

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 the CRSS is to provide annual, nationally representative estimates of police-reported motor vehicle crashes as well as characteristics of these motor vehicle crashes. The police crash report (PCR) 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 trafficway (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 (NTS) 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%
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%

CRSS Analysis Stratum	Analysis Stratum Description	Target Percent of Sample Allocation	Estimated Population (GES 2011)	Population Percent
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%

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

** : 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) was about 6.5 million in 2017. CRSS selects a sample from the population through a stratified multi-stage cluster scheme as follows:

(1) PSU Sample Selection

The country is divided 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 will be sampled from each PSU and weights are approximately equal within each PCR strata. The CRSS PSU MOS was defined as:

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

where

s = the PCR strata defined in Table 1.

n = the desired total sample size of PCRs

n_{++s} = the desired sample size of PCRs in the PCR stratum s

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

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

In the formula, n_{++s}/n is the desired PCR 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 PCR stratum s . In this way, a PSU with a larger desirable combination of estimated population counts of all PCR strata has a larger MOS.

PSU formation respects US Census region and urbanicity boundaries. While 23 outlying

counties in Alaska and three counties in Hawaii were excluded, the rest of the country is included in the PSU frame. There are 707 CRSS PSUs in the PSU frame.

The PSU frame was then stratified into eight 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. Secondary strata groupings were also based on minimizing the between-PSU variance within a stratum. As the result, 50 PSU strata were formed as indicated in Table 2.

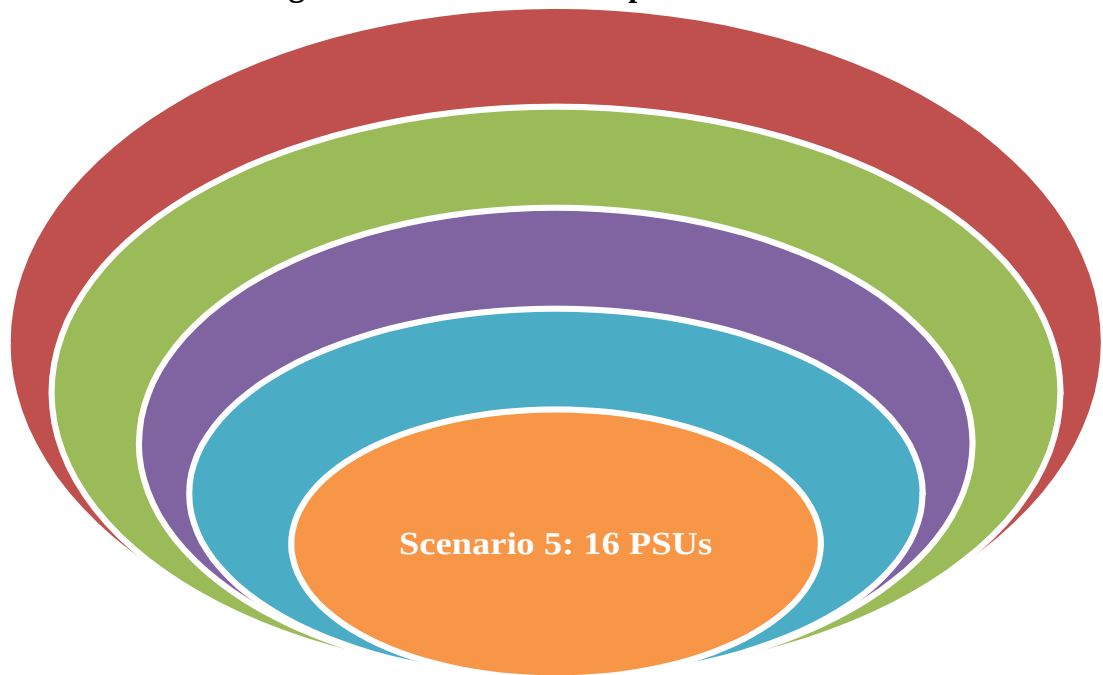
Supporting Statement for the Crash Report Sampling System (Part B)
 OMB Clearance Number: 2127-0714

