

## SUPPORTING STATEMENT

### FOR

#### P.L. 89-663, Title 1, Section 106, 108, 112. - COLLECTION OF CRASH DATA

OMB Control Number: None

#### B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

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

The purpose of Crash Report Sampling System (CRSS) is to provide annual, nationally representative estimates of the number, types and characteristics of police-reported motor vehicle crashes. The police crash report (PAR) is the sole source of data for CRSS. The CRSS universe, or sample frame, is the set of police-reported motor vehicle crashes on a traffic way (strata 2 – 10 of Table 1).

Table 1 - CRSS Analysis Strata, Target Sample Allocation, and Population Sizes

CRSS Analysis Stratum	Analysis Stratum Description	Target Percent of Sample Allocation	Estimated Population (GES 2011)	Population Percent
1	An in-scope Not-in-Traffic Surveillance (NiTS) crash (take all)*.			
2	Crashes not in Stratum 1 in which: Involves a killed or injured (includes injury severity unknown) non-motorist	9%	119,579	2.2%
3	Crashes not in Stratum 1 or 2 in which: Involves a killed or injured (includes injury severity unknown) motorcycle or moped rider	6%	76,513	1.4%
4	Crashes not in Stratum 1-3 in which: At least one occupant of a late model year** passenger vehicle is killed or incapacitated	4%	22,272	.42%

CRSS Analysis Stratum	Analysis Stratum Description	Target Percent of Sample Allocation	Estimated Population (GES 2011)	Population Percent
5	Crashes not in Stratum 1-4 in which: At least one occupant of an older** passenger vehicle is killed or incapacitated	7%	84,659	1.6%
6	Crashes not in Stratum 1-5 in which: at least one occupant of a late model year passenger vehicle is injured (including injury severity unknown)	14%	330,619	6.2%
7	Crashes not in Stratum 1-6 in which: involved at least one medium or heavy truck or bus (includes school bus, transit bus, and motor coach) with GVWR 10,000 lbs. or more	6%	302,781	5.7%
8	Crashes not in Stratum 1-7 in which: at least one occupant of an older passenger vehicle is injured (including injury severity unknown)	12%	800,390	15.0%
9	Crashes not in Stratum 1-8 in which: involved at least one late model year passenger vehicle, AND No person in the crash is killed or injured	22%	1,511,371	28.4%
10	Crashes not in Stratum 1-9: * This includes mostly PDO crashes involving a non-motorist, MC , moped, and passenger vehicles that are not late model year and any crashes not classified in strata 1-9.	20%	2,078,263	39.0%

\*: NiTS cases are not in the scope of CRSS. They are set aside for NiTS analysis.

\*\*\*: Note: Late model year passenger vehicle: passenger vehicle that are ≤4 years old

Older passenger vehicle: passenger vehicle that are 5 years old and older

The estimated CRSS population size (strata 2 – 10 of Table 1) is about 5.3 million a year. CRSS samples this population through a stratified multi-stage cluster scheme as follows:

(1) PSU Sample Selection

Divide the country into geographic units called Primary Sampling Units (PSUs). A PSU is a county or group of counties. PSUs were formed as groups of adjacent counties subject to a minimum measure of size (MOS) condition to ensure enough cases to be sampled from each PSU and approximately equal weights. The CRSS PSU MOS was defined as:

$$MOS_i = \sum_{s=2}^{10} \frac{n_{++s}}{n} \frac{N_{i+s}}{N_{++s}}$$

Here

$s$  = the PAR strata defined in Table 1.

$n$  = the desired total sample size of PARs

$n_{++s}$  = the desired sample size of PARs in the PAR stratum  $s$

$N_{++s}$  = the estimated population counts in the PAR stratum  $s$

$N_{i+s}$  = the estimated population counts in the PAR stratum  $s$  and PSU  $i$ .

In the formula,  $n_{++s}/n$  is the desired PAR strata sample allocation (the “Target Percent of Sample Allocation” column in Table 1), and  $N_{i+s}/N_{++s}$  is the relative estimated population counts of PSU  $i$  for PAR stratum  $s$ . In this way, a PSU with larger desirable combination of estimated population counts of all PAR strata has larger MOS.

PSU formation respects Census region and urbanicity boundary. Some outlying areas of Alaska and small islands of Hawaii were excluded. There are total of 707 CRSS PSUs in the PSU frame.

