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

NHTSA
www.nhtsa.gov

DOT HS 811 807

July 2013

NASS-CDS:
Sample Design and Weights

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.

Suggested APA Citation:

Zhang, F., & Chen, C-L. (2013, July). NASS-CDS: Sample Design and Weights. (Report No. DOT HS 811 807). Washington, DC: National Highway Traffic Safety Administration.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS 811 807		

4. Title and Subtitle	5. Report Date
NASS-CDS: Sample Design and Weights	July 2013
	6. Performing Organization Code: NVS-421
7. Author(s): Fan Zhang, Chou-Lin Chen	8. Performing Organization Report No.
9. Performing Organization Name Mathematical Analysis Division, National Center for Statistics and Analysis National Highway Traffic Safety Administration U.S. Department of Transportation 1200 New Jersey Avenue SE. Washington, D.C. 20590	10. Work Unit No. (TRAIS) 11. Contract or Grant No.
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 1200 New Jersey Avenue SE. Washington, D.C. 20590	13. Type of Report and Period Covered NHTSA Technical Report 14. Sponsoring Agency Code

15. Supplementary Notes

The authors would like to thank Charles (Mike) Fleming for his work on sampling and estimation methodologies of NASS-CDS. The authors would also like to thank Nancy Bondy and Ruby Li for sharing their knowledge of NASS-CDS.

Abstract

This report presents the sample design and weighting methodology used in NCSA's Crashworthiness Data System.

17. Key Words	18. Distribution Statement
National Automotive Sampling System, Crashworthiness Data System, NASS-CDS Sample Design, CDS Weights.	Document is available to the public from the National Technical Information Service www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	19	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorize

I. Introduction

The National Automotive Sampling System (NASS) is a nationwide motor vehicle crash data collection program operated by the National Center for Statistics and Analysis (NCSA) at the National Highway Traffic Safety Administration. The purpose of NASS is to produce statistics on national motor vehicle crashes and to provide a database for the development and the evaluation of motor vehicle and highway safety standards. See Web page www.nhtsa.gov/NASS for an overview of NASS.

NASS consists of two subsystems: the General Estimates System (GES) and the Crashworthiness Data System (CDS) – both probability sample surveys of police-reported motor vehicle traffic crashes across the country providing national estimates on crash, vehicle, and occupant characteristics. The GES collects information on motor vehicle traffic crashes from an annual sample of approximately 50,000 police crash reports (PARs). Shelton (1991) provided a detailed discussion of GES sampling and weighting procedures. The CDS, on the other hand, collects more detailed crash, vehicle, and occupant information, in addition to PARs, from a smaller annual sample of approximately 5,000 crashes by investigating motor vehicle crashes through interviews, medical records, vehicle inspection, and scene inspection. Fleming (2010) provided a detailed account of CDS sampling and weight calculation procedure as they were followed in CDS weighting production computer programs.

The purpose of this document is to explain how CDS weights are developed from its sample design. Section II is about CDS target population. Section III gives a description of CDS sample design. Section IV discusses the weight calculation procedure in CDS.

II. CDS Target Population

The NASS Crashworthiness Data System objectives are:

- Assess the overall state of traffic safety and identify the existing and potential traffic safety problems;
- Obtain detailed data on the crash performance of passenger cars, light trucks, vans, and sport utility vehicles with GVWRs less than 10,000 lbs.;
- Evaluate vehicle safety systems and designs;
- Increase knowledge about the nature of crash injuries as well as the relationship between the type and seriousness of a crash and the resultant injuries; and
- Assess the effectiveness of motor vehicle and traffic safety program standards.

From these objectives, CDS target population is defined as police-reported motor vehicle traffic crashes involving at least one passenger car, pickup, van, or SUV (also called CDS applicable vehicles) that was towed from the scene due to damage.

A PAR is a report filled out by a police officer at the scene of a motor vehicle traffic crash. PARs make up the national sampling frame for NASS.

NASS CDS applicable vehicles are passenger cars, pickups, vans, or SUVs with gross vehicle weight ratings less than 10,000 lbs.

A NASS CDS-eligible crash must have at least one in-transport CDS-applicable vehicle involved in the crash that was towed away from the crash scene due to damage as indicated by the police report of the crash.

