AN OVERVIEW OF THE 100-CAR NATURALISTIC STUDY AND FINDINGS

Vicki L. Neale
Thomas A. Dingus
Sheila G. Klauer
Jeremy Sudweeks
Virginia Tech Transportation Institute
Michael Goodman
National Highway Traffic Safety Administration
United States
Paper Number 05-0400

ABSTRACT

A key to the development of effective crash countermeasures is an understanding of pre-crash causal and contributing factors. This research effort was initiated to provide an unprecedented level of detail concerning driver performance, behavior, environment, driving context and other factors that were associated with critical incidents, near crashes and crashes for 100 drivers across a period of one year. A primary goal was to provide vital exposure and pre-crash data necessary for understanding causes of crashes, supporting the development and refinement of crash avoidance countermeasures, and estimating the potential of these countermeasures to reduce crashes and their consequences.

The 100-Car Naturalistic Driving Study database contains many extreme cases of driving behavior and performance, including severe fatigue, impairment, judgment error, risk taking, willingness to engage in secondary tasks, aggressive driving, and traffic violations. The data set includes approximately 2,000,000 vehicle miles, almost 43,000 hours of data, 241 primary and secondary drivers, 12 to 13 months of data collection for each vehicle, and data from a highly capable instrumentation system including five channels of video and vehicle kinematics. From the data, an “event” database was created, similar in classification structure to an epidemiological crash database, but with video and electronic driver and vehicle performance data. The events are crashes, near crashes and other “incidents.” Data was classified by pre-event maneuver, precipitating factor, event type, contributing factors, and the avoidance maneuver exhibited. Parameters such as vehicle speed, vehicle headway, time-to-collision, and driver reaction time are also recorded.

This paper presents the 100-Car Naturalistic Driving Study method, including instrumentation and vehicle characteristics, and a sample of study results. Presented analyses address the driver characteristics, the role of inattention and distraction in rear-end and lane change events. In addition, the methodological attributes of naturalistic data collection and the implications for a larger-scale naturalistic data collection effort are provided.

INTRODUCTION

Although the crash rate is declining, the number of driving related deaths is approximately 43,000 per year. While the development of mechanistic safety features, such as seat belts, air bags, and collapsible steering wheels, have been extremely important in lowering the vehicle-related death rate, it is plausible that the next significant decrease in roadway fatalities will require systems to assist drivers in preventing crashes. However, driver assistance systems require a more precise understanding of the driver behaviors prior to an adverse driving event to be more effective.

Data collected to study driver behavior have historically relied on epidemiological, simulator, and test track studies. While these are valuable techniques that certainly have their place in the study of driver behavior, they are not well suited to explain the combination of factors leading to an adverse driving event. For example, a police crash report form might list the cause of a rear-end collision as “following too close.” However, contributing factors might be fatigue, distraction, traffic backed up from the intersection, and/or a blind corner leading up to the same intersection. For this hypothetical case, there are both driver and infrastructure related causes of the event. Likewise, simulator and test track studies cannot mimic the combination of complex driving environments and the simultaneous array of driver behaviors that lead to many events.

As demonstrated in only a small handful of studies, naturalistic data collection fills the gap in current data collection methods. “Naturalistic” data includes data from a suite of vehicle sensors and unobtrusively placed video cameras. The drivers are given no special instructions, no experimenter is present, and the data collection instrumentation is unobtrusive. This naturalistic data collection method was applied to study fatigue and resulting driver performance in truck drivers making local/short haul deliveries [1] In this study, 42 drivers drove 4 instrumented vehicles while they made deliveries. The study resulted in approximately 1000 hours of data that included five video views and a host of vehicle sensor data.

In a long-haul truck driving study, naturalistic data was collected from 56 single and team drivers who drove one of two instrumented vehicles [2]. Data was collected to assess sleep quality, driver alertness, and driver performance on normal revenue-producing trips averaging up to eight days in length. This data collection effort resulted in 250 hours of data that was triggered based upon vehicle sensor data. The results showed that single drivers suffered the worst bouts of fatigue and had the most severe critical incidents (by about 4 to 1).

A key to the development of effective crash countermeasures is an understanding of pre-crash causal and contributing factors. This research effort was initiated to provide an unprecedented level of detail concerning driver performance, behavior, environment, driving context and other factors that were associated with critical incidents, near crashes and crashes for 100 drivers across a period of one year. A primary goal was to provide vital exposure and pre-crash data necessary for understanding causes of crashes, supporting the development and refinement of crash avoidance countermeasures, and estimating the potential of these countermeasures to reduce crashes and their consequences.