The PSU frame was then stratified into 8 primary PSU strata by two variables – region (Northeast, West, South, and Midwest) and urbanicity (urban and rural). Within each primary stratum, PSUs were further stratified by other secondary stratification variables such as vehicle miles traveled, crash rate, truck miles traveled, and crash rate by road type. PSUs with similar characteristics were grouped into secondary strata with approximately equal MOS sizes and minimum within stratum variances. As the result, total 50 PSU strata are formed.

Table 2: CRSS PSU Strata, PSU Population Counts, and Sample Size

PRIMARY STRATA	STRATID	VMT_RATE_IMP		TOT_CRASH_RATE		TRK_MI_RATE		ROAD_TYPE_RATE		Number of PSUs	PSU Sample Size
		Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower		
1	101	1800.660	0.000					358.504	0.000	5	2
1	102	4064.065	1800.660					358.504	0.000	5	2

1	103	7159.04 4	4064.0 65					358.50 4	0.000	8	2
1	104	5791.03 4	0.000	0.028	0.000	153756. 114	0.000	2175.0 24	358.50 4	6	2
1	105	8040.03 1	5791.0 34	0.028	0.000	153756. 114	0.000	2175.0 24	358.50 4	7	2
1	106			0.028	0.000	249917. 616	153756. 114	2175.0 24	358.50 4	7	2
1	107			0.028	0.000	591240. 550	249917. 616	2175.0 24	358.50 4	7	2
1	108			0.039	0.028			2175.0 24	358.50 4	11	2
2	201					236700. 660	0.000			22	2
2	202					1027525 695	236700. 660			22	2
3	301	4134.62 2	0.000			45708.7 32	0.000			3	2
3	302	7465.06 0	4134.6 22			45708.7 32	0.000			8	2
3	303	9897.83 4	7465.0 60			45708.7 32	0.000			10	2
3	304					102553. 858	45708.7 32			11	2
3	305	4443.52 9	0.000			339758. 109	102553. 858			13	2
3	306	6002.75 8	4443.5 29			339758. 109	102553. 858			11	2
3	307	11617.9 75	6002.7 58			339758. 109	102553. 858			10	2
4	401					66170.8 91	0.000	4344.5 84	0.000	28	2
4	402	6045.03 2	0.000			565024. 725	66170.8 91	4344.5 84	0.000	27	2
4	403	11623.1 51	6045.0 32			565024. 725	66170.8 91	4344.5 84	0.000	25	2
4	404							17641 397	4344.5 84	30	2
5	501	3619.86 6	0.000	0.048	0.000	125590. 275	0.000			5	2
5	502	4529.72 8	3619.8 66	0.048	0.000	125590. 275	0.000			8	2
5	503	4951.02 1	4529.7 28	0.048	0.000	125590. 275	0.000			6	2
5	504	5016.20 3	4951.0 21	0.048	0.000	125590. 275	0.000			3	2
5	505	5277.18 0	5016.2 03	0.048	0.000	125590. 275	0.000			5	2
5	506	5745.57 6	5277.1 80	0.048	0.000	125590. 275	0.000			6	2
5	507	6399.20 1	5745.5 76	0.048	0.000	125590. 275	0.000			5	2
5	508	12826.2 84	6399.2 01	0.048	0.000	125590. 275	0.000			8	2
5	509	5641.16 1	0.000	0.048	0.000	210430. 008	125590. 275			6	2
5	510	8347.64 0	5641.1 61	0.048	0.000	210430. 008	125590. 275			7	2
5	511	13892.0 20	8347.6 40	0.048	0.000	210430. 008	125590. 275			10	2
5	512			0.048	0.000	358684. 226	210430. 008			8	2
5	513			0.048	0.000	877545. 577	358684. 226			13	2
5	514			0.085	0.048					17	2
6	601					49853.6 75	0.000			35	2

6	602	6353.41 9	0.000			162415. 067	49853.6 75			34	2
6	603	14414.9 22	6353.4 19			162415. 067	49853.6 75			35	2
6	604					250189. 692	162415. 067			33	2
6	605	5693.12 1				1156242 .173	250189. 692			35	2
6	606	16270.5 33	5693.1 21			1156242 .173	250189. 692			35	2
7	700									1	1
7	701	6477.24 9	0.000	0.027	0.000	104521. 554	0.000			7	2
7	702	6920.87 4	6477.2 49	0.027	0.000	104521. 554	0.000			4	2
7	703	7860.80 5	6920.8 74	0.027	0.000	104521. 554	0.000			5	2
7	704	5137.29 3	0.000	0.027	0.000	249358. 333	104521. 554			3	2
7	705	8069.65 7	5137.2 93	0.027	0.000	249358. 333	104521. 554			10	2
7	706			0.048	0.027	92715.8 11	0.000			9	2
7	707			0.048	0.027	186409. 133	92715.8 11			7	2
8	801							3938.4 60	0.000	30	2
8	802							18292. 498	3938. 460	41	2

