SUPPORTING STATEMENT FOR THE NATIONAL SURVEY OF THE USE OF BOOSTER SEATS

OMB Clearance Number: 2127-0644

Part B: COLLECTIONS OF INFORMATON EMPLOYING STATISTICAL METHODS

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

The purpose of the survey is to gather information on the restraint use, and in particular, the use of booster seats, among children ages 4-7. The survey will visit gas stations, recreation centers, and five specific fast food restaurants (McDonald's, Taco Bell, Burger King, Wendy's, and Kentucky Fried Chicken). Data collectors will approach as many vehicles as possible that appear to have at least one child occupant under the age of 13 in order to allow for data collector visual misestimating of ages and try to ensure that as many children ages 4-7 are captured.

In this sense, the potential respondent universe consists of all child motorists (age 12 and under) who frequent gas stations, recreation centers, and five specific fast food restaurants (not located in shopping centers).

These site types (gas stations, recreation centers, and fast food restaurants (those not located in shopping centers)) were chosen because they are frequented by child motorists and because their parking lots are usually sufficiently small that data collectors can likely approach vehicles as they are parking, before child restraints have been unfastened.

Data collectors will approach as many motorists as possible who appear to have at least one child under the age of 13 in their vehicle for possible participation in the survey.

2. Describe the collection of information procedures.

The sample design information presented below is summarized from the publication: *The 2006 National Survey of the Use of Booster Seats—Methodology Report*, NHTSA Technical Report DOT HS 811111, 2009.

Sampling Frame
The sampling frame for the first stage of the NSUBS design consists of the 50 sample Primary Sampling Units (PSUs) used by the NOPUS in 2005, the time of the NSUBS design. The NSUBS PSUs were drawn from the available NOPUS sample in 2005. It was envisioned at the beginning to select 24 PSUs, which were considered affordable and needed to meet the survey goals. However, the PSU sample size was cut back to 16 due to further restriction in the survey budget. NOPUS PSUs were re-selected in 2006, and the multi-year transition was completed in 2010 but NSUBS was not redesigned

1

due to budgetary reasons. Nevertheless, it is NHTSA's opinion that the old PSU sample on which NSUBS is based has served well NHTSA's data needs given the limited budget available to the agency. NSUBS is an observational study, and requires building a substantial infrastructure of planning and survey site cooperation to conduct the survey. Reselection of the PSUs has an enormous implication to the survey budget. NHTSA has no choice under the current circumstances and intends to use the same sample until enough funds can be secured for reselection of PSUs from the current refreshed NOPUS PSUs.

For documentation on the NOPUS PSUs and how they were selected, see Glassbrenner, September 2002 listed in the reference section. In essence, the NOPUS PSUs, which consist of counties and groups thereof, were selected as a stratified PPS (probability proportional to size) sample, using vehicle miles traveled (VMT) as the measure of size. The strata used in the selection were based on four geographic regions (Northeast, Midwest, South, and West), and whether or not the county or group of counties comprises a Metropolitan Statistical Area (or MSA, as defined by the Office of Management and Budget) (OMB, 2005).

Primary Sampling Unit Selection
Sixteen PSUs were chosen from the sampling frame via the following three-step process:

1. Two NOPUS PSUs in the frame were identified with certainty because of their population density (total population density). [*Note: These two PSUs have a very high population density, so they also have a much larger number of booster-seat-age children compared to other PSUs. If a PPS sampling design with the number of booster-seat-age children as the measure of size had been used, they would have been selected as certainty anyway.*] This is the main reason why they were selected as certainty PSUs. An additional 22 NOPUS PSUs were selected from the remaining 48 NOPUS PSUs as an equal-probability systematic sample, with the 48 NOPUS PSUs sorted by the following three variables: whether or not the State containing the NOPUS PSU had (in 2005 at the time of the NSUBS design) a law requiring some children to be restrained in booster seats in at least some circumstances; whether the PSU lies in a Metropolitan Statistical Area; and the census region. (Each of the 48 NOPUS PSUs lies entirely within a single MSA, a single census region, and a single State or the District of Columbia). The PSU sample was selected in 2005, and the same sample has been used and is proposed for use in this application. Therefore, sorting by the booster seat law status was based on the 2005 information, not current information. The sorting order was first by the booster seat law status, then by MSA status within each of the booster seat law statuses, and lastly by the census region within each category of the first two sorts. Please note that each unit has the same probability of selection no matter how the list is sorted.

