Survey of Reported and Unreported Motor Vehicle Crashes Supporting Statement for Information Collection Request

Prepared for:

Jonathan Walker
National Highway Traffic Safety Administration

Prepared By:

Morris Davis and Company
1520 Locust Street, 3rd Floor
Philadelphia, PA 19102

Abt SRBI
8403 Colesville Road, Ste 820
Silver Spring, MD 20910

January 29, 2009

Changes By:
Jonathan Walker
National Highway Traffic Safety Administration
September 30, 2009

OMB Package NHTSA Unreported Crashes

TABLE OF CONTENTS

SUPPORTING STATEMENT

A. JUSTIFICATION

Survey of Reported and Unreported Motor Vehicle Crashes
Supporting Statement for Information Collection Request

Approval is requested to conduct a <u>National Telephone Survey of Reported and Unreported Motor Vehicle Crashes</u>.

A. Justification

The National Highway Traffic Safety Administration (NHTSA) of the U.S. Department of Transportation (DOT) has statutory authorities under Title 49, Section 105 of the U.S. Code to regulate the safety of motor vehicles and their occupants. The National Telephone Survey of Reported and Unreported Motor Vehicle Crashes will provide information that will assist NHTSA in evaluating the safety of motor vehicles and this information will ultimately assist NHTSA in the performance of its duties as a regulatory agency. NHTSA's primary mandate is to protect human life and reduce the property damage associated with automobile crashes. Each year, thousands of lives are lost, even more people are injured and billions are lost in destroyed property. Providing data that can be used to reduce the loss of life, reduce injuries, and reduce property loss is ultimate goal of all of NHTSA's work.

NHTSA wishes to conduct a nationally representative survey that will estimate the number of reported and unreported motor vehicle crashes. NHTSA also wishes to ascertain the financial costs of these reported and unreported accidents. One of the goals of this research is to improve the accuracy and timeliness of crash data and the costs associated with crashes.

In 1981, NHTSA conducted the National Accident Sampling System, the Survey of Nonreported Accidents. The 1981 study showed that many unreported crashes were minor and entailed damage only to the vehicle. In instances where the vehicle occupants were injured, these injuries were most often minor and did not require professional medical attention. These minor injuries nevertheless often entailed costs to the occupants in terms of self-care and time lost from work. Drivers may fail to report a crash for a number of reasons, for instance: lack of valid driving permits and insurance; desire to protect driving record; driving under illegal conditions, such as, intoxication or warrants for arrest or deportment; lack of awareness of the obligation to report a crash; or some other desire not to become involved with the authorities.

A.1. **Explain the circumstances that make the collection of information necessary. Identify any legal or administrative requirements that necessitate the collection. Attach a copy of the appropriate section of each statute and regulation mandating or authorizing the collection of information.**

a. Circumstances necessitating the data collection.

The last data collection on reported and unreported crashes occurred in 1981, that is, 28 years ago. During the past 28 years, the safety of cars, the safety of roads, the cost relationship of repairs and insurance, and the habits and expectations of drivers have changed dramatically. It is

obvious that the 1981 data no longer speak to current conditions or current concerns.

Furthermore, there is also the question of how well the data on reported crashes adequately represent all crashes. There is some concern that unreported crashes could be fundamentally different from reported crashes in some essential characteristic. Without accurate and up-to-date information on unreported crashes, we are unable to say the extent to which information on reported crashes adequately represents all crashes. Additionally, there is the constant concern with the crashworthiness of vehicles. Without accurate information about *all* crashes, it is impossible to say whether the standards for crashworthiness are correct and sufficient.

The National Automotive Sampling System (NASS) is an ongoing survey system that collects data from a sample of police crash reports. Two sets of data and estimates are produced from this system: Crashworthiness Data and General Estimates. The Crashworthiness data focus on passenger vehicle crashes and are primarily used to examine resultant injuries. The primary use of crash and injury data is to identify potential improvements in vehicle design. The General Estimates provides overall data on crashes and are used to determine the level, magnitude, and trend of passenger accidents. These reporting systems provide excellent data and national-level statistics, but unfortunately cover only vehicle crashes reported to the police.

There is a growing concern over the incidence of unreported crashes and the severity of injury and property damage that they entail. In fact, recent research, the 100-Car Naturalistic Driving Study (NHTSA and Virginia Tech, 2005)[1] suggests that the actual number of passenger vehicle crashes may be five times that of the number reported to the police. This is far beyond the previous estimates of approximately a one to one relationship between reported and unreported crashes. This presents a great need for updated and accurate data on unreported crashes.

b. Statute authorizing the collection of information

The National Traffic and Motor Vehicle Safety Act of 1966, Title 15 United States Code 1395, Section 106 (b), gives the Secretary authorization to conduct research, testing, development, and training as authorized to be carried out by subsections of this title. The Vehicle Safety Act was subsequently re-codified under Title 49 of the U.S. Code in Chapter 301, Motor Vehicle Safety. Section 30168 of Title 49, Chapter 301, gives the Secretary authorization to conduct research, testing, development, and training to carry out this chapter. (See Attachment A for full text).

A.2. **Indicate how, by whom, and for what purpose the information is to be used. Except for a new collection, indicate the actual use the agency has made of the information received from the current collection.**

The purpose of this survey is to provide critical information needed by NHTSA to evaluate, monitor, and address injury and property loss from passenger vehicle crashes. The data

---

[1] 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., Jermeland, J., and Knipling, R.R. The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment. NHTSA, April 2006, Report No. DOT HS 810593.

collected in the survey will be used to assist NHTSA in its ongoing responsibilities for: (a) planning program activity that addresses occupant protection issues; and (b) providing support to groups involved in improving public safety.

