



U.S. Department
of Transportation
**National Highway
Traffic Safety
Administration**



DOT HS 811 111

April 2009

NCSA Technical Report: The 2006 National Survey of the Use of Booster Seats — Methodology Report

DISCLAIMER

This publication is distributed by the U.S. Department of Transportation, National Highway Traffic Safety Administration, in the interest of information exchange. The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Department of Transportation or the National Highway Traffic Safety Administration. The United States Government assumes no liability for its contents or use thereof. If trade names, manufacturers' names, or specific products are mentioned, it is because they are considered essential to the object of the publication and should not be construed as an endorsement. The United States Government does not endorse products or manufacturers.

1. Report No. DOT HS 811 111		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle The 2006 National Survey of the Use of Booster Seats – Methodology Report				5. Report Date April 2009	
				6. Performing Organization Code NVS-421	
7. Author(s) Glassbrenner, Donna, Ph.D.				8. Performing Organization Report No.	
9. Performing Organization Name and Address Mathematical Analysis Division, National Center for Statistics and Analysis National Highway Traffic Safety Administration U.S. Department of Transportation, NVS-421 1200 New Jersey Avenue SE. Washington, DC 20590				10. Work Unit No. (TRAVIS)	
				11. Contract or Grant No. DTNH22-05-D-01002	
12. Sponsoring Agency Name and Address Mathematical Analysis Division, National Center for Statistics and Analysis National Highway Traffic Safety Administration U.S. Department of Transportation, NVS-421 1200 New Jersey Avenue SE. Washington, DC 20590				13. Type of Report and Period Covered NHTSA Technical Report	
				14. Sponsoring Agency Code	
15. Supplementary Notes The 2006 National Survey of the Use of Booster Seats was conducted by WESTAT, Inc., under Federal contract with NHTSA. A number of portions of this report are paraphrasings of WESTAT's training material and final report to NHTSA from the 2006 survey. We thank Dr. Eun Young Noh of URC Enterprises, Inc., for her assistance in preparing this report.					
Abstract The purpose of this report is to document the survey design used for the initial data collection of the National Survey of the Use of Booster Seats (NSUBS). The initial data collection occurred in 2006. Although this report is being published after the second data collection (which occurred in 2007) and as NHTSA is preparing for the third data collection (in 2008), this report serves as important documentation as the design used for subsequent NSUBS data collections were based on the design documented in this report, incorporating relatively minor changes in methodology. NHTSA expects to publish annual methodology reports that document subsequent changes in methodology.					
17. Key Words Child restraints, car seats, booster seats, survey methodology			18. Distribution Statement This document is available for free download at www.nhtsa.gov . Printed copies can also be purchased by contacting the National Technical Information Service, Springfield, VA 22161, or visiting www.ntis.gov .		
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 73	22. Price

TABLE of CONTENTS

1.	Summary	1
2.	The Circumstances That Gave Rise to the NSUBS	2
2.1	<i>Booster Age Children – An Area of Particular Concern</i>	2
2.2	<i>The Transportation Recall Enhancement, Accountability, and Documentation Act of 2000 (TREAD)</i>	2
2.3	<i>A Data Need</i>	2
3.	The Prior State of Knowledge - A History of Booster Seat Use Estimates	4
3.1	<i>What’s Difficult About Getting a Reliable Estimate of Booster Seat Use</i>	4
3.2	<i>Estimates of Booster Seat Use Prior to NSUBS</i>	4
3.3	<i>Considering an Optimal Survey</i>	5
4.	Sample Design	6
4.1	<i>Selection of Primary Sampling Units</i>	6
4.1.1	PSU Sampling Frame	6
4.1.2	Selection of PSUs	6
4.2	<i>Selection of Sites Within PSUs</i>	7
4.2.1	Site Sampling Frame.....	7
4.2.2	Selection of the Probability Sample of Sites	10
4.2.3	Post-Selection Refinements to the Probability Sample: Obtaining the “Refined Sample”	11
4.2.4	Post-Selection Process to Identify Duplicates in the Second Stage Frame	12
4.3	<i>Site Selection Probabilities</i>	13
4.4	<i>Sample Size Determination</i>	16
5.	Data Collection Protocols	19
5.1	<i>Obtaining Site Cooperation</i>	19
5.1.1	How Cooperation Was Obtained	19
5.1.2	Site Participation Rates for the 2006 Survey	19
5.2	<i>The Data Collection Schedule</i>	20
5.3	<i>Number, Gender, and Positioning of Data Collectors</i>	21
5.3.1	Number and Gender of Data Collectors	21
5.3.2	Positioning Data Collectors at the Site to Best Collect Data.....	21
5.4	<i>The Definition of Restraint Use Used by the Survey</i>	23

5.5	<i>The Categories of Race and Ethnicity Used by the Survey</i>	24
5.6	<i>The Survey Variables</i>	24
5.7	<i>The Wording of Interview Questions</i>	27
5.8	<i>Other Assorted Data Collection Topics and Definitions</i>	28
	5.8.1 <i>Items Provided to Data Collectors</i>	28
	5.8.2 <i>The Types of Vehicles Surveyed</i>	29
	5.8.3 <i>Keeping Track of Nonparticipating Vehicles</i>	29
	5.8.4 <i>Miscellaneous</i>	29
5.9	<i>Data Collection Protocols</i>	30
6.	<i>Quality Control Procedures</i>	34
	6.1 <i>Pilot Testing of Data Collection Protocols</i>	34
	6.2 <i>Recruitment of Field Staff</i>	34
	6.3 <i>Training</i>	34
	6.4 <i>Pre-Collection Test of Data Collectors</i>	35
	6.5 <i>Contact Information for Questions</i>	35
	6.6 <i>Unannounced Site Visits</i>	35
7.	<i>Data Entry, Editing, and Imputation</i>	36
	7.1 <i>Data Entry</i>	36
	7.2 <i>Editing</i>	37
	7.3 <i>Imputation</i>	37
	7.3.1 <i>Variables Imputed as Special Cases</i>	37
	7.3.2 <i>Variables Imputed by Logical Imputation</i>	37
	7.3.3 <i>Variables Not Imputed</i>	38
	7.3.4 <i>Variables Imputed by Hot-deck Imputation</i>	38
8.	<i>Estimation</i>	39
	8.1 <i>Estimator Design</i>	39
	8.2 <i>Adjustment for Variation in Duration of Data Collection</i>	41
	8.3 <i>Nonresponse Adjustment Factors</i>	41
	8.4 <i>Weight Trimming</i>	43
	8.5 <i>The Estimation Formula</i>	44
	8.6 <i>Estimates Computed</i>	45
	8.7 <i>Definitions of Categories Used in Estimates</i>	45
	8.8 <i>A Note on the Race/Ethnicity Estimates</i>	46

9.	Variance Estimation.....	47
10.	Rules for Suppressing Estimates in Publications.....	50
11.	Glossary of Terms.....	51
12.	Glossary of Notation.....	54
13.	References.....	58
14.	Appendix.....	60
14.1	<i>Data Collection Forms</i>	60
14.2	<i>Letter of Authorization</i>	61
14.3	<i>Incentives</i>	63
14.4	<i>Frequently Asked Questions</i>	63
14.5	<i>Card for Race/Ethnicity Questions</i>	65

TABLE of FIGURES

A Collection of Booster Seat Use Estimates.....	5
The NSUBS Probability Sample and Refined Sample	12
Desired Margins of Error	16
Achieved Margins of Error	17
Business Recruitment Results by Site Type	20
Location of Data Collectors at the Sites	22
Definitions Used for the Survey Variable “Restraint Used”	23
Survey Variables Collected.....	24
Wording of Interview Questions.....	27
Variables Imputed as Special Cases	37
Estimates Whose Variances Were Estimated Directly	47
Booster Seat Recording Form.....	60
Incentives	62

1. Summary

In 2006, NHTSA conducted the first-ever nationwide survey of booster seat use in the United States based on the observation of children in vehicles – the National Survey of the Use of Booster Seats, or NSUBS. The survey presented challenges in developing an appropriate sample design, data collection protocols, and statistical estimation.

The purpose of this publication is to present the choices made to address these challenges and fully document the design of the survey used for the 2006 data collection. Although this report is being published after the second data collection (which occurred in 2007) and as NHTSA is preparing for the third data collection (in 2008), this report serves as important documentation as the design used for subsequent NSUBS data collections were based on the design documented in this report, incorporating relatively minor changes in methodology. NHTSA expects to publish annual methodology reports that document the design used for any given data collection and identify any design-related changes made since the prior data collection.

The portions of this report on sample design, editing, nonresponse adjustment, estimation, and variance estimation are written for a statistical audience. A simpler description of the sample design that leaves out many of the details for the statistical audience may be found in Glassbrenner and Ye, 2007, and its annual updates on www.nhtsa.gov.

The NSUBS is conducted by the National Center for Statistics and Analysis (NCSA), an office of the National Highway Traffic Safety Administration. The survey design, data collection, data editing, nonresponse adjustment, and calculation of estimates and variances were conducted by WESTAT, Inc., under the direction of NCSA, via NHTSA contract number DTNH22-05-D-01002.

OMB approval was obtained for the collection of data for this survey. NHTSA obtained approval to collect data for the 2006-2009 surveys under OMB clearance number 2127-0644. The notice of OMB review can be found in the Federal Register, Volume 71, Number 30, page 7824, February 14, 2006.

2. The Circumstances that Gave Rise to the NSUBS

2.1 *Booster-Age Children – An Area of Particular Concern*

Great strides have been made in recent years in protecting child passengers. Among infants and toddlers, restraint use remains near 90 percent (98% for infants and 89% for children 1 to 3 years old in 2006), and crash-related fatalities for 0 to 3-year-old occupants dropped by 13 percent in 2005, compared to 2004. (Glassbrenner & Ye, February 2007; NHTSA, 2006)

Unfortunately, similar progress has not been achieved where older child passengers are concerned. Booster seat use -- estimated at only 10 to 20 percent nationwide¹ when the survey that is the subject of this report was conducted -- remains unacceptably low. According to NHTSA's Fatality Analysis Reporting System (FARS) and National Automotive Sampling System General Estimates System (NASS GES), in 2005 there were 346 fatalities among booster-age child passengers -- children between the ages of 4 and 7, inclusive -- as well as 49,000 injuries in this age group (NHTSA, 2006) Only 78 percent of children 4 to 7 were restrained in 2005, a 5-percentage-point drop since 2002, according to NHTSA's National Occupant Protection Use Survey (NOPUS) (Glassbrenner & Ye, February 2007; Glassbrenner, February 2005).

2.2 *The Transportation Recall Enhancement, Accountability, and Documentation Act of 2000*

In 2000, Congress passed the Transportation Recall Enhancement, Accountability, and Documentation (TREAD) Act of 2000. Section 14(i) of the act directs the Department of Transportation to reduce the deaths and injuries among children in the 4- to 8-year-old age group that are caused by failure to use booster seats by 25 percent. Conducting the National Survey of the Use of Booster Seats provides the Department with invaluable information on who is and is not using booster seats, helping the Department better direct its outreach programs to ensure that children are protected to the greatest degree possible when they ride in motor vehicles. In particular, the information collected in this survey support the Department of Transportation goal to improve safety in motor vehicle transportation.

Also in 2002, Congress enacted Public Law 107-318, known as Anton's Law, which contains additional provisions to improve the safety of child restraints in passenger motor vehicles, especially for older-child passengers. [Public Law 107-318, Dec. 4, 2002]

In the TREAD Act and Anton's Law, NHTSA was directed to conduct a range of initiatives, including rulemaking, compliance testing, and consumer education programs, to enhance the safety of older child passengers.

2.3 *A Data Need*

In order to adequately address the TREAD requirements, DOT needed data on who is and who is not using booster seats in order to target outreach programs.

¹ NHTSA estimated booster seat use to be in this range based on estimates from the Children's Hospital of Philadelphia (CHOP) in 2002 and NHTSA's Motor Vehicle Occupant Safety Survey (MVOSS) in 2003. See Partners for Child Passenger Safety, 2004, and Boyle et al., 2005, for more information on these estimates.

Previous estimates of booster seat use were not sufficiently reliable to use to effectively direct limited outreach resources. These estimates were obtained either using non-probability samples (and so the results might not be representative and one cannot measure the error in the estimates), or were obtained via telephone surveys (which could be subject to respondents' potential reluctance to report that their child was not in a booster seat). (See the next section for further information on these prior estimates of booster seat use.) What one would desire to adequately allocate limited resources for outreach programs would be a probability-based survey in which booster seat use is obtained by observation. This is what the NSUBS was designed to achieve. (See the design section for how the survey was designed to meet these goals.)

Thus the National Survey of the Use of Booster Seats is being conducted to respond to the Section 14(i) of the Transportation Recall Enhancement, Accountability, and Documentation (TREAD) Act of 2000.

3. The Prior State of Knowledge - A History of Booster Seat Use Estimates

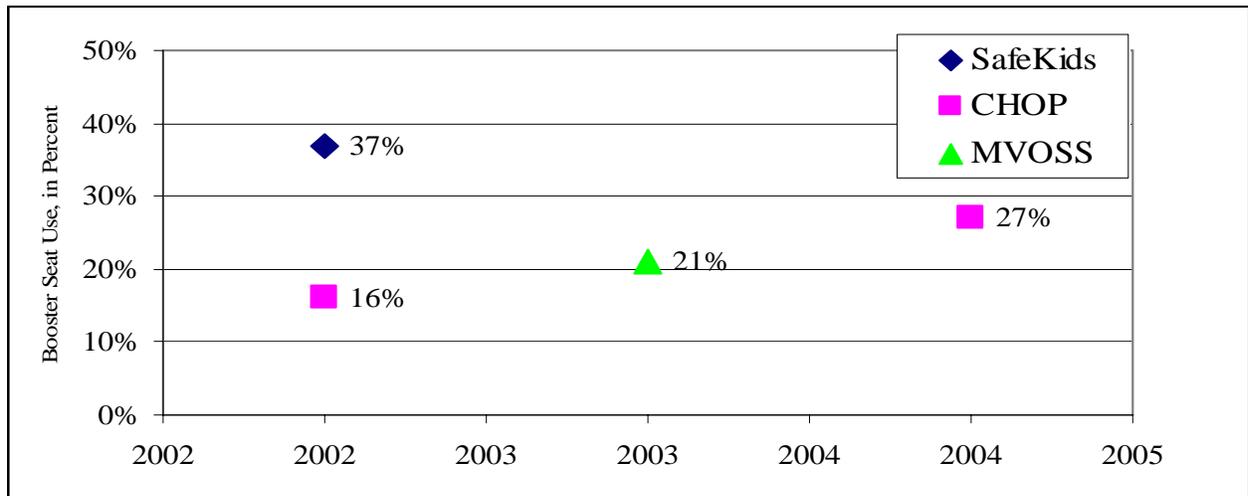
3.1 *What's Difficult About Getting a Reliable Estimate of Booster Seat Use*

Because of differences in reported use rates versus observed use rates (e.g., NHTSA's Motor Vehicle Occupant Safety Survey consistently finds reported belt use rates higher than the observed use rates in NHTSA's National Occupant Protection Use Survey), it is preferable to estimate booster seat use from a survey that observes vehicles on the road, rather than one that obtains its data from telephone interviews of drivers. However, observing booster seats presents a special challenge not encountered with, e.g., seat belts. Namely, one type of booster seats – backless boosters – cannot be reliably observed from the roadside. One can – and NHTSA's National Occupant Protection Use Survey does – produce observed estimates of high-backed booster seats of vehicles on roadways, but one cannot do the same for booster seat use per se, i.e., the percent of children using any type of booster seat. Thus, to estimate booster use, we are forced to take a different approach.

3.2 *Estimates of Booster Seat Use Prior to NSUBS*

Various approaches are possible, each with associated limitations. One can estimate use from crash data – as Children's Hospital of Philadelphia did using crashes of State-Farm insured vehicles. As mentioned above, one can estimate use from telephone surveys, as NHTSA has done in its Motor Vehicle Occupant Safety Survey (Boyle et al., 2005). However because crash-based estimates tend to underestimate use (as drivers in crashes might disproportionately engage in risk-taking behaviors) and because of the bias of telephone survey data discussed above, the best means of obtaining data with which to estimate booster use would be to observe usage up close through doors and windows of stopped vehicles (with the occupants' consent). Doing so at a probability sample of roadways and stopping vehicles via police checkpoints is prohibitively expensive - NHTSA estimates that such a survey would cost at least \$1,300,000 each time it is conducted. SafeKids handled the cost issues by conducting a survey at a convenience sample of sites where vehicles containing children tend to be, such as at fast food restaurants (Cody et al., 2002). The NSUBS does, in a sense, one better, by conducting such a survey at a probability sample of such sites.

A Collection of Booster Seat Use Estimates



Notes and Sources: SafeKids estimated the percent of 4- to-7-year-olds over 40 pounds in booster seats; source: Cody et al., 2002; CHOP estimated percent of 4- to-7-year-olds in crashes who were in boosters; source: Partners for Child Passenger Safety, 2004, 2005. MVOSS estimated the percent of 4- to-7-year-olds in boosters at least on occasion, via telephone interviews; source: Boyle et al., 2005.

3.3 *Considering an Optimal Survey*

In some sense the optimal survey would be one that captured vehicles in traffic at a probability sample of roadway sites. Setting aside the challenges of how to capture the vehicles without incurring an unsatisfactory degree of bias (e.g., considering using police checkpoints), we feel that the number of roadway sites one would need to collect data from would be cost-prohibitively high, given the relative incidence of booster-age children in general roadway traffic.

4. Sample Design

4.1 Selection of Primary Sampling Units

4.1.1 PSU 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.

For documentation on the NOPUS PSUs and how they were selected, see Glassbrenner, September 2002. 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).

Note in this report the term “PSU” without further modification shall refer to the NSUBS PSUs. When referring to the NOPUS PSUs, we shall say “NOPUS PSUs.”

4.1.2 Selection of PSUs

Sixteen PSUs were chosen from the sampling frame via the following three-step process. (The decision to choose 16 PSUs was motivated by variance constraints - See the below section on sample size determination for more information.)