2. Fourteen NOPUS PSUs were selected from the 22 NOPUS PSUs not chosen with certainty in Step 1 as an equal-probability systematic sample, with the 22 NOPUS PSUs sorted by the first two sort variables from Step 1 (namely, whether or not the State containing the NOPUS PSU had a booster seat law; and whether the NOPUS PSU lies in an MSA). Selection of 14 PSUs out of 22 (cocn-certainty) NOPUS PSUs was operationalized as follows:

    1. *The 22 NOPUS PSUs were sorted by the sort variables - let them be numbered 1, 2, …, 22 in the sorted list.*
    2. *Since the systematic sampling interval is 22/14 (= 1.57143), select a random number (R1) between 0 and 1.57143. Let S1 be the upper integer (ceiling) of the random number R1 (e.g., S1 = 1 if 0 < R1 <= 1 or S1 = 2 if 1 < R1 <= 1.57143). Then select S1.*

3. Let R2 = R1 + 1.57143 and S2 be the ceiling of R2. Then select S2 into the sample.

4. Let R3 = R2 + 1.57143 and S3 be the ceiling of R3. Then select S3 into the sample.

5. Continue this until the 14th PSU (i.e., S14) is selected.

6. Under this sampling scheme, every unit has the same probability of selection of 14/22.

3. Each of the 14 NOPUS PSUs from Step 2 and the 2 NOPUS PSUs selected with certainty in Step 1 was partitioned into county groups, where each county group consisted of a single county or two neighboring counties. The partitioning was conducted subjectively, motivated by reducing data collection costs in NOPUS PSUs that cover a wide geographic area.   In total, 43 county groups resulted from the partitioning of the 16 NOPUS PSUs. A single county group was selected from each of the 16 partitioned NOPUS PSUs via PPS sampling, with the population of children under age 5 according to the 2000 Census (again, the 2000 Census was used rather than the 2010 Census because the sample was selected in 2005 and it has not changed) as the measure of size. The 16 county groups resulting from these selections are the sample PSUs for the NSUBS survey. The design was completed in 2005 and was focused on the 4-7 year olds as the target population. We assumed that many of the children in the 0-5 age group identified in the 2000 census would be within the target population five years later.

Thus a total of 16 PSUs was selected for the NSUBS, with each PSU consisting of a single county or two neighboring counties that lie geographically within a NOPUS sample PSU.

Please note that consistent with our use of the phrase "PSU" in this report, the phrase "sample PSU" (e.g., "the 16 sample PSUs") shall refer to the 16 NSUBS PSUs selected in Step 3 above, and not the NOPUS sample PSUs.

The reason Step 2 was implemented instead of simply selecting 16 PSUs from the NSUBS sampling frame via systematic sampling, is because NHTSA initially envisioned using 24 PSUs, a decision later changed because of budget constraints.   (Alternatively, and roughly equivalently, we could have disregarded the 24-PSU result of Step 1 and re-applied Step 1 to select 2 certainty and 14 noncertainty PSUs.)

To best ensure that the data collected at the sites reflects the actual behavior of motorists, NHTSA does not release the locations of the 16 NSUBS (or even the NOPUS) PSUs.

Note that there is an implicit first stage of selection in the selection of the NSUBS PSUs, namely in the selection of the NOPUS sample PSUs.   As mentioned above, please see Glassbrenner, 2002, for documentation on the selection of the NOPUS PSUs.   The site selection probabilities for the NSUBS sample sites will contain a term reflecting the NOPUS PSU selection.