The 100-Car Naturalistic Driving Study (100-Car Study) was the first instrumented vehicle study undertaken with the primary purpose of collecting large-scale naturalistic driving data. Unique to the 100-Car Study was that the majority of the drivers drove their own vehicles (78 out of 100 vehicles). There is every indication that the drivers rapidly disregarded the presence of the instrumentation, as is indicated by the resulting database containing many extreme cases of driving behavior and performance including: severe fatigue, impairment, judgment error, risk taking, willingness to engage, aggressive driving, and traffic violations (just to name a few). These types of driving events have been heretofore greatly attenuated by other empirical techniques.

Due to the scale of the 100-Car Study and the fact that private vehicles were instrumented, new techniques had to be created and existing methods modified to make the study successful. The data collection effort resulted in the following data set contents:

- Approximately 2,000,000 vehicle miles
- Almost 43,000 hours of data
- 241 primary and secondary drivers participated
- 12 to 13 month data collection period for each vehicle
- Five channels of video and many vehicle state and kinematic variables

This paper presents a sample of the analysis results from the 100-Car Study data collected. The full study report is available through the National Highway Traffic Safety Administration [3].

METHOD

Instrumentation

The 100-Car instrumentation package was engineered by VTTI to be rugged, durable, expandable, and unobtrusive. It constituted the seventh generation of hardware and software, developed over a 15 year period that has been deployed for a variety of purposes. The system consisted of a Pentium-based computer that received and stored data from a network of sensors distributed around the vehicle. Data storage was achieved via the system's hard drive, which was large enough to store data for several weeks of driving before requiring data downloading.

Each of the sensing subsystems in the car was independent, so that any failures that occurred were constrained to a single sensor type. Sensors included a vehicle network box that interacted with the vehicle network, an accelerometer box that obtained longitudinal and lateral kinematic information, a headway detection system to provide information on leading or following vehicles, side obstacle detection to detect lateral conflicts, an incident box to allow drivers to flag incidents for the research team, a video-based lane tracking system to measure lane keeping behavior, and video to validate any sensor-based findings. The video subsystem was particularly important as it provided a continuous window into the happenings in and around the vehicle. This subsystem included five camera views monitoring the driver's face and driver side of the vehicle, the forward view, the rear view, the passenger side of the vehicle, and an over-the-shoulder view for the driver's hands and surrounding areas. An important feature of the video system is that it was digital, with software-controllable video compression capability. This allowed synchronization, simultaneous display, and efficient archiving and retrieval of 100-Car data. A frame of compressed 100-Car video data is shown in Figure 1.

The modular aspect of the data collection system allowed for integration of instrumentation that was not essential for data collection, but which provided the research team with additional and important information. These subsystems included automatic collision notification that informed the research team of the possibility of a collision; cellular communications that were used by the research team to communicate with vehicles on the road to determine system status and position; system initialization equipment that automatically controlled system status; and a GPS positioning subsystem that collected information on vehicle position. The GPS positioning subsystem and the cellular communications were often used in concert to allow for vehicle localization and tracking.

Figure 1. A compressed video image from the 100-Car data. The driver's face (upper left quadrant) is distorted to protect the driver's identity. The lower right quadrant is split with the left-side (top) and the rear (bottom) views.

The system included several major components and subsystems that were installed on each vehicle. These included the main Data Acquisition System (DAS) unit that was mounted under the package shelf for the sedans (Figure 2) and behind the rear seat in the SUVs.

Doppler radar antennas were mounted behind special plastic license plates on the front and rear of the vehicle (Figure 3). The location behind the plates allowed the vehicle instrumentation to remain inconspicuous to other drivers.

Figure 2. The main Data Acquisition System (DAS) unit mounted under the "package shelf" of the trunk.

Figure 3. Doppler radar antenna mounted on the front of a vehicle, covered by one of the plastic license plates used for this study.

The final major components in the 100-Car hardware installation were mounted above and in front of the center rear-view mirror. These components included an "incident" pushbutton box which housed a momentary pushbutton that the subject could press whenever an unusual event happened in the driving environment. Also contained in the housing was an unobtrusive miniature camera that provided the driver face view. The camera was invisible to the driver since it was mounted behind a "smoked" Plexiglas cover.

