Pre-Event Movement
- Critical Pre-crash Category / Event
- Attempted Avoidance Maneuver
- Pre-Impact Stability
- Pre-Impact Location
- Crash Type
Two programs focusing on crash causation, the Large Truck Crash Causation Study (LTCCS) 2001-2003 and NMVCCS 2005-2007, expanded precrash coding to include more components of Perchonok’s causal methodology such as the Critical Reason for the Critical Event and Associated Factors. The foundation of his approach is that there is no single specific cause of a given crash; rather, it views crashes as a process consisting of interrelated events and conditions [12]. This series of events leading to a crash is commonly referred to as the causal chain. Remove any one of the links in the chain, and a crash may not have occurred.

Due to the researchers’ unique on-scene perspective in LTCCS and NMVCCS, the core precrash elements, as well as other important components of the causal chain, Critical Reason, and Associated Factors could be coded with a high degree of accuracy [13]. The Critical Reason is the immediate reason for the Critical Event and describes why the Critical Event occurred [14]. Although the Critical Event and Critical Reason are principal parts of the description of the crash, it should be noted that they do not imply the cause of the crash or assignment of fault. The primary purpose of these elements is to enhance the description of events and allow analysts to better categorize similar events [15]. Additionally, the two causation programs collected a multitude of information on the Associated Factors in the crash including the Driver’s Condition, Recognition, Decision, Performance, and Emotional Factors. Factors associated with the vehicle, highway, and environment were collected as well.

Figure 2 shows the chronological order of the elements used to describe a single crash envelope in LTCCS and NMVCCS.

Here’s a simple example to demonstrate precrash coding of the PAR-based and follow-on programs:

A teenage driver was on his way home from school after finals. He was tired from studying late the night before and decided to call a friend to see how they did on the test. When he looked down momentarily to begin dialing a cell phone the vehicle drifted off the right side of the road. The driver realized he departed the road and attempted to steer left just prior to striking a tree.

<table>
<thead>
<tr>
<th>Driver Distraction</th>
<th>Distracted/Dialing cell phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Event Movement</td>
<td>Going straight</td>
</tr>
<tr>
<td>Critical Event</td>
<td>Off edge of road on right side</td>
</tr>
<tr>
<td>Avoidance Maneuver</td>
<td>Steering left</td>
</tr>
<tr>
<td>Pre-Impact Stability</td>
<td>Tracking</td>
</tr>
<tr>
<td>Pre-Impact Location</td>
<td>Departed roadway</td>
</tr>
<tr>
<td>Crash Type</td>
<td>Single driver/Right roadside departure/Drive off road</td>
</tr>
</tbody>
</table>

Using the same example, in addition to the seven basic precrash elements, the on-scene programs would include the critical reason and associated factors in the crash.

<table>
<thead>
<tr>
<th>Critical reason</th>
<th>Internal distraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Associated Factors</td>
<td>Fatigue</td>
</tr>
<tr>
<td></td>
<td>Inexperience</td>
</tr>
<tr>
<td></td>
<td>Internal distraction</td>
</tr>
</tbody>
</table>

Figure 2

Chronological Order of a Single Crash Envelope
In Causation Studies
The LTCCS and NMVCCS approaches to causation data collection and analysis were discussed in depth by Dan Blower and Kenneth Campbell in a LTCCS analysis series paper sponsored by the Federal Motor Carrier Safety Administration (FMCSA). They describe that the methodology used in these two programs is based around the view of traffic crashes as probabilistic events and how the causation information collected is best analyzed using statistical associations of the aggregate data and not clinical methods. In the statistical method, causation is not determined by researchers at the data collection stage. In fact, the causes of specific crashes are not assigned at any point. Instead, crash cause is identified in terms of changes in risk. However, Blower and Campbell go on to mention that due to the extensive amount of objective information about precrash events collected, these types of programs can support clinical methods of assessing causation [16]. One such clinical review of NMVCCS fatalities was conducted by a NHTSA multi-disciplinary team in 2011. They found that the data available in the NMVCCS cases was sufficient to determine critical and secondary factors which contributed to the crash and possible crash prevention measures [17].

As a testament to their sound foundation, the precrash elements have seen very little change since their introduction in 1992. A history of the core precrash elements from NASS-GES, NASS-CDS, and NMVCCS is included in appendix A.