The survey will consist of one questionnaire, administered to all respondents. The questionnaire will be developed by the contractor based on categories agreed upon by NHTSA. The questions will represent issues raised by NHTSA program and operations staff. NHTSA staff are aware of this survey and will contribute their technical expertise to the development of the questions. The 1981 questionnaire will be reviewed and discussed prior to the development of the current questionnaire.

The survey will collect detailed information important to developing effective programs, including data addressing the following questions:

- What is the current annual level of reported crashes and how does this level compare to other reporting systems? For instance, reported to the police, to insurance companies, to both?
- What is the current annual level of unreported crashes and how does this level compare to other estimates of unreported crashes, for instance, that based on the 100-Car Naturalistic Driving Study?
- How many people were injured in reported and unreported crashes? This includes estimates of people injured per crash and annual totals.
- What proportion of reported and unreported crashes required medical attention? This includes visits to: emergency rooms, urgent care clinics, physician offices, and other medical providers.
- Are reported and unreported crashes different in some fundamental and important way? For instance, are they different in intensity, magnitude, or consequences?
- What proportion of reported and unreported crashes required hospitalization and for how long?
- How many days of work were lost due to reported and unreported crashes?
- What are the main reasons that crashes are unreported?
- What is the distribution of single-vehicle/multi-vehicle by reported and unreported crashes?
- In multi-vehicle crashes, what is the distribution of type of crash (e.g., rear end, head-on, right angle, sideswipe, backing up, etc.) for reported and unreported crashes?
- What proportion of unreported crashes occurs on public roadways, private property, or in parking lots?
- Do unreported crashes cluster among specific demographic groups (i.e., age, income, or gender)?
- What are the financial consequences of unreported crashes, both at the individual and societal levels?
- Is there a particular type of vehicle or vehicle characteristic that is involved more often in unreported crashes?

The proposed survey will provide sorely needed data on the circumstances under which unreported crashes occur; the consequences to those individuals involved and society at large;

and will suggest potential measures to remediate the occurrence and effect of unreported crashes. The results will also be disseminated to others for use in their research and program development activities. If the survey is not conducted, NHTSA program efforts would lack direction due to inadequate information upon which to base program decisions, severely limiting the agency's effectiveness in reducing fatalities and injuries.

A.3.   Describe whether, and to what extent, the collection of information involves the use of automated, electronic, mechanical, or other technological collection techniques or other forms of information technology. Also describe any consideration of using information technology to reduce burden.

The data collection will be accomplished using Computer Assisted Telephone Interviewing (CATI). CATI systems collect responses 100 percent electronically. They also perform a number of functions prone to error when done manually by interviewers, including:

- Providing correct question sequence;
- Automatically executing skip patterns based on answers to prior questions (which decreases overall interview time and consequently the burden on respondents);
- Recalling answers to prior questions and displaying the information in the text of later questions;
- Providing random rotation of specified questions or response categories (to avoid bias);
- Ensuring that questions cannot be skipped;
- Rejecting invalid responses or data entries.

The CATI system lists questions and corresponding response categories automatically on the screen, eliminating the need for interviewers to track skip patterns and flip pages. Moreover, the interviewers enter responses directly from their keyboards, and the information is automatically recorded in the computer's memory.

CATI allows the computer to perform a number of critical assurance routines that are monitored by survey supervisors, including tracking average interview length, refusal rate, and termination rate by interviewer, and performing consistency checks for inappropriate combination of answers.

A.4.   Describe efforts to identify duplication. Show specifically why any similar information already available cannot be used or modified for use for the purposes described in Item 2 above.

This data collection entails no duplication. There is no similar information currently available. In almost 30 years, there has been only one other similar data collection, and that was performed by NHTSA in 1981. Without this data collection, there will be no current data available on unreported crashes whatsoever.

Overall, the following criteria were applied to determine whether existing information may be duplicative:

- Currency of information - The data must be current in order to have utility for making sound strategic decisions concerning future programmatic activity.
- National basis - The safety efforts of NHTSA are national in scope. NHTSA therefore requires national-level data for its planning.
- Focus on NHTSA program concerns - The items within the proposed survey instruments concern issues crucial to developing appropriate strategies for improving vehicle safety.

We reiterate that there is not a current data collection that deals with both reported and unreported crashes. The underlying forces that characterize unreported crashes have changed considerably since the previous data collection almost 30 years ago. There is a tremendous need for this data collection.

A.5. If the collection of information impacts small businesses or other small entities, describe methods used to minimize burden.

There will be no impact on small businesses or other small entities. The collection of information involves randomly selected individuals in their residences, not small businesses.

A.6. Describe the consequence to Federal program or policy activities if the collection is not conducted or is conducted less frequently, as well as any technical or legal obstacles to reducing burden.

The information provided by this data collection is necessary for NHTSA to make strategic planning decisions for vehicle safety and other crash-related issues. This is particularly important with the priority assigned to safety issues by the Administration. This supports the need for NHTSA to have up-to-date data with which to help guide programmatic decisions in these critical areas.