Mounted behind the center mirror were the forward-view camera and the glare sensor (Figure 4). This location was selected to be as unobtrusive as possible and did not occlude any of the driver's normal field of view.

Figure 4. The incident push button box mounted above the rearview mirror. The portion on the right contains the driver face/left vehicle side camera hidden by a smoked plexiglass cover.

Subjects

One-hundred drivers who commuted into or out of the Northern Virginia/Washington, DC metropolitan area were initially recruited as primary drivers to have their vehicles instrumented or receive a leased vehicle for this study. Drivers were recruited by placing flyers on vehicles as well as by placing newspaper announcements in the classified section. Drivers who had their private vehicles instrumented (78) received $125.00 per month and a bonus at the end of the study for completing necessary paperwork. Drivers who received a leased vehicle (22) received free use of the vehicle, including standard maintenance, and the same bonus at the end of the study for completing necessary paperwork. Drivers of leased vehicles were insured under the Commonwealth of Virginia policy.

As some drivers had to be replaced for various reasons (for example, a move from the study area or repeated crashes in leased vehicles), 109 primary drivers were included in the study. Since other family members and friends would occasionally drive the instrumented vehicles, data were collected on 132 additional drivers.

A goal of this study was to maximize the potential to record crash and near-crash events through the selection of subjects with higher than average crash- or near-crash risk exposure. Exposure was manipulated through the selection of a larger sample of drivers below the age of 25, and by the selection of a sample that drove more than the average number of miles. The age by gender distribution of the primary drivers is shown in Table 1. The distribution of miles driven by the subjects during the study appears as Table 2. As presented, the data are somewhat biased compared to the national averages in each case, based on TransStats, 2001 [4]. Nevertheless, the distribution was generally representative of national averages when viewed across the distribution of mileages within the TransStats data.

One demographic issue with the 100-Car data sample that needs to be understood is that the data were collected in only one area (i.e., Northern Virginia/Metro Washington, DC). This area represents primarily urban- and suburban driving conditions, often in moderate to heavy traffic. Thus, rural driving, as well as differing demographics within the U.S., are not well represented.

Table 1. Driver age and gender distributions.

	Gender		
Age N % of total	Female	Male	Grand Total
18-20	9 8.3%	7 6.4%	16 14.7%
21-24	11 10.1%	10 9.2%	21 19.3%
25-34	7 6.4%	12 11.0%	19 17.4%
35-44	4 3.7%	16 14.7%	20 18.3%
45-54	7 6.4%	13 11.9%	20 18.3%
55+	5 4.6%	8 7.3%	13 11.9%
Total N Total Percent	43 39.4%	66 60.6%	109 100.0%

Table 2. Actual miles driven during the study.

Actual miles driven	**Number of Drivers**	**Percent of Drivers**
0-9,000	29	26.6%
9,001-12,000	22	20.2%
12,001-15,000	26	23.9%
15,001-18,000	11	10.1%
18,001-21,000	8	7.3%
More than 21,000	13	11.9%

A goal of the recruitment process was to attempt to avoid extreme drivers in either direction (i.e., very safe or very unsafe). Self reported historical data indicate that a reasonably diverse distribution of drivers was obtained.

Vehicles

Since 100 vehicles had to be instrumented with a number of sensors and data collection hardware, and since the complexity of the hardware required a number of custom mounting brackets to be manufactured, the number of vehicle types had to be limited for this study. Six different vehicle models were selected based upon their prevalence in the Northern Virginia area. These included five sedan models (Chevrolet Malibu and Cavalier, Toyota Camry and Corolla, and Ford Taurus) and one SUV model (Ford Explorer). The model years were limited to those with common body types and accessible vehicle networks (generally 1995 to 2003). The distribution of these vehicle types was:

- Toyota Camry – 17%
- Toyota Corolla – 18%
- Chevy Cavalier – 17%
- Chevy Malibu – 21%
- Ford Taurus – 12%
- Ford Explorer – 15%

Classification of events

Table 3 provides definitions of traffic “events” that served as a basis for the classifications that follow. The distinction between *near crashes* and *incidents* was based on the subjective assessment of reviewers in concert with kinematic and proximity data associated with adjacent vehicles or objects.

