SUPPORTING STATEMENT

FOR

National Automotive Sampling System Law Enforcement Information

OMB Control Number 2127-XXXX

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. <u>Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.</u>

The potential respondent universe
The collection would gather information from law enforcement agencies who respond to motor vehicle crashes in one or more of the 197 county clusters selected as Primary Sampling Units (PSUs) for the successor to NASS, for which the crash count data we seek is not available from other sources.

The Potential Respondent Universe and Sample

the universe of potential respondents	1,450 law enforcement agencies
the sample of potential respondents	1,450 law enforcement agencies (census selection)
estimated response rate	90% of agencies estimated to provide full or partial response

We derive these figures in detail in the subsequent text.

Sampling methods used: how the PSUs were selected
Although we would take a census of PJs in the selected PSUs, the PSUs themselves were selected through probability samples. Of the 197 PSUs, 101 were selected for the General Estimates System (GES) crash module, and the remaining 96 for the

Follow-On Passenger Vehicle (FOPV) module.[1] Following the structure of the current NASS, we are planning that the GES would continue to collect basic information on a large sample of crashes in order to provide broad crash estimates at a national level. The FOPV would collect more detailed information from a smaller sample of crashes, in order to provide the type of detailed information needed to form and evaluate vehicle regulations and serve more specialized needs. Their sample designs are as follows.

For GES, NHTSA identified crash strata based on the types of vehicles involved in the crash, the ages of these vehicles, and the injuries sustained by crash victim. Following the methods of Folsom et al. (1987)[2], we formed a composite measure of size from these strata and defined PSUs by grouping counties together to achieve a threshold measure of size that would guarantee approximately equal weighting within each crash stratum.[3] One PSU was much larger than the others and identified as a certainty PSU, and we estimated that we would be able to afford to operate at most 100 additional GES PSUs in any given year. To ensure that we collect data in urban and rural areas from each region of the country, we stratified the (non-certainty) PSU frame by Census region and whether or not the PSU contains a Metropolitan Statistical Area (MSA). To achieve the greatest precision possible, we maximally substratified these into a total of 50 non-certainty strata that optimized the precision of key estimates, and selected two PSUs from each non-certainty stratum with probability proportional to the composite measure of size. These 100 non-certainty PSUs together with the certainty PSU make up GES' 101-PSU sample.

For FOPV, the primary interest is in newer vehicles and injury crashes. A driving constraint in the design is to select ample numbers of these cases. With fatal crashes being the rarest "type" of crash, NHTSA formed PSUs by grouping counties together to achieve a minimum threshold of fatal crashes, so that they might have ample numbers of (the less rare) newer vehicle crashes and injury crashes. Again following Folsom, we formed a composite measure of size based on ten crash strata that NHTSA identified from user needs. As with GES, we identified a maximum sample size of 101 PSUs, stratified by Census region and MSA/non-MSA, and substratified to reduce the variance of key variables. In order to guarantee enough crashes of primary interest, we selected PSUs in proportion to the composite measure of size formed from the crash strata involving these crash types (newer vehicle crashes and injury crashes). The end result of this process is our 96-PSU sample for FOPV.[4]

_____

1 The names ("GES" and "FOPV") of the system modules might be temporary, as we are considering new names for the new crash data system.
2 Folsom, R.E., Potter, F., and William, S.R. (1987). Notes on a composite size measure for self-weighting samples in multiple domains. *Proceedings of the ASA Survey Methods Research Section*, 792-796.
3 Equal weighting is desirable within the crash strata because the strata comprise common analysis domains.
4 The final FOPV sample involved 96 PSUs instead of 101, because the most propitious stratification (achieving the greatest variance reduction) involved 24 strata, so we chose 4 PSUs from each stratum.

The estimated size of the potential respondent universe
There are about 18,000 law enforcement agencies nationwide[5], not all of which respond to motor vehicle crashes.[6] Examining data from NHTSA's Fatality Analysis Reporting System (FARS)[7], about 10,000 agencies have reported at least one fatal crash in the past five years. Tabulating that there are 455 counties in our 197 Primary Sampling Units and there being 3,144 county-equivalents nationwide, we thus estimate that about 1,450 law enforcement agencies respond to crashes in one or more of the PSUs (455 x 10,000 / 3,144 = 1,447).

The number of potential respondents in the sample, overall and by stratum
Because the information we seek is for use in the next stage of sampling in the NASS design (the sampling of law enforcement agencies), we would need to contact each of the estimated 1,450 agencies in the universe. Thus, 1,450 law enforcement agencies is also the number of potential respondents in the sample. (As we are employing census selection, the sample is effectively drawn from one stratum, and so there are no additional per-stratum figures.)

Response rates achieved previously
NHTSA conducted a similar exercise in the 1970s, when the original NASS was designed. Our records from that time did not record the response rate, but the recollection of the few remaining staff that were involved in this collection of 30 years ago recall that the response rate was as high as 95 percent.

Estimated response rates for the proposed collection
Circumstances today certainly differ from 30 years ago, but we are hopeful that we will be able to get at least partial crash count information from at least 90% of the agencies contacted. We base this figure on the high degree of success of the previous collection, and the limited burden per respondent (i.e. that we are effectively seeking six crash count figures, together with the calendar year of their occurrence).

2. Describe the procedures for the collection of information including:
    - Statistical methodology for stratification and sample selection,
    - Estimation procedure,
    - Degree of accuracy needed for the purpose described in the justification,
    - Unusual problems requiring specialized sampling procedures, and
    - Any use of periodic (less frequent than annual) data collection cycles to reduce burden.