The traffic safety environment has changed substantially in recent years. In general, motor vehicles have become safer with the adoption of newer technologies and regulations. On the other hand, traffic congestion has increased and densely populated areas have experienced increases in aggressive driving. It is important to learn how these two separate trends interact and whether they offset each other or whether one outweighs the other. Examples of increased safety include passage of seat belt laws with provisions for primary enforcement, installation of new types of air bags in motor vehicles, and introduction of LATCH technology for child car seats. Examples of increased risks include the proliferation of cell phone and their role in distracting drivers and causing accidents, the increasing disregard for traffic lights and other regulations in densely populated areas, etc. Without up-to-date information, NHTSA will not be able to detect or address shifts in driver characteristics and behaviors, which are critical to decreasing crashes.

The burden to respondents is kept to a minimum. Each respondent participates in only one survey, and each household is limited to one respondent.

A.7. **Explain any special circumstances that would cause the information collection to be conducted in a manner inconsistent with the guidelines set forth in 5 CFR 1320.6.**

No special circumstances require the collection to be conducted in a manner inconsistent with the guidelines in 5 CFR 1320.6.

A.8. **Provide a copy and identify the date and page number of publication in the Federal Register of the agency's notice, required by 5 CFR 1320.8 (d), soliciting comments on the information collection prior to submission to OMB. Summarize public comments received in response to that notice and describe actions taken by the agency in response to these comments. Describe efforts to consult with persons outside the agency to obtain their views.**

FEDERAL REGISTER NOTICE: A copy of the Federal Register Notice is provided in Attachment B. The Notice appeared in the Federal Register, Volume 73, Number 86, pages 24349-24350, Friday, May 2, 2008. The closing date for comments was June 30, 2008. No comments were received.

EXPERT CONSULTATION: Experts both within and outside NHTSA played vital roles in the design of the baseline survey instrument. Prior to any development work, offices within NHTSA's Traffic Safety Programs (TSP) section were asked to submit issues, topics, and specific questions they considered important to include on the survey.

A.9. **Explain any decision to provide any payment or gift to respondents, other than remuneration of contractors or grantees.**

No payment will be made to respondents in the main survey or in the pretest.

A.10. **Describe any assurance of confidentiality provided to respondents.**

Respondents are informed in the survey introduction that their answers will be kept private and used only for statistical purposes. Participation in the survey is voluntary. The Contractor will not include any identifying information such as names, addresses, telephone numbers, or social security numbers in the database delivered to NHTSA.

A.11. **Provide additional justification for any questions of a sensitive nature, such as sexual behavior and attitudes, religious beliefs, and other matters that are commonly considered private.**

The survey does contain questions related to matters that are considered sensitive or private. Asking someone if they have had an unreported crash could be sensitive if they believe they should have reported the crash to the police. This information is necessary however to address the core purposes of this study.

A.12. **Provide estimates of the hour burden of the collection of information on the respondents.**

Please refer to Table 1. NHTSA estimates that 10% of the general population has been involved in a motor vehicle crash in the past year. In order to achieve a sample size of 1,800 landline respondents who have been in a motor vehicle crash in the past twelve months, a total of 18,000 individuals must be contacted and screened. The 16,200 individuals who are contacted in the landline survey but have not been involved in a motor vehicle crash in the past twelve months would require an average of 3 minutes to complete the screener questionnaire for a total of 810 hours. Each respondent in the final landline survey sample of 1,800 crash victims would require an average of 15 minutes to complete the telephone interview or a total of 450 hours. Thus, the number of estimated reporting burden hours on the landline respondents would be 1,260 for the proposed landline survey (810 for the screener questionnaire, and 450 for the full survey administration).

In order to achieve a sample size of 200 cell phone respondents who have been in a motor vehicle crash in the past twelve months, a total of 2,000 individuals must be contacted and screened. The 1,800 individuals who are contacted in the cell phone survey but have not been involved in a motor vehicle crash in the past twelve months would require an average of 4 minutes[2] to complete the screener questionnaire for a total of 120 hours. Each respondent in the final cell phone survey sample of 200 crash victims would require an average of 16 minutes to complete the telephone interview or a total of 53.3 hours. Thus, the number of estimated reporting burden hours on the cell phone respondents would be 173.3 for the proposed cell phone survey (120 for the screener questionnaire, and 53.3 for the full survey administration).

In total, the number of estimated reporting burden hours on all respondents would be 1,433.3 (1,260 in the proposed landline survey and 173.3 in the proposed cell phone survey).

---

[2] The additional minute of screening time in the cell phone survey is required to ask safety questions, such as "Are you driving a motor vehicle now?"

Table 1: Hour and Cost Burden on Respondents

	Landline Survey	Landline Calls	Qx Length (Hours)	Total Hours	Total Cost	Cost per Respondent
	No Crash	16,200	0.05	810	$15,844	$0.98
	Crash	1,800	0.25	450	$8,802	$4.89
Landline Subtotals		**18,000**	**0.07**	**1260**	**$24,646**	**$1.37**
	Cell Phone Survey	**Cell Phone Calls**	**Qx Length (Hours)**	**Total Hours**	**Total Cost**	**Cost per Respondent**
	No Crash	1,800	0.067	120	$2,347	$1.30
	Crash	200	0.267	53.3	$1,043	$5.22
Cell Phone Subtotals		**2,000**	**0.087**	**173.3**	**$3,390**	**$1.70**
	Combined Surveys	**Combined Calls**	**Qx Length (Hours)**	**Total Hours**	**Total Cost**	**Cost per Respondent**
	No Crash	18,000	0.052	930	$18,191	$1.01
	Crash	2,000	0.25	503.3	$9,845	$4.92
Grand Totals		**20,000**	**0.0717**	**1433.3**	**$28,036**	**$1.40**

A.13. Provide an estimate of the total annual cost burden to respondents or record keepers resulting from the collection of information.