RESULTS

Table 4 shows the relative frequency of crashes, near-crashes, and incidents for each conflicts type. Of the 82 crashes, 13 either occurred while the system was initializing after the vehicle ignition was started (approximately 90 seconds), or has incomplete data for other reasons (e.g., camera failure), leaving a total of 69 crashes for which data could be completely reduced. These data also included 761 near-crashes and 8,295 incidents. The first eight conflict types shown in Table 4 accounted for all of the crashes, 87 percent of the near-crashes and 93 percent of the incidents.

Table 3. Classification of Events.

Event Category	Definition
Crashes	Any contact between the subject vehicle and another vehicle, fixed object, pedestrian pedacyclist, animal
Near Crashes	Defined as a conflict situation requiring a rapid, severe evasive maneuver to avoid a crash.
Incidents	Conflict requiring an evasive maneuver, but of lesser magnitude than a near crash

It is important to note that all of the crashes, including low speed collisions that were not police reported, are shown in Table 5. A “crash” was operationally defined as “any measurable dissipation or transfer of energy due to the contact of the subject vehicle with another vehicle or object.” A benefit of the naturalistic approach is that it was possible to record all of these events; however the severity of the crashes must be delineated to better understand the data. Thus, the 69 crashes are parsed into the following four crash categories. Note that 75 percent of the single vehicle crashes were low-g force physical contact or tire strikes; in other words, most of the crashes involved very minor physical contact.

- Level I: Police-reported air bag deployment and/or injury
- Level II: Police-reported property damage only
- Level III: Non-police-reported property damage only
- Level IV: Non-police-reported low-g physical contact or tire strike (greater than 10 mph)

Since it was possible to detect all crashes regardless of severity, it is interesting to note the large number of drivers who experienced one or more collisions during the 12 to 13 month data collection period. Of all drivers, 7.5% of drivers never experienced an event of any severity. In contrast, 7.4% of the drivers experienced many incidents and 3 or 4 crashes. Thus, a handful of subjects were either very risky drivers or very safe, with the majority of drivers demonstrating a relatively normal distribution of events across the data collection period.

Table 4. Number of crashes, near-crashes, and incidents for each conflict type.

Conflict Type	Crash	Near-crash	Incident
Single vehicle	24	48	191
Lead-vehicle	15	380	5783
Following vehicle	12	70	766
Object/obstacle	9	6	394
Parked vehicle	4	5	83
Animal	2	10	56
Vehicle turning across subject vehicle path in opposite direction	2	27	79
Adjacent vehicle	1	115	342
Other	0	2	13
Oncoming traffic	0	27	184
Vehicle turning across subject vehicle path in same direction	0	3	10
Vehicle turning into subject vehicle path in same direction	0	28	90
Vehicle turning into subject vehicle path in opposite direction	0	0	1
Vehicle moving across subject vehicle path through intersection	0	27	158
Merging vehicle	0	6	18
Pedestrian	0	6	108
Pedalcyclist	0	0	16
Unknown	0	1	3

Table 5. Crash type by crash severity level.

Conflict Type	Total	Level I	Level II	Level III	Level IV
Single vehicle	24	1	0	5	18
Lead-vehicle	15	1	3	5	6
Following vehicle	12	2	2	5	3
Object/obstacle	9	0	1	3	5
Parked vehicle	4	0	0	2	2
Animal	2	0	0	0	2
Oncoming vehicle turning across subject vehicle path	2	1	1	0	0
Adjacent vehicle	1	0	0	1	0

Characterization of Driver Inattention

Historically, driver distraction has been typically discussed as a secondary task engagement. Fatigue has also been described as relating to driver inattention. In this study, it became clear that the definition of driver distraction needed to be expanded to a more encompassing 'driver inattention' construct that includes *secondary task engagement* and *fatigue* as well as two new categories, '*Driving-related inattention to the forward roadway*' and '*non-specific eye glance*'. '*Driving-related inattention to the forward roadway*' involves the driver checking rear-view mirrors or their blind spots. This new category was added after viewing multiple crashes, near-crashes, and incidents for which the driver was clearly paying attention to the driving task, but was not paying attention to the *critical aspect* of the driving task (i.e., forward roadway) at an inopportune moment involving a precipitating factor.