----

5 U.S. Department of Justice, Census of State and Local Law Enforcement Agencies, 2008
6 For instance, agencies associated with correctional facilities and juvenile justice typically do not respond to motor vehicle crashes.
7 The FARS is a census of police-reported motor vehicle trafficway crashes in the U.S. in which an involved person dies within 30 days.

PRA Application for OMB Number 2127-XXXX, NASS Law Enforcement Information

Statistical methodology for stratification and sample selection
As described in our response to B1, the sample of potential respondents would comprise a census of approximately 1,450 law enforcement agencies from the 197 PSUs selected for the new crash data system. This section also provides a detailed description of the probabilistic process through which the 197 PSUs were selected.

Estimation procedure
Because the purpose of this collection is to gather information for the next stage of the NASS redesign (this next stage being the probabilistic selection of law enforcement agencies within the PSUs), no estimates will be formed from the information we collect. That is, the data from the proposed collection would be used solely for the formation of the LEA sampling frame and as a measure of size for the selection of law enforcement agencies in the future NASS sample.

Degree of accuracy needed for the collection's responses
Although getting more accurate information in this collection will better inform the subsequent design stages, there is no threshold degree of accuracy required to produce an improved crash data system. That said, the information we seek is limited in scope and basic in nature (e.g. the number of motorcycle crashes in the most recent year for which the agency has data) and so we expect we will generally be provided with accurate figures.

Unusual problems requiring specialized sampling procedures
We foresee no such problems, and expect to be able to contact all of the estimated 1,450 agencies in the potential respondent universe within the requested 6-month collection period.

Use of periodic data collection cycles to reduce burden
As the burden per respondent is fairly minimal (six crash counts, plus the year in which they occurred) and the proposed collection is a one-time collection for sampling frame formation, we do not foresee the need for periodic collection cycles.

3. <u>Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.</u>

Plans for maximizing response rates
Our plan to collect the data is for Law Enforcement Liaisons working under existing arrangements with NHTSA's regional offices to contact the law enforcement agencies in their region via telephone.[8] One or two follow-up calls will be made in the case of non-response. Every effort will be made to make the process as convenient as

8 NHTSA has 10 regional offices, as described at: http://www.nhtsa.gov/nhtsa/whatis/regions/

PRA Application for OMB Number 2127-XXXX, NASS Law Enforcement Information

possible for the contacted agencies. For instance, if a contacted agency would prefer to refer us to a website or online query tool, rather than look up the information themselves, we will be happy to do so.

Plans for dealing with non-response
We plan to impute missing data (whether item or unit nonresponse) from other available data (such as population counts).

PJ information is mainly needed to estimate the Police Accident Report (PAR) strata counts in the sample. We use GES as an example, but the FOPV is similar. GES PJ information is mainly needed to estimate the 9 PAR strata counts. GES measure of size is a composite variable defined by the PJ's 9 PAR strata counts. To impute PJ level PAR strata counts, we'll proceed in two steps: first estimate county level PAR strata population counts, then allocate the county level PAR population counts to the PJs with missing data. We use GES as an example, FOPV is similar.

To estimate the county level PAR strata 4 and 5 for example, we shall combine the State Data System (SDS) imputed incapacitating (A) injuries with the fatal counts from PARS (a census of fatal crashes). These counts are at the county level. The Polk vehicle registration data is then used to divide the fatal crashes into those involving a killed or incapacitated occupant in a vehicle 0-4 years old, and those involving a killed or incapacitated occupant only in vehicles 5+ years old by multiplying the proportion of Polk data for occupant vehicles in those two model year categories. The method is PAR strata specific and all PAR strata counts will be estimated.

Once the raw county level PAR strata totals are estimated, they need to be post-stratified to agree with the corresponding estimated total crashes based on the 2011 GES. The post-stratified PAR strata totals are calculated by multiplying a post-stratification factor to each of the raw PAR stratum count.

After the county level PAR strata total are estimated, we shall use DOJ's PJ frame and USACOPS.COM PJ frame to identify the PJs in the PSU/county with missing PJ frame information. Geocoding will also be used to locate each of the PJs into the county, city or an area. We can then link the ACS (American Community Survey) data to link the population data to the county, city or area. And finally we can allocate the county level PAR strata totals proportionally to the PJ level using the population distribution.

Accuracy and reliability for intended use
As mentioned previously, the information collected would be used solely for subsequent stages of the NASS redesign, without any intention to generalize to larger universes. As such, there is no threshold level of accuracy or reliability required. That said, we do not expect accuracy or reliability to be problematic, due to the basic nature of the sought counts.

4. <u>Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information.</u>

Because we propose to collect the information through a conversation between two law enforcement officers (the Law Enforcement Liaison and an officer at the contacted agency) and the manner in which records are kept can differ from precinct to precinct, the Liaisons will not be reading a script. As such, we do not see benefit in testing such conversations in advance.

5. <u>Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.</u>

Statistical design
The following persons were consulted on statistical aspects of the design:

Staff from NHTSA, 1200 New Jersey Ave SE, Washington DC 20590
- Fan Zhang, COTR, (202) 366-0183, fan.zhang@dot.gov
- Donna Glassbrenner, (202) 366-3962, donna.glassbrenner@dot.gov
- Rajesh Subramanian, Chief, Mathematical Analysis Division, (202) 366-3365. rajesh.subramanian@dot.gov
- Chou-Lin Chen, Director, Office of Traffic Records and Analysis, (202) 366-1048, chou-lin.chen@dot.gov

Data collection
The data would be collected by the Law Enforcement Liaisons associated with NHTSA's ten regional offices.

Use of the data
As mentioned previously, the information from this collection would be used for the sole purpose of informing subsequent design stages in the NASS redesign. NHTSA has contracted Westat, Inc. of Rockville, Maryland, for this effort.

PRA Application for OMB Number 2127-XXXX, NASS Law Enforcement Information