A major challenge of the CRSS sample design is the uncertainty over the future operational budget. Unknown future funding levels and the need for a stable PSU sample require NHTSA to select a scalable PSU sample in which the PSU sample size can be decreased or increased with minimum impact to the existing PSU sample and the selection probabilities can be tracked. To this end, a multi-phase sampling method was used to select the CRSS PSU sample by selecting a sequence of nested PSU samples. In this method, a PSU sample larger than actually needed is first selected as the first phase PSU sample. From the first phase PSU sample, a smaller subset of PSU sample is selected as the second phase PSU sample. From the second phase PSU sample, another smaller third phase PSU sample is selected. This process is continued until the PSU sample size reaches unacceptable levels. In this way, a sequence of nested PSU samples is obtained. Each of these PSU samples is a probability sample and can be used for data collection (see Figure 1). According to the prevailing budget level, a sample with the appropriate sample size is picked from the nested sequence. This allows us to easily track the selection probabilities and minimizes changes to the PSU sample.

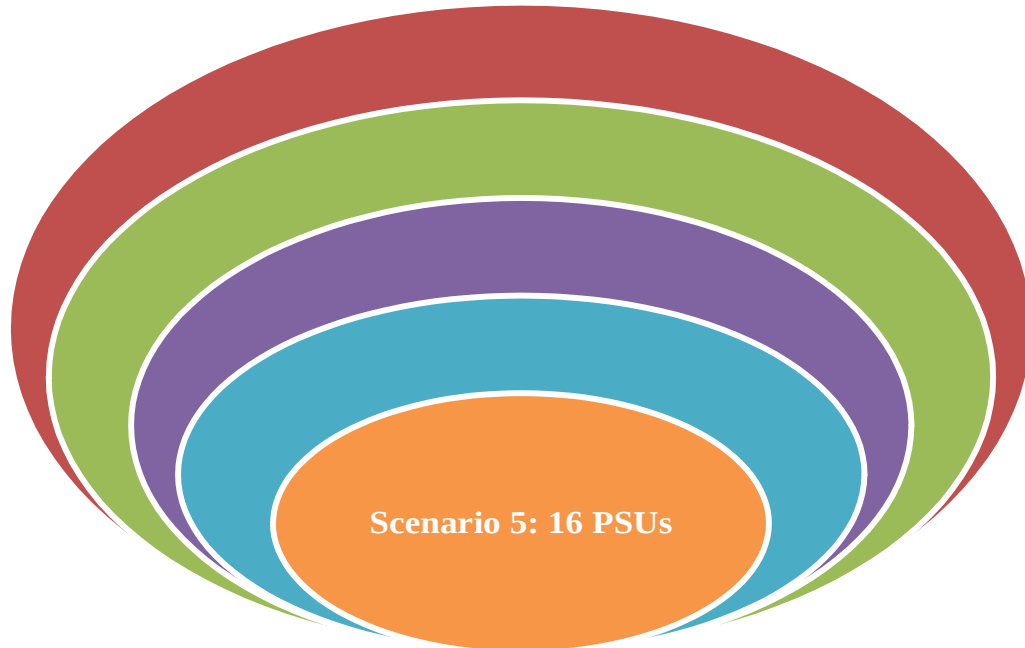


Figure 1: Nested PSU Samples

For CRSS, 5 PSU samples were selected under the 5 scenarios of number of PSU strata and PSU sample sizes. Table 3 summarizes the CRSS PSU sample scenarios.