The NOPUS PSUs were used to select the PSUs for the National Survey of the Use of Booster Seats, motivated by greater comparability of the results of the two surveys. We note that the NOPUS has adopted a new sample since the time the NSUBS PSUs were selected, and thus NHTSA may wish at some point in the future to reselect the NSUBS sample from the current NOPUS sample for the same reason.

A note on terminology: The reader will note that the NSUBS sample design is technically a three-stage design, as the NSUBS "PSUs" are selected in two stages, Step 3 consisting of the second stage. However as a matter of terminology, we find it convenient to call the county groups from which the

NSUBS sites were selected "PSUs" instead of "SSUs."

Site Selection within PSUs

Site Sampling Frame
The sampling frame for the second stage of sampling consists of:

- the daycare centers in the 16 sample PSUs (i.e., the 16 PSUs selected in Step 3 of Section 4.1.2), together with

- the recreation centers, gas stations, and restaurants in five fast food chains in the collection of

ZIP Codes contained in whole or in part in the 16 sample PSUs

that were found in a process described below to meet the following four restrictions:
1) the establishment was not on a military base and not in an office building;
2) if the establishment was not a gas station, the establishment was not located in a shopping center;
3) the recreation centers did not merely contain a park, climbing wall, or senior center; and
4) the daycare centers were licensed for at least 20 children.

We call the above four restrictions the *site sampling frame restrictions*. The site sampling frame restrictions are geared toward reasonably high traffic volume and a parking area conducive to intercept sampling used for NSUBS. We used Google Earth to review the location and parking availability for sampled data collection sites. Urban locations without dedicated parking facilities are not eligible, because there are no locations at which restraint use can be observed.

We recognize that ZIP code is not a particularly well-defined geography that doesn't necessarily correspond to neat, contiguous areas nor does it necessarily respect county or other political boundaries. The 2009 NSUBS methodology document acknowledges that NHTSA experienced some problems of ZIP/PSU mismatch. NHTSA proposes to address possible ZIP/PSU mismatches for future NSUBS efforts by geocoding the sites. Sites falling outside the county boundaries will be removed from the sample.

Selection of Probability Sample of Sites

Selection of the Probability Sample of Sites
Sites were selected in a three-step process. Initially a sample of 323 sites was selected via stratified systematic sampling (described in detail in the following). However in anticipation of businesses declining allowing the survey to be conducted on their premises, an additional 302 sites were selected from the remaining sampling frame. Finally, two sites were added for reasons specified below, yielding a total of 627 sites.

Step 1: The selection of 323 sites
Initially, a target sample size of 20 sites per PSU was set, except in one PSU that was set to have 23 sites. (See the Section 4.4 for how the target sample sizes were developed.)

The target sample size of 20 or 23 sites per PSU was allocated across strata as follows. The designated stratum sample sizes for daycare centers and recreation centers were in all but 5 PSUs set to be 2 for each. The numbers of daycare and recreation centers were generally significantly smaller than those of fast food restaurants and gas stations, thus a proportional allocation would have

resulted in very small sample sizes.[3] A sample size of 2 was decided upon in these cases. The remaining sample size in the PSU (generally 16) was allocated to gas stations and fast food restaurants in proportion to their frame counts.

The stratum sample sizes in each of the 16 PSUs having been determined, the 323 sites were chosen as a stratified systematic sample in each PSU, with the sites in a given stratum of a given PSU sorted as follows:

- Fast food strata in which more than 20 percent of the stratum members straddle two adjacent counties and that have more than 25 members were sorted by chain name;

- Gas station strata in which more than 20 percent of the stratum members straddle two adjacent counties and that have more than 25 members were sorted in random order; and

- All other strata were sorted by ZIP Code.