There are no record keeping costs to the respondents. Respondents will be contacted randomly and asked specific questions about their involvement in reported and unreported crashes during the last 12 months. All responses are provided spontaneously. Each respondent only participates once in the data collection. Thus, there is no preparation of data required or expected of respondents. Respondents do not incur capital costs, start up costs, or operation, maintenance, and purchase costs as a result of participating in the survey. As shown in Table 1, at $19.56[3] per hour, the total annual estimated cost associated with the burden hours is $19.56 x 1,433.3 hours for a total of $28,036 ($24,646 in the landline survey plus $3,390 in the cell phone survey). This amounts to $1.40 per respondent ($1.37 in the landline survey and $1.70 in the cell phone survey). Respondents would not incur any other reporting cost from the information collection.

---

[3] *From http://www.bls.gov/oes/current/oes_nat.htm#b00-0000, All occupations, Mean Hourly Wage Estimate; viewed December 1, 2008.

A.14. Provide estimates of annualized cost to the Federal government.

Total estimated cost to the government for conducting the survey is as follows:

Number of completed interviews (10 Pre-test)	2,010
Total estimated cost of conducting survey	$724,583.67
Cost per completed interview	$360.49

This estimate is based on the total cost of the awarded survey contract divided by the specified number of completed pretest/survey interviews.

A.15. Explain the reasons for any program changes or adjustments reported in Items 13 or 14 of the OMB Form 83-1.

Since this is new survey, there will be a program change of increased burden hours of 1433.3 to NHTSA's overall burden hour amount.

A.16. For collections of information whose results will be published, outline plans for tabulation, and publication.

NHTSA plans to publish results of the survey in two volumes:
• Volume I: Findings
• Volume II: Methodology Report

The Findings Report will consist of figures and tables, with limited accompanying text. The data presentations will be largely made up of percentage distributions and cross-tabulations. The data will be segmented by the following characteristics:
• Age
• Race
• Gender
• Household income
• Reported crash/Unreported crash
• Severity of injury

The final sample size within each cell will determine the granularity with which we will analyze each characteristic. Only cells that have sufficient observations to support reliable estimates will be used in the analysis and reported. No complex analytical techniques are planned at this time.

The Methodology Report will include information on the sampling frame, the survey participation rate, the weighting procedures, and copies of the questionnaires in both English and Spanish.

The planned data collection period is September/October 2009 with reports and summary sheets being published January/February 2010.

A.17. If seeking approval to not display the expiration date for OMB approval of the

information collection, explain the reasons that display would be inappropriate.

Approval is not sought to not display the expiration date.

A.18. Explain each exception to the certification statement identified in Item 19, Certification for Paperwork Reduction Act Submissions," of OMB Form 83-1.

No exceptions to the certification statement are made.

B.   Collections of Information Employing Statistical Methods

The proposed study will employ statistical methods to analyze the information collected from respondents. The following sections describe the procedures for respondent sampling and data tabulation.

B.1.   Describe the potential respondent universe and any sampling or other respondent selection method to be used.

The potential respondent universe for this survey consists of all non-institutionalized persons residing in the United States who are 16 years and older. This population age-eligible to drive motor vehicles. Reported and unreported crashes will include both drivers and passengers of these crashes.

A total of 2,000 interviews are planned for this universe of more than 200 million persons. A single questionnaire will be administered to all respondents. We will conduct 1,800 of these interviews by contacting the respondent on a landline telephone. We will conduct the remaining 200 of these interviews by contacting the respondent on a cell phone. These respondents will be screened to ensure that they live in a cell phone-only household. This will enable us to obtain a reasonably representative sample of the target population that will require only a limited degree of post-stratification weighting.

Random Digit Dialing (RDD) will be used to select and screen the landline sample. The cell phone sample will be randomly generated from 1,000 series blocks that are dedicated to providing cellular service and dialed manually. Given that an automobile crash is a relatively rare event, we expect the screening effort necessary to identify individuals who have experienced crashes in the last 12 months to be considerable.

There are no up-to-date statistics regarding the breakdown of who is involved in unreported crashes (hence the impetus for the survey). However, from Traffic Safety Facts 2006, there were 202,810,428 registered drivers and 10,584,000 police-reported crashes in that year. If we assume a 1:1 ratio of unreported crashes to reported crashes, as was found in a study by Greenblatt et al. in 1981,[4] we need to add another 10,584,000 crashes to the reported crashes. The total yearly crashes for both reported and unreported crashes therefore is 21,168,000 crashes. Dividing the total crashes by the number of registered drivers (21,168,000/202,810,428) we find that approximately 10% of the drivers experienced a crash in 2006. A study by Dingus in 2006[5] suggests unreported crashes may have increased from an equal number of crashes to reported crashes (1:1 ratio), to as much as four times the level of reported crashes (4:1 ratio). If this is correct, then our estimate of 10% of drivers having an accident in a year is a conservatively low estimate. The result is that a sample size of 20,000 drivers should be more than adequate to produce 2,000 completed interviews. Table 2 presents this summary.

---

[4] Greenblatt, J, Merrin, M.B., Morganstein, D., Schwartz, S. National Accident Sampling System Nonreported Accident Survey. NHTSA, November 1981, Report No. DOT HS-806 198.

[5] 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., Jermeland, J., and Knipling, R.R. The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment. NHTSA, April 2006, Report No. DOT HS 810593.

