



## REWEIGHTING OF THE PRIMARY SAMPLING UNITS IN THE NATIONAL AUTOMOTIVE SAMPLING SYSTEM

Nancy Bondy and Barbara Rhea

The National Automotive Sampling System (NASS) is based on a multi-stage probability sample which requires periodic maintenance to preserve the probability basis and minimize any bias which can creep in over time. Since the NASS was first implemented in 1979, the first stage Primary Sampling Units (PSUs) have been reselected once. In an ideal world, the PSUs should be selected periodically to allow for any shifts in the number and type of crashes occurring across the country. This research note describes the process recently used to account for shifts in the number of crashes without a reselection.

### BACKGROUND

There are three stages to the NASS sample design:

- The first stage -- the selection of the PSUs;
- The second stage -- the selection of the police jurisdictions within a PSU; and,
- The third stage -- the selection of crashes for investigation.

The subject of this research note involves only the first stage.

The original design of the National Automotive Sampling System - formerly the National Accident Sampling System - called for 75 PSUs randomly selected from PSUs which were grouped into various strata across the U.S. The implementation of the PSU sample occurred gradually with the first ten implemented in 1979, the second ten in 1980, the next ten in 1981 and the final twenty in 1982/83 (this last group was done in two phases). Due to budget constraints in 1984, the system was to remain at 50 PSUs. In 1986, an agency decision was made to redesign the NASS into two different systems that would begin operations in January 1988. The NASS was split into two systems to fulfill the needs of the users both in-house within NHTSA and outside.

The two data collection systems implemented in 1988 are: the General Estimates System (GES) and the Crashworthiness Data System (CDS). The GES provides national estimates on crash, vehicle and occupant

characteristics obtained from the police crash report for all police-reported crashes. The CDS provides national estimates on very detailed crash, vehicle and occupant characteristics obtained through vehicle and scene inspections, interviews, and medical records for police-reported crashes involving at least one passenger car, pickup, van, or sport utility vehicle that was towed from the scene due to crash damage.

The first stage sample of NASS PSUs was re-selected in mid-1986 to allow the field operations side of the CDS to prepare for the 1988 data collection year. The U.S. was grouped into 1,195 Primary Sampling Units (PSUs) made up of either a central city, a large county without a central city, a large county, or a group of counties. To the extent possible, the same PSUs as were in the original sample were maintained. To account for the many differences in the U.S., PSUs were grouped into four geographic areas: Northeast, South, Midwest, and West; and three urban classes: Central City, Suburban, and Other. Within each of the twelve strata (four geographic areas and three urban classes), the PSUs were selected with probability proportional to size. That is, the PSUs were selected based on the number of fatal and injury crashes occurring within its boundaries during a calendar year. The more fatal and injury crashes occurring in a PSU, the greater the likelihood that PSU will be selected. Of the 50 original NASS PSUs, thirty were reselected. Six new PSUs were added. To obtain this result, a procedure based on work done by Nathan Keyfitz<sup>1</sup> was used to maximize the overlap between this sample and the sample selected for the 1979 implementation. After the CDS sample PSUs were selected, an additional 24 PSUs were selected to complete the GES sample of PSUs. In 1991, the CDS PSU sample was reduced to 24 PSUs by randomly subsampling within strata with more than 2 CDS PSUs. The 60 PSU GES sample remained the same.

<sup>1</sup>Keyfitz, N. (1951). Sampling with probabilities proportional to size: adjustments for changes in the probabilities. *Journal American Statistical Association*, 46, 105-109.

## ISSUE AT HAND

PSUs in the CDS/GES have not been reselected since the 1986 redesign because of the cost and time required to do so. It is costly because data must be collected from both the old and the new PSU samples simultaneously for at least one year. This is necessary to permit estimates to be made from both samples so that the time series is not interrupted. It is time-consuming because of the planning and implementation of new contractual arrangements required to complete the old sample of PSUs and implement the new one.

Recently, however, it was decided that changes in the distribution of crashes was large enough to warrant some type of a change in procedures. Exhibit 1 shows the

changes in the number of fatal and injury crashes from 1984 to 1992 by geographic location and urbanization. Compared to 1984, the number of fatal and injury crashes greatly increased in the South-Suburban and South-Other strata which resulted in the under-counting of the CDS estimates. To account for these shifts, the procedures used to stratify and select the PSUs in 1979 and 1986 would be followed, without actually resampling the PSUs. Rather, the weights of the current PSUs would be adjusted to reflect the changes. This process was referred to as 'pseudo-reselection'. Because some of the changes were so dramatic, NHTSA decided to make adjustments to PSU weights every three years. Currently, 1995 data are being collected to prepare for the PSU weight revision to be effective in 1998.

Exhibit 1 Total Number of Fatal and Injury Crashes by Geographic Location and Urbanization Comparison by 1984 and 1992								
Geographic Area	Urbanization						Total	
	Central City		Suburban		Other		1984	1992
	1984	1992	1984	1992	1984	1992		
Northeast	103,059	125,177	240,687	252,929	99,880	103,908	443,626	482,014
Midwest	106,240	112,943	205,625	211,911	195,605	216,658	507,470	541,512
South	124,641	144,915	245,117	301,897	307,241	442,189	676,999	889,001
West	83,093	97,159	178,426	212,978	105,808	115,320	367,327	425,457
<b>Total</b>	<b>417,033</b>	<b>480,194</b>	<b>869,855</b>	<b>979,715</b>	<b>708,534</b>	<b>878,075</b>	<b>1,995,422</b>	<b>2,337,984</b>

## METHODOLOGY

The pseudo-reselection was performed using 1992 state data available through state offices. The procedures used to perform the pseudo-reselection are as follows:

- Counts of fatal and injury crashes were obtained for each of the 1,195 PSUs.
- For all PSUs within each stratum, the number of fatal and injury crashes were summed.
- The probability of a PSU being selected was determined by dividing the number of fatal and injury crashes occurring within a PSU by the total number of fatal and injury crashes occurring in all of the PSUs within that stratum. The weight of the PSU is simply the inverse of the probability of the PSU being selected.