Damage is road vehicle damage that precludes departure of the vehicle from the scene of the accident in its usual operating manner by daylight after simple repairs (see ANSI D16.1, Section 2.3.11).

For more detailed discussion of CDS target population, see “Sampling Manual for the National Automotive Sampling System” (NHTSA, 2011).

III. CDS Sample Design

CDS sampling has a stratified, multiphase, unequal selection probability design. In this section, we give a brief description of the current CDS sample design.

Primary Sample Selection

The current CDS primary sampling unit (PSU) sample is related to the original NASS PSU sample first implemented in 1979. The original NASS was one system instead of two. In the original NASS, PSU was defined as a city, a county, or a group of counties with population size as close to 50,000 as possible. The PSU frame was stratified into 10 strata by urbanization and per capita service station sales. The PSU sample was selected by a stratified probability proportional to size (PPS) sampling method using estimated 1973 PSU population size as the measure of size (MOS). By 1986, NASS had 50 PSUs. See Kahane (1976) for more detail about the 1979 NASS PSU sample selection.

In 1988, NASS PSU sample was reselected as part of the effort to divide NASS into two components: the GES and the CDS. For the first stage of NASS sample selection, the United States was partitioned into 1,195 PSUs. A NASS PSU can be a county, a group of counties, a central city, or a portion of a large county excluding a central city. The PSUs were then stratified into 12 strata consisted of four geographic regions (Northeast, South, Central, and West) and three urbanization types (Large Central City, Large Suburban Area, and All Others). In 1988, a total of 36 PSUs were selected as both NASS-CDS and NASS-GES PSU sample from the 12 strata with at least two PSUs per stratum. The selection probabilities are proportional to the PSUs' 1983 injury and fatal counts. No certainty PSUs were selected. To reduce the cost, a Keyfitz-based procedure (Keyfitz, 1951) was used to maximize the overlap of the 36 PSUs in the new CDS PSU sample with the 50 PSUs in the existing NASS PSU sample. As the result, among the 36 selected PSUs, 30 were from the existing NASS PSU sample, 6 were new PSUs. After the NASS-CDS PSU sample was selected, an additional 24 PSUs were selected to supplement NASS-GES PSU sample. The current GES has total 60 PSUs. See Bondy and Rhea (1997) for more detail about CDS and GES PSU sample selection.

In 1991, because of budget reduction, the 36 CDS PSU sample was reduced to 24 PSUs with 2 PSUs per stratum by a PPS sub-sampling from strata with more than 2 CDS PSUs using 1983 injury and fatality count. These 24 PSUs are the currently used CDS PSUs. Therefore, the current CDS PSU sample can be viewed as the result of a two-phase PPS sampling. Table 1 lists the 1991 CDS PSU frame and sample sizes by strata.

The selection probabilities for the 36 PSUs selected in 1988 can be expressed as:

$$\pi_{hi}^{(36)} = n_h \frac{m_{hi}}{\sum_{all\ i\ in\ h} m_{hi}}.$$

Here hi refers to PSU i in stratum h, $\pi_{hi}^{(36)}$ is the selection probability of the PSU hi for the 1988's 36 PSU sample, n_h is PSU sample size for stratum h, m_{hi} is 1983 injury and fatal count for PSU hi.

The 24 PSU sample of 1991 were selected as following: if a PSU stratum has 2 PSUs in the 1988 PSU sample, both PSUs were selected with certainty; if a PSU stratum has more than 2 PSUs in the 1988 PSU sample, 2 PSUs were selected from the 1988 PSU sample using PPS sampling with m_{hi} as MOS (Bondy, N., personal communication, 2012). Specifically, the PPS selection of PSU sample was carried out in the following manner. First, if PSU i of PSU stratum h has a measure of size $m_{hi} \geq \sum_{all\ 1988\ PSUs\ in\ h} m_{hi}/2$ (here 2 is the PSU sample size of PSU stratum h), then PSU hi is selected with certainty. And PSU sample size for PSU stratum h is reduced from 2 to 1. Then another PSU was selected with PPS sampling. The corresponding selection probabilities are:

$$\pi_{hi}^{(24)} = \begin{cases} 1, & if\ hi\ is\ a\ certainty\ PSU; \\ \dfrac{m_{hi}}{\sum_{non\ certainty\ 1988\ PSUs\ in\ h} m_{hi}}, & otherwise. \end{cases}$$