Table 3 – CRSS PSU Sample Scenarios: Number of Strata and Sample Size

Scenario	# of PSU Strata	# of Sampled Non-certainty PSU	# of Sampled Certainty PSU	Total # of Sampled PSU
1	50	97	4	101
2	37	74	1	75
3	25	50	1	51
4	12	24	0	24
5	8	16	0	16

With sample size of 100 and without stratification, one PSU was identified as certainty PSU by condition:

$$\pi_i = \frac{100 * MOS_i}{\sum_{i=1}^N MOS_i} \geq 1$$

Here N is the total number of PSUs in in the PSU frame. This certainty PSU was set aside and selected with certainty. Then 2 PSUs were selected using probability proportional to size (PPS) sampling from each of the 50 scenario-1 strata. With sample size 2 for each PSU stratum, total 3 PSUs were identified as certainty PSUs from 3 of the 50 scenario-1 strata by condition:

$$\pi_{hi} = \frac{2 * MOS_{hi}}{\sum_{i=1}^{N_h} MOS_{hi}} \geq 1$$

Here  $N_h$  is the total number of PSUs in stratum  $h$ . The certainty PSUs were aside and selected with certainty. The corresponding stratum PSU sample size was reduced by 1. Then a PPS sample of non-certainty PSUs was selected using the revised PSU stratum sample size.

Scenario-1 sample has total 101 PSUs. For a non-certainty PSU, the selection probability is:

$$\pi_{hi} = \frac{n_{hi} * MOS_{hi}}{\sum_{i=1}^{N_h} MOS_{hi}}$$

Here  $n_h$  is the non-certainty PSU sample size for PSU stratum  $h$ .

For scenario-2, with sample size of 74 and without stratification, one PSU was identified as certainty PSU and was set aside. Then 13 of the scenario-1 strata were collapsed with other strata to form total 37 scenario-2 strata. The collapsing of strata follows the following rule:

- Only the secondary strata in the same primary stratum can be collapsed;
- Only the contiguous secondary strata can be collapsed;
- The resulting strata have similar stratum total MOS within each primary stratum.

In each of the scenario-2 stratum, the sampled scenario-1 PSUs were treated as the sampling frame. Each PSU was assigned a new MOS equal to its scenario-1 stratum total MOS. Then 2 PSUs were selected from each scenario-2 stratum with PPS sampling using the new MOS. In this way, the resulting selection probability of the scenario-2 PSU is still PPS selection probability.

Other scenario samples were selected in similar way.

## (2) SSU Sample Selection

The secondary sampling units (SSU) of CRSS are police jurisdictions. Within each PSU, PARs are stratified by the police jurisdictions (PJ) where PARs are available and PJs become the second stage sampling units. A composite MOS is assigned to each PJ in the selected PSUs. Similar to PSU MOS definition, it is sensible to assign larger selection probability to PJs with desirable crash composition. To this end, crash counts of the 9 PAR strata in

Table 1 for each PJ in the selected PSUs were estimated from the information collected from the PJs in the selected PSUs. For the PJ  $j$  in the PJ frame within the sampled PSU  $i$ , the composite SSU MOS is defined as the following:

$$MOS_{j \vee i} = \sum_{k=2}^{10} \frac{n_{++s}}{n} \frac{N_{ijs}}{N_{++s}}$$

where

$n$  = the desired total sample size of crashes

$n_{++s}$  = the desired sample size of crashes in the PAR stratum  $s$

$N_{++s}$  = the estimated population number of crashes in PAR stratum  $s$

$N_{ijk}$  = the estimated population number of crashes in PAR stratum  $s$ , PJ  $j$  and PSU  $i$

PJs are then stratified into two PJ strata by their MOS (largest 50% vs the rest) in addition to certainty PJs. A PJ sample is then selected from each PJ stratum using sequential Poisson sampling.

Sequential Poisson sampling method (see Ohlsson, Esbjörn (1998): Sequential Poisson Sampling, Journal of Official Statistics, Vol.14, No.2, pp. 149–162) produces an approximate PPS sample, handle the frame changes and minimize the changes to the existing sample at the same time.

Sequential Poisson sampling method was applied to the PJ sample selection for each of non-certainty PJ strata (large MOS or small MOS stratum) within the sampled PSU  $i$ , as following:

Generate a permanent uniform random number  $r_{ij} \sim U(0,1)$  for each PJ  $j$  in the PJ frame.

Identify certainty PJs by the condition:

$$\frac{m_i * MOS_{ij}}{\sum_{j=1}^{M_i} MOS_{ij}} \geq 1$$

Here  $m_i$  is the PJ sample size and  $M_i$  is the PJ frame size for a PJ stratum within PSU  $i$ .  $MOS_{ij}$  is the PJ MOS. The identified certainty PJs are set aside. And this process is repeated to the remaining PJs based on the reduced PJ sample size until there is no more certainty PJs. Let the total number of certainty PJs be  $m_c$ .