Step 1: Two NOPUS PSUs in the frame were identified with certainty because of their population density. 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).

Step 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).

Step 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 as the measure of size. The 16 county groups resulting from these selections are the sample PSUs for the NSUBS survey.

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. For more information on the current NOPUS sample, see Glassbrenner, to appear.

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.”

4.2 *Selection of Sites Within PSUs*

4.2.1 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 chains² 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:

² The NSUBS includes among its sites restaurants in five fast food restaurant chains. In the interest of retaining these chains in future surveys, the names of the chains (which are known to staff working on the survey) are kept confidential in this report.

- 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*.

Formation of the Site Sampling Frame

The sampling frame was formed as follows:

- The daycare centers in the sampling frame were compiled from State and county lists of licensed daycare centers.
- The gas stations were obtained by searching for all gas stations in the PSU's ZIP Codes using the Web sites yellowpages.com and superpages.com. The search was implemented by typing the phrase "gas station" in the business type field and typing each of the PSU's ZIP Codes in the ZIP Code field of these Internet sites.
- The fast food restaurants were obtained by searching for all such restaurants in the PSU's ZIP Codes using the Web sites yellowpages.com and superpages.com. The search was implemented by typing the names of each of the five fast chains in the business type field and typing each of the PSU's ZIP Codes in the ZIP Code field of these Internet sites.
- The recreation centers were obtained by compiling State and county lists of recreation centers with the list obtained by searching for all recreation centers in the PSU's ZIP Codes using the Web sites yellowpages.com and superpages.com. The search was implemented by typing the phrase "recreation center" in the business type field and typing each of the PSU's ZIP Codes in the ZIP Code field of these Internet sites.

Please note that the manner in which the sampling frame was formed can result in gas stations, fast food restaurants, and recreation centers that are in the sampling frame but not in any of the sample PSUs. This situation arises when a ZIP Code lies partly within and partly outside of a sample PSU (a phenomenon that did occur among the 16 sample PSUs). It was impractical to address this deficiency during sampling frame formation. We intended to remedy this deficiency after sample selection (see the section "A Post-Selection Substitution We Planned to Make But Didn't" in Section 4.2.2), but through an oversight this was not implemented.

Pre-Selection Process to Remove Duplicates From the Frame

Note that the manner in which the sampling frame was formed can result in two or more members of the sampling frame that identify the same establishment. This can result in a number of ways:

- *Establishments with multiple phone numbers:* E.g., a recreation center with two phone numbers might appear as two listings in an Internet search, once for each phone number (e.g., the listings "Peoria Recreation Center, 10 Main St, Peoria IL 61602, (309) 555-1000" and "Peoria Recreation Center, 10 Main St, Peoria IL 61602, (309) 555-1001" might appear as distinct results on superpages.com, or one might appear on superpages.com and the other on yellowbook.com, or one might appear on the State list of recreation centers and the other on the county list of recreation centers).
- *Shorthand for street addresses:* A fast food restaurant at 10 Main Street might appear once under "10 Main Street" and once under "10 Main St."

- *Establishments that have changed names:* E.g., a gas station that changed names from “Bob’s Gas” to “Steve’s Gas” might appear as two listings in an Internet search, once for each name.
- *Establishments on street corners:* E.g., a McDonald’s located at the intersection of Main St. and 1st St. might appear once under Main St. and once under 1st St.

We shall call a member of the sampling frame that identifies the same establishment as another member of the frame a *duplicate* (or *duplicate site*).

Some duplicates were identified prior to sample selection by printing the name, addresses, and phone numbers of the establishments in the sampling frame for a given sample PSU and site type, and visually scanning each printout to identify instances of establishments having the same address. Duplicates identified in this manner were removed from the sampling frame.

Note that this process will not identify all instances in which a given establishment is listed multiple times in the sampling frame. We note that we could have automated and/or refined this process of duplicate identification using record linkage software.

We also note that we will engage in two subsequent processes for identifying duplicate sites: These processes are described in the sections “Post-Selection Exclusions from the Sample” in Section 4.2.3 and “Post-Selection Process to Identify Duplicates in the Frame” in Section 4.2.4.

Process to Apply the Four Site Sampling Frame Restrictions

Restrictions 1 to 3 were applied by examining the sites’ addresses. (Addresses that identified businesses as being located in shopping centers, on military bases, or in office buildings were eliminated. E.g., a daycare center identified as “Happy Kids Daycare, Parklawn Building” would have been eliminated. Addresses of recreation centers that suggested the presence of merely a park, climbing wall, or senior center were eliminated. E.g., an address identifying the center as “Rockville Climbing Wall” or “Rock Creek Park” or “Golden Oldies Senior Center” was eliminated.) Restriction 4 was applied from licensing information contained on the county lists.

The choice of site types and Restrictions #3 and 4 were motivated by the desire to capture large numbers of children, particularly in the 4- to 7-year-old age range. Restriction #1 was necessitated by the ability to access the site. Restriction #2 was motivated by the practical consideration of data collectors being able to approach vehicles before its occupants have exited the vehicle.

There were additional sampling frame restrictions that we would have liked to apply, but were impractical and so were effectively applied after site selection. E.g., we would have liked to restrict the sampling frame of recreation centers to those that contain programs for children under age 12, but this was impractical to implement as a frame restriction. Instead we implemented such restrictions through information ascertained in phone calls to the sample sites. We will describe this process further in Section 5.1.2.

The sampling frame was stratified by the four site types: gas stations, recreation centers, daycare centers, and fast food restaurants. Instances of the word “strata” (or “stratum” or “stratification,” etc.) in this report refer to this stratification.

In this report, the term *site sampling frame* shall refer to the sampling frame formed in this section. Thus the site sampling frame only contains sites in the ZIP Codes of the 16 selected NSUBS PSUs. As

mentioned above, in Section 4.2.3 we shall identify duplicate sites in the site sampling frame, and thus the site sampling frame (as defined in this report) does not contain distinct members.

4.2.2 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.

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.

³ There were a few exceptions to this. There were two noncertainty PSUs in which there were many (400 or more) daycare centers on the frame, and a sample size of 4 or 5 was assigned in these cases. There were two other PSUs where there was one frame unit for recreation centers in the PSU, and in this case the one frame unit was taken. There was one PSU with no recreation centers.

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).

Following are 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.

We shall call the set of 627 sites resulting from the above three steps the *probability sample* (or for emphasis, the *NSUBS probability sample*). Please note that we shall make some post-selection refinements to the probability sample in Section 4.2.3 so that the probability sample will not consist of the set of sites on which the survey will attempt to collect data. (See Section 4.2.3 for details.)

4.2.3 Post-Selection Refinements to the Probability Sample: Obtaining the “Refined Sample”

In the previous section we selected 627 sites in a probabilistic manner, and called this set of sites the *probability sample*. In this section, we make a number of post-selection refinements to the probability sample. We shall call the set of sites resulting from the operations described in this section the *refined sample* (or for emphasis, the *NSUBS refined sample*).

Post-Selection Substitutions in the Sample

One site in the probability sample was excluded via the following process. The addresses of each of the 627 sites in the probability sample were entered into the software ArcView Geographic Information System, version 3.2 (manufactured by ESRI Corporation). This software identifies the latitude and longitude of the addresses, through which it was discovered that one site (a gas station among the 323 sites selected in Step 1 above) was listed twice in the sample (i.e., two of the Step 1 sites had the same latitude and longitude). Recall that the “next” member of the stratum containing this site was selected into the probability sample in Step 2 above. The “next” member following this “Step 2” site was selected as the substitute for the duplicate gas station.

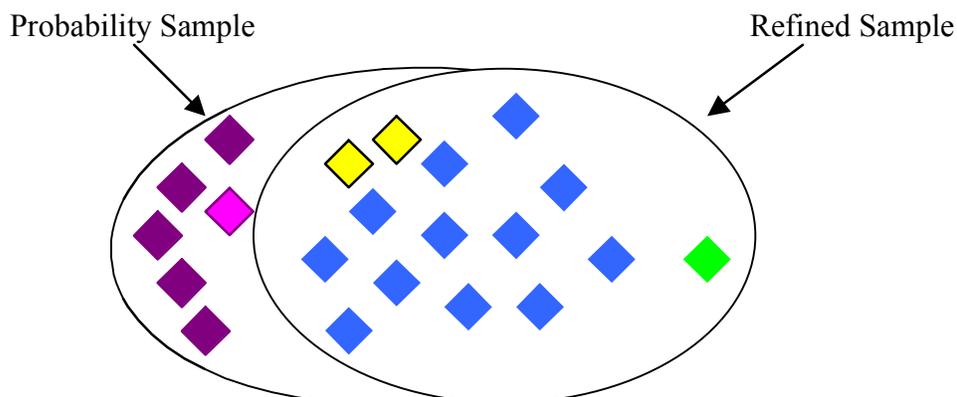
Post-Selection Exclusions from the Sample

A total of 68 sites were excluded based on information obtained upon the data collectors visiting the site to conduct the survey. Among these, 58 sites were found to have gone out of business or to not comply with the four frame restriction criteria from Section 4.2.1. (E.g., a site selected as a gas station found to not sell gas, or a site selected as one of the five fast food chains that was found to be no longer a member of one of these chains.) In addition, 5 sites were excluded because they had no parking lots (and thus there was no location at which to effectively conduct the survey). Finally, 5 other sites were excluded for other reasons, such as being deemed by the data collectors as unsafe. (For breakouts of these numbers by site type, please see the table “Business Recruitment by Site Type” in Section 5.1.2.)

As a reminder, we call the set of sites resulting from the above post-selection exclusions the *refined sample*. Thus the refined sample contains 559 members (i.e., 559 sites). This is the collection of sites at which the survey will attempt to collect data in Section 5.

Following is a depiction of the relationship between the probability sample and the refined sample:

The NSUBS Probability Sample and Refined Sample



Legend	
	one of the 68 sites excluded in Section 4.2.3 for which no substitute was chosen (not all 68 depicted)
	one of the 556 sites selected in Steps 1 or 2 of Section 4.2.2 that is not excluded in Section 4.2.3 (not all 556 depicted)
	one of the two sites selected in Step 3 of Section 4.2.2
	the site excluded in Section 4.2.3 for which a substitute was chosen
	the substitute site selected in Section 4.2.3
Probability sample	   
Refined sample	  

A Post-Selection Substitution We Planned to Make But Didn't

Recall from Section 4.2.1 that the sampling frame contains sites not in any of the selected PSUs. This arose only for gas stations and fast food restaurants in cases where a selected PSU contained part of a ZIP Code. We had planned to enter location information for each of the selected fast food restaurants and gas stations into the ArcView software. This software would have allowed us to identify the selected sites that did not lie in any of the selected PSUs. We had planned to choose substitutes for these sites by taking the next member in the sorted frame (or in the case in which the initially selected member is the last member of a stratum, we would have chosen the penultimate member of the stratum). However through an oversight this was not implemented in the 2006 sample and due to the date of discovery of the oversight. We plan to address it in the 2008 survey.

As alluded to earlier, in the interest of data quality NHTSA does not publicly release the locations of the observation sites or even the States in which they lie.

4.2.4 Post-Selection Process to Identify Duplicates in the Second Stage Frame

The addresses and phone numbers of the sampling frame in the selected PSUs were entered into MS Access. MS Access identified 19 distinct duplicate sites in the frame (i.e., 19 distinct sites, each of which

had multiple occurrences on the frame), by looking for sites with the same address or same phone number.

We note that we could have incorporated this process upon forming the sampling frame, or could have refined this process through the use of address-matching software.

Recall from Section 4.2.1 that in this report the term *site sampling frame* is defined as the sampling frame formed in Section 4.2.1. Thus the members of the sampling frame are not distinct and the process identified in the current section identifies duplicates in the sampling frame.

4.3 Site Selection Probabilities

Because of the identification of frame duplicates in Section 4.2.4, the calculation of the site selection probabilities is nontrivial. We shall first calculate what the site selection probabilities would have been had the sampling frame constructed in Section 4.2.2 contained no duplication. We shall then develop an adjustment that approximately adjusts the site selection probabilities for the frame duplication. What shall result will be an approximate but not exact calculation of the true site selection probabilities, which would have been prohibitively complex to calculate.

Some Notation

It is at this point that we shall need to establish some notation. We shall use the following notation throughout the report. For the reader's convenience, a glossary of the definitions of all mathematical notation used in this report appears in Section 11.

Consider the NSUBS site sampling frame (which only contains sites in the 16 sample PSUs). Order the sample PSUs and the four strata within the sample PSUs in some manner fixed for the duration of this report.

Let $1 \leq i \leq 16$ and $1 \leq j \leq 4$ denote integers. Let m_{ij} denote the sample size (i.e., the number of members in the probability sample) in the j^{th} stratum (which we shall also call "stratum j ") of the i^{th} PSU (i.e., PSU i).

Let m_{1ij} (respectively, m_{2ij}) denote the sample size for stratum j of the NSUBS PSU i in the selection of 323 sites in the probability sample selected in Step 1 of Section 4.1.2 (respectively, the 302 sites selected in Step 2), and let m_{3ij} take the value 1 if stratum j of PSU i contains one of the two sites selected in Step 3 of Section 4.2.1, and 0 otherwise. Thus $m_{ij} = m_{1ij} + m_{2ij} + m_{3ij}$.

Let M_{ij} denote the number of sites in the site sampling frame for stratum j of PSU i .

List the M_{ij} sites in stratum j of PSU i of the site sampling frame in a manner such that:

- the first m_{1ij} sites in the list consist of the sites (in this stratum and PSU) that were selected in Step 1 of Section 4.2.2, followed by
- the m_{2ij} sites selected in Step 2 of Section 4.2.2, followed by
- the m_{3ij} sites selected in Step 3 of Section 4.2.2, followed by
- the $M_{ij} - m_{ij}$ sites that are not in the probability sample.

Note that the sublist described by the third bullet is nonempty only if stratum j of PSU i contains one of the two sites selected in Step 3. Again, this sort order shall be fixed for the duration of this report.

Let $1 \leq k \leq M_{ij}$ and consider site k of stratum j in PSU i of the site sampling frame.

The goal of this section is to develop the formula for the probability that site k of stratum j in PSU i is selected in the NSUBS probability sample. Thus our formula will be defined for the members of the site sampling frame. We shall not develop this formula for the sites in the non-sample PSUs, as the partitioning of the NOPUS PSUs into the NSUBS PSUs involved a subjective process that was only performed for some of the NOPUS PSUs (namely, those described in Step 3 of Section 4.1.2.)

We shall also define further notation as needed throughout the report. Notation defined anywhere in this report appears in Section 11, together with its definition.

The Site Selection Probabilities Had There Been No Duplicates on the Frame

Had the frame formed in Section 4.2.1 contained no duplicate sites, the selection probability for site k of stratum j in PSU i would have been:

$$p'_{ijk} := \begin{cases} q_i \delta_i \frac{Pop_i}{TotPop_i} \frac{m_{1ij} + m_{2ij}}{M_{ij}} & \text{for } 1 \leq k \leq m_{1ij} + m_{2ij} \text{ and } m_{ij} < k \leq M_{ij} \\ q_i \delta_i \frac{Pop_i}{TotPop_i} & \text{for } m_{1ij} + m_{2ij} < k \leq m_{ij} \end{cases}$$

where

q_i denotes the probability of selection of the NOPUS PSU containing (the NSUBS) PSU i ,

$\delta_i := 1$ if the NOPUS PSU containing PSU i is one of the two certainty PSUs identified in Step 1 of Section 4.1.2, and 14/48 otherwise,

$Pop_i :=$ the population in 2000 of children under age 5 in PSU i ,

$TotPop_i :=$ the population in 2000 of children under age 5 in the NOPUS PSU containing PSU i ,

M_{ij} denotes the number of sites in the sampling frame for stratum j of PSU i ,

We note that the nonunity value of the term δ_i reflects the combination of Steps 1 and 2 from Section 4.1.2, in which 22 PSUs were selected from 48 via systematic sampling, followed by a further systematic subsampling of 14 PSUs from the selected 22.

Adjusting for Duplicates on the Frame: The (Approximate) Site Selection Probabilities

Recall that in Section 4.2.4 we identified 19 instances on the sampling frame in the 16 selected PSUs that had duplicates. Thus p'_{ijk} does not truly reflect the site selection probability, as duplicated sites had multiple chances to be selected.

A first idea for approximating the true selection probability (in a manner that reflects the fact that duplicates exist on the sampling frame) would be to replace the portion of p'_{ijk} that reflects the second stage selection with the probability that a site is selected at least once from the given PSU i and stratum j .

Letting r_{ijk} denote the number of occurrences of site k of stratum j of PSU i in the site sampling frame, we could estimate this probability as

$$p''_{k|ij} := 1 - \left(1 - \frac{p'_{ijk}}{p_i}\right)^{r_{ijk}} \quad \text{for all } 1 \leq i \leq 16, 1 \leq j \leq 4, 1 \leq k \leq M_{ij}$$

where p_i denotes the probability that PSU i is selected (i.e., $p_i := q_i \delta_i \frac{Pop_i}{TotPop_i}$). (We are viewing the

second stage selection as a Poisson selection in which members of the site sampling frame (which was created in Section 4.2.2 and contains duplicate sites) are selected independently, each with probability p'_{ijk} .) (Neither of the two sites selected in Step 3 of Section 4.2.2 were duplicate sites and thus $p''_{k|ij}$ is well defined.)