Table 2. Sample Size Estimate Assuming a 1:1 Ratio of Reported to Unreported Crashes			
Category	**Population**	**Percentage**	**Sample**
Drivers	202,810,428	100%	20,000
Reported Crash in past year	10,584,000	5%	1,000
Unreported Crash in past year	10,584,000	5%	1,000
Total Crashes in past year	21,168,000	10%	2,000

We intend to determine estimates of underreporting by comparing the number of reported crashes to unreported crashes. This ratio then will be compared to the ratios found in the Greenblatt (1:1) and Dingus (4:1) studies.

In Table 3 we show the Census Bureau Expected Population and Sample Distribution by age for the 2,000 respondent sample. Although we expect our final sample to include a higher proportion of younger drivers relative to the US population, we are unable to make an educated guess about the age proportions of the final sample.

TABLE 3
EXPECTED POPULATION AND SAMPLE DISTRIBUTION BY AGE
BASED ON JUNE 1, 2008 CENSUS BUREAU ESTIMATES

Age Category	**Population**		**Sample**
	Target Population (in thousands)	Percentage of Target Population	Population-Based Distribution of Sample
Total (16+)	**233,627**		**2,000**
16-24	37,476	16.04%	321
25-34	39,960	17.10%	342
35-44	41,735	17.86%	357
45-64	77,397	33.13%	663
65+	37,060	15.86%	317
U.S. Bureau of the Census, Population Estimates, Age Category Estimates, 6/01/2008			

The response rate target for this survey will be 50% based on the subject matter, interview length, callback procedures and government sponsorship. In order to maximize response rates the strategy will focus on more limited sample release to reduce cases in callback status. The methods proposed to maximize response rates on this survey are described in Section B.3.

Cell phones are a growing problem for telephone surveys. The coverage error that can result from excluding cell phones from the sample will introduce bias into the estimates. In 2008

more than 1 out of 6 American households (17.5%) were cell phone only, did not have a landline, and could not be reached via traditional RDD survey methods. An additional 13.3% of American households were cell phone mostly meaning the residents received all or almost all of their calls on their cell phone, despite the fact that there is a landline in the household (Blumberg, 2008[6]). Many studies have started to include cell phones in their sample in order to account for the cell only households and cell mostly households. However, the expense to include cell phone only/mostly households in the sample is as much as 2.5 times greater per case, than those of traditional landline households (Keeter, 2007[7]).

The bias that results from not including cell phone only households in RDD studies is related to the differences between the households with landlines and those that are cell phone only. Cell phone only households are much more likely to be younger, renters, minority, and lower income.

Another concern for the current data collection is the social desirability effect. Simply stated, the social desirability effect warns that when asking about circumstances that could potentially be incriminating or unethical in nature the respondent may not report past behavior truthfully in order to conform to social expectations. This could be an issue for those who were in a crash and did not report it to the police even though they are supposed to do so. We do expect respondents to underreport such behavior – such as leaving the scene of an accident without notifying the police. However, benchmark data exist which will help us determine the extent to which underreporting has taken place. The initial Unreported Crashes Survey conducted by NHTSA (Greenblatt, 1981), gives us an indication that people are willing to report such behavior, although this tells us nothing regarding the rate of underreporting. More importantly, NHTSA conducted an in-vehicle study (Dingus, 2006) which provides recent, and accurate, estimates of unreported crashes. The estimates from the survey will be compared to both studies in order to generate a data-based estimate of the social desirability effect.

B.2. Describe the procedures for the collection of information.

The most important elements of the study design of the Survey of Reported and Unreported Motor Vehicle Crashes are:

- Survey population is defined as total non-institutionalized population, age 16 and older, of the United States residing in households having telephones.

- Survey will be conducted by telephone, using computer-assisted telephone interviewing.

- There will be one questionnaire delivered to every respondent. This questionnaire will take approximately 15 minutes to complete.

---

[6] Blumberg SJ, Luke JV. Wireless substi-tution: Early release of estimates from the National Health Interview Survey, January-June 2008. National Center for Health Statistics. Available from: **http://www.cdc.gov/nchs/nhis.htm**. December 17, 2008.

[7] Keeter, S. (2006, May 15). *The cell phone challenge to survey research*. Retrieved from http://people-press.org/reports/display.php3?ReportID=276

- One eligible adult (16 years or older) will be selected in each sampled household, using the "most recent/next birthday" for systematic selection within household.

- A total sample size of 2,000 persons ages 16 and older will be interviewed (1,800 on a landline telephone and 200 on a cell phone) using the questionnaire. The questionnaire focuses on whether or not the respondent was ever involved in an unreported crash, and the details of the incident.

- The survey will include a Spanish language version of the questionnaires used by bilingual interviewers to minimize language barriers.

- Professional interviewers who are experienced in conducting interviews will carry out the interviews, using computer assisted telephone interviewing.

- The completed data sets will be weighted to correct for disproportionate sampling and non-response bias.

B.2.a. Sampling Frame

The purpose of this study is to generate accurate population estimates of the non-institutionalized population age 16 and older currently living in the United States and the District of Columbia. Traditionally this would require the use of a sample of randomly generated landline telephone numbers (RDD). However, increasing numbers of households have been dropping their landline in favor of exclusively using their cell phone. The result is the lowering of covering in the RDD sample. Because of this, we will now incorporate randomly generated cell phone numbers in the study.

An efficient method of constructing a national area probability landline survey involves multi-stage geographic sampling. The sample is first stratified by geographic region, with sample allocation proportionate to population distribution. The second stage involves random selection of residential telephone exchanges within the geographic strata. The third stage involves random digit dialing (RDD) selection of a telephone number within selected exchanges for household contact. The fourth stage requires the random selection of one age-eligible respondent within each household as the designated respondent.