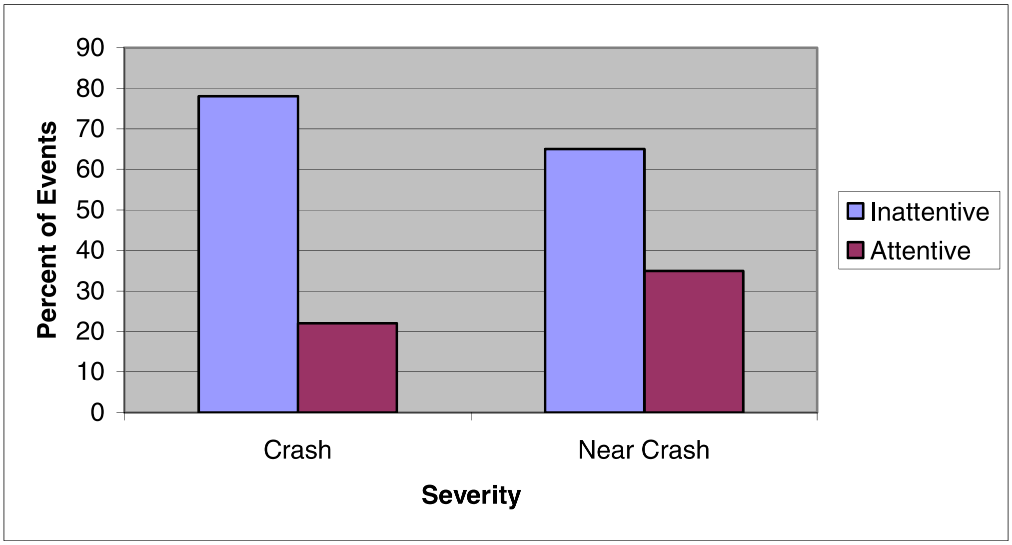

Figure 5. Percentage of events for attention by severity level.

A second analysis of the crashes and near-crashes in the 100-Car database was also conducted using the eye glance analysis performed manually by data reductionists. The '*non-specific eyeglance away from the forward roadway'* describes cases for which drivers glanced, usually momentarily, away from the roadway, but at no discernable object or person. For this project, eye glance reduction was accomplished for crash and near-crash events only, so this category can only be used for the more severe events. The four inattention categories identified above and considered together, suggested that driver's glances away from the forward roadway potentially contribute to a much greater percentage of events than has been previously thought. As shown in Figure 5, 78 percent of the crashes and 65 percent of the near crashes had one of these four inattention categories as a contributing factor.

An analysis of these types of inattention revealed that secondary task distraction was the largest of the four categories. The sources of inattention that generally contributed to the highest percentages of events (Figure 6) were wireless devices (primarily cell phones) internal distractions, and passenger-related secondary tasks (primarily conversations). It is important to note that "exposure," the frequency and duration of inattention associated with each source of inattention, is not considered in these data. Since it is exposure that determines the overall risk of a distraction source, an analysis of frequency of device use is currently being conducted for a future report that will allow calculations of event rates to determine estimates of the relative risk associated with these tasks.

Figure 7 shows a breakdown of the wireless device tasks and associated events. For these data, all of the crashes (about 8.7 percent of total study crashes) and a majority of the near crashes and incidents occurred during a cell phone conversation, although the dialing task was relatively high in term of total conflicts and was associated with the largest number of near crashes for this source of inattention. Although these data are important in that they represent the factors that contribute to events, they also highlight the need for the exposure data described above to establish the degree of risk.

Inattention for Rear End Lead-Vehicle Scenarios

Of particular interest in the analyses of rear-end conflict contributing factors was the prevalence of distraction. An important aspect in rear-end crash countermeasure development is the degree to which an un-alerted driver can be warned and make a proper response. Of course, the 100-Car data can provide great insight into the degree to which distraction is an issue in such conflicts. The important finding in this regard is that 93 percent of all lead vehicle crashes (13 out of 14) involved *inattention to the forward roadway* as a contributing factor (Figure 8). Note also that a majority (68 percent) of the near crashes have inattention identified as a contributing factor.

