National Child Restraint Use - Special Study (NCRUSS)
Supporting Statement for Information Collection Request
(Part B)

Prepared By:

Charlene T. Doyle
Robert Sivinski
&
Angela Eichelberger
National Highway Traffic Safety Administration

October 19, 2010

TABLE OF CONTENTS

SUPPORTING STATEMENT

National Child Restraint Use – Special Study (NCRUSS)
Supporting Statement for Information Collection Request

A. **Justification**: Please see separate Part A section.

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 consists of all child passengers aged 0-8 and their drivers in passenger vehicles. Although researchers will not be able to identify the target population with perfect accuracy before approaching the vehicle, the age of the child passenger(s) will be determined early in the interview to minimize time spent on subjects outside of the respondent universe. Should additional funds become available, the respondent universe may be expanded to children age 0-12.

Sampling will be conducted at four separate levels. First is the PSU, second is the site within each PSU, third is the vehicle at each site, and the fourth stage is specific children within the vehicle.

Sample Size:
Design effects are estimated using results from the 2009 NSUBS survey. This survey used similar site types and sampling methods, but only 16 PSU's. Design effect was computed for a simple definition of misuse; the use of the incorrect type of restraint for a child based on age, size and weight. The design-based estimate of the design effect (the ratio of the true variance of estimate of misuse to the variance of an estimate derived from a simple random selection) for this survey was 13.55. The model-based estimate was 20.78. In determining the likely design effect for the NCRUSS survey, the larger model-based estimate will be considered so that it can be adjusted for differences in proposed number of PSUs and overall sample size. The model-based design effect is given by:

$$¿\, F_m = \left(1 + c\,v_w^2\right)\left\{1 + \left(\bar{b} - 1\right)\hat{\rho}\right\} \qquad (1)$$

Where $c\,v_w^2$ is the squared coefficient of variation of the sampling weight, $\hat{\rho}$ is the estimated intraclass correlation within PSUs computed using the weighted means and $\bar{b}$ is the simple average of the cluster sizes. The results from the 2009 NSUBS give:

$$¿\, F_m = \left(1 + 1.465\right)\left(1 + \left(398.375 - 1\right)0.0187\right) = 20.78$$

It is difficult to predict changes to $c\,v_w^2$ and $\hat{\rho}$ resulting from the differences in the NSUBS and NCRUSS methodologies and it is assumed in this calculation that they will be equal. However it is possible to adjust $\bar{b}$ for the NCRUSS design. Since the number of PSUs is fixed at 24 and there is an upper limit of 3,000 on overall sample size imposed by available funds, the adjusted average cluster size $\bar{b}^{¿}$ is 3000/24, or 125. Substituting gives:

$$¿\,F_m^{¿} = \left(1 + c\,v_w^2\right)\left\{1 + \left(\bar{b}^{¿} - 1\right)\hat{\rho}\right\}$$

$$¿\,F_m^{¿} = \left(1 + 1.465\right)\left[1 + \left(125 - 1\right)0.0187\right] = 8.18$$

To find the desired sample size to estimate a population proportion with a margin of error of 5% and 95% certainty we begin calculate:

$$n = p\left(1 - p\right)\left(\frac{z_c}{E}\right)^2 * ¿\,F_M^{¿} \qquad (2)$$

Where p, the rate of misuse, is estimated by the 2009 NSUBS as .413, z_c is the z-score associated with our desired 95% level of certainty, and E is the desired margin of error. The overall sample size is therefore:

$$n = .413\left(.587\right)\left(\frac{1.96}{.05}\right)^2 * 8.18$$

$$n = 373 * 8.18 = 3047$$

There are two factors that suggest that a smaller sample size will be sufficient to obtain the desired level of accuracy. It has been asserted that Kish's formula for design effect (1) gives the upper bound of the design effect, and in the case of the NSUBS this seems plausible given the large difference between the model-based design effect given by Kish's formula (20.78) and the design-based effect (13.55).

The second factor is the estimate of misuse used in these calculations. The NSUBS definition is a simplified version of the definition to be used by the NCRUSS survey, which is similar to the misuse definition used in a 2003 NHTSA misuse study that found the rate of misuse to be 0.726. If this rate of misuse is substituted for p in (2), then the resulting desired sample size is 2503.

In order to ensure that an overall rate of misuses can be estimated with the desired accuracy, this study proposes to collect the maximum number of observations allowed by funding; 3,000 child passengers from age 0-8.

S***ampling:*** The first stage of sampling (PSU level) has been completed, and details of the process can be found in the National Automotive Sampling System – General Estimates System (NASS-GES) Analytical User's Manual (http://www-nrd.nhtsa.dot.gov/Pubs/AUM08.pdf). Briefly, the country is divided into 1195 PSU's (geographic areas) which are stratified by type (large central city, large suburban area, and all others). 24 PSU's are then selected using PPS with the number

of reported car accidents within each PSU as the measure of size. PSU weights are available in the NASS GES data files.