For the remaining  $M_i - m_c$  non-certainty PJs in the frame, divide their permanent random number by the MOS to obtain the transformed random number:  $r_{ij} / MOS_{ij}$ . Then, sort the transformed random number from the smallest to the largest as following:

$$\frac{r_{i1}}{MOS_{i1}} \leq \frac{r_{i2}}{MOS_{i2}} \leq \dots \leq \frac{r_{i(M_i - m_c)}}{MOS_{i(M_i - m_c)}}$$

Thus, the  $m_c$  certainty PJs plus the first  $m_i - m_c$  non-certainty PJs on the above list are the PJ sample for a PJ stratum within PSU  $i$ .



Sequential Poisson sampling is approximately PPS. The PJ selection probability is:

$$\pi_{j \vee hi} = \frac{m_{hi} MOS_{hij}}{\sum_j MOS_{hij}}$$

Here  $j$  is for PJ,  $m_{hi}$  is the PJ sample size for PSU  $i$ ,  $MOS_{hij}$  is the PJ MOS. The summation is over all non-certainty PJs in the selected PSU.

### (3) TSU Sample Selection

The tertiary sampling units (TSU) of CRSS are police crash reports (PAR). The CRSS PAR sample is selected by a stratified systematic sampling. For each selected SSU (PJ), PARs are periodically obtained by either technician's visit to the PJ or electronic transmission. All the PARs are listed in the order they become available, and stratified by the PAR strata identified in Table 1. Through this listing process, PAR sampling frame in each selected PJ are prepared for PAR sample selection.

For a large PJ with too many PARs to be listed, PARs are sub-listed by systematic sampling. For example, only PARs with even PAR number may be listed if a sub-listing factor is 2. Or 1 of every 5 PARs is listed if a sub-listing factor is 5. If one of every  $L$  PARs is sub-listed in PJ  $j$ , PSU  $i$ , the sub-listing probability for all sub-listed PARs are:

$$\pi_{l \vee hij} = \frac{1}{L}$$

After PARs are listed, a PAR sample is selected by systematic sampling from the listed (or sub-listed) PARs by PAR stratum within each selected PJ. PAR  $k$  selection probability is:

$$\pi_{ijlsk} = \frac{n_{ijls}}{N_{ijls}}$$

Here  $n_{ijls}$  is the number of PARs selected from PAR stratum  $s$ ,  $N_{ijls}$  is the number of PARs listed in PSU  $i$ , PJ  $j$ , PAR stratum  $s$ .

The overall selection probability is:

$$\pi_{hijkl} = \pi_{hi} * \pi_{j \vee hi} * \pi_{l \vee hij} * \pi_{k \vee hijl}$$

The design weight is the inverse of  $\pi_{hijkl}$ .

### (4) Sample Allocation

CRSS PSU, PJ and PAR sample sizes are estimated using optimization by minimizing variance subject to cost assuming three stage simple random sampling without replacement.

The optimization model consists of the objective function, cost constraint, and variance constraints as following.

$$\text{Minimize: } \sum_{g=1}^G V_{CRSS}(\bar{\bar{y}}_g) = \sum_{g=1}^G \left\{ \frac{S_{1,g}^2}{n} \left(1 - \frac{n}{N}\right) + \frac{S_{2,g}^2}{nm} \left(1 - \frac{m}{M}\right) + \frac{S_{3,g}^2}{nmk} \left(1 - \frac{k}{K}\right) \right\}$$

$$\text{Subject to: } C = C_0 + nC_1 + nmC_2 + nmkC_3,$$

$$V_{CRSS}(\bar{\bar{y}}_g) = \frac{S_{1,g}^2}{n} \left(1 - \frac{n}{N}\right) + \frac{S_{2,g}^2}{nm} \left(1 - \frac{m}{M}\right) + \frac{S_{3,g}^2}{nmk} \left(1 - \frac{k}{K}\right)$$