METHOD

During a three year period between 2005 and 2007 there were three nationally representative NHTSA programs with precrash information available being conducted concurrently at 24 sites across the country, NASS-GES, NASS-CDS, and NMVCCS. Although the three programs have very different purposes - NASS-GES (large scale statistical trends), NASS-CDS (detailed crashworthiness), and NMVCCS (causation data) – they are similar in that each collected the same basic precrash data elements. However, due to their separate objectives, a major difference in the programs is the sources with which to code the precrash elements. NASS-GES is limited to the police crash report. NASS-CDS uses follow-on vehicle and crash scene inspections and driver interviews along with the police report. NMVCCS collected data from the scene, vehicle, and drivers on-scene, usually within minutes of the crash. This paper compares the precrash data coding in common cases from each of the three programs to examine differences that may occur due to additional sources of information.

A query of NHTSA’s internal databases was conducted to identify crashes common to the three programs. Cases were identified by matching the following:
1. Sites or Primary Sampling Units (PSUs),
2. Date and time,
3. Vehicle Year/Make/Model,
4. Vehicle identification number (VIN)

Each case was then evaluated individually to determine applicability to this review. There were unknown VINs from certain states in NASS-GES, but the cases were included in the study if all other matching criteria were met.

Five of the basic precrash elements were compared among all three programs: Distraction/Inattention, Attempted Avoidance Maneuver, Pre-Impact Stability, Pre-Impact Location, Crash Type

Two of the precrash elements - Pre-Event Movement and Critical Event - could only be compared between NASS-GES and NASS-CDS due to subtle differences in their coding convention in NMVCCS. The different coding convention was tied to timing nuances in the crash configurations. In NMVCCS the Movement Prior to the Critical Crash Envelope was typically described as two stages prior to crash occurrence [18].

RESULTS

A total of 379 crashes involving 653 vehicles were eventually determined to be present in all three programs.

The first precrash element compared was Driver Distraction. In recent years distracted driving has been one of the areas most emphasized by the Department of Transportation and NHTSA with the development of guidelines, public awareness campaigns, and increased enforcement. In this review the elements and attributes used to determine the presence of a distraction were the same ones utilized in the NHTSA Traffic Safety Facts Research Note on distracted driving published in 2010 [19]. Table 1 shows the percentage of the common vehicles with a coded Distraction in each of the programs.
Table 1
Common Vehicles with a Distraction Present
(Percentages rounded)

<table>
<thead>
<tr>
<th>Distraction</th>
<th>NASS-GES</th>
<th>NASS-CDS</th>
<th>NMVCCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>11%</td>
<td>14%</td>
<td>28%</td>
</tr>
<tr>
<td>No</td>
<td>60%</td>
<td>46%</td>
<td>48%</td>
</tr>
<tr>
<td>Unknown</td>
<td>30%</td>
<td>40%</td>
<td>24%</td>
</tr>
</tbody>
</table>

As Table 1 indicates, in these same vehicles a Distraction was coded in the on-scene program twice as often as in the follow-on program; and 2½ times more often than in the PAR-based program. The on-scene based program also had a lower percentage of Unknown Distraction coding. Figure 3 displays the comparability of Distraction coding between the programs when grouping the various attributes in each data collection program into Yes, No, or Unknown.

Table 2 presents the percentage of the vehicles common to the three programs with a coded Avoidance/Corrective Maneuver.

Table 2
Common Vehicles with an Avoidance Action Present (Percentages rounded)

<table>
<thead>
<tr>
<th>Avoidance</th>
<th>NASS-GES</th>
<th>NASS-CDS</th>
<th>NMVCCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>16%</td>
<td>35%</td>
<td>40%</td>
</tr>
<tr>
<td>No</td>
<td>17%</td>
<td>40%</td>
<td>51%</td>
</tr>
<tr>
<td>Unknown</td>
<td>67%</td>
<td>25%</td>
<td>9%</td>
</tr>
</tbody>
</table>

As Table 2 indicates, in these same vehicles an Avoidance Maneuver or Corrective Action was coded in the on-scene program 2½ times more often than in the PAR-based program, but only 5% higher than in the follow-on program. The on-scene based program had a much lower percentage of Unknown Avoidance coding than the other programs due their quick response to the scene and contact with the drivers. Figure 4 shows the percentage of cases between programs with the same Avoidance Maneuvers coded.

Figure 3
Distraction Coding Agreement
Between Programs
(Yes, No, or Unknown)

Attested Avoidance Maneuver is another key precrash element in the NHTSA data collection programs that will be widely used in the future as the focus on crash avoidance becomes more prevalent in the industry. This element provides important information towards identifying the driver’s actions leading up to a crash that are essential in development of new technologies such as auto/assisted braking and forward collision warning.