Here $\pi_{hi}^{(24)}$ are the second phase selection probabilities of the 1991 PSUs. On the other hand, if $m_{hi} < \sum_{all\ 1988\ PSUs\ in\ h} m_{hi}/2$ for all PSUs in stratum h, then no certainty PSU is selected. Then the second phase PSU selection probability becomes:

$$\pi_{hi}^{(24)} = \frac{2 * m_{hi}}{\sum_{all\ 1988\ PSUs\ in\ h} m_{hi}}.$$

The overall selection probabilities for the current 24 PSUs are the result of a two-phase sampling. The overall PSU selection probability is the product of the two-phase selection probabilities: (see for example, Särndal et al., 1992, Chapter 9):

$$\pi_{hi} = \pi_{hi}^{(24)} * \pi_{hi}^{(36)}.$$

CDS PSU design weight w_{hi} is the inverse of π_{hi}: $w_{hi} = 1/\pi_{hi}$.

Table 1: 1991 CDS Frame and Sample Sizes by Strata

Strata	PSU Frame Size	PSU Sample Size
Northeast		
Large Central City	19	2
Large Suburban	69	2
All Other	94	2
Midwest		
Large Central City	15	2
Large Suburban	79	2
All Other	280	2
South		
Large Central City	21	2
Large Suburban	87	2
All Other	368	2
West		
Large Central City	12	2
Large Suburban	42	2
All Other	109	2
Total	**1,195**	**24**

Secondary Sample Selection

The secondary sample selection in CDS is the selection of police jurisdictions (PJs). Within each selected PSU, PARs were assembled and stored by police jurisdictions. Therefore, it is operationally more efficient to treat PJs as the next stage sampling units. All PJs within the selected PSU constitute the sampling frame of the secondary sampling. For each PJ, the number of instances in which at least one death, or at least one incapacitating injury, or at least one non-incapacitating injury had occurred (abbreviated by KAB) was calculated using 1992 data for each PJ. Table 2 lists NASS defined injury types. PJs were then sorted by KAB in a descending order. A probability sample of PJs within each selected PSU was selected using PPS sampling with PJ's 1992 KAB as measure of size. An average of 7 PJs were selected within each PSU. Total about 170 PJs were selected. The current PJ sample was selected in 1995 using 1992 KAB data. PJ sample selection was conducted independently for each selected PSU. Therefore, CDS secondary sample selection resembles a one stage PPS sample selection of PJs. Specifically, the PPS selection of PJ sample was carried out in the following manner. First, if PJ j of PSU hi has a measure of size $KAB_{hij} \geq \sum_{all\ j\ in\ hi} KAB_{hij}/n_{hi}$ (here n_{hi} is the original PJ sample size in PSU hi), then PJ hij is selected with certainty. And PJ sample size for PSU hi is reduced from the original size n_{hi} to $n_{hi} - 1$. This process is performed repeatedly until there is no more certainty PJ. Assuming the total number of certainty PJs is c_{hi}, a sample of non-certainty PJs of size $n_{hi} - c_{hi}$ was then selected from the remaining non-certainty PJs using PPS with MOS=KAB_{hij}. The PJ selection probability can be expressed as:

$$\pi_{j|hi} = \begin{cases} 1, & if\ hij\ is\ a\ certainty\ PJ; \\ (n_{hi} - c_{hi})\ \dfrac{KAB_{hij}}{\sum_{all\ noncertainty\ PJs\ in\ hi} KAB_{hij}}, & otherwise. \end{cases}$$

The inverse of $\pi_{j|hi}$ is the PJ design weight: $w_{j|hi} = 1/\pi_{j|hi}$.

Table 2: NASS-Defined Injury Types

NASS Injury Type	Description
K	Fatal Injury
A	Incapacitating Injury
B	Non-incapacitating Injury
C	Possible Injury
O	No Injury
U	Injured but Severity Unknown

Source: National Automotive Sampling System – NASS Sampling, Part One, 2000. Transportation Safety Institute, Highway Safety Division, NASS Training Branch DTI-71.