However the inverses of these first attempts at selection probabilities do not sum over the sample to (and might not even be close to summing to) the frame count for stratum j of PSU i . That is,

$$\sum_{k=1}^{m_{ij}} \frac{1}{p''_{k|ij}} \neq M_{ij} - \sum_{k=1}^{M_{ij}} (r_{ijk} - 1)$$

Thus we shall multiply the $p''_{k|ij}$ by a factor to arrange for this. Doing so gives that our best estimate of the selection probability for site k of stratum j of PSU i is:

$$p_{ijk} := p_i p''_{k|ij} \frac{\sum_{l=1}^{m_{ij}} \frac{1}{p''_{l|ij}}}{2M_{ij} - \sum_{k=1}^{M_{ij}} r_{ijk}} \quad \text{for all } 1 \leq i \leq 16, 1 \leq j \leq 4, 1 \leq k \leq M_{ij}$$

(For the pure convenience of making p_{ijk} well defined we have utilized the same multiplicative factor $p_{ijk} / p_i p''_{k|ij}$ for the members of the site sampling frame that are not in the probability sample. The definition of the site selection probabilities for these members will only matter for the purpose of estimation for the members of the refined sample that are not in the probability sample. Alternatively, we could have restricted the definition of p_{ijk} to those values of i, j , and k that refer to members of the refined sample.)

We shall call p_{ijk} the *site selection probability* for site k of stratum j of PSU i , and note that it is only approximate. In order to truly calculate the site selection probability, we would need to calculate the actual (and not just approximate) probability that at least one occurrence of a duplicate is selected, which under the site selection method of stratified PPS sampling is prohibitively complex.

We define the *sampling weight* of a site in the site sampling frame to be the inverse of its site selection probability. As the site selection probability is only an approximation of the true selection probability, the sampling weight is only an approximation of the true sampling weight. Letting w_{ijk} denote the sampling weight for site k of stratum j of PSU i , we have:

$$w_{ijk} = \frac{1}{p_{ijk}} = \frac{2M_{ij} - \sum_{k=1}^{M_{ij}} r_{ijk}}{p_i p''_{k|ij} \sum_{l=1}^{m_{ij}} \frac{1}{p''_{l|ij}}} \quad \text{for all } 1 \leq i \leq 16, 1 \leq j \leq 4, 1 \leq k \leq M_{ij}$$

Thus note that in this section we have established an approximate site selection probability, which we call the *site selection probability* (and corresponding approximate sampling weight, which we call the *sampling weight*) for all sites in the site sampling frame. In particular, these terms are defined for the members of the refined sample and the members of the probability sample.

Some Final Notes

We note that even if the site sampling frame had no duplicates, each site X that is not a daycare center and that lies in a ZIP Code that straddles PSUs Y and Z would have two chances of being selected. Namely, X could be selected because PSU Y was selected and X was found in the Internet search in the given ZIP Code, or X could be selected because PSU Z was selected and X was found in the Internet search in the given ZIP Code. We do not have the means to take this into account in the calculation of X’s selection probability, as we do not have available the selection probabilities of the NOPUS frame PSUs that were not selected in the NOPUS sample (and thus could not handle the case in which the ZIP Code lies partly in a non-selected NOPUS PSU).

We note that every gas station, restaurant in the five fast food chains, daycare center, and recreation center in the 50 States and District of Columbia that could be found by an Internet search or, for recreation centers and daycare center, is on a county list of such establishments, and satisfies the four site sampling frame restrictions from Section 4.2.1 had a nonzero probability of being selected for the NSUBS probability sample. This is because the sampling frame for the NOPUS PSUs contained a partitioning of the combined region formed by the 50 States and the District of Columbia (Glassbrenner, 2002).

The sampling frame from which the current NOPUS PSUs were chosen employs some modest sampling frame exclusions, namely, the exclusion of 37 counties with very low vehicle miles traveled (VMT) (Glassbrenner, to appear) Thus if the NSUBS sample is redesigned so that its PSUs are selected from the current NOPUS PSUs (or partitionings thereof), all sites in the 50 States and DC that are outside of these 37 counties could be found by an Internet search and satisfy the four site sampling frame restrictions in Section 4.2.1 would have a nonzero selection probability.

4.4 Sample Size Determination

In designing the sample for the NSUBS in 2005, we wished to achieve (if possible) the following cost constraint and the variance constraints in the table “Desired Margins of Error”:

Cost Constraint

The survey shall (in 2005) cost no more than \$300,000 in total survey costs, including costs for survey preparation, training, data collection costs, quality control procedures, and production of estimates.

Desired Margins of Error

Estimate (Nationwide)	Desired Margin of Error, Using 90% Confidence
Booster Age Children (Age 4-7)	
Restraint use	2 percentage points
Booster seat use	2 percentage points
Child safety seat use ¹	2 percentage points
Seat belt use ²	2 percentage points

Toddlers (Age 1-3)	
Restraint use	2 percentage points
Booster seat use	2 percentage points
Child safety seat use ¹	2 percentage points
Seat belt use ²	2 percentage points

¹ Use of a rear-facing or front-facing child safety seat.

² Use of a lap/shoulder or lap only seat belt.

If adequate frame information were available, one would like to approach this problem as one of determining sample sizes (i.e., the number of PSUs, and the number of sites per PSU) to minimize variances for a fixed cost. (This is the approach taken to design the NOPUS sample; see Glassbrenner, to appear.)

However we did not have adequate information with which to model variances of booster seat use, and thus the sample sizes had to be made on an intuitive basis.

Based on information from a 2005 pilot study of data collection protocols (described in Section 6.1), we set the duration of the data collection period for the NSUBS to be 2 hours per site. Based on sample sizes used in the National Occupant Protection Use Survey, we set the number of PSUs to be 16 and the number of sites per PSU to be 20.

As noted in Section 4.2 we set a target sample size of 40 sites per PSU in order to ensure at least 20 participating sites per PSU.

The margins of error in the 2006 survey

The target margins of error from the table “Desired Margins of Error” do not appear to have been achieved in the 2006 survey, which yielded the following estimated margins of error:

Achieved Margins of Error

Estimate (Nationwide)	Achieved Margin of Error, Using 90% Confidence
Booster Age Children (Age 4-7)	
Restraint use	4 percentage points
Booster seat use	9 percentage points
Child safety seat use ¹	NA ³
Seat belt use ²	6 percentage points
Toddlers (Age 1-3)	
Restraint use	3 percentage points
Booster seat use	6 percentage points
Child safety seat use ¹	NA ³
Seat belt use ²	3 percentage points

¹ Use of a rear-facing or front-facing child safety seat.

² Use of a lap/shoulder or lap only seat belt.

³ Not computed.

There could be a variety of reasons why the target margins of error were not achieved. Perhaps the intuitive setting of sample sizes (numbers of PSUs, sites per PSUs, and duration of data collection) for the fixed cost differed greatly from the optimal determinations. In particular, the intuitive determination of the sample sizes was based on sample sizes from NOPUS, but the NOPUS estimator differs from the NSUBS estimator. (See Section 8 for details on these estimators.) We also note that the amount of funds available for the survey (\$300,000) was fairly small, and had we been able to collect more data, the achieved margins of error might have been markedly lower.

5. Data Collection Protocols

Many preliminary items are required in order to specify the instructions data collectors were given with regards to how they were to collect data. We present these preliminary items in Sections 5.1 – 5.8, followed by the data collection protocols in Section 5.9. Section 5.6 provides a list of all data collection variables.

The preliminary topics covered are as follows:

- Techniques used to obtain cooperation from the data collection sites,
- The data collection schedule,
- The number, gender, and positioning of the data collectors,
- Definitions of restraint use used by the survey,
- The categories of race and ethnicity used by the survey,
- The wording of the interview questions, and
- Other assorted data collection topics and definitions.

5.1 *Obtaining Site Cooperation*

5.1.1 How Cooperation Was Obtained

Cooperation with recreation centers and daycare centers was obtained in advance of visiting these sites to collect data via sending letters requesting cooperation, followed by phone calls to secure cooperation. At times, it was also necessary to provide “hold harmless” agreements and certificates of insurance to certain locations. In some localities, permission to use county recreation facilities was subject to the approval of county commissioners and similar governing bodies.

Data collectors and quality control monitors approached individual fast food and gas station establishments in person to secure cooperation. These staff received training in recruiting techniques to try to maximize the participation rates of business establishments.

We note that in some cases, it was discovered during the process of attempting to secure cooperation that the site was either no longer in business or had changed to an ineligible site type. An example seen of the latter is a gas station that had changed to a car repair shop. In a few other cases it was discovered upon visiting the site to secure cooperation (i.e., for gas stations and fast food restaurants) that collecting data at the site could pose a safety risk to the data collectors because of vagrants congregating in the parking lot, and these few sites were dropped from the survey. More details are provided in the next section, on site participation rates.

5.1.2 Site Participation Rates for the 2006 Survey

In total 383 of the 559 sites in the refined sample gave permission for the survey to be conducted on their premises.

Business Recruitment Results by Site Type

Site Type	Number of Sites in the NSUBS Probability Sample Which for the 2006 NSUBS Survey...				
	Participated in the Survey	Expressly Declined Survey Participation	Were Ineligible for Survey Participation*	Did Not Specifically Decline Survey Participation But Neither Granted Permission	Total
Daycare Centers	28	47	0	0	75
Fast Food	107	51	23	20	201
Gas Stations	205	29	44	20	298
Recreation Centers	43	9	1	0	53
Total	383	136	68	40	627

*These establishments were found during business recruitment to have gone out of business or have changed business to an ineligible site type (58 cases); have no parking lot (5 cases); or were not suitable for some other reason (5 cases).

5.2 The Data Collection Schedule

This section describes the dates and times of day that the 2006 survey was conducted, and how the data collection schedule was determined.

Data collection for the 2006 survey was conducted during the period July 17–29, 2006. Data was collected on all days of the week and during all daylight hours (7 a.m. to 6 p.m.).

Because children tend to be at certain site types at certain times of day, the times of day during which data was collected varied by site type. In nearly all cases, data was collected at daycare centers in the mornings (7 a.m. to 10 a.m.), while data at recreation centers was collected in the morning and midday (8 a.m. to 2 p.m.). At fast food restaurants, data was collected at breakfast, lunch, and dinner mealtimes (8 a.m.–10 a.m., noon–2 p.m., and 4 p.m.–6 p.m.). Gas stations were visited throughout the day (8 a.m. to 6 p.m.).

Surveys that involve visits to sites often use probabilistic algorithms to determine the schedule with which the sites will be visited, in order to avoid bias in the times of day or days of week on which various types of sites are visited. (These probabilistic designs are often clustered for efficient data collection, e.g., the desired days of data collection might be subdivided into weeks, the PSUs might be randomly assigned to weeks; the weeks subdivided into time slots for data collection, and the sites in each PSU randomly assigned to the time slots in the assigned week.)

However the challenge of securing site cooperation in the NSUBS made it impractical to utilize a probabilistic assignment of the data collection schedule. Rather the NSUBS sites were scheduled for data collection as follows: Each PSU was assigned a string of consecutive days during the period July 17 – 29, during which data for the PSU would be collected. E.g., PSU 1 might have been assigned to have its data collected during July 17 – 23. Appointments at daycare centers and recreation centers were scheduled for times recommended by the managers of these centers as prime drop-off periods for children. The remaining eligible time period for data collection were filled in by soliciting the cooperation of gas station and fast food restaurant managers.

The schedule determined in the previous paragraph is called the *site visitation schedule*.

We note that there are conditions, described in Section 5.9, under which a site visit may be rescheduled.

5.3 *Number, Gender, and Positioning of Data Collectors*

5.3.1 Number and Gender of Data Collectors

Although the survey could have been conducted using a single data collector at each site, it was decided to use two per site. At many high volume daycare and recreation centers, having two data collectors increases the number of observations and interviews that can be obtained during the two-hour data collection period. In addition, data collectors are required to monitor all vehicles that enter the data collection sites (see Section 5.8.3). This would be challenging for an individual, especially when there are numerous entrances to the parking lots. Previous experience from the NSUBS pilot study and the NOPUS survey suggested that data collectors preferred to work in pairs rather than alone, particularly when attempting to locate a site in an unfamiliar area and when collecting data at sites where they might be questioned by members of the public about the authority to conduct the survey. Collecting data in teams can also provide some measure of protection against data falsification.

The data collection was divided between the data collectors by having each data collector independently collect data on different vehicles. Information from the NSUBS pilot study suggested that this was a more efficient means of data collection than having the data collectors collect different survey variables on the same vehicles. Also, having one data collector (rather than two) approach a vehicle was considered to be potentially less threatening to the drivers.

It was also decided as a result of the NSUBS pilot study and a focus group conducted to explore views of potential survey respondents to use female data collectors to the greatest extent possible. This was because the information from the focus group and pilot indicated that the use of male data collectors could decrease response rates. (Some focus group participants expressed that they would be hesitant to talk to a male approaching their vehicle when they had their children with them.)

Two data collectors were assigned to collect all data in a given PSU. Thus there was a total of 32 data collectors used for the 2006 survey.

5.3.2 Positioning Data Collectors at the Site to Best Collect Data

The question of where to position data collectors at the sites to best collect data turns out to be a nontrivial one.

At some of the sites where children will be exiting the vehicle to go into the establishment (such as daycare centers, recreation centers, and fast food restaurants where the vehicle is not going to a drive-thru lane), children might in eager anticipation (or for other reasons) unbuckle their seat belts and car seat harnesses after the vehicle enters the parking lot but before it parks. Since the NSUBS desires to reflect child restraint usage on the road, we wish to observe the restraint use status before these children unbuckle.

On the other hand it is also vastly preferable for data quality purposes to observe restraint use in a stopped vehicle, when possible. E.g., it is difficult to record restraint use for five children in a moving vehicle, even a slow-moving one, and it may be difficult to see whether a child in the vehicle is on a backless booster.

Thus for sites (such as daycare centers) at which we generally expect children to exit the vehicle to go into the establishment, it would seem best to station data collectors near the entrance to the parking lot to collect “initial” observations for a small number of variables (specified below) that focus on capturing child restraint use, follow the vehicle until it parks, and then conduct (with the driver’s consent) the

interview portion of the survey and make certain types of corrections (also specified below) that pertain to restraint type but not restraint status.

For other sites, i.e., those for which we generally expect children to remain in the vehicle (such as gas stations), it would seem best for data collectors to approach the vehicle as it is parking, and collect all survey items (with the driver’s permission) when the vehicle is parked.

To that end, data collectors were instructed (unless this poses a safety risk or the business manager objects) to position themselves at the following locations to begin data collection on a vehicle.

Location of Data Collectors at the Sites

Site Type	Location of Data Collectors to Collect Data on Vehicles and Occupants ¹
Gas Station	All data on vehicles and occupants are collected at a gas pump island.
Fast Food Restaurant	Data on vehicles and occupants at a given fast food site are collected using one or both of the following two paradigms: <u>Drive-Thru Paradigm</u> All data on vehicles and occupants is collected at a drive-thru lane. <u>Lot Entrance Paradigm</u> Data from a limited number of observational variables ² is collected from an entrance to the parking lot, with the remainder of the variables and certain types of corrections ³ to observational data collected from the parked vehicle’s parking space. The determination of which paradigms to use is made the by data collectors upon arriving at the site. If there is a drive-thru lane and the business manager does not object to the survey being conducted there, then one data collector uses the Drive-Thru Paradigm and the other uses the Lot Entrance Paradigm. Otherwise, both data collectors use the Lot Entrance Paradigm, and if the site’s parking lot has multiple entrances, the data collectors station themselves at different entrances.
Daycare Center	Data from a limited number of observational variables ² is collected from an entrance to the parking lot, with the remainder of the variables and certain types of corrections ³ to observational data collected from the parked vehicle’s parking space.
Recreation Center	Data from a limited number of observational variables ² is collected from an entrance to the parking lot, with the remainder of the variables and certain types of corrections ³ to observational data collected from the parked vehicle’s parking space.

¹ We allowed data collectors to choose a location other than the specified location if the specified location posed a safety risk or was not permitted by the business manager. There is a limited amount of data collected that does not pertain to vehicles or occupants (such as the times at which data collection began and ended). See Section 5.6 for the survey variables.

² The variables Restraint Used and Seating Position of as many occupants who appear to be under age 13 as possible, followed if possible (i.e., if the data collector can record accurately for this moving vehicle) by the variables Restraint Used, Age, and Gender of the driver, followed if possible by Restraint Used, Age, Gender, and Seating Position of other occupants. See Section 5.9 for more information.

³ Data collectors were instructed to make corrections to restraint types (e.g., whether a child is in a backless booster seat or a seat belt) but not to restraint status (i.e., whether a child is restrained according to the definition of “restrained” provided in the table “Definitions Used for the Survey Variable ‘Restraint Used’”). See Section 5.9 for more information.

We note that the two data collectors at a given site might be working at some distance from each other, even when both data collectors are conducting observations at a parking lot entrance, since the site's parking lot might have more than one entrance.

Note that the seemingly necessary differential treatment of sites where children are expected to exit the vehicles (versus other sites) could result in some amount of a corresponding differential population of vehicles on which observational data is collected. As we will see in Section 5.9, data collectors collect the observational survey variables prior to the interview survey variables. At gas station islands and in fast food drive-thru lanes, it is possible that a potential respondent would effectively terminate the collection of at least some, including perhaps some of the observational, survey variables (e.g., ask a data collector what s/he is doing and to stop doing this, in which case the data collector would cease collecting further data on the vehicle). At sites in which data collection was begun at the parking lot entrance (i.e., all daycare centers, all recreation centers, and the fast food restaurants using the Lot Entrance Paradigm), potential respondents would seem to be much less likely to terminate the collection of the observational variables, as they are driving a vehicle past the data collector at the time.

At the time of this publication we have not studied whether this possible differential population data impacted the survey results. We only note that there were a relatively small number of vehicles (namely, 570 vehicles with 908 child occupants in the 2006 survey) that were observed and declined participation in the survey when asked.

5.4 *The Definition of Restraint Use Used by the Survey*

The survey utilized the following definitions of restraint use.