Figure 4
Avoidance Coding Agreement
Between Programs
(Specific Maneuvers/Actions)

While the correlation between the programs in the first two elements presented was generally low, some of the precrash elements do have a very high degree of consistency between programs. Figures 5 and 6 respectively show the Pre-Impact Stability and Pre-Impact Location coding of the vehicles common to all three data collection programs.
The Crash Type element will also be very important to data users as crash avoidance countermeasures and intelligent transportation system development continues. The element is particularly valuable when estimating potential safety benefits of technologies such as lane departure warning, forward collision warning, auto/assisted braking, blind spot detection, or electronic stability control. Figure 7 illustrates the Crash Type coding of the vehicles common to all three programs.
DISCUSSION

The results suggest there are some differences in the precrash coding of the crashes common to all three data programs. Further examination revealed several issues likely responsible for these discrepancies. The most significant factor is the amount of information on the precrash phase of the crash available given the data collection method. PAR-based data collection used in NASS-GES has the least amount of data available to make precrash assessments. As discussed in the introduction, some of the information necessary to determine crash envelopes and other important precrash facts can be difficult to determine, particularly with limited materials. Although many states have elected to adopt NHTSA’s Model Minimum Uniform Crash Criteria (MMUCC) for their police reports, only a handful of the MMUCC coded elements help with precrash coding [20]. Most of the precrash elements must be determined by NASS-GES coders using information from the coded boxes on a police report in conjunction with a crash narrative that may or may not have significant detail. The primary purpose of a police report is not an in-depth look at the actions of the vehicle and driver leading up to a crash.

Another difficulty faced by NASS-GES, and NASS-CDS to a lesser extent, is that several of the precrash elements depend on knowing actions taken by the driver and their intent. Not coincidently, these data elements - Distraction, Avoidance Maneuver, Critical Event and Crash Type - are the elements with the lowest amount of consistency between the programs. Driver actions and intentions are typically described on a limited basis in police reports; instead they normally require a driver interview to assess with a high confidence level. Experience has shown that the sooner a driver interview is conducted and the scene is inspected, the more reliable the information.

It’s very important to acknowledge that due to extensive quality control efforts, in-depth training, and investigator/coder experience, each of the NHTSA data collection programs discussed have shown over the years to have very high coding reliability rates. In general, the precrash data in the cases is coded correctly based on the information they have available to them.

Of the three data collection methods discussed, there is no debate precrash data from on-scene causation-based programs give the most accurate portrayal of the events and actions of the driver leading up to a crash. This study also demonstrates that investigation-based programs utilizing follow-on methods achieve good results. However, these investigation-based methods have a much higher cost per case, require elevated degrees of local law enforcement cooperation, and need highly trained personnel to collect the data. Herein lies the advantage of PAR-based programs like NASS-GES. NASS-GES is able to produce a large number of cases at a fraction of the cost of more in-depth data collection methods. In addition, due to the larger sample size, statistically analyzing the NASS-GES data more closely approximates the represented population. The drawback to PAR-based programs is that some of the precrash elements could dramatically underestimate important details necessary for countermeasure development such as driver distraction and avoidance maneuver. In the 653 vehicles examined in this study Distraction and Avoidance Maneuvers were underreported by 2½ times compared to NMVCCS. There was also a large discrepancy in Critical Event coding, the cornerstone of the precrash methodology. As one would expect, unknown values in NASS-GES were much higher as well. To reiterate, precrash data is not a primary focus of police crash reporting.

Note that this study can only compare crashes common among the three systems, which are mostly severe in nature. Conclusions only represent the most stringent qualification criteria shared in these systems (towed light vehicle, etc.). Other, less severe crashes cannot be evaluated in this study. Therefore, conclusions made in this paper may not be representative of any of the three data programs as a whole.

CONCLUSION

NHTSA’s data collection programs have been collecting information on the events and factors leading up to a crash for many years. The precrash data available in the various data systems is based on sound methodologies widely accepted in the auto safety community. In the future the precrash elements will be used even more extensively as emphasis in the industry shifts from vehicle crashworthiness toward crash avoidance and intelligent transportation systems. The data collected by NHTSA will be crucial in identifying crash scenarios a given technology could potentially prevent or mitigate, as well as in the evaluation of their effectiveness once implemented.