Tertiary Sample Selection

CDS tertiary sample selection is the selection of PARs. In addition to the information on PARs, CDS also collects more detailed crash, vehicle and occupant information. Researchers conduct vehicle and scene inspections and interview people involved in the crashes. CDS PAR selection is performed weekly so that the evidence is still intact and the memory of the people involved in the crash is still fresh. Once a week, researchers visit all the selected PJs to obtain all new PARs accumulated since last visit. In this sense, the PARs in the selected PJs are first stratified by the weeks of the year.

Within each week stratum k, all PARs accumulated since last visit are then reviewed and classified into 14 "PAR strata": A, B, C, D, E, F, G, H, J, K, X, Y, Z, N (see Appendix A for a narrative definition). Among them, ten PAR strata are CDS applicable PARs (A, B, C, D, E, F, G, H, J, K). See Table 3 for detailed definitions. This process is called PAR listing. The CDS sample is selected from the CDS applicable PARs. The classified PARs are entered into a computer system for further sub-sampling.

In some large PJs there are too many PARs to be examined and listed by the researchers working at the PSU. For these PJs, only 1 of every 2, or 1 of every 5 PARs are examined and listed. These PJs are called "sub-listing PJs." Sub-listing PJs may vary from year to year. See Table 4 for a list of sub-listing PJs in 2010.

In the PAR listing process, only the information about the time of the collision and the PAR strata identification information are recorded for classification in order to select subsample of PARs. Therefore, PAR listing can be viewed as one phase of PAR selection. For the sub-listing PJs, the listed PARs are a systematic sample of all PARs (Li, R., personal communication, 2011). For the non-sub-listing PJs, the listed PARs are selected with certainty. Therefore, each listed PAR is selected with probability 1, 0.5 or 0.2. The inclusion probability of PAR r in PJ hij for this first phase of PAR sampling is:

$$\pi_{kr|hij}^{(1st)} = \begin{cases} 1, & non\ sub-listing\ PJs; \\ 0.2, & PSU\ 72, PJ\ 1-6; \\ 0.5, & all\ other\ sub-listing\ PJs. \end{cases}$$

The corresponding weight for the first phase of PAR sampling is: $w_{kr|hij}^{(1st)} = \left(\pi_{kr|hij}^{(1st)}\right)^{-1}$.

Every quarter, researchers visit the non-sampled PJs in the selected PSU (PJs that are not included in the weekly visits schedule) to collect PAR counts by PAR strata. All new PARs since last visit will be listed by PAR strata. The total PAR counts by PAR strata will be used later in post-stratification.

Table 3: NASS-CDS PAR Strata (NASS 2011)

Late Model Year (LMY) Motor Vehicle Involvement	MOST SEVERE POLICE REPORTED INJURY IN A TOWED CDS-APPLICABLE MOTOR VEHICLE								
	FATAL INJURY	TRANSPORTED						NOT TRANSPORTED	
		Serious Injury "A"					Injury Or Unknown If Injured "B," "C" or "U"	Injured, Not Injured or Unknown, If Injured "A," "B," "C," "O" or "U"	
		Single CDS Applicable Vehicle		Multiple CDS Applicable Vehicles					
		TOWED		AT LEAST TWO TOWED		ONLY ONE TOWED		At Least One CDS Applicable Motor Vehicle Was Towed	No CDS-applicable Motor Vehicles Were Towed
		Hospitalized	Not Hospitalized	Hospitalized	Not Hospitalized				
Injury in Towed, LMY CDS-applicable Motor Vehicle	A	J	C	J	C	E		G	NOT IN SCOPE
Injury in Towed, Non-LMY CDS-applicable Motor Vehicle	B	K	D	K	D	F		H	

Source: National Automotive Sampling System – NASS Sampling, Part One, 2000. Transportation Safety Institute, Highway Safety Division, NASS Training Branch DTI-71.