Definitions Used for the Survey Variable “Restraint Used”

Restraint Type	Definition Used for the National Survey of the Use of Booster Seats ¹
Lap/Shoulder Belt	The occupant (of any age) is not in a child safety seat or booster seat. A seat belt is across the front of his/her body and a seat belt is across his/her lap. The belt may have slack in it and the shoulder belt may be under his/her arm.
Lap Belt	The occupant (of any age) is not in a booster seat, has a seat belt across his/her lap, and has no seat belt across the front of his/her body.
Rear-Facing Child Safety Seat	The occupant is a child in a seat that sits on top of the vehicle seat in such a way that s/he faces the rear of the vehicle, and the harness straps are across his/her front. The harness straps might be secured or not.
Forward-Facing Child Safety Seat	The occupant is a child in a seat that sits on top of the vehicle seat in such a way that the occupant faces the front of the vehicle, and with harness straps that are across his/her front.
High-Backed Booster Seat	The occupant is a child in a seat with a seat back that sits on top of the vehicle seat, and has a seat belt across the front of his/her body, whether lap or lap/shoulder. No harness is in use.
Backless Booster Seat	The occupant is a child sitting on a platform with no seat back that sits on top of the vehicle seat, and has a seat belt across the front of his/her body, whether lap or lap/shoulder. No harness is in use.

¹ These definitions were developed to provide characterizations that can be reliably implemented by data collectors. They are not meant to convey any notion of what constitutes proper use or misuse of a particular restraint type.

5.5 The Categories of Race and Ethnicity Used by the Survey

In accordance with the standards for collection of race and ethnicity in Federal surveys established by the Office of Management and Budget in the October 30, 1997, Federal Register Notice, Volume 62, Number 210, pages 58781-58790, the NSUBS uses the following categories of ethnicity:

- Hispanic or Latino,
- Neither Hispanic nor Latino,

and the following categories of race:

- White,
- Black or African-American,
- Asian,
- Native Hawaiian or Other Pacific Islander, and
- American Indian or Alaska Native.

Because the survey collects data on children and in order to not burden multiple adult occupants when there may be restless children in the vehicle, the race/ethnicity of all occupants in a given vehicle is obtained by asking the driver, with the data collectors being given the discretion to take into account answers offered by other occupants. E.g., if a passenger (of any age) offers information regarding his/her race and/or ethnicity in response to the driver being questioned about him/her, the data collector uses his/her judgment as to whether to record the passenger's answer or any offered by the driver.

Respondents are allowed to choose more than one race, in which case (c.f., Section 5.6) the data collector records that the occupant has chosen more than one race, but does not record the particular races chosen. Respondents are not allowed to choose more than one ethnicity.

5.6 The Survey Variables

The survey collects the following variables.

The survey variables fall into three categories: those pertaining to the site or data collection conditions, those pertaining to vehicles or occupants collected by observation, and those pertaining to occupants collected by interview.

Survey Variables Collected

Variable	Definition ¹
Variables Pertaining to the Data Collection Site or Data Collection Conditions (All Obtained via Observation)	
Observer Name	The name of the data collector recording information at a given site
Booklet Number ²	The number of the booklet on which the data collector records information about the site. Each data collector records the booklet number as "1" for the first booklet s/he uses at a site, and increases this value by 1 for each subsequent booklet used.
Site Identification Number ³	The identification number assigned to the sample site
Site Type	Whether the site is a gas station, recreation center, daycare center, or a fast food restaurant

Variable	Definition ¹
Site Name	The name of the site, e.g., “Frank’s Gas Station”
Street Address	The site’s street address, e.g., “10 Main Street, Peoria IL 61601”
Start Time	The time that the data collector began collecting the data in the given booklet at the given site, e.g., 9:12 a.m.
End Time	The time the data collector finished collecting data in the given booklet at the given site, e.g., 11 a.m.
Urbanization	Whether in the consensus assessment of the data collectors the site is located in an urban, suburban, or rural location
Weather Conditions	Whether the conditions at the time of data collection are clear, foggy, or have light rain ⁴
Number of Refusals	The number of vehicles whose drivers declined participation in the survey
Number of Missed Vehicles	The number of passenger vehicles that appeared to have at least one occupant under age 13 for which the data collectors were not able to collect data because they were busy collecting data for other vehicles
Variable Pertaining to Vehicles (Collected via Observation)	
Vehicle Type	Whether a vehicle appears to be a car, van/SUV, or pickup truck
Variables Pertaining to Occupants Collected via Observation	
Gender	Whether an occupant appears to be male or female
Seating Position	The seating position of the occupant ⁵
Restraint Used	Whether an occupant is in a rear-facing safety seat, front-facing safety seat, high-backed booster seat, backless booster seat, or seat belt, or is unrestrained, as defined by the Table “Definitions Used for the Survey Variable ‘Restraint Used’”
Variables Pertaining to Occupants Collected via Interview⁶	
Ethnicity	Whether an occupant is of Hispanic or Latino origin, or not ⁷
Race	Whether an occupant is White; Black or African-American; Asian; Native Hawaiian or Other Pacific Islander; American Indian or Alaska Native; or more than one of these categories ⁷
Weight	The weight, in pounds, of an occupant who appears to be less than 13 years old
Height	The height, in inches, of an occupant who appears to be less than 13 years old
Time Spent in Vehicle ⁸	The number of hours (or approximate) that an occupant who appears to be under age 13 spent in the observed vehicle in the past week with that driver
Number of Visits to Gas Stations ⁸	The number of times in the past week that an occupant who appears to be under age 13 has visited a gas station with that driver
Number of Visits to Fast Food ⁸	The number of times in the past week that an occupant who appears to be under age 13 has visited a fast food restaurant in the five restaurant chains used by the survey with that driver ⁹
Number of Visits to Recreation Centers ⁸	The number of times in the past week that an occupant who appears to be under age 13 has visited a recreation center in the sample, with that driver
Number of Visits to	The number of times in the past week that an occupant who appears to be under age 13 has visited a daycare center in the sample, with that driver

Variable	Definition ¹
Daycare Centers ⁸	
Variables Pertaining to Occupants Collected via Interview or Observation, Depending on the Occupant's Age	
Age	<p><i>For occupants who appears to be at least 13 years old:</i> whether an occupant appears to be 13 to 15, 16 to 24, 25 to 69, or at least 70 years old (collected via observation).</p> <p><i>For occupants who appears at most 12 years old:</i> the age of a child who appears to be under age 13, in years and months (collected by asking the driver).</p>

¹ See Section 5.9 for information on the protocols used to collect these variables. The Appendix displays the data collection forms on which the values of the survey variables were recorded.

² The data collection forms are given to the data collectors in “booklets”. See Appendix 12.1 for a description of the booklets and the forms they contain.

³ Each of the 559 sites in the NSUBS refined sample was given an identification number when the sample was selected. The data collectors copy this number to the data collection form from a printed schedule given to them listing the sites they are to visit on a given day and when they are to visit them.

⁴ In the interest of data quality, data collectors did not conduct the survey under other weather conditions.

⁵ The survey recorded up to three occupants in each of the first three rows of seats and none in any other rows (if there were any).

⁶ Data collectors ask the driver whether a child occupants is under age 13 when they are not sure whether this is the case. See the Table “Wording of Interview Questions” for the specific wording of the interview questions used to collect these variables. At the data collector’s discretion, data collectors could also take into account answers offered by occupants other than the driver.

⁷ Race/ethnicity categories as specified by the Office of Management and Budget in the October 30, 1997, Federal Register Notice, Volume 62, Number 210, pages 58781-58790.

⁸ These variables were collected on the initial idea that they might be used in estimation, but ultimately they were not used.

⁹ The NSUBS includes among its sites restaurants in five fast food restaurant chains. In the interest of retaining these chains in future surveys, the names of the chains (which are known to the data collectors) are kept confidential in this report.

We will refer to the variables listed above under “Variables Pertaining to the Data Collection Site or Data Collection Conditions” as *site variables*, those (the sole variable) under “Variable Pertaining to Vehicles” as the *vehicle variables*, those under “Variables Pertaining to Occupants Collected via Observation” as *observed occupant variables*, and those under “Variables Pertaining to Occupants Collected via Interviewing the Driver” as *interviewed occupant variables*. Note that the first three types of variables are collected via observation, while the last is collected via interview. Note that the variable “Age” is collected by interview or observation, depending on the age of the occupant as assessed by the data collector collecting the interviewed occupant variables. Thus there are 12 site variables, one vehicle variable, 4 observed occupant variables, and 10 interviewed occupant variables, for a total of 26 survey variables (noting that the variable Age is both an observed occupant variable and an interviewed occupant variable).

The survey recorded only one occupant per seating position. If a child occupant was sitting on the lap of an adult, the data collectors collected data on the child and did not collect data on the adult. If more than three persons occupied a row of a vehicle, the data collector recorded data on three of them and favored children in determining which occupants to include.

Recording of Unknowns Strongly Discouraged

Data collectors were strongly discouraged from recording a value of “unknown”. For the nonoccupant variables and the observed occupant variables, data collectors were instructed to choose among the values allowed in the table above as best they could. (We note however that there were a relatively small number of data collection forms with no values recorded for one or more of these variables, indicating that the data collector either forgot to record the variable or was sufficiently uncertain about its value that s/he chose to record no value.) For the interviewed occupant variables, data collectors were instructed to record “DK” (don’t know) for any variables for which no response was provided. (We note that all item nonresponses will be imputed prior to estimation. See Section 7.3)

5.7 *The Wording of Interview Questions*

The survey utilized the following wordings for its interview questions.

Because some questions pertain to children and in order not to burden multiple adult occupants when there may be restless children in the vehicle, all questions were directed to the driver, but data collectors were allowed to take into account responses offered by other occupants. E.g., when a data collector asks a driver about a passenger’s race, the data collector was allowed to take into account any answer offered by the passengers, as well as by the driver, in deciding which race to mark on the data collection form for this occupant.

Wording of Interview Questions

Variable	Wording of Interview Question Used to Collect This Variable ¹
Age	Can you tell me the age of this child?
Ethnicity ²	Are you (or this occupant) of Hispanic or Latino origin?
Race ²	What is your race (the race of this occupant)? Please select one or more: White Black or African-American Asian Native Hawaiian or Other Pacific Islander American Indian or Alaska Native
Weight ³	Can you tell me the weight of this child?
Height ³	Can you tell me the height of this child?
Time Spent in Vehicle ³	In the past week, how many minutes did this child spend in a vehicle driven by you?
Number of Visits to Gas Stations ³	In the past week how many times did you visit a gas station when this child was in a vehicle with you?
Number of Visits to Fast Food ^{3,4}	In the past week how many times did you go to a (Chain 1) with this child in a vehicle with you? (Chain 2)? (Chain 3)? (Chain 4)? (Chain 5)?
Number of Visits to Recreation Centers ³	In the past week how many times did you go to a recreation center when this child was in a vehicle with you? Which one(s)? ⁵
Number of Visits to Daycare Centers ³	In the past week, how many times did you go to a daycare center with this child in a vehicle with you? Which one(s)? ⁵

¹All interview questions are directed to the driver, and the data collectors were permitted to take into account answers volunteered by other occupants, subject to the data collector’s judgment. Drivers were asked to report answers to the questions for themselves and up to 8 other occupants of the vehicle seated in the first three rows.

² Race/ethnicity categories as specified by the Office of Management and Budget in the October 30, 1997, Federal Register Notice, Volume 62, Number 210, pages 58781-58790.

³ These questions were only asked concerning occupants who appeared to be under age 13.

⁴ The NSUBS includes among its sites restaurants in five fast food restaurant chains. In the interest of retaining these chains in future surveys, the names of the chains (which are known to the data collectors) are kept confidential in this report. The question as written here is worded with “Chain 1,” “Chain 2,” etc., substituted for the actual names of the chains. The data collector uses the actual names of the chains in asking the question. The driver is asked about the number of visits to each restaurant chain to aid him/her in his/her recollection. The data collector records the total number of visits to each chain as the value of the variable “Number of Visits to Fast Food.”

⁵ The question “Which one(s)?” is used to determine whether the child visited daycare centers/recreation centers in the survey sample. Visits to daycare centers/recreation centers not in the sample were not included in the count for this variable.

5.8 *Other Assorted Data Collection Topics and Definitions*

This section presents assorted additional topics needed to describe the data collection protocols in the next section. We cover items provided to the data collectors, the types of vehicles surveyed, the protocols used to keep track of vehicles that do not participate in the survey, and assorted definitions.

5.8.1 Items Provided to Data Collectors

Each data collector was provided with the following items:

- Booklets of data collection forms;
- A badge with photo identification identifying him/her as an employee of WESTAT authorized to collect data for the survey;
- A sheet (the *Site Assignment Sheet*) identifying the sites they are to visit, their locations, and when they are to visit them;
- An authorization letter from NHTSA;
- A set of maps used to find their assigned data collection sites;
- A set of children’s stickers, to be used as incentives for participation;
- Copies of the NHTSA brochure *A Parents Guide to Buying and Using Booster Seats* in English and Spanish; and
- A card listing the race and ethnicity categorizations used by the survey.

The Site Assignment Sheets, one for each PSU, are generated from the site visitation schedule determined in Section 5.2. The Site Assignment Sheet for a given PSU provides the name and address of each site in the PSU and the date and time on which it is to be visited. As there are 16 PSUs, there are 16 distinct Site Assignment Sheets. Each data collector is provided a copy of the Site Assignment Sheet for his/her PSU.

The maps provided to the data collectors consist of both commercial street maps and customized maps generated by WESTAT.

The badge was worn by the data collectors during data collection.

The stickers were offered to potential respondents as a small incentive to increase response rates. The NHTSA brochure was also offered to respondents.

The authorization letter explains that WESTAT is authorized by NHTSA to collect this data and provides contact information should a member of the public have questions about the survey. The data collectors were instructed to offer this letter if their authority to conduct the survey is questioned (e.g., by a member of the public or a police officer).

The data collection forms, letter of authorization, stickers, and race/ethnicity card are depicted in the Appendix.

5.8.2 The Types of Vehicles Surveyed

The survey collects data on passenger vehicles (i.e., cars, vans, minivans, SUVs, and pickup trucks) that have at least one occupant who appeared to be under age 13 (as many such vehicles as data collectors were able to collect data on). This age restriction was motivated by the desire to capture data on children, particularly 4- to 7-year-old children.

We shall call these vehicles (passenger vehicles appearing to have at least one occupant under age 13) *eligible vehicles*.

5.8.3 Keeping Track of Nonparticipating Vehicles

The survey will adjust its estimates for eligible vehicles that were at the site but did not participate in the survey.

We define two types of nonparticipating vehicles. One type we shall call *refusals*, which we define to be eligible vehicles whose driver declined to answer interview questions in Step 6.6 of Section 5.9.

The other type we shall call *missed vehicles*, which we define to be eligible vehicles at the site during the assigned data collection period for the site, on which no data is collected by either data collector.

One task for the data collectors (described in Section 5.9) is to record the numbers of refusals and missed vehicles at each site. As we shall see in Section 5.9, keeping track of refusals is an easy matter. However keeping track of missed vehicles is nontrivial.

Recall from Section 5.3.1 that the two data collectors at a given site work independently with regards to collecting data on vehicles. Since eligible vehicles may enter and exit the site while one or both data collectors is busy collecting data on a vehicle, and since the data collectors may be stationed at different locations (e.g., one in the fast food drive-thru lane and one at the parking lot entrance, or at two different parking lot entrances to the same establishment) there is no clear way to keep track of the missed vehicles accurately.

The NSUBS allows the data collectors to decide how they can best keep track of missed vehicles at a given site, choosing between the following two protocols:

- The data collectors choose one person between the two of them who will attempt as well as possible to keep track (by notating a hash mark on a designated area of the booklet for each missed vehicle) of all missed vehicles at the site, or
- Both data collectors will keep track (again by noting hash marks on the booklets) of the missed vehicles and will communicate with each other as frequently as seems needed to ensure that each of them is counting different missed vehicles (so that the number of missed vehicles at the site will be the sum of their counts).

5.8.4 Miscellaneous

We note that the data collection will involve personnel called *quality control monitors*. There is one quality control monitor for each pair of PSUs. This person supervises the data collectors in these two PSUs and conducts quality control activities described in Section 6.

Finally, we note two definitions (of the terms Seating Row Limits and Suspension Conditions) that are used in describing the data collection protocols in Section 5.9.

Seating Row Limits

Data is to be collected from no more than three occupants per row of seats in the vehicle, with no more than two non-driving occupants in the first row, and from only the first three rows of the vehicle. If a child is on a lap, then the child is coded for that seating position.

Suspension Conditions

- The vehicle is sufficiently out of view that observations can no longer be conducted (in the case where the data collector is stationed at an entrance to the site's parking lot)
- The vehicle reaches the fast food drive-thru window (if the data collector is at a fast food drive-thru)
- An occupant initiates contact of some sort with the data collector (e.g., asks "What are you doing?", "Do I know you?", "Can I help you?", engages eye contact with the data collector, etc.)

The Seating Row Limit is motivated by efficient data collection. The only passenger vehicles with a fourth row of seats are 15-passenger vans, and a pilot study conducted before the NSUBS found only a very small minority of 15-passenger vans at the site types. It was also found in the pilot study to be relatively rare that more than three occupants occupy a given row of seats. Data collectors are instructed that in the case in which more than three occupants are in a given row of seats, they should only record the data of three of these occupants and should give preference to children under age 13 in deciding which occupant(s) to not collect data on.

The Suspension Conditions are so named because the existence of (at least one of) the conditions will suspend data collection until (and terminate it unless) permission is secured from the driver to pursue data collection on his/her vehicle further.

5.9 Data Collection Protocols

We are now ready to describe the protocols that the data collectors were instructed to follow to collect the survey data.

Each data collector conducts Steps 1 to 7 below for each site on the Site Assignment Sheet provided to them from Section 5.8.1. If at any point during these steps the authority of the data collector to collect data is questioned (e.g., by a member of the public or a police officer), the data collector is instructed to offer the authorization letter from Section 5.8.1 to the inquirer.

Step 1: The data collector attempts with his/her partner to travel to the site and arrive at the time scheduled on the Site Assignment Sheet. The data collectors use the maps provided to them in Section 5.8.1 to locate the site. If they cannot locate the site, they contact their quality control monitor, who may in turn revise their remaining Site Visitation Schedule to decrease unutilized time. Otherwise the data collectors proceed to Step 2.