Sorting by ZIP Code ensures good geographic dispersion, and is preferred for this reason. However because of our frame sources and sampling methods for fast food restaurants and gas stations, sorting these strata by ZIP Codes could result in selecting an undesirably large number of sites that lie outside the 16 PSUs, and thus the alternative sorts were used. The specific sampling locations and types of fast food restaurants (restaurant chain names) are not included in the published NSUBS reports in order to protect these data sources.

Step 2: The selection of an additional 302 sites
The supplemental sample was formed by taking the next member in the sorted frame following each of the selected 323 sites in the initial sample (or in the case in which the initially selected member is the last member of a stratum, we chose the penultimate member of the stratum). The supplemental sample contained fewer than 323 members because in some cases the "next member" was a member of the initial sample.

Step 3: The selection of two additional sites
Two sites were inadvertently included in the sample. As we will document in Section 4.3, these sites were treated as second-stage certainties in weighting.

We note that although Step 3 in the selection of PSUs from Section 4.1.2 involves a subjective process, the sample of 627 sites (with the exception of the two additional sites from the previous paragraph) is a probability sample, since the subjective process involved only the sampling frame formation, not the selection of PSUs (or sites).

The number of sites by site type in the NSUBS sample: 53 recreation centers, 75 daycare centers, 201 fast food locations, and 298 gas stations.

The site types used by the National Survey of the Use of Booster Seats (gas stations, recreation centers, and the five fast food restaurants) were chosen because they are frequented by child motorists and because their parking lots are sufficiently small that data collectors can likely approach vehicles as they are parking, before child restraints have been unfastened. NHTSA acknowledges that the elimination of fast food restaurants in shopping centers (those without parking lots) does introduce bias, but this is necessary to interview children and observe their restraint use. In recent months, sampling sites were reselected PPS, including provisions for overlap. This was done to reduce the discrepancy in weighting among site types.

Systematic sampling was done by selecting a random number between 0 and the upper limit of the sampling interval, which is the number of sites in the sorted list divided by the allocated sample size. Determining which sites are selected was done in the same way as was explained in the response to Question 1. This sampling scheme is an equal probability sampling method however the list is sorted. Therefore, any fast food site sorted by chain name had the same chance of selection. However, the systematic sampling plan creates implicit strata by chain with proportional allocation, that is, the number of sites selected from a particular chain is proportional to the size of the chain within the PSU. In systematic sampling, sorting by the sort variables creates implicit stratification (which is the main purpose of sorting), so if ZIP Code is used as a sort variable, it is like stratifying the sampling list by ZIP Code and selecting a sample with proportional allocation. If simple random sampling is used instead of systematic sampling, all sites could be selected from one ZIP Code, or a few ZIP Codes receive a much larger or smaller proportion of the sample by chance but in systematic sampling, the sample is proportionally disperse over the ZIP Codes (i.e., geographically disperse because ZIP Codes are geographic codes.).

In systematic sampling, any sample set is a systematic sample. For NSUBS, two sets of systematic samples were selected, one for the regular sample and the other as a supplemental sample. The sampling scheme can be explained by an example more easily. Suppose that there are 10 sites, out of which 2 sites are selected, and the regular sample consists of sites 2 and 6. Then the supplemental sample consists of sites 3 and 7 (next members of the initial sample). However, if the initial sample is 5 and 10, then the supplemental sample is 4 and 9 (penultimate members of the initial sample). Occasionally two adjacent sites can be selected into the initial sample, and the "next" member is already included in the initial sample, so the supplemental sample may get a smaller number of sites than the initial sample. This happens especially when the frame size is small and the sampling rate is high. For example, suppose that 2 sites are selected out of 3, and the initial sample consists 1 and 2. Then the supplemental sample is supposed to contain 2 and 3 but 2 is already in the initial sample, so the supplemental sample gets only one site. This happened mainly for daycare centers and recreation centers. The site sampling plan explained in Part B was used for the first NSUBS survey. Subsequent surveys used cleaned and refreshed site frames (within PSUs), so there was no need of selecting supplemental samples.