Cellular samples are generated randomly within 1,000 series blocks that are dedicated to providing cellular service (these blocks are excluded from the RDD sample frames). Unlike landline RDD sample, cellular RDD sample is generated with the absence of working bank information. This is because there is no white page listings of cell phones as there is for listed landlines. Given this fact and the extreme mobility of cellular telephone numbers, it is necessary to limit cellular sample generation to area code and state geographies.

The specific steps that will be used to construct the sample are outlined in detail below.

B.2.b. National Population Sample Design

National samples of assigned telephone banks will be randomly selected from the active landline and cell phone telephone exchanges within the United States. Phone numbers in the United States are 10 digits long - where the first three numbers are the area code, the next three are the exchange, and the last four numbers are the number within the exchange. The first two digits of the four-digit number define a cluster, with each cluster containing 100 numbers or a "100-bank" (the last two digits). Selection will be made from those 100-banks containing residential listings for the landline survey and from the 1,000 series blocks for the cell phone survey.

In the second sampling stage, a two-digit number is randomly generated for each 100-bank with one or more listed residential number (landline survey) or 1,000 series block (cell phone survey) selected in the first stage. This second stage sampling technique is known as random digit dialing (RDD). Every telephone number within the 100-block has an equal probability of selection, regardless of whether it is listed or unlisted. The use of RDD sampling eliminates the otherwise serious problem of unlisted telephone numbers. The use of cell phone sampling eliminates the undercoverage in the landline frame due to the cell phone only households.

Landline RDD sample is typically put through a purging process to eliminate non-working and business numbers. These purging processes typically utilize an automated dialer. Cell phone samples will not be purged of non-working or business numbers. There are two reasons for this: there is no database of business cell phone numbers for purging and it is illegal to use an automated dialer to call cell phone numbers. The cell phone sample will be manually dialed during the fielding period of the study due to this law.

B.2.c. Selection of Respondent Within Households (Landline Survey)

The sample construction described above yields a population-based, random digit dialing sample of telephone numbers. The systematic dialing of those numbers to obtain a residential contact should yield a random sample of telephone households. Next, a random selection procedure will be used to select one designated respondent for each household sampled. The "most recent/next birthday method" will be used for within household selection among multiple eligible respondents. Salmon and Nichols (1983[8]) proposed the birthday selection method as a less obtrusive method of selection than the traditional grid selections of Kish, et al. In theory, birthday selection methods represent true random selection (Lavrakas, 1987[9]). Empirical studies indicate that the birthday method produces shorter interviews with higher response rates than grid selection (Tarnai, Rosa and Scott, 1987[10]).

---

[8] Salmon, C. and Nichols, J. *The Next-Birthday Method of Respondent Selection*. Public Opinion Quarterly, 1983, Vol. 47, pp. 270-276.

[9] Lavrakas, P. *Telephone Survey Methods: Sampling, Selection and Supervision*. Beverly Hills: Sage Publications, 1987.

[10] Tarnai, J., Rosa, E. and Scott, L. *An Empirical Comparison of the Kish and the Most Recent Birthday Method for Selecting a Random Household Respondent in Telephone Surveys*. Presented at the Annual Meeting of the American Association for Public Opinion Research. Hershey, PA, 1987.

Upon contacting the household, interviewers will briefly state the purpose of their call (including noting the respondents will remain anonymous), and then request to speak to the person in the household within the eligible age range who has had the most recent birthday, or will have the next birthday. The CATI system will randomly select whether the interviewer asks for the most recent or next birthday. If the person who answered the phone is the selected respondent, then the interviewer will proceed with the interview. If the selected respondent is someone else who then comes to the phone, then the interviewer will again introduce the survey and proceed with the interview. If the selected respondent is not available, then the interviewer will arrange a callback.

B.2.d. Precision of Sample Estimates

The objective of the sampling procedures described above is to produce a random sample of the target population. A random sample shares the same properties and characteristics of the total population from which it is drawn, subject to a certain level of sampling error. This means that with a properly drawn sample one can make statements about the properties and characteristics of the total population within certain specified limits of certainty and sampling variability.

The 95% confidence interval for a sample estimate of a population proportion, using simple random sampling without replacement, is calculated by the following formulas:

$$x = (1.96) SE$$

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

Where:
  x = half-width of the confidence interval
  SE = the standard error of the sample estimate for a population proportion
  p = the proportion of the sample displaying a certain characteristic or attribute
  n = the size of the sample

Using this formula, the maximum expected sampling error at the 95% confidence level (i.e., in 95 out of 100 repeated samples) for a total sample of 2,000 is $\pm$ 2.2 percentage points. It should be noted that the maximum sampling error is based upon the conservative estimate that p = 0.5.

The total sample size for the survey is large enough to permit robust estimates for sub-samples of particular interest to this study, such as reported vs unreported crashes, injury severity, age, gender and income. Subsamples will consist of at least 384 respondents to allow for a $\pm$ 5 sampling error. Respondents will be grouped into categories based on these characteristics and cell size. The analysis will ensure that cell size is large enough for these subsamples to produce reliable estimates.

B.2.e Data Collection

Data collection for the National Telephone Survey of Reported and Unreported Motor Vehicle Crashes will be conducted from the telephone call centers of organizations experienced in conducting large-scale national surveys. The contractor will administer the survey using computer-assisted telephone interviewing (CATI), and will have sufficient numbers of CATI stations to conduct large numbers of interviews in short periods of time.