Step 2: The data collectors jointly determine whether data can be collected at the site. Data collectors did not collect data at a given site if any of the following occurred:

- the site could not be located;
- the site was found upon visit to be of an ineligible site type;
- the manager on duty declined to allow the survey to be conducted at that time;
- the data collectors felt uncomfortable collecting data at the site due to a matter of personal safety (e.g., vagrants congregating in the parking lot);

- weather conditions (e.g., moderate to heavy rain) precluded data collection or would have caused a very low response rate; or
- (for sites other than gas stations) the site does not have a dedicated parking lot.

In these cases, data collectors notified their data collection monitor to await further instructions. The quality control monitor might reschedule given site or other sites as a consequence.

If data can be collected at the site, the data collectors proceed to Step 3.

Step 3: The data collectors jointly agree on values for the following survey variables, and record their values on page 1 of the data collection form. These variables are collected via observation.

- Observer Name
- Booklet Number
- Site Identification Number
- Site Type
- Site Name
- Street Address
- Start Time
- End Time
- Urbanization
- Weather Conditions

Step 4: Data collectors identify the location at which they will position themselves at the given site in order to collect data on vehicles and occupants, according to the instructions given in Section 5.3.2. E.g., if the given site is a fast food restaurant, the data collectors jointly determine whether one or both of them will utilize the Drive-Thru Paradigm for collecting data on the vehicles they will observe.

Step 5: Data collectors decide at this time who (possibly one data collector or both) will count the missed vehicles according to the guidance given in Section 5.8.3. The data collector(s) chosen for this task will tally vehicles according to the procedures specified in Section 5.8.3 as well as they can while simultaneously conducting Step 6.

Step 6: In the time remaining in the assigned two-hour time block for data collection at the site, each data collector conducts Steps 6.1 – 6.11 repeatedly and independently of his/her partner.

Step 6.1: The data collector goes to the location identified in Step 4.

Step 6.2: The data collector identifies (according to his/her subjective assessment) the closest passenger vehicle appearing to have at least one occupant under age 13, excluding any vehicle on which the partner data collector is collecting data and excluding vehicles on which the data collector or partner data collector (to the knowledge of the data collector following these instructions) has collected data.

Step 6.3: The data collector records by observation the following variables regarding the vehicle from Step 6.2, in the order listed below, until a Suspension Condition (defined in Section 5.8.4) arises or all of these variables have been collected:

- the variables Restraint Used and Seating Position of all occupants appearing to be under age 13, subject to the Seating Row Limits (defined in Section 5.8.4)
- Restraint Used, Age, and Gender of the driver

- Restraint Used, Seating Position, Age, and Gender of the remaining occupants, subject to the Seating Row Limits

The data collector uses the definitions from Section 5.4 in recording the variable Restraint Used.

Step 6.4: If the data collector is situated at a parking lot entrance, s/he follows the vehicle until it parks and the driver exits the vehicle.

Step 6.5: The data collector approaches the driver of the vehicle and recites the following text verbatim:

“Hi, my name is _____ from Westat, a national research organization. We are conducting a Booster Seat Survey for the National Highway Traffic Safety Administration. We would simply like to record the restraint use of everyone in your vehicle and ask some simple questions. All your responses and any observations I make are completely confidential.”

The data collector also offers the stickers from Section 5.8.1 to the driver for his/her participation in the survey.

Step 6.6: If the driver declines participation, the data collector terminates the data collection on this vehicle, records a hash mark in the “Refusals” section of the data collection form, and returns to Step 6.1.

If the driver consents to participate, the data collector proceeds to Step 6.7.

Step 6.7: The data collector collects the variable Vehicle Type.

Step 6.8: The data collector re-examines each vehicle occupant on whom data was collected in Step 6.3 from a closer standpoint to the vehicle if possible, without entering or reaching inside of the vehicle, and make corrections to the survey data collected in Step 6.3 on these occupants on the data collection form, with the following exception: Values of the variable Restraint Type recorded in Step 6.3 as other than “Unrestrained” cannot be changed in this step to “Unrestrained” even that the occupant in question appears now (or clearly is) unrestrained.

If the data collector is not sure whether one or more occupants is less than 13 years old, the data collector asks the driver whether this is the case. If a child the data collector thought in Step 6.3 was under 13 years turns out not to be so, the data collector revises the value of the variable Age accordingly and records the child’s gender (as they would have done if they had guessed the age correctly in Step 6.3). If a child whom the data collector thought in Step 6.3 was over 12 turns out not to be so, the data collector revises the variable Age accordingly.

Step 6.9: The data collector records the variables listed but not collected in Step 6.3. E.g., if in Step 6.3, a Suspension Condition arose while the data collector was collecting data on the last occupant appearing to be under age 13, she or he would record the remaining data for this occupant, and record the variables listed in Step 6.3 for the driver and occupants appearing to be over age 12. As in Step 6.8, the data collector asks the driver in any case where the data collector is not sure whether an occupant is under age 13.

Step 6.10: The data collector obtains the values of the following variables for all occupants, subject to the Seating Row Limits, by interviewing the driver using the questions from Section 5.7:

- Race;
- Ethnicity;
- Weight (for occupants appearing to be under age 13 years, only);

- Height (for occupants appearing to be under age 13 years, only);
- Age (for occupants appearing to be under age 13 years, only);
- Time Spent in Vehicle (for occupants appearing to be under age 13, only);
- Number of Visits to Gas Stations (for occupants appearing to be under age 13, only);
- Number of Visits to Daycare Centers (for occupants appearing to be under age 13, only);
- Number of Visits to Recreation Centers (for occupants appearing to be under age 13, only); and
- Number of Visits to Fast Food Restaurants (for occupants appearing to under age 13, only).

In obtaining the Race and Ethnicity variables, the data collector is instructed to show the driver the race/ethnicity card from Section 5.8.1.

If a person other than the driver volunteers an answer to a particular question, the data collector is instructed to use his/her judgment regarding the accuracy of the answer as to whether the non-driver's response appears to be more accurate. (E.g., a passenger might offer a response to the question on his/her ethnicity and the data collector uses his/her judgment as to whether this response is more accurate than any offered by the driver.)

Step 6.11: The data collector thanks the driver for his/her time and offers him/her the stickers and brochure from Section 5.8.1.

Step 7: The data collector tallies the total number of refusals and records this on the data collection form. If the data collector was recording the missed vehicles, s/he does likewise for the missed vehicles.

Note that in Step 6.8, the data collector is allowed to correct errors in the type of restraint used. E.g., if a child appeared to be in a front-facing child seat in Step 6.3, but the data collector discovers upon the closer inspection of Step 6.8 that the restraint is actually a high-backed booster seat, the data collector changes the value of the variable Restraint Used for this occupant to "High-Backed Booster Seat." What the data collector is not permitted to do in Step 6.8 is change a value other than "Unrestrained" in Step 6.3 to "Unrestrained" in Step 6.8. The reason for this is we wish to capture as accurately as possible the restraint use when the vehicle was on the road, and the child (or adult, in the case of seat belts) may have unfastened the restraint between the times that Steps 6.3 and 6.8 occurred.

Note that potential survey respondents are asked, in accordance with OMB requirements, for their voluntary participation in the survey and are assured of the confidentiality of their responses in Step 6.5.

At least some of survey variables (namely, some of the variables obtained by observation) were collected in Step 6.3, prior to the data collectors asking for cooperation in Step 6.5. As we noted in Section 5.3.2, this was done to capture restraint use before restraints are unfastened. However another advantage of this approach is that it allows us to examine the response bias, and to reduce the response bias of the estimates involving only observed data (as we will see in Chapter 8).

In order to increase response rates, the NSUBS uses a number of bilingual Spanish/English-speaking data collectors. Drivers who could not participate in the interview portion of the survey due to other language barriers were recorded as "Refusals" in Step 6.6.

6. Quality Control Procedures

6.1 *Pilot Testing of Data Collection Protocols*

All data collection protocols were rigorously tested in a pilot study conducted at voluntarily participating sites in Florida in 2005.

6.2 *Recruitment of Field Staff*

The contractor that conducted the 2006 NSUBS, WESTAT, Inc., has a field staff of thousands of data collectors who conduct numerous surveys. For the 2006 NSUBS, WESTAT selected staff who had prior experience conducting occupant restraint use surveys (such as the NOPUS and restraint use surveys conducted for the Insurance Institute for Highway Safety) and who had interviewing experience.

In all, the 2006 NSUBS used 32 data collectors and 8 quality control monitors. An additional 4 backup personnel who could serve as substitute data collectors or quality control monitors were also hired and trained for the study. The data collectors were paired into teams of two and assigned to collect data in a PSU relatively close to the part of the country in which they lived. Each data collection monitor was assigned to monitor the data collection in 2 geographically proximate PSUs.

The purpose of the data collectors is to collect all survey data. The purpose of the quality control monitors is to monitor data collection through unannounced site visits, and help as needed in securing cooperation from sites, answering questions from data collectors during data collection, and coordinate the rescheduling of sites for which reliable data could not be collected at their originally scheduled date and time (e.g., due to inclement weather).

6.3 *Training*

Training was conducted during the period July 12–14, 2006, ending just prior to the start of data collection on July 17, 2006.

All data collectors and quality control monitors received extensive training in protocols for interviewing motorists and observing restraint use in a manner that is professional and as unobtrusive as possible.

Training was conducted in two components, classroom training and field training.

Classroom training comprising the following topics was transmitted via PowerPoint presentations given by senior contractor employees who participated extensively in the pilot study and who answered questions posed by attendees during training:

- the data collection protocols from Section 5.9;
- tips for and sample scripts illustrating the successful recruitment of businesses and motorists to participate in the survey;
- techniques for conducting successful interviews;
- a list of questions (Frequently Asked Questions) that potential respondents might ask about the survey, together with answers;

- the authorization letters and photo identification cards; and
- the laminated card for the race/ethnicity questions.

Please see the Appendix for the Frequently Asked Questions and race/ethnicity cards used for the 2006 survey.

All data collectors and quality controls monitors also participated in role playing, in which they practiced recruiting and interviewing each other, with instructor feedback.

Actual child seats were used in training in order to facilitate training data collectors in the definition of the restraint use survey variable.

In the field training portion of training, data collectors and quality control monitors practiced the data collection protocols at sites near the training site at which prior cooperation had been secured, again with feedback from the instructors.

6.4 Pre-Collection Test of Data Collectors

In order to give the highest quality results in identifying restraint use for the various types of child restraints, training concluded with a written test on this topic given to all data collectors and quality control monitors. In the test, test takers were to identify the restraint use of all occupants as they could best ascertain in a series of photographs.

The results of the tests were as follows:

- The average score among all 44 test takers (the 32 data collectors, 8 quality control monitors, and 4 backup field personnel) was 95 percent.
- The average among the 8 quality control monitors was 97 percent.
- The average among the 36 data collectors and the 4 backup personnel was 92 percent.

Note that we trained 4 backup personnel, to allow for the possibility that data collectors could not conduct the survey during the scheduled data collection period for personal reasons, did not pass the pre-collection test, or were found to be unsuitable for the survey for whatever reason.

Of the 44 test takers, 5 scored lower than 75 percent. These people received additional training and their data collection was monitored during the first days of data collection to ensure high quality performance.

6.5 Contact Information for Questions

Data collectors were also provided with contact information (phone numbers) for their quality control monitor and other WESTAT staff, whom they could call with any questions that arise during data collection.

6.6 Unannounced Site Visits

The quality control monitors conducted unannounced site visits to monitor the quality of data collection their assigned PSUs.

7. Data Entry, Editing, and Imputation

7.1 Data Entry

Data entry and formation of the booklet-level and page-level files

Data from the paper forms used in the survey were entered manually into a Microsoft Access database, and contractor staff verified the correct entry of 25 percent of data forms and of all outlier values. The records in this database are defined at the level of the data collection form. (See Section 12.1 for the data collection forms.) That is, there is one and only one record for each form filled out by a data collector.

The booklet-level file

As there are two data collection forms, the database consists of two collections of records. One collection is produced from the form “Booster Seat Survey Recording Form” and contains one record for each booklet turned in by a data collector in which at least some information is recorded. Since there are at least two booklets for each of the two data collectors and each of the 383 sites that participated in the 2006 survey, this collection contains at least 766 records and exactly 12 variables (the 12 site variables).

The page-level file

The other collection of records contains all recorded data from the untitled data collection form (i.e., the form a copy of which appears as pages 1-20 of each booklet from Appendix 12.1.) This collection contains one record for each page of a data collection booklet (other than the cover page) on which information was recorded. Since a “page” records information on one vehicle from one of the two data collectors, and data was collected on 3,489 vehicles in the 2006 survey, this collection consists of 3,489 records. As the survey records 13 occupant variables and records data on up to 9 occupants per vehicle, this database contains 117 occupant variables. It also contains the Site Identification Number from the cover page of the booklet containing the page, the booklet number (again from the cover page of the booklet), the page number, and the vehicle variable, for a total of 121 variables.

These two collections of records (the booklet-level and page-level files), which existed as Microsoft Access data tables, were imported to SAS (Statistical Analysis Software) as SAS data sets.

Data reconciliation and formation of the Master File

Since the survey estimates are occupant-related (e.g., the percentage of 4- to 7-year-old children who were restrained in booster seats), we desire to compile the information in the above records to produce a single file (the *Master File*) containing one and only one record for each occupant on which at least some survey data was recorded and containing 28 variables (a PSU identifier, the 12 site variables in the site-level file above, and the page number and 14 vehicle and occupant variables from the page-level file above).

In order to produce such a file, we have to perform some data reconciliation on the site variables when the data collectors reported different values for a site variable (e.g. when the two data collectors at a site reported different weather conditions), create an occupant-level file from the page-level file, and then merge the occupant-level and booklet-level files by Site Identification Number. We perform the data reconciliation in the basically obvious way (e.g., choosing the non-missing value when one data collector doesn't record a value and his/her partner does, and arbitrarily choosing one data collector's value when both recorded a value and they differ).

Since the 2006 survey collected data on 9,955 occupants, the Master File contains 9,955 records and (as mentioned above) 28 variables.

7.2 Editing

No statistical editing was performed to alter the recorded values of outliers.

7.3 Imputation

The following provides a basic description of the imputation procedures used by the survey. We plan to provide additional detail in future methodology reports.

The survey used logical imputation and hot-deck imputation, except for a few variables that were not imputed and a few variables imputed as special cases.

7.3.1 Variables Imputed as Special Cases

The following 10 variables were imputed to have the following values.

Variables Imputed as Special Cases

Variable	Imputed Value ¹
Site Name	The site name listed in the file created when the sample was drawn
Street Address	The street address listed in the file created when the sample was drawn
Site Type	The site type listed in the sampling frame
Start Time	The time that data collection was scheduled to begin at the site, as listed on the Site Assignment Sheet
End Time	The time that data collection was scheduled to end at the site, as listed on the Site Assignment Sheet
Weather Conditions	Clear conditions
PSU Identification Number	The PSU Identification Number listed in the drawn sample
Site Identification Number	The Site Identification Number listed in the drawn sample
Booklet Number	Assigned in a logical manner to produce consecutive booklet numbers at each site
Page Number	Assigned in a logical manner to produce consecutive page numbers in each booklet for each site

¹ Recall from Section 7.1 that the Master File only contains a missing value for a given nonoccupant variable when neither of the two data collectors at a given site recorded a value.

7.3.2 Variables Imputed by Logical Imputation

The following 6 variables, which are all occupant variables, were imputed via logical imputation based on information from other occupants in the vehicle:

- Restraint Used;
- Race;
- Ethnicity;
- Age;
- Height (for occupants who appeared to be under age 13);
- Weight (for occupants who appeared to be under age 13);

E.g., if the races of some but not all occupants of a given vehicle were known than the races of the occupants with unknown race were randomly chosen from those of the occupants with known race.

7.3.3 Variables Not Imputed

The following variables, which are the vehicle variable, 4 of the site variables, and one of the occupant variables, were not imputed:

- Observer Name;
- Urbanization;
- Number of Refusals;
- Number of Missed Vehicles;
- Vehicle Type; and
- Seating Position.

Note that the Seating Position variable can be determined from the data collection form and so no imputation is required.

7.3.4 Variables Imputed by Hot-Deck Imputation

All remaining variables (namely, the 6 remaining occupant variables) were imputed via hot-deck imputation. These variables are:

- Gender;
- Time Spent in Vehicle;
- Number of Visits to Gas Stations;
- Number of Visits to Fast Food;
- Number of Visits to Recreation Centers; and
- Number of Visits to Daycare Centers.

The circumstances under which these variables were imputed

We imputed for the missing values of these survey variables except in the following case:

- (NI) For a given record R with a missing value for a given variable V among these 6 variables, we did not impute for V in record R precisely when the values of all 10 interviewed occupant variables were missing in record R. (This occurred precisely when the driver either declined the entire interview portion of the survey or declined to provide any responses for a given occupant.)

Following we describe the imputation of all variables when imputation occurred (i.e., when the condition [NI] was not met).

Imputation for these 6 occupant variables

Missing values for these 6 occupant variables, except in the case where condition (NI) above applies, were imputed via hot-deck imputation. Donor groups for each variable were formed using some combination of the following variables: county group, age, sex, height, driver ethnicity, child ethnicity, driver race, and child race.

8. Estimation

8.1 Estimator Design

In general one can estimate at least three types of parameters that in some way measure the use of a device in vehicles, such as booster seats. One could estimate the percent of travel time occupants spend using the device (we shall call this the “time-based estimate”), the percent of miles occupants travel using the device (an estimate we are not interested in for this publication and so shall not name), or one could estimate the percent of occupants using the device at a random time. We shall call the latter the “snapshot estimate.” One could visualize the snapshot estimator as representing what we would see if we placed an all-seeing camera over the entire United States and took a photograph of all vehicular occupants at a randomly chosen time.

It is easily seen that these parameters are different. For instance, consider the simple example of 2 drivers on a block of a particular street (which we shall call Main Street) during a particular time period (e.g., 8–10 a.m. on Monday, November 20, 2006). Driver 1 is driving on (this block of) Main Street during 8–9 a.m. and is not belted. (Driver 1 then exits Main Street at 9 a.m., not to return.) Driver 2 is on Main Street during the entire 8–10 a.m. period and is belted the entire time. The snapshot estimate of use on Main Street between 8–10 a.m. is 75 percent, while the time-based estimate is 66 percent (as 2 of the 3 person-hours of driving were spent belted).