Data Collection Schedule
Data collectors will visit each site for approximately 4 hours. The data collection schedule was set to take advantage of time periods during which child motorists are more likely to visit the sites. Fast food site visits will only be scheduled between the hours of 10 am and 2 pm on weekdays and 10 am – 6 pm on weekends. Recreation centers will only be visited between 10 am and 4 pm on weekdays and weekends. Gas stations will only be visited during morning and evening rush hours (7-9 am and 4-6 pm) on weekdays, and between 10 am and 6 pm on weekends. The specific schedule of site visits, in terms of which team will visit which site at which time, will be set to yield efficient collection of data given the site locations.

NHTSA acknowledges that the clustering of the observation periods according to the schedule described above will introduce time bias, but this is done in the interest of data collection efficiency to maximize the number of observed occupants and to satisfy the budget constraint of the survey. Gas stations are visited throughout the day by the researchers. However, traffic volumes, and thus gas

station visits are more frequent during rush hours. Parents may be transporting children for day care, meals, recreation activities, etc.

Estimated Yield
We estimate that based on the number of hours of data collection, data collectors will approach approximately 4,800 vehicles. Based on previous iterations of the NSUBS conducted in 2006, 2007, 2008, 2009, and 2011 there are an average of 179 interviews in any four-hour observation period.

Information Collected
If an adult in the approached vehicle agrees to participate in the survey, data collectors will ask the adult for the following information:

- Age of each occupant
- Children's heights
- Children's weights
- Child's race and ethnicity
- How many hours in the last week has each child spent in the vehicle?
- How many times in the last week has each child been to this type of site (e.g. gas stations)?

In addition, data collectors will collect the following information by observation only, and not by interview:

- Date
- Time
- Survey site
- Site type (e.g. gas station, fast food restaurant, etc)
- Vehicle type
- Seating position of each occupant
- Restraint use for each occupant, specifying the types of child restraints used
- Gender of each occupant

Data Collection Form
The data collection form to be used by the survey is attached at the end of this document.

Data collectors for this survey will fill out page 1 of the form (site type, weather, etc) when they arrive at each data collection site. Data collectors will then fill out one copy of pages 2 and 3 of the form for each vehicle whose occupants agree to participate in the survey.

On page 1 of the form, the PSU and site numbers are identification numbers for the survey site assigned by the survey. The "booklet" consists of the entire package of forms filled out by the data collector on a given day. For information on "misses and refusals", please see the explanation below explaining page 4 of the form.

Data collectors will recite the text on page 2 of the form "Hi, my name is ___ ..." to each potential respondent to ensure them that their participation is voluntary..

For motorists who agree to participate, data collectors will interview an adult motorist in a vehicle containing children for answers to the questions on the form, and fill out the form's information on restraint use based on observing the children in the vehicle.   One member of each team of two data collectors will conduct and record the interview data and the other data collector will observe and

record the observed data on restraint use. The part of the form with the pictures of signs from fast food restaurants where data collectors will record information from the last question - on the number of times the motorists have visited each of those establishments in the past week.

Information on page 2 on the vehicle type, time, and vehicle number will be filled out by the data collector. (S/he will not ask the motorist for this information.) The vehicle number simply reflects the number of vehicles the data collector has observed so far, i.e. the first vehicle, second vehicle, etc. It does not reflect any identifying information for the vehicle, such as the Vehicle Identification Number or license plate.

Data collectors will keep track of the number of vehicles that they missed and the number of vehicles whose occupants declined to participate in the survey and record these counts on page 4 of the form when they leave the site.

Data collectors will receive extensive training in protocols for interviewing motorists and observing restraint use in a manner that is professional and as unobtrusive as possible.

The PRA disclosure statement is printed on the data collection booklet as a convenience for the data collector. Respondents are not given flyers and no assurances of confidentiality are made to respondents. However, data collectors may inform respondents that no personally identifying information is collected in the survey.