All interviewers on the project will have been previously trained in effective interviewing techniques as a condition of their employment. The contractor will develop an interview manual for this survey, and conduct a training session specific to this study with all interviewers prior to their conducting any interviews.

Interviewing will be conducted on a schedule designed to facilitate successful contact with targeted households (concentrating on weekends and weekday evenings). Interviewers will make up to fifteen call attempts to ring no answer numbers before the number is classified as a permanent no answer. These call attempts shall be made at different times, on different days over a number of weeks, according to a standard call attempt strategy. However, the fifteen-call attempt protocol shall only apply to telephone numbers where nobody picks up the telephone. If someone picks up the phone but terminates the contact before in-house selection of a subject can be made, then the Contractor shall apply an alternative protocol of five additional contact attempts designed not to repeatedly call and annoy the prospective household participant.

The Contractor will implement a plan for gaining participation from persons in households where telephone contact is made with an answering machine. The Contractor will leave a message on the machine encouraging survey participation and providing information that the household member can use to verify the legitimacy of the survey and contact the Contractor. The Contractor will set up a toll free number that the prospective survey participant can call. In addition, NHTSA shall place a statement on its Web site that the prospective survey participant can access by computer to verify the legitimacy of the survey. The Contractor may also note these sources of verification to persons directly contacted on the phone if that would be deemed helpful in getting their participation.

For landline respondents when the household is reached, the interviewer will use a systematic procedure to randomly select one respondent from the household. If the respondent is reached but an interview at that time is inconvenient or inappropriate, the interviewer will set up an appointment with the respondent. If contact is made with the household, but not the designated respondent, the interviewer will probe for an appropriate callback time to set up an appointment. Once a household is reached, a minimum of five additional contact attempts will be made to identify and interview the designated respondents. A total of 15 contact attempts represent a tradeoff between the agency goals of achieving a high response rate and minimizing public complaints about harassment. If contact is made with the eligible respondent, but the respondent refuses to participate, then the interviewer will record information for use in refusal conversion to be conducted at a later time (see Section B.3.).

For cell phone respondents the interviewer will follow a special protocol. This protocol includes:

- Hand dialing the number. Due to most state regulations, auto-dialers are not permitted for the conduct of the cell phone sample.
- Using a cell phone screener which asks the respondent if they are in a safe place to conduct the interview (i.e. not driving while holding the phone, out to dinner, or socializing in bar, etc).
- Offering to schedule a time to callback later, if they are not in a safe place to conduct the interview.
- Offering a toll-free number which the respondent can call to complete the interview.
- Asking for a landline number they can be reached at, if they would prefer not to use their cell phone to speak with our interviewers. If the respondent does use their cell phone to complete the screener questions, we will pay the respondent a five dollar incentive.

The cell phone procedures are necessary to ensure that the respondent is comfortable speaking with us on his/her cell phone and to follow the regulations regarding calling cell phones for research purposes.

If the interviewer encounters a language barrier, either with the person answering the phone or with the designated respondent, the interviewer will thank the person and terminate the call. If the case is designated as Spanish language, it will be turned over to a Spanish-speaking interviewer.

All interviewers on the study will be supervised in a manner designed to maintain high quality control. One component of this supervision will entail periodic monitoring of interviewers while they are working. Through computer and phone technology, supervisors can silently monitor an interviewer's work without the awareness of either the interviewer or respondent. Second, supervisors will check interviewers' completed work for accuracy and completeness.

B.2.f. Sample Weighting

The Contractor will carry out a multi-stage sequential process of weighting. The first stage weight will correct for known selection biases in the sampling procedures. Since random digit dialing gives households with more than one telephone number an unequal likelihood of selection in the landline sample, the first stage weight will be equal to the inverse of the number of different telephone numbers in the household. The second stage weight will correct for unequal probability of selection within sampled households in the landline sample. Since a respondent's probability for selection within a household is inverse to the number of eligible adults in the household, the second stage weight equals the number of eligible respondents within the household. No such weights are required within the cell phone only sample. The next step corrects the study design for disproportionate selection of population subsets in the sample design. The third stage weight corrects the achieved sample for disproportionate sampling by dividing the expected population distribution, based on Census projections or estimates, by the (weighted) achieved sample distribution on the stratification variables. Finally, we must weight the landline and cell phone only respondents to their estimated population sizes based on the most recent National Center for Health Statistics (NCHS) estimates. These steps correct the achieved sample for known biases in sample selection.

B.2.g. Statistical Analysis

The sample estimates from which population projections are derived initially involve univariate frequencies of survey responses. The frequency of all responses to each question will be tallied. Where appropriate, measures of central tendency will be calculated, such as means and medians. NHTSA personnel will use SAS$^{©}$ to test for significant differences between the previous telephone survey and the present one, and to test for significant differences among the demographic variables listed below. Given that many such tests will be made, differences will not be considered significant unless they exceed the 99% confidence level.

It is important for NHTSA's strategic planning needs that the status of awareness, knowledge, attitudes, and behaviors related to traffic safety be determined for identifiable subgroups of the total respondent sample. Thus, questionnaire items will be cross-tabulated with key demographic variables including:

    o Age;
    o Sex;
    o Race/Ethnicity;
    o Education;

These cross-tabulations will be scanned to determine the most important bivariate relationships. This data will be extracted, and tables suitable for survey reports and research notes will be constructed. These tables will present response distributions for respondents as a whole (total) and within specific subgroups of interest.

Where appropriate, multivariate analyses will be conducted to assess the predictors of key dependent measures within the survey.

B.3. Describe methods to maximize response rates and to deal with issues of non-response.