Site Sampling: The second stage of sampling (site) will be accomplished through joint effort by the statistical consultant, the data collection consultant, and the NHTSA research team. The sites for survey will be stratified by type of establishment, allowing researchers to focus on sites that are likely to provide access to child passengers. The types of sites that will be eligible for survey are: large discount stores, national chain fast food restaurants, gas stations, daycare centers, and recreation centers. A sampling frame will be constructed that contains all of the eligible sites within each PSU. This will be done by the statistical contractor using a NAVTEQ software package that is continually updated from several sources.

The types of sites selected for the study were based on several factors including those relating to safety of the data collectors, candidate drivers and child occupants, site or community cooperation and efficiency of data collection (i.e., volume of child passengers). This selection was made based on information in a contracted report from Westat, a statistical contractor with experience in child safety seat surveys that described benefits, impediments, and past experiences of different site types.

Site or community cooperation is critical as well. Permission will be gained from candidate site property managers or owners. For sites under community jurisdiction, government and police agencies will be notified.

Assuming 1.58 children per interviewed driver (results from 2009 NSUBS) and a 50% vehicle response rate (NSUBS had a 75% rate, but this study is predicted to show a lower response rate due to longer interview times and the need to manually inspect child seats), we will need approximately 3,800 interview attempts to reach the desired sample size of 3,000 child passengers. Each individual site will be used for one four-hour data collection period. NSUBS was able to administer 14 vehicles interview attempts per two-hour site observation period, and assuming that the NCRUSS observations will take about half again as long to conduct (based on pilot data) we will need approximately 270 participating sites each with three hour observation blocks in order to reach our target sample size (270 sites * 14 vehicle attempts * 0.5 response rate * 1.58 children per vehicle = 2986).

The 270 sites will be distributed equally across the PSUs, resulting in 11 sites being selected for 18 of the 24 PSUs and 12 sites being selected for the remaining six PSUs. The six PSUs with 12 sites will be randomly selected. The sites selected within each PSU will be a stratified random sample to ensure inclusion of less common site types. The proposed site type strata are:

- Large discount stores
- Gas stations
- National chain fast food restaurants
- Daycare centers and recreation centers (combined to ensure an adequate frame to sample from)

Other site types are still under consideration for inclusion and may be added to the design. The site type strata will be sampled using the following sampling plan to ensure inclusion of less common site types:

- Two large discount stores will be selected randomly per PSU
- Two daycare centers or recreation centers will be randomly selected per PSU
- The remaining seven to eight sites will be selected randomly from the fast food restaurants and gas stations with relative frequency determined by proportion of frame count.

The site selection probabilities will require the establishment of some notation. Let $1 \leq i \leq 24$ and $1 \leq j \leq 4$ denote integers. Let M_{ij} denote the total number of sites in the sampling frame for stratum j of PSU i, and let $1 \leq k \leq M_{ij}$ be the k^{th} site in the j^{th} stratum of the i^{th} PSU.

The initial sample of 270 sites will be selected as described above, and a supplemental sample will be taken to account for site refusals by taking the next member of the sorted sampling frame following the sites selected in stage 1. Let the number of initially sampled sites that are included in the final sample from the j^{th} stratum of the i^{th} PSU be denoted by m_{1ij}, and let the number of supplemental sites included in the final sample from the j^{th} stratum of the i^{th} PSU be denoted by m_{2ij}

Using this notation we can represent the probability that site k in stratum j of PSU i is included in the NCRUSS probability sample using the following formula:

$$P'_{ijk} = \delta_i \frac{m_{1ij} + m_{2ij}}{M_{ij}} \; for \; 1 \leq k \leq M_{ij}$$

where δ_i denotes the selection probability for the i^{th} PSU (inverse of the PSU weight from the NASS database).

Vehicle Sampling: The third stage of sampling (vehicle) will be conducted on site by the data collection researchers. It is not likely that the researchers will be able to approach every vehicle containing a child occupant in the population of interest. This level of selection will be largely based on convenience of the data collection researchers, although it will be considered pseudo-random and unlikely to bias any estimates. There will be a non-response bias adjustment at this level to account for missed vehicles. A census of vehicles will be taken at during the data collection period to provide estimates of total traffic volume during the collection period.

Let N_{ijk} be the total number of vehicles containing at least one child passenger under the age of 12 that enters the k^{th} site in the j^{th} stratum of the i^{th} PSU during the four-hour observation period and let $1 \leq l \leq N_{ijk}$ be an integer. Let n_{ijk} be the subset of N_{ijk} that is successfully sampled by the research staff. The vehicle (driver interview) weight of the l^{th} vehicle can then be given by the following formula:

$$P'_{ijkl} = \delta_i \frac{m_{1ij} + m_{2ij}}{M_{ij}} \frac{n_{ijk}}{N_{ijk}} \; for \; 1 \leq k \leq M_{ij} \wedge 1 \leq l \leq N_{ijk}$$

Child Sampling: The fourth stage of sampling will be conducted in cases of multiple children within a single vehicle and will not affect the weighting of the driver interview data. Therefore, two sets of weights will be applied to the final data, one for the driver and one for the individual child response.