We naturally desire the NSUBS estimator to be consistent with that for NOPUS. The NOPUS estimator is as follows: (For simplicity, we present only the estimator of belt use nationwide, as the subnational estimators and estimators of other restraint types are, of course, similar.)

$$\text{Belt use} = \frac{\sum_k w_k F_k S_k B_k}{\sum_k w_k F_k S_k O_k}$$

where k runs over the observation sites; w_k denotes the inverse of the selection probability for site k ; F_k denotes the product of various nonresponse adjustment factors (see Glassbrenner, 2002, for more information); B_k denotes the number of belted occupants observed at site k ; O_k denotes the total number of occupants observed at site k ; and $S_k := L_k/s_k t_k$, where L_k (respectively, s_k , t_k) denotes the length of the road segment corresponding to site k in the selection of the NOPUS sample (respectively, the estimated speed of the vehicles observed at site k , the duration of the observation period at site k). (One might restrict the terms B_k and O_k to occupants who appear to be over the age of 7, since this restriction, although it is immaterial for our point on estimation, is used in NOPUS.)

Although it may not be initially obvious, the NOPUS estimator produces a snapshot of use, namely the percent of occupants on U.S. roadways who are belted at a random (daylight, as NOPUS of course observes during daytime) time. To see this, it may be useful to consider the analogy of balls traveling the length of a chute. As the NOPUS observers are observing at a point on a road segment, we shall consider balls observed at some point (for convenience, balls coming out the end of the chute). If x balls are observed at the end of a chute of length L during t minutes, and the balls are uniformly spaced traveling at the same, constant speed s , the number of balls on the chute at a randomly chosen time is Lx/st . Thus the numerator of the NOPUS estimator is the number of “belted balls” on the chute (i.e.,

belted occupants on the road segment) at a random time, and the denominator similarly estimates the (total) number of balls on the chute in a snapshot. Thus, NOPUS is a snapshot estimator. Its reference population is the set of vehicles on all roads (subject to certain modest frame exclusions employed by the survey) in the United States at a given point in time.

Thus we wish NSUBS to have a snapshot estimator as well. However, note that in NSUBS, there are no “road segments” as NOPUS has, only sites. Thus there is no factor corresponding to the term “L/st” from the previous paragraph, and the NSUBS estimator is simply as follows (expressed below using our “i-j-k” notation for PSUs, strata, and sites from Section 4.3 and Section 11). (Again we present only the estimator of booster seat use nationwide among 4- to 7-year-old children, as the other survey estimators are similar.)

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

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 defined below (which do not include a “travel time” factor); 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 .

The next sections define the various adjustment factors, of which there are several, used by the NSUBS. In Section 8.5, we will apply these to give the formula for the survey’s estimates. All adjustment factors are defined for members of the refined sample.

Some Notation

It will be useful to establish the following notation. Let C denote a characteristic of occupants (e.g., C might denote being of the age 4-7 years) and let R denote a restraint type (e.g., a booster seat). Let Use_{CR} denote the percentage of occupants restrained in restraint type R among those occupants having characteristic C . (All survey estimates produced by the NSUBS are restraint use rates and thus have this form.)

Some Terminology

The adjustment factors used in estimation will utilize the following terminology.

For a given occupant age 0-12 observed at a given site, we define the child to have a *complete interview* if interview variables were obtained for the child in Step 6.10 of Section 5.9, a *partial interview* if the variable age was collected (for the child) but at least one interview variable was not collected (for the child), and *no interview* if age was not collected (for the child).

Note that a child for whom only the ethnicity (or only the ethnicity and restraint use) was collected is considered to have no interview. The reason for this curious definition is that we will apply a simplified nonresponse adjustment to adjust for interview data not obtained, as we found in the 2006 survey that there were very few children on which some interview variables were collected, but not age. So, e.g., when estimating the restraint use of 4- to 7-year-old Hispanic children, rather than troubling to adjusting for children on which ethnicity or age was not obtained, we shall for simplicity only adjust for the age nonresponse.

A site is considered *eligible* if it is a member of the refined sample and is determined in Step 2 of Section 5.9 to be safe to collect data at, and, for sites other than gas stations, to have a dedicated parking lot. Note that the eligibility of some members of the refined sample is unknown. A recreation center that did not respond to our requests in Section 5.1.1 to conduct the survey at their establishment is an example of a site with unknown eligibility.

A site is considered to have *participated* in the survey, if it is a member of the refined sample and Step 3 of Section 5.9 was performed at the site.

Note that all sites that participated in the survey are eligible, as Step 3 of Section 5.9 is only performed at sites found in Step 2 of Section 5.9 to be eligible.

We define a vehicle to be *eligible* if it is a passenger vehicle containing at least one child occupant under the age of 13.

We define a vehicle to be *observed* if in Step 6.3 of Section 5.9 the data collector recorded the restraint use of at least one occupant assessed to be under the age of 13.

8.2 Adjustment for Variation in Duration of Data Collection

The data collectors may have for a variety of reasons collected data at a given site for a period of time that is longer or shorter than the scheduled 2 hours. If site k of stratum j in PSU i is a member of the refined sample, we define its *duration adjustment factor* $DurAdj_{ijk}$ to be

$$DurAdj_{ijk} := \frac{120}{Dur_{ijk}}$$

where $Dur_{ijk} :=$ the duration in minutes that the data collectors collected data from site k of stratum j of PSU i .

8.3 Nonresponse Adjustment Factors

The NSUBS employs standard unit nonresponse adjustment, i.e. applying the ratio of total cases to known cases. The survey has three types of response S_1, S_2, S_3 to adjust for, each defined on a different type of unit u_1, u_2, u_3 . Namely,

$S_1 :=$ participated, $u_1 :=$ eligible site

$S_2 :=$ observed, $u_2 :=$ vehicle

$S_3 :=$ have complete or partial interview, $u_3 :=$ observed child occupant

E.g., the first type of nonresponse is defined for eligible sites, and an eligible site is considered to “respond” if it participated in the survey.

For each type of response (i.e. for each pair (S_t, u_t) for $1 \leq t \leq 3$), we define some number of nonresponse cells, that together partition the refined sample. We shall denote the nonresponse cells for the pair (S_t, u_t) as

$$NRC(S)_1^{(t)}, \dots, NRC(S)_{Q_t}^{(t)}$$

where Q_t denotes the number of nonresponse cells for the response type pair (S_t, u_t) .

In each case (i.e., for all three values of t), the nonresponse cells consists of PSU strata (the portions thereof in the refined sample) or unions thereof. We shall denote the members of the nonresponse cells as the ordered triples (i,j,k) corresponding to the site k of stratum j of PSU i that lies in the nonresponse cell.

We note that the nonresponse cells for the 2006 NSUBS were defined in such a way that $Q_1 = 52$, $Q_2 = 559$ (i.e., the size of the refined sample), and $Q_3 = 54$. (That is, the nonresponse cells used to adjust for unobserved vehicles are precisely the members of the refined sample.)

Due to the hierarchical nature of the unit types u_1, u_2, u_3 , we shall need to define the nonresponse adjustment factors $NRAdj(S)_s^{(t)}$ for $1 \leq s \leq Q_t$ recursively.

For each $1 \leq t \leq 3$, $1 \leq i \leq 16$, $1 \leq j \leq 4$, and each value of k for which site $k \in \text{RefSamp}_{ij}$, let $\text{Tot}_{ijk}^{(t)}$ denote the number of units of type u_t at site k of stratum j of PSU i , and let $\text{Resp}_{ijk}^{(t)}$ denote the number of units of type u_t at site k of stratum j of PSU i for which we have a response for S_t .

E.g., if $j=2$ denotes the gas station stratum and the 3rd gas station in the sampling frame for PSU 1 is in the refined sample, then $\text{Tot}_{123}^{(1)}$ equals 1 or 0 depending on whether this gas station was eligible (i.e., was considered by the data collectors to be safe to collect data at), while $\text{Resp}_{123}^{(1)}$ equals 1 or 0 depending on whether this gas station participated in the survey.

Note that if $t > 1$ and if site k did not participate in the survey, then both $\text{Tot}_{ijk}^{(t)}$ and $\text{Resp}_{ijk}^{(t)}$ are zero.

The case of $t=1$ will be a special case, because we can only estimate the numerator of the nonresponse adjustment factors, as the eligibility of some sites is unknown. We define the nonresponse factors for $t=1$ (i.e., for $S_1 = \text{participate}$ and $u_1 = \text{eligible site}$) to be:

$$NRAdj(S)_s^{(1)} := \frac{\sum_{(i,j,k) \in NRC(S)_s^1 \cap Elig} w_{ijk} + e_s}{\sum_{(i,j,k) \in NRC(S)_s^1 \cap Elig} \text{Resp}_{ijk}^{(1)} w_{ijk}} \quad \text{for } 1 \leq s \leq Q_1$$

where $Elig := \{(i,j,k): \text{site } k \text{ in stratum } j \text{ of PSU } i \text{ is known to be eligible}\}$, $UnknownElig := \{(i,j,k): \text{the eligibility of site } k \text{ in stratum } j \text{ of PSU } i \text{ is unknown}\}$, and

$$e_s := \frac{\sum_{(i,j,k) \in NRC(S)_s^1 \cap Elig} w_{ijk}}{\sum_{(i,j,k) \in NRC(S)_s^1 \setminus UnknownElig} w_{ijk}} \quad \text{for } 1 \leq s \leq Q_1$$

Since vehicles are a smaller unit than sites, we shall need to incorporate the nonresponse factors for $t=1$ into those for $t=2$. That is, we define the nonresponse factors for $t=2$ (i.e., for $S_2 = \text{observed}$ and $u_2 = \text{vehicle}$) to be:

$$NRAdj(S)_s^{(2)} := \frac{\sum_{(i,j,k) \in NRC(S)_s^2} Tot_{ijk}^{(2)} NRAdj(S)_{s_1(i,j,k)}^{(1)} w_{ijk}}{\sum_{(i,j,k) \in NRC(S)_s^2} Resp_{ijk}^{(2)} NRAdj(S)_{s_1(i,j,k)}^{(1)} w_{ijk}} \quad \text{for } 1 \leq s \leq Q_2$$

where for $1 \leq v \leq 3$, $s_v(i,j,k)$ denotes the value a for which $(i,j,k) \in NRC(S)_a^{(v)}$

Similarly the nonresponse factors for $t=3$ incorporates those for $t=1$ and $t=2$. That is, we define the nonresponse factors for $t=3$ (i.e., for $S_3 =$ have complete or partial interview and $u_3=$ observed child occupant) to be:

$$NRAdj(S)_s^{(3)} := \frac{\sum_{(i,j,k) \in NRC(S)_s^3} Tot_{ijk}^{(3)} \prod_{v=1}^2 NRAdj(S)_{s_v(i,j,k)}^{(v)} w_{ijk}}{\sum_{(i,j,k) \in NRC(S)_s^3} Resp_{ijk}^{(3)} \prod_{v=1}^2 NRAdj(S)_{s_v(i,j,k)}^{(v)} w_{ijk}} \quad \text{for } 1 \leq s \leq Q_3$$

8.4 Weight Trimming

Consider the product of the sampling weight w_{ijk} together with the adjustment factors from Sections 8.2 and 8.3, i.e. consider:

$$w'_{ijk} := w_{ijk} \prod_{v=1}^3 NRAdj(S)_{s_v(i,j,k)}^{(v)} DurAdj_{ijk}$$

Considering this as a “weight”, we shall define two associated trimmed weights, one used when the characteristic C involves at least one interview variable and one when it doesn't.

To define the two trimmed trimming factors, define

$$\bar{w}^{OBS} := \frac{1}{|RefSamp|} \sum_{(i,j,k) \in RefSamp} ObsChild_{ijk} \frac{w'_{ijk}}{NRAdj(S)_{s_3(i,j,k)}^{(3)}}, \text{ and}$$

$$\bar{w}^{INT} := \frac{1}{|RefSamp|} \sum_{(i,j,k) \in RefSamp} IntChild_{ijk} w'_{ijk}$$

where $RefSamp := \bigcup_{i=1}^{16} \bigcup_{j=1}^4 RefSamp_{ij}$ denotes the refined sample from Section 4.2.3, and for each i,j,k

for which $k \in RefSamp_{ij}$, $ObsChild_{ijk}$ (respectively, $IntChild_{ijk}$) denotes the number of children observed (respectively, have a complete or partial interview) at site k of stratum j of PSU i . (Note that if site k did not participate in the survey, then both $ObsChild_{ijk}$ and $IntChild_{ijk}$ are zero.)

We shall trim the “weight” back to 4.5 times the corresponding weighted mean of these “weights”. Namely, we define the weight-trimming factors $Trim_{ijk}^{OBS}$ and $Trim_{ijk}^{INT}$ as:

$$Trim_{ijk}^{OBS} := \begin{cases} 1, & \text{if } w'_{ijk} \leq 4.5 \bar{w}^{OBS} \\ \frac{4.5 \bar{w}^{OBS}}{w'_{ijk}}, & \text{otherwise} \end{cases}$$

$$Trim_{ijk}^{INT} := \begin{cases} 1, & \text{if } w'_{ijk} \leq 4.5 \bar{w}^{INT} \\ \frac{4.5 \bar{w}^{INT}}{w'_{ijk}}, & \text{otherwise} \end{cases}$$

8.5 The Estimation Formula

We are now ready to write down the estimation formula. Let C denote a characteristic of occupants (e.g., C might denote being of age 4 to 7) and let R denote a restraint type (e.g., a booster seat). We shall next provide the formula for the survey’s estimate Use_{CR} of the percentage of occupants restrained in restraint type R among those occupants having characteristic C. (All survey estimates produced by the NSUBS are restraint use rates and thus have this form.)

The NSUBS estimator of Use_{CR} is as follows:

Estimates Involving Interview Data

If C involves at least one variable obtained by interview, then we estimate Use_{CR} by the following formula:

$$Use_{CR} := \frac{\sum_{(i,j,k) \in Re fSamp} Trim_{ijk}^{INT} w_{ijk} \prod_{v=1}^3 NRAdj(S)_{s_v(i,j,k)}^{(v)} DurAdj_{ijk} B_{ijk}}{\sum_{(i,j,k) \in Re fSamp} Trim_{ijk}^{INT} w_{ijk} \prod_{v=1}^3 NRAdj(S)_{s_v(i,j,k)}^{(v)} DurAdj_{ijk} O_{ijk}}$$

where for each (i,j,k) in RefSamp, B_{ijk} denotes the number of occupants of characteristic C restrained in restraint type R at site k of stratum j in PSU i, and O_{ijk} denotes the number of occupants of characteristic C at site k of stratum j in PSU i.

Examples of survey estimates that would be computed using this formula would be:

- booster seat use among 4- to 7-year-old children (since the ages of children are obtained by interview),
- the use rate for front-facing child safety seats among children weighing 20 to 40 pounds who appear to be under age 13 (as weight is obtained by interview), and
- restraint use among Hispanic children (as ethnicity is obtained by interview).

Estimates Involving No Interview Data

For all other values of C (i.e., when all variables involved in C are obtained by observation), then we estimate Use_{CR} to be:

$$Use_{CR} := \frac{\sum_{(i,j,k) \in Re\ fSamp} Trim_{ijk}^{OBS} w_{ijk} \prod_{v=1}^2 NRAj(S)_{s_v(i,j,k)}^{(v)} DurAdj_{ijk} B_{ijk}}{\sum_{(i,j,k) \in Re\ fSamp} Trim_{ijk}^{OBS} w_{ijk} \prod_{v=1}^2 NRAj(S)_{s_v(i,j,k)}^{(v)} DurAdj_{ijk} O_{ijk}}$$

Examples of survey estimates that would be computed using this formula would be:

- restraint use among children who appear to be under the age of 13,
- belt use among 16- to 24-year-olds, and
- restraint use among children taken to gas stations who appear to be under age 13 (since restraint use, the ages of occupants who appear to be over age 12, and whether or not an occupant is under 13 are assessed by observation).

8.6 *Estimates Computed*

The survey computes estimates of restraint use by a variety of characteristics derived from the survey variables listed in Section 5.4. For instance the survey estimates restraint use by age and restraint type, and by height and restraint type. The following publications present the major estimates from the 2006 survey: Glassbrenner and Ye, DOT HS 810 796, August 2007; Glassbrenner and Ye, DOT HS 810 797, August 2007; Glassbrenner and Ye, DOT HS 810 798, August 2007.

8.7 *Definitions of Categories Used in Estimates*

Although the NSUBS collects children's individual ages, heights, and weights, we combine these results into categories in order to produce reliable estimates.

Age categories

The NSUBS uses the following age categories: 0, 1-3, 4-7, 8-12, 13-15, 16-24, 25-69, and 70 and above. The choice of these age groups is motivated by consistency with the NOPUS survey, which uses the age groups 0, 1-3, 4-7, 8-12, 13-15, 16-24, 25-69, and 70 and above, combined with taking into account that the NSUBS collects interview data on children ages 0-12.

Height and weight categories

The NSUBS uses the following height categories: under 36 inches tall, 37-53 inches, 54-56 inches, and 57 inches or taller. The survey uses the weight categories 0-19 pounds, 20-40 pounds, 41-60 pounds, and 61 pounds or heavier. These categories were chosen because they are used in NHTSA's recommendation for the choice of restraint use for children.

Regional categories

The 16 PSUs selected in the NSUBS constitute a probability sample of PSUs (counties and groups thereof) in the United States. The data is not sufficient to produce State-by-State results. However NSUBS can and does produce regional estimates using the following categories:

Northeast: ME, VT, NH, MA, RI, CT, NY, PA, NJ
Midwest: MI, OH, IN, IL, WI, MN, IA, MO, KS, NE, SD, ND
South: WV, MD, DE, VA, KY, TN, NC, SC, GA, FL, AL, MS, AR, LA, OK, TX, DC
West: AK, WA, OR, CA, NV, ID, UT, AZ, NM, CO, WY, MT, HI

These definitions of the four NSUBS regions are the same regional definitions used in the NOPUS. The NSUBS regional categories were chosen to be the same as the NOPUS categories for the purpose of consistency.