$$\leq V_{GES}(\bar{\bar{y}}_g), \text{ for } g=1, \dots, G.$$

$$mk \geq l.$$

- $g$ : Subscript of the identified key estimate,  $g=1, \dots, G$ . Here  $G=13$ .
- $\bar{\bar{y}}_g$ : Identified key proportion estimate.
- $n, m, k$ : Optimal sample sizes of PSUs, PJs, and cases (PARs) to be determined.
- $N$ : Population size of PSUs
- $M$ : Average population size of PJs.
- $K$ : Average population size of PARs
- $V(\bar{\bar{y}}_g)$ : Variance of the identified key estimate  $\bar{\bar{y}}_g$ .
- $S_{1,g}^2, S_{2,g}^2, S_{3,g}^2$ : Variance component at PSU-, PJ-, and case-level.
- $C, C_0, C_1, C_2, C_3$ : Total, fixed, PSU-, PJ-, and crash-level cost coefficients.
- $V_{GES}(\bar{\bar{y}}_g)$ : Variance of the identified key estimate  $\bar{\bar{y}}_g$  in the current system (NASS GES).
- $l$ : known case load.
- Standard errors for seven key estimates under current General Estimates System (GES) were used as constraints in the above optimization model to ensure the corresponding degree of accuracy under CRSS will be at least as good as GES.

Under the current GES budget and the current GES cost components, NHTSA determined the sample allocation is about 60 PSUs, 6 PJs per PSU, and 140 PARs per PJ.

#### (5) Weighting Adjustments, Imputation and variance estimation

After design weights are calculated, the weights need to be adjusted for the following reasons:

- Refusal/non-respondent adjustments;
- Frame coverage bias correction;
- Matching marginal totals to other data sources – for example, total fatality to FARS;
- Large weight trimming;

Calibration technique will be used as the adjustment method. The potential auxiliary information to be used for calibration includes FARS, Census population counts, and PSU level total crash counts.

The calibration adjustment method that handle all the above have been implemented in SUDAAN 11 WTADJX procedure. SUDAAN WTADJX procedure will be used to create the final analysis weights.

Some key item missing values will be imputed. Several imputation methods will be considered and used for imputation, depending on the missing variable and available information. The imputation methods include but not restricted to: logical imputation, regression imputation, and hot deck imputation.

The resulting CRSS PSU sampling rate is quite low. We expect the PSU sample selection can be approximated treated as with-replacement sample selection. The standard specialized software such as SAS SURVEY procedures and SUDAAN procedures can be used for CRSS data analysis.

1. Describe collection of information procedures.

Once a PAR has been selected for data collection, data coders record information in the PAR into the computer system as appropriate. CRSS data collection is solely based on PARs.

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

CRSS has a three stage sample design. The first stage sampling units are counties or groups of counties. A PSU becomes a non-responding PSU only if all selected police jurisdictions (PJs) within the PSU are non-responding PJs. In CRSS, PJ samples are selected using sequential Poisson sampling method. The whole PJ frame can be used as replacement sample. Therefore, a PSU becomes non-responding PSU only if all PJs in the frame are non-responding PJs. In the current GES, the PJ cooperation rate is about 100%. Therefore, by design, it is unlikely there will be any non-responding PSUs in CRSS.

The second stage sampling units of CRSS are PJs. A sampled PJ becomes non-responding PJ if it refuses to cooperate. To improve PJ cooperation rate, NHTSA plan to visit each selected PJ and meet with local law enforcement officers to gain cooperation. In the current GES, NHTSA obtained almost 100% cooperation from sampled PJs. Therefore, we expect only a few non-responding PJs in the CRSS PSUs.

The third stage sampling units of CRSS are PARs. First all police accident reports (PARs) in the selected PJs are listed. Then a systematic sample of PARs is selected and coded. Because the sampling units are police reports and a PAR must be available before it is listed, therefore by design, there is no unit non-response at the third stage of sampling for CRSS.

The CRSS quality control system will be designed to produce the most accurate, reliable, and complete database possible within the limits of available resources. A sample of all PARs will be given a thorough review by experienced data quality control personnel. Quality control personnel will also visit each PSU regularly to observe the data collection activities and to discuss systematic problems revealed in edit and reviews of the data collector's cases.

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

NHTSA will test new data collection procedures for six (6) months. The test will include gathering police crash reports and identifying qualified crashes, analyzing data, and monitoring for quality control.

The attached PAR example (Attachment 4) shows the data elements to be collected from the selected PARs. The electronic forms and protocols are being developed to collect this information on tablet computers.

5. 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.

Ms. Chou-Lin Chen, National Center for Statistics and Analysis, NHTSA, 202-366-1048 is responsible for CRSS survey design.

NHTSA has decided to undertake a basic redesign of the National Automotive Sampling System that will attempt to meet new and diverse requirements through expanding its scope and making it more responsive to changing needs. Accordingly, NHTSA has contracted with Westat (contract DTNH22-12-F-00389) to help the CRSS survey design effort. The contract award date for CRSS data collection and coding is estimated to be March 16, 2015.