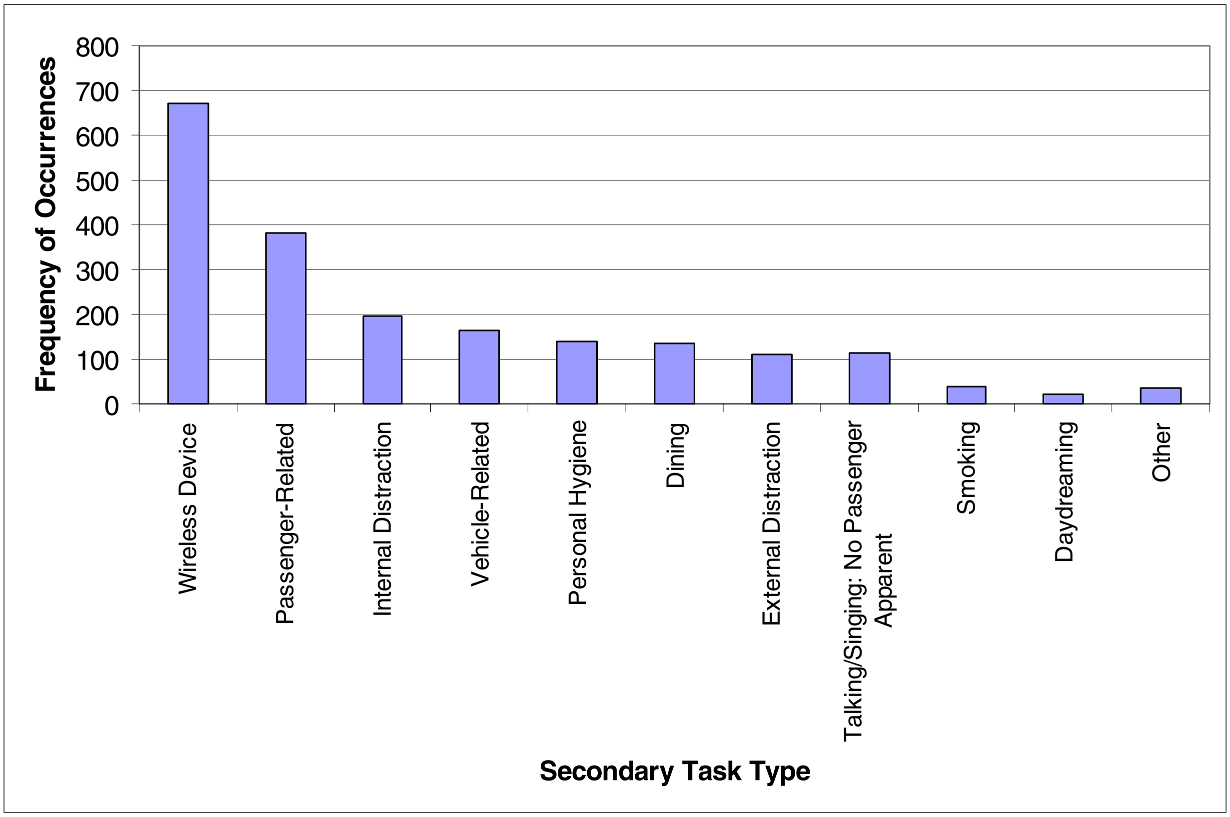

Figure 6. Frequency of occurrence of secondary tasks for crashes, near crashes and incidents.