Time of day and day of week categories

The NSUBS uses the following day of week and time of day categories, which are the same used for the 2006 NOPUS:

Weekday Rush Hour: 8-10 a.m. and 3:30-6 p.m. on Monday-Friday

Weekday Outside of Rush Hour: 10 a.m.–3:30 p.m. on Monday-Friday

Weekend: 8 am–6 pm on Saturday and Sunday

8.8 *A Note on the Race/Ethnicity Estimates*

When computing estimates by race and/or race/ethnicity, multiracial occupants are excluded (i.e., we did not impute a single race for persons reporting they are multiracial). Also we had to collapse some race and race/ethnicity categories in order to comply with NHTSA standards for reliability in publishing estimates. (See Section 10 for more information on these standards.) A common situation in which collapsing categories was necessary was in estimating use rates among Hispanic non-Whites.

9. Variance Estimation

This section documents the variance estimation procedures utilized by the survey, without providing the motivation for these procedures. We plan to provide the motivation for these procedures in a subsequent methodology report.

Two methods for calculating variances were employed, depending on whether the estimator whose variance is being estimated is based on relatively few observations (and thus is such that direct estimates of variances are not reliable).

Variance estimates for estimates in the following table were computed directly via jackknife variance estimation:

Estimates Whose Variances Were Estimated Directly

Restraint Use Among Children Age 0-12
Restraint Use Among Children Age 0-12 in Passenger Cars
Restraint Use Among Children Age 0-12 in Vans and Sport Utility Vehicles
Restraint Use Among Children Age 0-12 in Pickup Trucks
Restraint Use Among Children Age 0-12 in the Front Seat
Restraint Use Among Children Age 0-12 in the Second Row of Seats
Restraint Use Among Children Age 0-12 in Pickup Trucks in the Third Row of Seats
Restraint Use Among Children Age 0-12 Driven by a Driver Age 16 to 24
Restraint Use Among Children Age 0-12 Driven by a Driver Age 25 to 69
Restraint Use Among Children Age 0-12 Driven by a Driver Age 70 or Older ³
Restraint Use Among Children Age 0-12 Driven by a Driver Whose Ethnicity Is Hispanic or Latino and Whose Race Is American Indian or Alaska Native ³
Restraint Use Among Children Age 0-12 Driven by a Driver Whose Ethnicity Is Hispanic or Latino and Whose Race Is Asian ³
Restraint Use Among Children Age 0-12 Driven by a Driver Whose Ethnicity Is Hispanic or Latino and Whose Race Is Black or African-American ³
Restraint Use Among Children Age 0-12 Driven by a Driver Whose Ethnicity Is Hispanic or Latino and Whose Race Is Native Hawaiian or Other Pacific Islander ³
Restraint Use Among Children Age 0-12 Driven by a Driver Whose Ethnicity Is Hispanic or Latino and Whose Race Is White
Restraint Use Among Children Age 0-12 Driven by a Driver Whose Ethnicity Is not Hispanic nor Latino and Whose Race Is American Indian or Alaska Native ³
Restraint Use Among Children Age 0-12 Driven by a Driver Whose Ethnicity Is not Hispanic nor Latino and Whose Race Is Asian ³
Restraint Use Among Children Age 0-12 Driven by a Driver Whose Ethnicity Is not Hispanic nor Latino and Whose Race Is Black or African-American
Restraint Use Among Children Age 0-12 Driven by a Driver Whose Ethnicity Is not Hispanic nor Latino and Whose Race Is Native Hawaiian or Other Pacific Islander ³
Restraint Use Among Children Age 0-12 Driven by a Driver Whose Ethnicity Is not Hispanic nor Latino and Whose Race Is White
Restraint Use Among Children Age 0-12 Driven by a Male Driver
Restraint Use Among Children Age 0-12 Driven by a Female Driver
Restraint Use Among Children Age 0-12 Driven by a Belted Driver

Restraint Use Among Children Age 0-12 Driven by an Unbelted Driver
Restraint Use Among Children Age 0-12 Months ³
Restraint Use Among Children Age 1-3 Years
Restraint Use Among Children Age 4-7 Years
Restraint Use Among Children Age 8-12 Years
Restraint Use Among Boys Age 0-12
Restraint Use Among Girls Age 0-12
Restraint Use Among Children Age 0-12 Whose Ethnicity Is Hispanic or Latino and Whose Race Is American Indian or Alaska Native ³
Restraint Use Among Children Age 0-12 Whose Ethnicity Is Hispanic or Latino and Whose Race Is Asian ³
Restraint Use Among Children Age 0-12 Whose Ethnicity Is Hispanic or Latino and Whose Race Is Black or African-American ³
Restraint Use Among Children Age 0-12 Whose Ethnicity Is Hispanic or Latino and Whose Race Is Native Hawaiian or Other Pacific Islander ³
Restraint Use Among Children Age 0-12 Whose Ethnicity Is Hispanic or Latino and Whose Race Is White
Restraint Use Among Children Age 0-12 Whose Ethnicity Is not Hispanic nor Latino and Whose Race Is American Indian or Alaska Native ³
Restraint Use Among Children Age 0-12 Whose Ethnicity Is not Hispanic nor Latino and Whose Race Is Asian ³
Restraint Use Among Children Age 0-12 Whose Ethnicity Is not Hispanic nor Latino and Whose Race Is Black or African-American
Restraint Use Among Children Age 0-12 Whose Ethnicity Is not Hispanic nor Latino and Whose Race is Native Hawaiian or Other Pacific Islander ³
Restraint Use Among Children Age 0-12 Whose Ethnicity Is not Hispanic nor Latino and Whose Race Is White
Restraint Use Among Children Age 0-12 Whose Height Is Under 36 Inches
Restraint Use Among Children Age 0-12 Whose Height Is 37-53 Inches
Restraint Use Among Children Age 0-12 Whose Height Is 54-56 Inches
Restraint Use Among Children Age 0-12 Whose Height Is 57 Inches or More
Restraint Use Among Children Age 0-12 Whose Weight Is Under 19 Pounds ³
Restraint Use Among Children Age 0-12 Whose Weight Is 20-40 Pounds
Restraint Use Among Children Age 0-12 Whose Weight Is 41-60 Pounds
Restraint Use Among Children Age 0-12 Whose Weight Is At Least 61 Pounds
Restraint Use Among Children Age 0-12 in Vehicles in Light Precipitation
Restraint Use Among Children Age 0-12 in Vehicles in Fog ³
Restraint Use Among Children Age 0-12 in Vehicles in Clear Weather Conditions
Restraint Use Among Children Age 0-12 in Vehicles in the Northeast ¹
Restraint Use Among Children Age 0-12 in Vehicles in the Midwest ¹
Restraint Use Among Children Age 0-12 in Vehicles in the South ¹
Restraint Use Among Children Age 0-12 in Vehicles in the West ¹
Restraint Use Among Children Age 0-12 in Vehicles in Urban Areas
Restraint Use Among Children Age 0-12 in Vehicles in Suburban Areas
Restraint Use Among Children Age 0-12 in Vehicles in Rural Areas
Restraint Use Among Children Age 0-12 in Vehicles During Weekday Rush Hour ²
Restraint Use Among Children Age 0-12 in Vehicles on Weekdays Outside of Rush Hour ²
Restraint Use Among Children Age 0-12 in Vehicles on Weekdays
Restraint Use Among Children Age 0-12 in Vehicles on Weekends
Restraint Use Among Children Age 0-12 in Vehicles at Gas Stations

Restraint Use Among Children Age 0-12 in Vehicles at Fast Food Restaurants
Restraint Use Among Children Age 0-12 in Vehicles at Daycare Centers
Restraint Use Among Children Age 0-12 in Vehicles at Recreation Centers

¹See Section 8.3 for the definitions of the NSUBS regional categories of Northwest, Midwest, South, and West.

²See Section 8.3 for the definition of rush hour.

³These variables had fewer than 200 observations in the 2006 survey.

For these estimates we also computed the within- and between-PSU variances (again directly, through jackknife variance estimation), and we calculated the average of the ratios of the total variance to the within-PSU variance for all estimates in this table having 200 or more observations.

$$R = \frac{1}{48} \sum_{i=1}^{48} \frac{Var(Y_i)}{WVar(Y_i)} \quad (1)$$

where Y_i denotes the i^{th} member of the table “Estimates Whose Variances Were Computed Directly” that have 200 or more observations, and for a random variable Y defined on the NSUBS sample, $Var(Y)$ (respectively, $WVar(Y)$) denotes the jackknife-calculated estimate of the variance of Y (respectively, the jackknife-calculated estimate of the within-PSU variance of Y).

The value of R from the 2006 survey data was 3.12.

The variances of all other estimates were computed by calculating the within-PSU variance of the variable via jackknife and multiplying by the ratio from (1), i.e.:

$$Var(Y) = R \times WVar(Y)$$

for each estimate Y other than those in the table “Estimates Whose Variances Were Estimated Directly”.

10. Rules for Suppressing Estimates in Publications

In order not to publish estimates that are not sufficiently reliable, NHTSA employs the following suppression rule for the NSUBS:

NSUBS Suppression Rule

Use estimates whose numerator is based on fewer than 5 persons observed, whose denominator is based on fewer than 30 persons observed, or that are not statistically different from 0 percent use (i.e. the standard error is at least half the point estimate) are to be suppressed. These should be reported as “NA” in publications, and any related estimates (i.e., change in use and confidence estimates) should also be suppressed.

This the same rule used for the NOPUS survey.

11. Glossary of Terms

The following records terms and acronyms that are defined in this report:

Complete interview: For a given occupant up to age 12 observed at a given site, we define the child to have a *complete interview* if values of all interview variables were obtained for the child in Step 6.10 of Section 5.9.

Duplicate (or *duplicate site*): a member of the site sampling frame formed in Section 4.2.1 identifying the same establishment as another member of this frame.

Eligible site: a site in the refined sample determined in Step 2 of Section 5.9 to be safe to collect data at, and, for sites other than gas stations, to have a dedicated parking lot.

Eligible site type: the following 4 types of establishments: fast food restaurant, gas station, daycare center, recreation center.

Eligible vehicle: a passenger vehicle containing at least one child occupant under age 13.

Interviewed occupant variables (or for short, *interview variables*): The 10 survey variables that pertain to occupants and whose values are obtained by interview for at least some occupants, namely Ethnicity, Race, Weight, Height, Time Spent in Vehicle, Number of Visits to Gas Stations, Number of Visits to Fast Food, Number of Visits to Recreation Centers, Number of Visits to Daycare Centers, and Age.

Interview variables (or *interviewed occupant variables*): The 10 survey variables that pertain to occupants and whose values are obtained by interview for at least some occupants, namely Ethnicity, Race, Weight, Height, Time Spent in Vehicle, Number of Visits to Gas Stations, Number of Visits to Fast Food, Number of Visits to Recreation Centers, Number of Visits to Daycare Centers, and Age.

Master File: The file created in Section 7.1, which contains one record for each occupant observed in the survey and all 28 survey variables.

Missed vehicle: An eligible vehicle observed according to the protocols in Section 5.8.3 on which no data is collected by either data collector in Step 6 of Section 5.9.

No interview: For a given occupant up to age 12 observed at a given site, we define the child to have *no interview* if the no value for the variable Age was obtained for the child in Step 6.10 of Section 5.9, or the driver declined to give an interview in Step 6.6 of Section 5.9.

NOPUS: The National Occupant Protection Use Survey.

NOPUS PSU: a primary sampling unit for the NOPUS.

NSUBS: The National Survey of the Use of Booster Seats.

The *NSUBS probability sample* (or for short, the *probability sample*): The collection of 627 gas stations, recreation centers, daycare centers, and fast food restaurants selected in Steps 1-3 of Section 4.2.2

The *NSUBS refined sample* (or for short, the *refined sample*): The collection of 559 gas stations, recreation centers, daycare centers, and fast food restaurants resulting from Section 4.2.3.

Observed occupant variables: The 4 survey variables that pertain to occupants and whose values are obtained by observation for at least some occupants, namely Gender, Seating Position, Restraint Used, and Age.

Observed vehicle: a vehicle on which a data collector recorded the restraint use of at least one occupant assessed to be under the age of 13 in Step 6.3 of Section 5.9.

Partial interview: For a given occupant up to age 12 observed at a given site, we define the child to have a *partial interview* if a value for the variable Age was obtained for the child in Step 6.10 of Section 5.9, and there was at least one interview variable whose value was not obtained for the child in Step 6.10 of Section 5.9.

Participating site: A member of the refined sample for which Step 3 of Section 5.9 was performed.

Passenger vehicle: a passenger car, van, sport utility vehicle, or pickup truck.

The *probability sample* (or for emphasis, the *NSUBS probability sample*): The collection of 627 gas stations, recreation centers, daycare centers, and fast food restaurants selected in Steps 1-3 of Section 4.2.2

PSU, or *NSUBS PSU*: a primary sampling unit for the NSUBS.

The *refined sample* (or for emphasis, the *NSUBS refined sample*): The collection of 559 gas stations, recreation centers, daycare centers, and fast food restaurants resulting from Section 4.2.3.

Refusal: An eligible vehicle whose driver declined to answer interview questions in Step 6.6 of Section 5.9.

Sample PSU (e.g., “the 16 sample PSUs”): one of the 16 NSUBS PSUs selected in Step 3 of Section 4.1.2.

Sampling weight: the inverse of the site selection probability.

Seating Row Limits: The condition, used in Section 5.9, that data is to be collected from no more than three occupants per row of seats in the vehicle, with no more than two non-driving occupants in the first row, and from only the first three rows of the vehicle.

Site Assignment Sheet: A sheet identifying the sites the data collectors are to visit, their locations, and when they are to visit them

Site sampling frame: the sampling frame of gas stations, recreation centers, daycare centers, and fast food restaurants in the 16 sample PSUs formed in Section 4.2.1.

Site sampling frame restrictions: The following set of four restrictions, which apply to establishments that are gas stations, fast food restaurants, daycare centers, and recreation centers:

- 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.

Site selection probability: the approximation to the site selection probability calculated in Section 4.3.

Site variables: The 12 survey variables that pertain to the data collection site or data collection conditions, namely Observer Name, Booklet Number, Site Identification Number, Site Type, Site Name, Street Address, Start Time, End Time, Urbanization, Weather Conditions, Number of Refusals, Number of Missed Vehicles.

Site visitation schedule: The schedule determined in Section 5.2, which specifies the dates and times at which the survey is to be conducted at each member of the refined sample.

The strata: The stratification of the NSUBS sampling frame by the four site types: gas stations, recreation centers, daycare centers, and fast food restaurants.

Suspension Conditions: The following three conditions, which are used in Section 5.9:

- the vehicle is sufficiently out of view that observations can no longer be conducted (in the case where the data collector is stationed at an entrance to the site's parking lot);
- the vehicle reaches the fast food drive-thru window (if the data collector is at a fast food drive-thru)
- an occupant initiates contact of some sort with the data collector (e.g. asks "What are you doing?", "Do I know you?", "Can I help you?", engages eye contact with the data collector, etc).

SUV: sport utility vehicle.

Vehicle variable: The sole survey variable that pertains to vehicles, namely Vehicle Type.

12. Glossary of Notation

For convenience, we record in the following notation that is defined in this report.

i : denotes a sample PSU, and takes an integer value between 1 and 16

j : denotes a stratum, and takes an integer value between 1 and 4

m_{ij} : denote the number of members in the probability sample in the j^{th} stratum of the i^{th} PSU.

m_{1ij} (respectively, m_{2ij}): denotes the number of members of the probability sample for stratum j of the NSUBS PSU i in the selection of 323 sites in the probability sample selected in Step 1 of Section 4.1.2 (respectively, the 302 sites selected in Step 2)

m_{3ij} : takes the value 1 if stratum j of PSU i contains one of the two sites selected in Step 3 of Section 4.2.1, and 0 otherwise.

M_{ij} : denotes the number of sites in the site sampling frame for stratum j of PSU i .

The members of the site sampling frame are sorted as follows. List the M_{ij} sites in stratum j of PSU i of the site sampling frame in a manner such that:

- the first m_{1ij} sites in the list consist of the sites (in this stratum and PSU) that were selected in Step 1 of Section 4.2.2, followed by
- the m_{2ij} sites selected in Step 2 of Section 4.2.2, followed by
- the m_{3ij} sites selected in Step 3 of Section 4.2.2, followed by
- the $M_{ij} - m_{ij}$ sites that are not in the probability sample.

q_i : denotes the probability of selection of the NOPUS PSU containing (the NSUBS) PSU i ,

δ_i : takes the value 1 if the NOPUS PSU containing PSU i is one of the two certainty PSUs identified in Step 1 of Section 4.1.2, and 14/48 otherwise,

Pop_i : denotes the population in 2000 of children under age 5 in PSU i ,

$TotPop_i$: denotes the population in 2000 of children under age 5 in the NOPUS PSU containing PSU i ,

$$P'_{ijk} := \begin{cases} q_i \delta_i \frac{Pop_i}{TotPop_i} \frac{m_{1ij} + m_{2ij}}{M_{ij}} & \text{for } 1 \leq k \leq m_{1ij} + m_{2ij} \text{ and } m_{ij} < k \leq M_{ij} \\ q_i \delta_i \frac{Pop_i}{TotPop_i} & \text{for } m_{1ij} + m_{2ij} < k \leq m_{ij} \end{cases}$$

r_{ijk} : denotes the number of occurrences of site k of stratum j of PSU i in the site sampling frame.

$$p_i := q_i \delta_i \frac{Pop_i}{TotPop_i}$$

$$p''_{k|ij} := 1 - \left(1 - \frac{P'_{ijk}}{p_i}\right)^{r_{ijk}} \text{ for all } 1 \leq i \leq 16, 1 \leq j \leq 4, 1 \leq k \leq M_{ij}$$

$$p''_{ijk} := p_i p''_{k|ij} \frac{\sum_{l=1}^{m_{ij}} \frac{1}{p''_{l|ij}}}{2M_{ij} - \sum_{k=1}^{M_{ij}} r_{ijk}} \text{ for all } 1 \leq i \leq 16, 1 \leq j \leq 4, 1 \leq k \leq M_{ij}$$

$$w_{ijk} = \frac{1}{p_{ijk}} = \frac{2M_{ij} - \sum_{k=1}^{M_{ij}} r_{ijk}}{p_i p''_{ijk} \sum_{l=1}^{m_{ij}} \frac{1}{p''_{l|ij}}} \text{ for all } 1 \leq i \leq 16, 1 \leq j \leq 4, 1 \leq k \leq M_{ij}$$

RefSamp_{ij}: denotes the collection of members of the refined sample that are in stratum j of PSU i

Dur_{ijk} := the duration in minutes that the data collectors collected data from site k of stratum j of PSU i.

$$DurAdj_{ijk} := \frac{120}{Dur_{ijk}}, \text{ defined for all } i, j, \text{ and for } k \in \text{RefSamp}_{ij}:$$

S₁, S₂, S₃: denote three types of response, each defined on a different type of unit u₁, u₂, u₃.