Table 4: 2010 Sub-listing PJs

PSU	PJ	Sub-listing Rate $\left(\pi_{kr\|hij}^{(1st)}\right)$	Weight $\left(w_{kr\|hij}^{(1st)}\right)$
3	All	1/2	2
49	1	1/2	2
72	1, 2, 3, 4, 5, 6	1/5	5
72	7	1/2	2
79	1, 7	1/2	2
81	1, 2	1/2	2

After PAR listing, each CDS eligible PAR stratum g was assigned a factor f_g ($g = 1, 2, \ldots 10$) as part of the measure of size for the next phase of PAR sample selection (see Table 5). All PARs in PAR stratum g of PJ hij are assigned a common measure of size m_{hijg} that is equal to its PJ weight $w_{j|hi}$ multiplied by factor f_g and the first phase PAR sampling weight $w_{kr|hij}^{(1st)}$: $m_{hijg} = f_g * w_{kr|hij}^{(1st)} * w_{j|hi}$. Then for each selected PSU, PARs listed in all selected PJs were pooled together and sorted by the four variables collected or generated in the PAR listing process: PAR strata, the minutes of the collision, the hours of the collision, and the sequence number the PAR was examined. From this pooled and sorted PAR list, a systematic PPS sample of average two PARs per researcher per week is selected from each PSU using m_{hijg} as the measure of size. The systematic PPS sampling was carried out as the following (Bondy, N., personal communication, 2012): the MOS m_{hijg} were laid out in the sorted order one after another on a horizontal axis from 0 to the total MOS of the week:

$$m_{hik} = \sum_{all\ listed\ PARs\ in\ week\ k\ in\ hi} m_{hijg}$$

The PAR sample size n_{hik}, or the "work load", selected for the week k at PSU hi normally equals to two times of the number of researches working for that PSU. Then n_{hik} intervals with equal length:

$$l_{hik} = m_{hik}/n_{hik}$$

were formed on the segment $[0,\ m_{hik}]$. Here l_{hik} is the length of the intervals. If any PAR's MOS was longer than the interval: $m_{hijg} > l_{hik}$, then this PAR was selected with certainty and it was set aside and the PAR sample n_{hik} was reduced by 1. This process was repeated until there was no certainty PAR. Assume the total number of certainty PARs was c_{hik}. The total MOS for all PARs in PSU hi for week k is calculated:

$$m'_{hij} = \sum_{all\ listed\ non-certainty\ PARs\ in\ week\ k\ in\ hi} m_{hijg}\ .$$

And equal intervals with common length:

$$l'_{hik} = m'_{hij}/(n_{hik} - c_{hik})$$

were formed on the segment $[0,\ m'_{hij}]$. Then a random number is generated in $[0,\ l'_{hik}]$ and a systematic sample of size $n_{hik} - c_{hik}$ was selected.

Table 5: Measure of Size Factor f_g by PAR Strata

PAR Strata	A	B	C	D	E	F	G	H	J	K
f_g	400	400	175	25	7	3	2	1	400	300

Source: Fleming, 2010.

The PAR strata created in the PAR listing process are actually used to ensure there will be enough severe crashes/injuries selected into the sample over the year by assigning a large measure of size factor f_g to corresponding PAR strata. Let n_g (g=1, 2, …, 10) be the known desired sample sizes for the PAR strata over the year, N_g (g=1, 2, …, 10) be the total number of PARs for the PAR strata estimated from historical data, then f_g (g=1, 2, …, 10) can be determined by solving the following linear equations:

$$\frac{f_g N_g}{\sum_{g=1}^{10} f_g N_g} = \frac{n_g}{\sum_{g=1}^{10} n_g}\ , (g = 1, 2, \dots, 10).$$

PAR stratum variable is also used as a sorting variable for PAR selection but it is not explicitly used as a stratification variable. And not all PAR strata have at least one PAR selected every week.

The above systematic sampling of PARs produces PPS selection probabilities. The second phase PAR selection probabilities are:

$$\pi_{kgr|hij}^{(2nd)} = \begin{cases} 1, & for\ certainty\ PARs; \\ (n_{hik} - c_{hik})\, m_{hijg}/m'_{hij}, & for\ non-certainty\ PARs. \end{cases}$$

The overall tertiary selection probability is the product of the selection probabilities of the two phases:

$$\pi_{kgr|hij} = \pi_{kr|hij}^{(1st)} \pi_{kgr|hij}^{(2nd)}\ .$$

And the tertiary sampling unit design weights are: $w_{kgr|hij} = \pi_{kgr|hij}^{-1}$.

IV. CDS Weights

CDS has a stratified multiphase sample design. The overall design weight (or base weight) d_{hijkgr} is the product of the design weights at all phases/stages:

$$d_{hijkgr} = w_{hi} * w_{j|hi} * w_{kgr|hij} .$$

Here r is PAR, g is PAR stratum, j is PJ, i is PSU, h is PSU stratum, k is week. $w_{hi}, w_{j|hi}, w_{kgr|hij}$ are the primary, secondary, and tertiary sampling design weights.