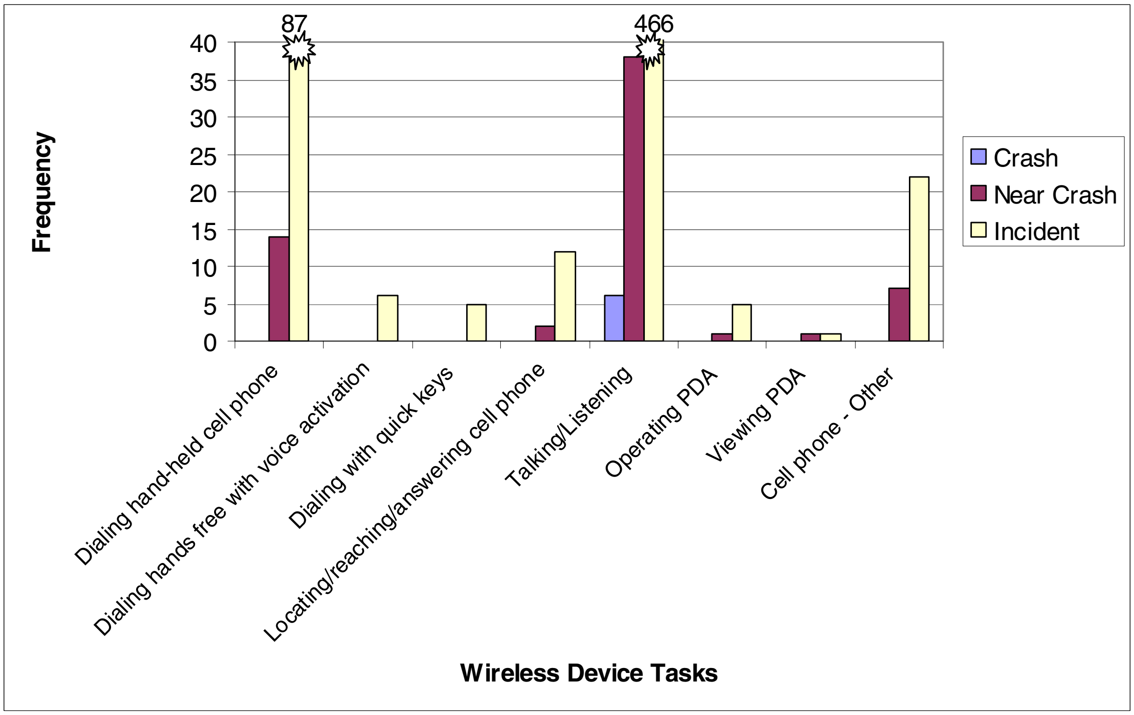

Figure 7. Frequency of occurrences in which the contributing factor was wireless device use by level of severity.

Figure 9 shows the frequency of each source of inattention for each of the secondary tasks. This allows comparison of the actual contribution of each of these sources of inattention to lead vehicle conflicts. Wireless devices (primarily cell phones, but also including PDAs) were the most frequent contributing factor for lead vehicle events, followed by passenger-related inattention. The trend was very similar for near-crashes. Interior distractions were the most frequent source of inattention for crashes.

While cell phone use contributed much more frequently to incidents and near-crashes than any other secondary task, cell phone use did not contribute to any lead vehicle conflict crashes. Nevertheless, cell phone use did contribute to other types of crashes, such as run off road, single vehicle conflict (driver ran into a barricade), and following vehicle conflict (subject vehicle was struck).

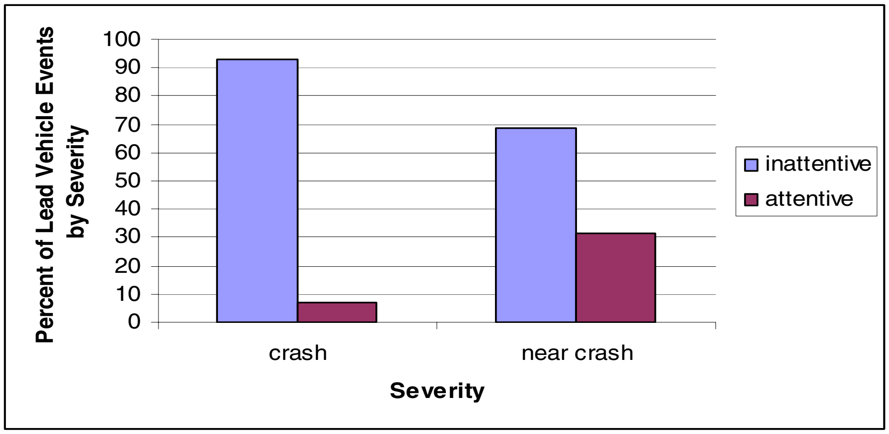

Figure 8. Percent of lead vehicle events for which inattention was listed as a contributing factor (includes the non-specific eye glance events for crashes and near crashes).