Q_t: denotes the number of nonresponse cells for the response type pair (S_t, u_t).

NRC(S)₁^(t), ..., NRC(S)_{Q_t}^(t): denote the nonresponse cells for the response type pair (S_t, u_t).

Tot_{ijk}^(t): denotes the number of units of type u_t at site k of stratum j of PSU i, defined for each 1 ≤ t ≤ 3, 1 ≤ i ≤ 16, 1 ≤ j ≤ 4, and each value of k for which site k ∈ RefSamp_{ij}.

Resp_{ijk}^(t): denote the number of units of type u_t at site k of stratum j of PSU i for which we have a response for S_t, defined for each 1 ≤ t ≤ 3, 1 ≤ i ≤ 16, 1 ≤ j ≤ 4, and each value of k for which site k ∈ RefSamp_{ij}.

Elig := {(i,j,k): site k in stratum j of PSU i is known to be eligible}

UnknownElig := {(i,j,k): the eligibility of site k in stratum j of PSU i is unknown}

$$e_s := \frac{\sum_{(i,j,k) \in NRC(S)_s^1 \cap Elig} w_{ijk}}{\sum_{(i,j,k) \in NRC(S)_s^1 \setminus UnknownElig} w_{ijk}} \text{ for } 1 \leq s \leq Q_1$$

$$NRAdj(S)_s^{(1)} := \frac{\sum_{(i,j,k) \in NRC(S)_s^1 \cap Elig} w_{ijk} + e_s}{\sum_{(i,j,k) \in NRC(S)_s^1 \cap Elig} Resp_{ijk}^{(1)} w_{ijk}} \quad \text{for } 1 \leq s \leq Q_1$$

$s_a(i,j,k)$: denotes the value v for which $(i,j,k) \in NRC(S)_v^{(a)}$, defined for $1 \leq a \leq 3$

$$NRAdj(S)_s^{(2)} := \frac{\sum_{(i,j,k) \in NRC(S)_s^2} Tot_{ijk}^{(2)} NRAdj(S)_{s_1(i,j,k)}^{(1)} w_{ijk}}{\sum_{(i,j,k) \in NRC(S)_s^2} Resp_{ijk}^{(2)} NRAdj(S)_{s_1(i,j,k)}^{(1)} w_{ijk}} \quad \text{for } 1 \leq s \leq Q_2$$

$$NRAdj(S)_s^{(3)} := \frac{\sum_{(i,j,k) \in NRC(S)_s^3} Tot_{ijk}^{(3)} \prod_{v=1}^2 NRAdj(S)_{s_v(i,j,k)}^{(v)} w_{ijk}}{\sum_{(i,j,k) \in NRC(S)_s^3} Resp_{ijk}^{(3)} \prod_{v=1}^2 NRAdj(S)_{s_v(i,j,k)}^{(v)} w_{ijk}} \quad \text{for } 1 \leq s \leq Q_3$$

$$w'_{ijk} := w_{ijk} \prod_{v=1}^3 NRAdj(S)_{s_v(i,j,k)}^{(v)} DurAdj_{ijk}$$

$$RefSamp := \bigcup_{i=1}^{16} \bigcup_{j=1}^4 RefSamp_{ij}$$

$ObsChild_{ijk}$ (respectively, $IntChild_{ijk}$): denotes the number of children observed (respectively, have a complete or partial interview) at site k of stratum j of PSU i , defined for all i,j,k for which $k \in RefSamp_{ij}$.

$$\bar{w}^{OBS} := \frac{1}{|RefSamp|} \sum_{(i,j,k) \in RefSamp} ObsChild_{ijk} \frac{w'_{ijk}}{NRAdj(S)_{s_3(i,j,k)}^{(3)}}$$

$$\bar{w}^{INT} := \frac{1}{|RefSamp|} \sum_{(i,j,k) \in RefSamp} IntChild_{ijk} w'_{ijk}$$

$$Trim_{ijk}^{OBS} := \begin{cases} 1, & \text{if } w'_{ijk} \leq 4.5 \bar{w}^{OBS} \\ \frac{4.5 \bar{w}^{OBS}}{w'_{ijk}}, & \text{otherwise} \end{cases}$$

$$Trim_{ijk}^{INT} := \begin{cases} 1, & \text{if } w'_{ijk} \leq 4.5 \bar{w}^{INT} \\ \frac{4.5 \bar{w}^{INT}}{w'_{ijk}}, & \text{otherwise} \end{cases}$$

C: denotes a characteristic of vehicle occupants

R: denotes a type of restraint

B_{ijk} denotes the number of occupants of characteristic C restrained in restraint type R at site k of stratum j in PSU i, defined for each $(i,j,k) \in \text{RefSamp}$.

O_{ijk} denotes the number of occupants of characteristic C at site k of stratum j in PSU i, defined for each $(i,j,k) \in \text{RefSamp}$.

$$Use_{CR} := \frac{\sum_{(i,j,k) \in \text{Re } fSamp} Trim_{ijk}^{INT} w_{ijk} \prod_{v=1}^3 NRAj(S)_{s_v(i,j,k)}^{(v)} DurAdj_{ijk} B_{ijk}}{\sum_{(i,j,k) \in \text{Re } fSamp} Trim_{ijk}^{INT} w_{ijk} \prod_{v=1}^3 NRAj(S)_{s_v(i,j,k)}^{(v)} DurAdj_{ijk} O_{ijk}}, \text{ defined for all R and for those C}$$

that involve at least one variable obtained by interview.

$$Use_{CR} := \frac{\sum_{(i,j,k) \in \text{Re } fSamp} Trim_{ijk}^{OBS} w_{ijk} \prod_{v=1}^2 NRAj(S)_{s_v(i,j,k)}^{(v)} DurAdj_{ijk} B_{ijk}}{\sum_{(i,j,k) \in \text{Re } fSamp} Trim_{ijk}^{OBS} w_{ijk} \prod_{v=1}^2 NRAj(S)_{s_v(i,j,k)}^{(v)} DurAdj_{ijk} O_{ijk}}, \text{ defined for all R and for those C}$$

that involve only variables obtained by observation.

13. References

Boyle, J., & Vanderwolf-Schulman, P. (2005, March). 2003 Motor Vehicle Occupant Safety Survey, Volume 5: Child Safety Seat Report. Publication no. DOT HS 809 858. Washington, DC: National Highway Traffic Safety Administration.

Cody, B. E., Mickalide, A. D., Paul, H. P., & Colella, J. M. (2002, February). Child Passengers at Risk in America: A National Study of Restraint Use. Washington, DC: National SAFE KIDS Campaign [now called Safe Kids Worldwide].

Reports, Forms, and Recordkeeping Requirements, Agency Information Collection Activity under OMB Review, Federal Register Notice, Volume 71, Number 30, page 7824, February 14, 2006.

Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity, Federal Register Notice, Volume 62, Number 210, pages 58781-58790, October 30, 1997.

Glassbrenner, D. (2002, September). Safety Belt and Helmet Use in 2002 – Overall Results. DOT HS 809 500. Washington, DC: National Highway Traffic Safety Administration.

Glassbrenner, D. (2005, February). Child Restraint Use in 2004 – Overall Results. DOT HS 809 845. Washington, DC: National Highway Traffic Safety Administration.

Glassbrenner, D. (To appear). The 2006 National Occupant Protection Use Survey – Methodology Report, Washington, DC: National Highway Traffic Safety Administration.

Glassbrenner, D. (To appear). The Sample Redesign for the National Occupant Protection Use Survey Washington, DC: National Highway Traffic Safety Administration.

Glassbrenner, D., & Ye, J. (2007, February). Child Restraint Use in 2006 – Overall Results, National Highway Traffic Safety Administration, DOT HS 810 737. Washington, DC: National Highway Traffic Safety Administration.

Glassbrenner, D., & Ye, J. (2007, August). Booster Seat Use in 2006. DOT HS 810 796. Washington, DC: National Highway traffic Safety Administration.

Glassbrenner, D., & Ye, J. (2007, August). Child Restraint Use in 2006 – Demographic Results DOT HS 810 797. Washington, DC: National Highway Traffic Safety Administration.

Glassbrenner, D., & Ye, J. (2007, August). Child Restraint Use in 2006 –Use of Correct Restraint Types. DOT HS 810 798. Washington, DC: National Highway Traffic Safety Administration.

NHTSA. (2006, December). Motor Vehicle Traffic Crash Fatality Counts and Estimates of People Injured for 2005. DOT HS 810 639. Washington, DC: National Highway Traffic Safety Administration.

Office of Management and Budget, Update of Statistical Area Definitions and Guidance on Their Uses, OMB BULLETIN NO. 05-02, February 2005.

Partners for Child Passenger Safety. (2004, July). *The Forgotten Child*. Philadelphia: The Children's Hospital of Philadelphia.

Partners for Child Passenger Safety. (2005, October). *Fact and Trend Report*. Philadelphia: The Children's Hospital of Philadelphia.

Public Law 106-414, 106th Congress, *Transportation Recall Enhancement, Accountability, and Documentation (TREAD) Act*, 114 STAT. 1800, November 1, 2000.

Public Law 107-318, 107th Congress, *Anton's Law*, 114 STAT. 1800, December 4, 2002.

14. Appendix

14.1 Data Collection Forms

The following forms were used by the data collectors to record the survey data. Each data collector was given one “booklet” of forms for each site visit, plus additional booklets that they could use at a given site if necessary (this will occur if they record data on more than 20 vehicles at a given site).

A “booklet” consisted of the form “Booster Seat Survey Recording Form” (displayed below, modified so as to keep anonymous the names of the five fast food chains that participated in the survey) as a cover page, followed by 20 copies of the untitled form displayed below following the Booster Seat Survey recording form (i.e., the form that displays the survey’s OMB number).

The pages of the booklet were numbered. The displayed second form below, which has also been modified to keep anonymous the participating fast food chains, shows Page 1.

Hi, My name is _____ from Westat, a national research organization. We are conducting a Booster Seat Survey for the National Highway Traffic Safety Administration. We would simply like to record the restraint use of everyone in your vehicle and ask some simple questions. All your responses and any observations I make are completely confidential.

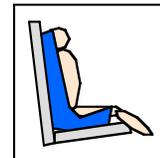
Can you tell me the Age of this child? Height? Weight?

Survey Questions

1. In the past week, how many minutes did this child spend in a vehicle driven by you?
2. In the past week how many times did you visit a gas station when this child was in a vehicle with you?
3. In the past week how many times did you go to a (Chain 1) with this child in a vehicle with you?
(Chain 2)? (Chain 3)? (Chain 4)? (Chain 5)?
4. In the past week how many times did you go to a recreation center when this child was in a vehicle with you?
Which one(s) ?
5. In the past week, how many times did you go to a day care center with this child in a vehicle with you?
Which one(s) ?

Booster Seat Survey Recording Form

Booklet # ____ of ____



Date: _ _ / _ _ / _ _	PSU #: _____	Site # : _____
Start Time for Booklet _____ AM PM		End Time for Booklet _____ AM PM

Site Type: Gas Station Fast Food Day Care Recreation Center
Site Name _____
Street Address _____

Weather : Light Precipitation Light Fog Clear
Area : Urban Suburban Rural

Total Misses: ____	Observer Name: _____
Total Refusals: ____	

PSU #	Site #	MM / DD / YY	Car Pickup Van/SUV	Booklet # ___ of ___	Page 1				
1st Row			Question Key Adult Age Group Restraint use Race [B W O] Gender [M F]	2nd Row			3rd Row		
Driver	Middle	Right		Driver Side	Middle	Right	Driver Side	Middle	Right
Teen (T) (13-15) Young (Y) (16-24) Adult (A) (25-69) Senior (S) (70+)	/	/	Children ONLY (≤ 12 yrs) Age	/	/	/	/	/	
			Height						
			Weight						
			1. Time in Vehicle						
			2. Gas station						
			(Chain 1)						
			(Chain 2)						
			(Chain 3)						
			(Chain 4)						
			(Chain 5)						
			3. FAST FOOD Places in the Front Vehicle						
			4. Rec. Ctr. (Name)						
			5. Day Care (Name)						

14.2 Letter of Authorization

The following letter of authorization was provided to the data collectors to show to people questioning their authority to conduct the survey.



U.S. Department
of Transportation
**National Highway
Traffic Safety
Administration**

400 Seventh Street, S.W.
Washington, D.C. 20590

March 22, 2006

To Whom It May Concern:

Westat is under contract to the National Highway Traffic Safety Administration, U.S. Department of Transportation, to conduct a survey on child transportation characteristics. The data collection phase of the survey will take place from July 16, 2006, through August 1, 2006. It will consist of identifying vehicles with child passengers and conducting brief interviews with those drivers at selected sites across the country.

This county has been selected as one of the 16 areas across the United States that will be surveyed. Information from this survey will be used to help design programs that improve the safety of child passengers in motor vehicles.

Please direct any questions you may have to the Westat Project Director, [redacted], at [redacted]. Thank you very much for your support of this important research program.

Sincerely,

Donna Glassbrenner, Ph.D.

Program Manager

14.3 Incentives

The following incentives, which were stickers, were offered to potential respondents as incentives to participate in the survey.



14.4 Frequently Asked Questions

The following list of questions and answers was provided to data collectors to help answer any questions that potential respondents or others might have about the survey.

How was I selected for the survey?

As you drove on to the property I noticed that there were children in the vehicle.

How many people are you interviewing?

We are interviewing approximately 4,800 drivers of vehicles with child passengers throughout the United States.

What is the purpose of this study? What is this survey about?

This survey will allow the government to compute national estimates of child safety seat use.

How long will this take?

The survey takes about 5 minutes to complete.

How will the study results be used? / What will you do with this information?

The survey will identify how children of different age groups are restrained when riding in a vehicle.

How do I know you will keep this information confidential?

We are not collecting any personal identifying information. We will not be asking for names or recording license plate information.

Why do you need to know the height and weight of my child(ren)?

The selection of appropriate child safety seats is dependent on age, height, and weight.

Why do you need to know how often my child goes to a gas station, fast food restaurant, daycare center, or recreation center?

Department of Transportation will use this information to generate national estimates.

How will the results be published? Will the results be made available?

The Department of Transportation will publish the results of the study. However, that can take up to a year or more. I can take your name and address and we can send you the study results when they are available.

Do I have to do this? / Do I have to answer this survey? / I don't want to do this.

You do not have to respond, but your help is very important to us. The information will provide better national estimates on child safety seat use. You may refuse to answer any question at any time.

Why don't you ask someone else?

We are attempting to stop every adult who is driving with children in the vehicle to better understand child safety seat use.

I had a bad experience recently with someone taking a survey, so I don't think I want to participate.

I'm sorry that your experience was unpleasant. We hope to make your contact with us a pleasant and interesting experience. This is a legitimate research effort, in which your responses will help us to learn about the use of child safety seats in passenger vehicles.

I think this whole business is stupid. The federal government could better spend my money. The money for this study could be spent more wisely, etc.

[Occasionally you will encounter an argumentative respondent. In spite of their statements, this is usually a person who is interested in the study, but wants to talk about what he feels before consenting to complete the survey. Bear with him and hear him out! As long as he keeps talking, he has not refused to do the survey. Do not argue: simply make short, neutral comments to let him know you are listening.]

What is the authority/sponsor for this study?

The Department of Transportation is sponsoring this study. This survey is authorized by the United States Code, Title 49, Section 111(c)(2).

Who do you work for?

I work for Westat, a survey research firm in the Washington, DC, area, and we have been contracted by the Department of Transportation to conduct this study.

Who can I call at the Department of Transportation?

If you have questions about your rights as a person who is part of this study, please call the DOT at: [redacted]. Please leave a short message with your name, phone number, and mention that you are calling about the Booster Seat Survey. Someone will return your call as soon as possible.

How do I know the survey is legitimate? / How do I know that you are really an interviewer for this study?

If you would like, you can speak to my supervisor, or I can give you a toll-free number to call at your convenience. The toll free number is: [redacted] and ask to speak with [redacted]. I also have an authorization letter with the toll-free number on it.

Does this survey have approval from the Office of Management and Budget (OMB)? / What is the OMB number?

Yes, the study has been approved by the Office of Management and Budget (OMB). The approval number assigned to the study is 2127-0644 (it is listed on the side of the recording forms).

14.5 Card for Race/Ethnicity Questions

The following information was provided to data collectors on a laminated card to show to interviewees to aid in answering the race and ethnicity questions.

Are you of Hispanic or Latino origin?

- 1) Yes
- 2) No

What is your race? Please select one or more.

- 1) White
- 2) Black or African-American
- 3) Asian
- 4) Native Hawaiian or other Pacific Islander
- 5) American Indian or Alaska Native

National Survey of the Use of Booster Seats
OMB Control No. 2127-0644

DOT HS 811 111
April 2009



U.S. Department
of Transportation
**National Highway
Traffic Safety
Administration**

★★★★★
NHTSA

www.nhtsa.gov