The current CDS PSU sample has been selected and used since 1991. The current CDS PJ sample has been selected and used since 1995. The PAR sample is selected weekly. Unit non-response is rare in CDS.

To correct potential non-response and coverage bias, a post-stratification adjustment is performed to the design weights d_{hijkgr}. The post-strata are cross tabulate of two variables: PSU strata and collapsed PAR strata. Here PSU strata are the 12 original PSU strata (h=1, 2, …, 12). The original 10 PAR strata (g=1, 2, …, 10) are collapsed into 7 larger PAR strata (c=1, 2, …, 7) because the population counts for the non-sampled PJs are collected quarterly therefore variables such as hospitalization cannot be determined (Bondy, N., personal communication, 2012). The definition for the collapsed PAR strata can be found in Table 6. There are 84 (12*7) possible post-strata.

Table 6: Collapsed PAR strata

Collapsed PAR Strata	AB	CJ	DK	E	F	G	H
Original PAR Strata	A & B	C & J	D & K	E	F	G	H

To calculate the post-stratum adjustment factor, two estimates of post-stratum population total are calculated. The first estimate is the summation of design weights of all selected PARs in each post-stratum hc:

$$D_{hc} = \sum_{all\ sampled\ i} \sum_{all\ sampled\ j} \sum_{all\ k} \sum_{all\ g\ in\ c} \sum_{all\ selected\ r} d_{hijkgr} .$$

Here c is collapsed PAR stratum. D_{hc} is a population total estimate of post-stratum hc using observed PARs therefore subject to non-response bias or coverage bias. The second post-stratum population total estimate uses PSU level listed PAR total estimates:

$$N_{hc} = \sum_{all\ sampled\ i\ in\ h} \sum_{all\ g\ in\ c} \left(R'_{hig} + R''_{hig}\right) w_{hi} .$$

Here w_{hi} is PSU weight. R'_{hig} is the summation of four quarters of PARs in PAR stratum g for all non-sampled PJs in PSU hi. R''_{hig} is an estimate of total PARs in PAR stratum g in all sampled PJs in PSU hi over the year:

$$R''_{hig} = \sum_{all\ sampled\ j} \sum_{all\ k} \sum_{all\ sampled\ r\ in\ g} w^{(1st)}_{kr|hij}\ .$$

Here $w^{(1st)}_{kr|hij}$ is the first phase PAR sampling weight (Table 4) for the sampled PJs. N_{hc} is the total listed and unlisted PARs from both sampled and non-sampled PJs and is considered a better estimate than D_{hc} for the total number of PARs in post-stratum hc.

The post-stratification adjustment factor for post-stratum hc is defined as:

$$p_{hc} = \frac{N_{hc}}{D_{hc}}\ .$$

And the CDS final weight is:

$$w_{hijkgr} = p_{hc} * d_{hijkgr}\ .$$

w_{hijkgr} corresponds to the case weight variable RATWGT in CDS files.

VIII. Reference

ANSI D16.1, Section 2.3.11: http://www-nrd.nhtsa.dot.gov/Pubs/07D16.pdf

Bondy, N., & Rhea, B. (1997, September). Reweighting of the Primary Sampling Units in the National Automotive Sampling System. (Research Note). Washington, DC: National Highway Traffic Safety Administration. Available at www-nrd.nhtsa.dot.gov/Pubs/97.845.pdf

Fleming, C. (2010, May). Sampling and Estimation Methodologies of CDS. (Report No. DOT HS 811 327). Washington, DC: National Highway Traffic Safety Administration. Available at www-nrd.nhtsa.dot.gov/Pubs/811327.pdf

Kahane, C. J. (1976). National Accident Sampling System Selection of Primary Sampling Units. (Report No. DOT HS 802 063). Washington, DC: National Highway Traffic Safety Administration.

Keyfitz, N. (1951). Sampling with Probabilities Proportionate to Size: Adjustment for Changes in Probabilities. *Journal of the American Statistician Association, 46*, pp. 105-109.

NHTSA. (2011). Sampling Manual for the National Automotive Sampling System. Washington, DC: National Highway Traffic Safety Administration.