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

PRIMARY STRATA	STRAT ID	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	1801	0					359	0	5	2
1	102	4064	1801					359	0	5	2
1	103	7159	4064					359	0	8	2
1	104	5791	0	0.028	0	153756	0	2175	359	6	2
1	105	8040	5791	0.028	0	153756	0	2175	359	7	2
1	106			0.028	0	249918	153756	2175	359	7	2
1	107			0.028	0	591241	249918	2175	359	7	2
1	108			0.039	0.028			2175	359	11	2
2	201					236701	0			22	2
2	202					1027526	236701			22	2
3	301	4135	0			45709	0			3	2
3	302	7465	4135			45709	0			8	2
3	303	9898	7465			45709	0			10	2
3	304					102554	45709			11	2
3	305	4444	0			339758	102554			13	2
3	306	6003	4444			339758	102554			11	2
3	307	11618	6003			339758	102554			10	2
4	401					66171	0	4345	0	28	2
4	402	6045	0			565025	66171	4345	0	27	2
4	403	11623	6045			565025	66171	4345	0	25	2
4	404							17641	4345	30	2
5	501	3620	0	0.048	0	125590	0			5	2
5	502	4530	3620	0.048	0	125590	0			8	2
5	503	4951	4530	0.048	0	125590	0			6	2
5	504	5016	4951	0.048	0	125590	0			3	2
5	505	5277	5016	0.048	0	125590	0			5	2
5	506	5746	5277	0.048	0	125590	0			6	2
5	507	6399	5746	0.048	0	125590	0			5	2
5	508	12826	6399	0.048	0	125590	0			8	2
5	509	5641	0	0.048	0	210430	125590			6	2
5	510	8348	5641	0.048	0	210430	125590			7	2
5	511	13892	8348	0.048	0	210430	125590			10	2
5	512			0.048	0	358684	210430			8	2
5	513			0.048	0	877546	358684			13	2
5	514			0.085	0.048					17	2
6	601					49854	0			35	2
6	602	6353	0			162415	49854			34	2
6	603	14415	6353			162415	49854			35	2
6	604					250190	162415			33	2
6	605	5693	0			1156242	250190			35	2
6	606	16271	5693			1156242	250190			35	2
7	700									1	1
7	701	6477	0	0.027	0	104522	0			7	2
7	702	6921	6477	0.027	0	104522	0			4	2
7	703	7861	6921	0.027	0	104522	0			5	2
7	704	5137	0	0.027	0	249358	104522			3	2
7	705	8070	5137	0.027	0	249358	104522			10	2
7	706			0.048	0.027	92716	0			9	2
7	707			0.048	0.027	186409	92716			7	2
8	801							3938	0	30	2
8	802							18292	3938	41	2

A major challenge of the CRSS sample design is the uncertainty of the future operational budget. Due to unknown future funding levels and the need for a stable PSU sample, NHTSA implemented a scalable PSU sample, which allows for the PSU sample size to be decreased or increased with minimum impact to the existing PSU sample and for the selection probabilities to 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 what is actually needed is selected during the first phase of the PSU sample. From the first phase of the PSU sample, a smaller subset of the PSU sample is selected as the second phase of the PSU sample. From the second phase of the PSU sample, another smaller third phase of the 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 existing PSU sample.

Figure 1. Nested PSU Samples



For CRSS, five PSU samples were selected under the five scenarios. Table 3 summarizes the number of PSU strata and sampled PSUs for the CRSS PSU sample scenarios.

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

Scenario	Number of PSU Strata	Number of Sampled Non-certainty PSU	Number of Sampled Certainty PSU	Total Number 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

For scenario-1, with a sample size of 100 and without stratification, one PSU was identified as a certainty PSU by the condition:

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

Let N be the total number of PSUs in the PSU frame. The certainty PSU was set aside and selected with certainty.¹ Then two PSUs were selected using proportional to size (PPS) sampling from each of the 50 scenario-1 strata. With a sample size of two for each PSU stratum, three PSUs were identified as certainty PSUs from three of the 50 scenario-1 strata by the condition:

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

Let N_h be the total number of PSUs in stratum h . The certainty PSUs were set aside and selected with certainty. The corresponding stratum PSU sample size was reduced by one. Then a PPS sample of non-certainty PSUs was selected using the revised PSU stratum sample size.

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

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

Let n_h be the non-certainty PSU sample size for PSU stratum h .

¹ In the probability proportional to size (PPS) sampling, a certainty PSU is identified when the selection probability is equal to or greater than one. If a PSU is identified as certainty, it must be in the sample and its selection probability is set to one. A non-certainty PSU is selected with its selection probability that is greater than 0 and less than 1. If a PSU has a selection probability closer to one, it has more chance to be in the sample. On the other hand, if a PSU has a selection probability closer to zero, it has less chance to be in the sample. For more details, please see Pages 13-28 in the published Technical Report <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812706>

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

- Only the secondary strata in the same primary stratum can be collapsed;
- Only the contiguous secondary strata can be collapsed;
- The resulting strata has a 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 two PSUs were selected from each scenario-2 stratum using PPS sampling based on 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 a similar way.