It is assumed that at least two children can be observed while the interview is taking place, but in cases where there are multiple children in a vehicle it will be necessary to sample them using a random number table to determine order of observation in anticipation of insufficient time to observe each child. Data collectors will be trained in the use of the random number tables to ensure a non-biased sample, and a new random number table will be generated for each observation site.

If we let q_{ijkl} denote the number of children in the l^{th} vehicle at the k^{th} site in the j^{th} stratum of the i^{th} PSU that are successfully observed out of the total Q_{ijkl} eligible children in the vehicle and let $1 \le r \le Q_{ijkl}$ be an integer, then the child (observation) weight can be given by the following formula:

$$P'_{ijklr} = \delta_i \frac{m_{1ij} + m_{2ij}}{M_{ij}} \frac{n_{ijk}}{N_{ijk}} \frac{q_{ijkl}}{Q_{ijkl}} for\ 1 \le k \le M_{ij}, 1 \le l \le N_{ijk}, 1 \le r \le Q_{ijkl}$$

There will be an implicit sampling level for the time of day and day of week of data collection. Collection will take place 7 days a week, likely in three hour intervals from 7:00 AM to 7:00 PM. Effort will be made to balance collection at the site type level across time of day and day of week, with due consideration of traffic at these sites. For example, nearly all of the traffic at a daycare site will be in the morning and late afternoon, and therefore data collection teams will not be sent to these sites during the middle of the day. Adjustments to the observation weights may be made to reflect any differences in length of collection intervals at individual sites.

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

Data will be collected by teams of two researchers consisting of a trained NASS field researcher and a certified Child Passenger Safety technician. The NASS researcher will approach the driver and administer the driver interview. In cases of refusal, the NASS researcher will record information about the vehicle and the driver for non-response analysis. The CPS technician will conduct the child passenger observations while the driver interview is taking place. Observations will take place in a designated area in the parking facilities of the site.

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

Several aspects of the data collection are designed to maximize the response rate. The design of the survey using two-person data collection teams allows it to be completed in approximately 15 minutes for each vehicle depending upon the type and number of child restraints that are used in the vehicle. Candidate drivers will be approached in a non-threatening way and directed to a

highly visible area of the parking lot by a professional data collection staff. All data collectors will be issued badges identifying them as Federal contractors.

The response rate will likely be a concern as a potential source of bias. Because of this concern, adjustments will be made to the final observation weights using standard unit nonresponse. The NCRUSS will have three type of response to adjust for (s_1, s_2, s_3) each defined over a different type of unit (u_1, u_2, u_3).

- u_1 = eligible site, s_1 = participated
- u_2 = eligible vehicle s_2 = observed
- u_3 = child occupant s_3 = observed

The data structure is hierarchical, and the nonresponse adjustment factors will be computed using the methodology of the NSUBS, which had similarly structured data. The details of the methods can be found in the NSUBS methodology report (http://www-nrd.nhtsa.dot.gov/Pubs/811111.pdf, Section 8.2: Nonresponse Adjustment Factors).

Logical imputation will be performed where needed and possible (for example, age may be imputed from birth date and collection date). Implicit or probabilistic imputation will be conducted by a statistical consultant if it is deemed necessary.

Variance estimation will be done using SAS PROC SURVEYFREQ. This procedure allows for detailed specification of the sampling structure.

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

Data collection forms and instructions are being developed by staff in NHTSA's Office of Regulatory Analysis and Evaluation and NHTSA's Office of Behavioral Safety Research with input from the NASS Field Staff. The data collection forms and procedures will be evaluated in a pilot test of all data collection procedures to be conducted at test sites in each PSU. At numerous locations, a two-person team consisting of contractor staff will pretest the data collection forms and methodology, as part of a one day pretest. This pretest should help refine procedures, forms, and cooperation for the start of the full study. Minor modifications to the data collection forms, as well as some changes in procedures may result. The data collection forms are included as Attachment C in Part A.

B.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.

Survey Design and Estimation: Bob Sivinski, National Center for Statistics and Analysis (NCSA), NHTSA (NVS-431; 202.366.2570)

Data Analysis:	Bob Sivinski, NCSA, NHTSA (NVS-431, 202.366.9308) Charlene Doyle, NCSA, NHTSA (NVS-431, 202.366.1276) Angela Eichelberger, NHTSA (NTI-132, 202.366.5586)
Data Collection:	Gary Toth, NCSA, NHTSA, COTR for Contracts (NVS-411, 202.366.5378)
Data Collection Contractors:	Calspan Corporation. (Contract DTNH22-06-C-00022) KLD Associates, Inc. (Contract DTNH22-06-C-00024).