Särndal, C-E., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Shelton, T. S. T. (1991, December). National Accident Sampling System, General Estimates System Technical Note, 1988 to 1990. (Report No. DOT HS 807 796). Washington, DC: National Highway Traffic Safety Administration. . Available at www-nrd.nhtsa.dot.gov/Pubs/807796.pdf

Appendix A: CDS Tertiary Sample Selection PAR Stratum Definition

The 10 PAR Strata (A, B, C, D, E, F, G, H, J, and K) for CDS-eligible PARs are listed below:

A. NASS crashes in which at least one occupant of a towed CDS-applicable late model year vehicle had a police-reported injury of "K" (fatal injury).
B. NASS crashes not qualifying for Stratum A in which at least one occupant of a towed CDS-applicable non-late model year vehicle had a police-reported injury of "K" (fatal injury).
C. NASS crashes not qualifying for Strata A, B, J or K in which at least one occupant of a towed CDS-applicable late model year vehicle had a police-reported injury of "A" (incapacitating injury) AND was transported to a treatment facility for treatment. If the crash involved more than one CDS-applicable vehicle, then at least two of the CDS-applicable vehicles must be towed.
D. NASS crashes not qualifying for Strata A, B, J, K or C in which at least one occupant of a towed CDS-applicable non-late model year vehicle had a police-reported injury of "A" (incapacitating injury) AND was transported to a treatment facility for treatment. If the crash involved more than one CDS-applicable vehicle, then at least two of the CDS-applicable vehicles must be towed.
E. NASS crashes not qualifying for Strata A, B, J, K, C or D in which at least one occupant of a towed CDS-applicable late model vehicle was transported from the scene to a treatment facility for treatment.
F. NASS crashes not qualifying for Strata A, B, J, K, C, D or E in which at least one occupant of a towed CDS-applicable non-late model vehicle was transported from the scene to a treatment facility for treatment.
G. NASS crashes not qualifying for Strata A, B, J, K, C, D, E or F that involve at least one CDS-applicable late model vehicle that was towed, according to the police report, from the scene due to damage.
H. NASS crashes not qualifying for Strata A, B, J, K, C, D, E, F or G that involve at least one CDS-applicable non-late model vehicle that was towed, according to the police report, from the scene due to damage.

J. NASS crashes not qualifying for Strata A or B in which at least one occupant of a towed CDS-applicable late model year vehicle had a police-reported injury of "A" (incapacitating injury) AND was transported to a treatment facility for treatment AND was admitted overnight to the hospital. If the crash involved more than one CDS-applicable vehicle, at least two of the CDS-applicable vehicles must be towed.
K. NASS crashes not qualifying for Strata A, B or J in which at least one occupant of a towed CDS-applicable non late model year vehicle had a police-reported injury of "A" (incapacitating injury) AND was transported to a treatment facility for treatment AND was admitted overnight to the hospital. If the crash involved more than one CDS-applicable vehicle, at least two of the CDS-applicable vehicles must be towed.

All CDS-applicable crashes are also GES applicable crashes. GES applicable but not CDS-applicable crashes are defined as strata X, Y, and Z:

X. NASS crashes not qualifying for the CDS strata, involving at least one medium or heavy truck in which a non-CDS-applicable motor vehicle was towed due to damage, or at least one involved person had a police-reported injury of “K” (fatal injury), “A” (incapacitating injury), “B” (non-incapacitating injury), or “C” (possible injury). Transported status is not a consideration.
Y. NASS crashes not qualifying for the CDS Strata or Stratum X in which none of the motor vehicles involved in the crash was a medium or heavy truck and at least one person involved in the crash had a police-reported injury of “K” (fatal injury), “A” (incapacitating injury), or “B” (non-incapacitating injury). Transported status is not a consideration.
Z. NASS crashes not qualifying for the CDS Strata, Stratum X or Stratum Y.

In 2008, NASS added a non-NASS (non-GES eligible and non-CDS eligible) stratum N:

N. Crashes not qualifying for NASS CDS or GES, involving motor vehicle crashes occurring outside the trafficway (non traffic crashes) that result in injury or death. Transport status is not a consideration. Cases in this stratum are NiTS cases.

DOT HS 811 807
July 2013

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

9550-071913-v3a