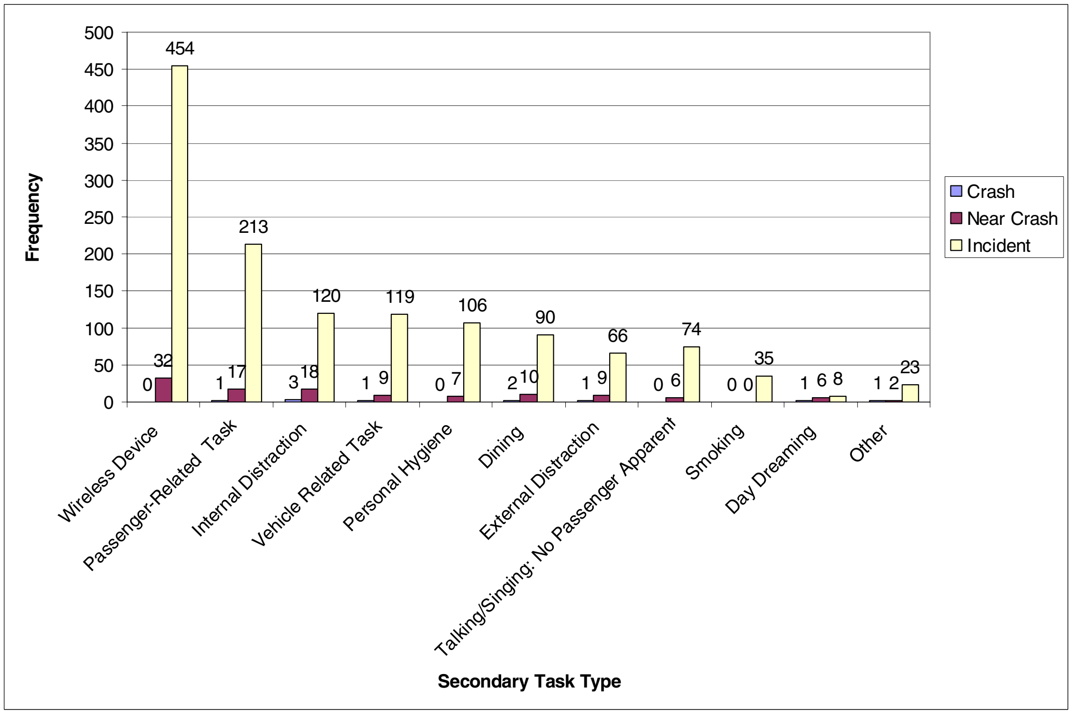

Figure 9. Total frequency of secondary task type by severity.

SUMMARY AND CONCLUSIONS

The event database that was created during the 100-Car Study can be useful for a variety of purposes; for example, evaluation of risky driving behavior and crash risk, calculation of relative risk of engaging in secondary tasks, and evaluation of driver response to lead vehicle brake lights. To facilitate this process, the initial event database will be made publicly accessible via the Internet. In addition, the initial event database can be expanded to address additional issues, since all of the video and electronic data for the entire study have been archived. The 100-Car Study contract specified ten objectives or goals that would be addressed through the initial analysis of the event database. However, as of the time of this writing, there are three additional data reduction and analysis efforts underway for the purpose of addressing another eight goals, and there is considerable interest in using the data for even more purposes. Progressing toward this potential for a multi-purpose, highly flexible and adaptable tool for driving safety may be the most important aspect of this study.

Despite the massive scope of the current effort, it was designed to serve as an exploratory study to a determine the feasibility, value, and methods for initiating a larger, more representative study. From an epidemiological viewpoint, the study was small with the presence of 15 police-reported and 82 total crashes, including minor collisions. Furthermore, drivers were represented from one area of the country (Northern Virginia/Washington, DC metro area). One purpose of a large-scale study would be to have a statistically representative sample of crashes (perhaps 2,000) and a more representative driver/environment sample.

The challenge of a large-scale study is not only the expense of such data collection but the management and analysis of such a large body of data. Nevertheless, it is believed that a large-scale database would be an enormous asset and would be used by transportation researchers for many years to gain insight and understanding into a wide array of driving behavior issues and potentially serve as a basis for decision making and program development within both the government and business sectors. This belief is based upon the robustness of the study results and the expectation that these data will continue to be analyzed and the results made available, from a variety of researchers and research organizations. Clearly, these data can provide unique insights into issues that have eluded the highway safety community for years.

REFERENCES

1. Hanowski, R. J., Wierwille, W. W., Garness, S. A. and Dingus, T. A. (2000). *Impact of local/short haul operations on driver fatigue: Final project report* (DOT-MC-00-203). Washington, DC: Federal Motor Carrier Safety Administration.

2. Dingus, T. A., Neale, V. L., Garness, S. A., Hanowski, R. J., Keisler, A. S., Lee, S. E., Perez, M. A., Robinson, G. S., Belz, S. M., Casali, J. G., Pace-Schott, E. F., Stickgold, R. A., Hobson, J. A. (2001). *Impact of sleeper berth usage on driver fatigue: Final project report* (PB2002-107930). Springfield, VA: NTIS.

3. Dingus, T. A., Klauer, S.G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., and Jermeland, J. (In press). *The 100-car naturalistic driving study, Phase II - results of the 100-car field experiment* (Contract No. DTNH22-00-C-07007). Washington, DC: National Highway Traffic Safety Administration.

4. Bureau of Transportation Statistics, TransStats 2001 (http://www.transtats.bts.gov/Databases.asp?Mode_ID=3&Mode_Desc=Highway&Subject_ID2=0)