Over the years precrash information has been collected by NHTSA’s data programs using three different methods:
1. Gathering data from the police report (NASS-GES),
2. Follow-on vehicle and crash scene inspections and driver interviews along with the police report (NASS-CDS),
3. On-scene vehicle and scene inspections and driver interviews (NMVCCS)

Each of the collection methods has advantages and disadvantages: number of cases available, statistical strength, cost, level of detail, and how accurately the crash circumstances are portrayed.

Examining a set of cases with data collected using all three of these methods on the same crashes revealed significant differences in coding for some of the precrash elements. The disparities were largely attributed to the varying amounts of information available to the investigators and coders based on the data collection method. When analyzing NHTSA’s precrash data, caution should be exercised to use data from the program that provides the level of detail and accuracy needed to achieve the objective.
References


APPENDIX A

History of Core Precrash Elements in NASS and NMVCCS
[SAS DATASET.ELEMENT in brackets]

NASS (prior to 1988 separation of GES and CDS)
1982-1984
Third to last action prior to avoidance maneuvers
[DRIVER.PRIORREAR]
Second to last action prior to avoidance maneuvers
[DRIVER.PRIORMID]
Last action prior to avoidance maneuvers
[DRIVER.PRIORLAT]
Attempted avoidance maneuvers
[DRIVER.AVOIDMAN]

1985-1987
Attempted avoidance maneuvers
[DRIVER.AVOIDMAN]
Accident type
[DRIVER.ACCTYPE]

NASS-GES (continued)
1995-2011
Driver distracted by
[VEHICLE.DR_DSTRD]
2002-[DISTRACT.MDRDRSTRD]
Movement prior to critical event
[VEHICLE.P_CRASH1]
Critical precrash event
[VEHICLE.P_CRASH2]
Corrective action attempted
[VEHICLE.P_CRASH3]
2002-[MANEUVER.MDRMANAV]
Pre crash vehicle control
[VEHICLE.P_CRASH4]
Pre crash location
[VEHICLE.P_CRASH5]
Accident type (Crash type Beginning 2011)
[VEHICLE.ACC_TYPE]

NASS-GES (continued)
1988-1991
Attempted avoidance maneuver
[GV.MANEUVER]
Accident type
[GV.ACCTYPE]

1992-1995
Pre event movement
[GV.PRMOVE]
Critical precrash event
[GV.PREEVENT]
Attempted avoidance maneuver
[GV.MANEUVER]
Pre crash stability after avoidance maneuver
[GV.PRESTAB]
Pre crash directional consequences of avoidance maneuver
[GV.CONSEQ]
Accident type
[GV.ACCTYPE]

1990-1991
Driver distracted by
[VEHICLE.DR_DSTRD]
Vehicle maneuver
[VEHICLE.MANEUVER]
Accident type
[VEHICLE.ACC_TYPE]

1992-1995
Driver distracted by
[VEHICLE.DR_DSTRD]
Movement prior to critical event
[VEHICLE.P_CRASH1]
Critical precrash event
[VEHICLE.P_CRASH2]
Corrective action attempted
[VEHICLE.P_CRASH3]
Vehicle control after corrective action
[VEHICLE.P_CRASH4]
Vehicle path after corrective action
[VEHICLE.P_CRASH5]
Accident type
[VEHICLE.ACC_TYPE]
NASS-CDS (continued)
1995-2010
Drivers distraction/inattention to driving
   [GV.DRIVDIST]
Pre event movement
   [GV.PRMOVE]
Critical precrash event
   [GV.PREEVENT]
Attempted avoidance maneuver
   [GV.MANEUVER]
Pre impact stability
   [GV.PREISTAB]
Pre impact location
   [GV.PREILOC]
Accident type
   [GV.ACCTYPE]

NMVCCS
2005-2007
Driver distraction (multiple associated factor elements are used)
   [PCA.OTDRACT]
   [PCA.CONVERSE]
   [PCA.EXTFA]
   [PCA.INATTEN]
Movement Prior To Critical Crash Envelope
   [PCA.PREMOVE]
Critical precrash event
   [PCA.PREEVENT]
Critical reason for the critical precrash event
   [PCA.CRITREASON]
Attempted avoidance maneuver
   [PCA.AVMAN]
Pre impact stability of vehicle
   [PCA.PREISTAB]
Pre impact location on trafficway
   [PCA.PREILOC]
First harmful event crash type
   [PCA.ACCTYPE]