(2) SSU Sample Selection

The secondary sampling units (SSU) of CRSS are police jurisdictions. Within each PSU, PCRs are stratified by the police jurisdictions (PJ) where PCRs are available and PJs become the second stage sampling units. A composite MOS is assigned to each PJ in the selected PSUs. Similar to the PSU MOS definition, it is sensible to assign larger selection probabilities to PJs with a desirable crash composition. For each PJ in the selected PSUs, crash counts from the 9 PCR strata in Table 1 were estimated from the information collected from the PJs in the selected PSUs. For PJ j in the PJ frame within the sampled PSU i , the composite SSU MOS is defined as the following:

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

where

n = the desired total sample size of crashes

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

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

N_{ijk} = the estimated population number of crashes in PCR 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 Pareto sampling. The Pareto sampling method (see Rosén (1997): On Sampling with Probability Proportional to Size, Journal of Statistical Planning and Inference, Vol. 62, pp. 159-191) produces an approximate PPS sample, handles the frame changes and minimizes the changes to the existing sample at the same time. Pareto sampling method was applied to the PJ sample selection for each of the non-certainty PJ strata (large MOS or small MOS stratum) within the sampled PSU i , as the 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:

$$\frac{m_i * MOS_{ij}}{M_i} \geq 1$$

$$\sum_{j=1} MOS_{ij}$$

Let m_i be the PJ sample size and M_i be the PJ frame size for a PJ stratum within PSU i . MOS_{ij} is the PJ MOS. The identified certainty PJs are set aside. This process is repeated for the remaining PJs based on the reduced PJ sample size until there are 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, calculate the PPS inclusion probability with the non-certainty PJ sample size ($m_i - m_c$):

$$p_{ij} = \frac{(m_i - m_c) MOS_{ij}}{\sum_{j=1} MOS_{ij}}$$

Then, calculate the transformed random numbers and sort the transformed random numbers from the smallest to the largest as following:

\dot{i}

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

Pareto sampling is approximately PPS, and the PJ selection probability is:

$$\pi_{j \vee i} \approx p_{ij}$$

(3) TSU Sample Selection

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

For a large PJ with too many PCRs to be listed, PCRs are sub-listed by systematic sampling. For example, only PCRs with a PCR number ending in 0 through 4 may be listed if the sub-listing factor is 2 (i.e., 5 PCRs among 10 PCRs are listed). Or only PCRs with a PCR number ending in 0 or 1 are listed if the sub-listing factor is 5 (i.e., 2 PCRs among 10 PCRs are listed). If L PCRs among 10 PCRs are sub-listed in PJ j , PSU i , the sub-listing probability for all sub-listed PCRs are:

$$\pi_{l \vee ij} = \frac{L}{10}$$

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

$$\pi_{k \vee ij} = \frac{n_{ijl}}{N_{ijl}}$$

Let n_{ijl} be the number of selected PCRs and N_{ijl} be the number of listed PCRs from each PCR stratum in PJ j of PSU i .

The overall selection probability is:

$$\pi_{ijk} = \pi_i * \pi_{j \vee i} * \pi_{l \vee ij} * \pi_{k \vee ij}$$

The design weight is the inverse of π_{ijk} .

(4) Sample Allocation

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

The optimization model consists of the objective function, cost constraint, and variance constrains as the following:

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

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

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

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

- 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 per PSU, and cases (PCRs) per PJ to be determined.
- N : Population size of PSUs
- M : Average population size of PJs.
- K : Average population size of PCRs
- $V(\bar{\bar{y}}_g)$: 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).
- Standard errors for thirteen key estimates under current 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 PCRs per PJ.

(5) Weighting Adjustments, Imputation and Variance Estimation

After design weights are calculated, the weights are adjusted in the following steps:

Non-response adjustments at all three stages;
Duplicate PCR adjustment;
Post-stratification (i.e., within PSU calibration);
Calibration of case weight;

Case weight calibration benchmarks crash counts from the Fatality Analysis Reporting System (FARS) and US Census population counts simultaneously, and is implemented with SUDAAN WTADJX procedure.

Some key item missing values are imputed. Several imputation methods are considered and used for imputation, depending on the missing variable and available information. The imputation methods include the sequential regression multivariate imputation method, the univariate imputation method, and logical imputation.