In order to attain the highest possible response rate, an interviewing strategy with the following major components will be followed. The initial contact script has been carefully developed and refined to be persuasive and appealing to the sample. Only thoroughly trained and experienced interviewers, highly motivated and carefully monitored, will conduct the interviewing. Interviewers will be trained on how to overcome initial reluctance, lack of interest or hostility during the contact phase of the interview. The interviewing corps will include Spanish-speaking interviewers to ensure that Spanish language is not a barrier to survey participation. The Contractor will make fifteen call attempts to ring-no-answer numbers, and shall leave an approved message on answering machines according to study protocol (see Section B.2.f.).

The CATI program will record all refusals and interview terminations in a permanent file, including the nature, reason, time, and the interviewer. This information will be reviewed on an ongoing basis to identify any problems with the contact script, interviewing procedures, questionnaire items, etc. Also, the refusal rate by interviewer will be closely monitored. Using these analyses, a "Conversion Script" will be developed. This script will provide interviewers with responses to the more common reasons given by persons for not wanting to participate in the survey. The responses are designed to allay concerns or problems expressed by the telephone contacts.

The contractor will implement a refusal conversion plan in which each person selected for the sample who refuses to participate will be re-contacted by the contractor approximately one-to-two weeks following the refusal. The contractor will use the Conversion Script in an attempt to convince the individual to reconsider and participate in the survey. Only the most experienced and skilled interviewers will conduct the refusal conversions. Exceptions to refusal conversion will be allowed on an individual basis if for some reason the refusal conversion effort is deemed inappropriate.

There will be maintenance and regular review of field outcome data in the sample reporting file, derived from both the sample control and CATI files, so that patterns and problems in both response rate and production rates can be detected and analyzed. Meetings will be held with the interviewing and field supervisory staff and the study management staff to discuss problems with contact and interviewing procedures and to share methods of successful persuasion and conversion.

The procedures described above resulted in a response rate for the cross-sectional sample of between 45% and 50% for past NHTSA surveys. The contractor will release less sample initially for the current survey and apply the 15 contact and callback strategy to improve response rates for this survey. The sampling methods described above will yield reliable data that can be generalized to the universe studied.

Non-Response Analysis

A comparison of the characteristics of the completed and non-completed cases from the random digit dialing sample will be conducted to determine whether there is any evidence of significant non-response bias in the completed sample. There are three points at which sample characteristics are collected in the course of the telephone survey: prior to household contact; at respondent selection for the survey; and after beginning the interview. Analysis will be conducted with the data available for respondents and non-respondents at each of these points.

The drawn sample of random digit dialing numbers that are released to telephone interviewers for contact attempts can be coded for certain contextual characteristics. The "hundreds bank" from which the RDD number is drawn (e.g., 301-608-38xx) is classified geographically according to the unit from which the majority of listed numbers in that bank are located. This permits us to classify all of the numbers drawn by Census region, Census district and urbanicity (central city, metro remainder, and non-metropolitan area). It also permits us to classify those numbers by percent African-American and Hispanic based on Census data. A comparison of the characteristics of all dialed numbers in which an interview had not been completed with those that had been completed can be done with these variables.

Once contact is achieved with a household, one adult is selected in each household as the designated respondent. If there is more than one eligible respondent per household, then the individual with the most recent/next birthday is selected. The age and gender of this individual is obtained so that they can be identified if callbacks are necessary. This information permits additional comparisons of any differences in the age and gender of respondents in the completed sample and those in the non-completed sample after household contact and respondent selection had been made.

Finally, we expect there will be a small percentage (~2.5%) of terminated interviews based on a past survey we conducted for NHTSA (MVOSS 2003[11]) throughout the course of data collection. A terminated interview is one in which the designated respondent begins the interview (answers the first question) but refuses to complete the interview at some point after the first question. Some key motor vehicle and driving characteristics are collected at the very beginning of the interview, including frequency of driving, type of vehicle and seat belt use. Difference in driver characteristics can be analyzed at this third stage between those who terminate the interview and those who complete the interview.

This three-step process of analyzing the characteristics of respondents and non-respondents to the survey should identify whether there is any evidence of significant non-response bias in the most likely areas: region, size of place, minority communities, age, gender, and driving characteristics. This analysis will suggest whether any weighting or other statistical adjustment needs to be made to correct for non-response bias in the completed sample.

B.4. Describe any tests of procedures or methods to be undertaken.

A pretest will be conducted among 10 people, to ensure the CATI script is functioning properly

---

[11] Boyle, John and Patricia Vanderwolf. 2003 Motor Vehicle Occupant Safety Survey. September, 2003.

and the data is being collected accurately. The pretest will consist of the entire survey process, from sample management to tabulation of results. Any problems encountered during the pretest of the questionnaire will be resolved before the survey is put into the field.

B.5. Provide the name and telephone number of individuals consulted on statistical aspects of the design.

The following individuals have reviewed technical and statistical aspects of procedures that will be used to conduct the Unreported Crashes Survey:

John Boyle, Ph.D.
Executive Vice President
Abt SRBI, Inc.
8403 Colesville Road, Suite 820
Silver Spring, MD 20910
(301) 608-3883

Paul Schroeder
Senior Analyst
Abt SRBI, Inc.
8403 Colesville Road, Suite 820
Silver Spring, MD 20910
(301) 608-3883

Thomas Sexton, Ph.D.
Statistician,
M. Davis & Co. Inc.,
Stony Brook University
317 Harriman Hall
College of Business
Stony Brook, NY 11794
(631) 632-7181 (office)