Some tests were conducted in 2005 with regard to incentives as part of the methodology development. No incentives are needed, nor are any provided to respondents.

Statistical Editing, Imputation, Estimation, and Variance Estimation
Simple range edits will be performed on the data to improve data quality. For instance the data will be edited to ensure that children's ages fall between 0 and 18 years. Data that fall out of range will be treated as missing.

Regarding the interview data, we do not expect many missing values for the ages of children, as the interviewed adult motorist will most likely know the ages.   We will not impute for missing values of the remaining interview variables (children's weights and heights and time spent in the vehicle or at the site type) as there would not seem to be a good basis for forming reasonable imputed values.

Restraint use will be estimated by the following:

$$\text{Booster Seat Use} = \frac{\sum_{i=1}^{16} \sum_{j=1}^{4} \sum_{k \in \text{RefSamp}_{ij}} w_{ijk} F_{ijk} B_{ijk}}{\sum_{i=1}^{16} \sum_{j=1}^{4} \sum_{k \in \text{RefSamp}_{ij}} w_{ijk} F_{ijk} O_{ijk}}$$

Where RefSamp_{ij} denotes the collection of members of the refined sample that are in stratum j of PSU i; F_{ijk} denotes the product of various adjustment factors; B_{ijk} denotes the number of children 4 to 7 in booster seats observed at site k in stratum j of PSU i, and O_{ijk} denotes the total number of children 4 to 7 observed at site k in stratum j of PSU i. For a complete discussion of estimation in the NSUBS, please refer to Glassbrenner 2009, specifically Chapter 8 (Estimation).

The estimates generated by this formula will be reported as estimates of restraint use by children (of the age range in question) at gas stations, recreation centers, and the five specific fast food restaurants (in all locations except those with shopping centers). NHTSA does not feel that these estimates can be extrapolated to estimates reflecting all child motorists (i.e. to account for children who do not frequent these site types).

Variance estimates will be computed using WesVar, which utilizes a jackknife variance methodology.

3. Describe the methods used to maximize response rates and to deal with issues of nonresponse.

The refusal rate by year is described in the following table:

Year	Total Number of Observed Vehicles	Total Number of Interviews	Total Number of Refusals	Refusal Rate
2006	3489	2920	548	15.7
2007	4828	4199	181	3.7
2008	6204	4899	224	3.6
2009	6033	4601	286	4.7
2011	6350	5191	300	4.7

We do not expect many missing values in the observed portion of the data (both the site information on page 1 of the data collection form and the observed motorist data on pages 2-3) because the data collectors will be well trained and they should have adequate time to record site information and restraint use.

Regarding the interview data, we do not expect many missing values for the ages of children, as the interviewed adult motorist will most likely know the ages. We will not impute for missing values of the remaining interview variables (children's weights and heights and time spent in the vehicle or at the site type) as there would not seem to be a good basis for forming reasonable imputed values.

Data collectors for the National Survey of the Use of Booster Seats will undergo extensive training in order to minimize errors that could arise from their categorizing or recording data incorrectly.

NHTSA does not believe that there is reliable information with which to adjust the survey results to account for inaccurate responses given by motorists, motorists who choose not to participate in the survey, motorists who do not frequent the site types, or motorists who frequent the site types outside of the observation period. The Agency's published report will clearly state that the results are based on motorists who visit the site types (again, except those 5 specific fast food restaurants located in shopping centers) and voluntarily chose to participate in the survey.

4. Describe any tests of procedures or methods to be undertaken.
No tests of procedures or methods are planned.

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.**

This survey was designed and will be conducted under Federal Contract with Westat, Inc. The Contracting Officer's Technical Representative is Mr. Timothy M. Pickrell and can be reached at (202) 366-2903. The program manager at Westat, Inc. is Ms. Fran Bents who can be reached at (240) 314-7557.