The resulting CRSS PSU sampling rate is quite low. We expect the PSU sample selection can be approximately 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.

CRSS data collection efforts are dependent on the method in which the crash reports are accessed. The crash reports are accessed through NHTSA's Electronic Data Transfer (EDT) program, data feeds, secure email, State websites, or manually by contract staff that physically visit the police jurisdiction.

The EDT program consists of a routine automated transfer of crash data from the State crash database to NHTSA. EDT reduces the level of effort required to share crash data because data is automatically shared nightly from the State to NHTSA. States may also provide crash reports to NHTSA through secure web service portals on a routine basis. The State designates the frequency with which they share data with NHTSA under this protocol. Crash report accessed via EDT and secure web portal are uploaded into the Police Accident Report Sampling Engine (PARSE). The PARSE application is a centralized, web-based repository in which CRSS applicable crash reports are listed, categorized, and selected for further coding.

Alternatively, States may provide access to their crash data collection websites. States provide log-in credentials to view crash reports for the sample PJs. The sampler would then list, categorize, and sample the crash reports for sample agencies within the PARSE application.

When States are not able to provide electronic access to crash reports, NHTSA seeks manual access to crash reports from individual police jurisdictions identified in the CRSS sample. Generally, this includes visiting the office to access paper or electronic files, uploading crash reports on an encrypted thumb drive, linking crash reports to a secure email or copying crash reports and sending the crash reports via mail courier service. These more manual processes are completed on a schedule established by the police agency. Once the schedule is agreed upon, then the CRSS sampler can view, list, categorize, and sample the crash reports within the PARSE application.

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, the PJ sample is selected using the Pareto sampling method. The whole PJ frame can be used as replacement sample. Therefore, a PSU becomes a non-responding PSU only if all PJs in the frame are non-responding PJs. In 2017 CRSS, one PSU was non-responding. Since the CRSS PSU sample is scalable, we increased the sample size from 60 to 61 and selected a replacement PSU without changing the original PSU sample. The weight of the non-responding PSU was adjusted.

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 visits each selected PJ and meets with local law enforcement officers to gain cooperation. In 2017 CRSS, four PJs among 397 sample PJs were non-responding. The weights of non-responding PJs were adjusted.

The third stage sampling units of CRSS are PCRs. First all police crash reports (PCRs) in the selected PJs are listed. Then a systematic sample of PCRs is selected and coded. A PCR is identified as non-responding if it has un-readable pages or missing pages. In 2017 CRSS, 17 PCRs among 55,274 sampled PCRs were non-responding. The weights of the non-responding PCRs were adjusted.

The CRSS quality control system is designed to produce the most accurate, reliable, and complete database possible within the limits of available resources. Each selected case, is reviewed by quality control personnel for accuracy before proceeding to coding. Additionally, the Police Accident Report Sampling Engine (PARSE) automatically selects five percent from each sample of non-selected cases to review. The findings from the listed cases review helps identify any quality control issues and additional training needs for the CRSS Sampler.

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

NHTSA transitioned from GES to CRSS in 2016; 2015 was the last year of GES and CRSS was launched in 2016. 2019 is the fourth year of data collection through CRSS. NHTSA tested the CRSS PCR sampling algorithm in PARSE before implementing CRSS. NHTSA simulated test data (1,033 PCRs from 5 PJs of 2 PSUs) from GES. The test was conducted by comparing the PCR sample selected from the PCRSE with the PCR sample selected from the independently developed SAS program to verify the successful import of sampling parameters and correct

sample selection.

A pilot was conducted to test how the various manual and electronic crash report acquisition methods impacted entering the data into PARSE. No changes were made to the data collection protocol as a result of the pilot. However, CRSS operations continues to assess for ways to increase efficiencies in data collection operations and the PARSE application was revamped in 2017. This updated PARSE application includes user friendly features that enabled faster listing, added a quality control module, and introduced a search feature to locate crash reports. The updated PARSE application underwent user acceptance testing for each of the various crash report access methods identified herein.

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 decided to undertake a basic redesign of the National Automotive Sampling System in order to meet the new and diverse requirements through expanding its scope and making it more responsive to changing data needs. Accordingly, NHTSA contracted with Westat (contract DTNH22-12-F-00389) on the CRSS survey design effort.

NHTSA has contracted with KLD Associates Inc. (contract DTNH2214D00366L/0002) for the data collection, coding and quality control for the CRSS data collection effort.