The CDS adopted the 'new weights' in 1995. The implementation of the 'new weights' resulted in a significant increase in the estimated number of motor vehicle crashes, vehicles, and occupants compared to the estimates for 1993 and 1994. To account for this difference, the 'new weights' were phased into 1993 and

1994. For 1993, the adjusted weights were calculated using one-third of the 'new weights' plus two-thirds of the old weights. For 1994, the adjusted weights equaled two-thirds of the 'new weights' plus one-third of the old weights.

The GES adopted the 'new weights' in 1996. The implementation of the 'new weights' resulted in slight increases in the estimated number of crashes, vehicles and persons compared to the estimates from 1993, 1994, and 1995. To account for these differences, the 'new weights' were directly applied to the 1993 through 1995 GES cases. Specifically for 1993 through 1995, the old PSU weights were replaced with the new PSU weights.

## EFFECT ON ESTIMATES

Exhibit 2 shows the effects of distributing the 'new weights' across the 1993, 1994, and 1995 CDS. The table shows about a 5 percent change for 1993, about a 10 percent change for 1994, and a 15 percent for 1995. These large differences may be due to the small number of PSUs in the CDS sample and the relative size of the estimates.

Exhibit 2 Comparison of the Published Estimates to the Revised Estimates for Crashes, Towed Passenger Vehicles, and Occupants in the NASS CDS, 1993-1995							
Year	Crashes	Towed Passenger Vehicles		Occupants			
		Cars	LTVs	Not Injured	Injured	Severely Injured	
1993	Published	2,002,525	1,955,038	591,178	2,120,811	1,746,290	85,718
	Revised	2,113,573	2,063,677	622,560	2,228,630	1,848,838	91,612
	Percent Change	5.5%	5.5%	5.3%	5.1%	5.9%	6.9%
1994	Published	2,135,183	2,153,128	671,752	2,260,123	1,957,335	110,834
	Revised	2,339,040	2,365,744	727,942	2,494,572	2,119,046	115,431
	Percent Change	10%	9.9%	8.4%	10.4%	8.3%	4.1%
1995	Published	2,267,923	2,233,553	659,285	2,236,640	2,155,425	101,766
	Revised	2,618,871	2,624,582	750,660	2,579,596	2,518,748	116,284
	Percent Change	15.5%	17.5%	13.1%	15.3%	16.9%	12.5%

Exhibit 3 shows motor vehicle crash estimates that were published in the "Traffic Safety Facts, A Compilation of Motor Vehicle Crash Data from the Fatal Analysis Reporting System and the General Estimates System" for 1993 - 1995 and the revised estimates which are based on the adjustments to the sampling weights. In 1993 and 1994, the revision of PSU weights resulted in very minor increases in the total number of motor vehicle crashes -- less than 1 percent. In 1995, the revision increased the number of motor vehicle crashes by 1.3 percent.

Exhibit 4 presents published and revised GES estimates for vehicles involved in injury and property-damage-only crashes. From 1993-1995, most revised estimates for the vehicles involved in injury crashes show a slight increase. In 1993 and 1994, the estimates for passenger cars involved in property-damage-only crashes decreased by about one-half of 1 percent.

Exhibit 5 depicts published and revised GES estimates of Persons Injured by Person Type and Vehicle Type for 1993-1995. Overall, the effect on the estimated number

of persons injured for 1993, 1994, and 1995 increased by 1, 1.6, and 2.3 percent, respectively, compared to the estimates based on the published GES data. In 1995, the number of occupants injured in passenger cars increased by 2.2 percent and light truck occupant injuries increased by 1.8 percent.

Exhibit 3 Total Motor Vehicle Crashes, 1993-1995 Published and Revised Estimates in the NASS GES			
	1993	1994	1995
Published Crash Estimates*	6,105,000	6,492,000	6,613,000
Revised Crash Estimates*	6,106,000	6,495,000	6,700,000
Percent Change from Published to Revised	0%	0%	1.3%

\*Note: Estimates include fatal crash counts from the FARS, and injury and property-damage-only crash estimates from the GES.

Exhibit 4 Vehicles Involved in Crashes by Vehicle Type and Crash Severity, 1993-1995 Published and Revised Estimates (in Thousands)									
Year		Injury Crashes				Property-Damage-Only Crashes			
		Passenger Cars	Light Trucks	Large Trucks	Motor-cycles	Passenger Cars	Light Trucks	Large Trucks	Motor-cycles
1993	*Published	2,610	829	97	54	4,812	1,879	294	16
	**Revised	2,631	843	97	56	4,789	1,884	296	17
	Percent Change	0.8%	1.7%	0%	3.7%	-0.5%	0.3%	0.7%	6.3%
1994	*Published	2,742	893	95	53	5,155	2,025	361	13
	**Revised	2,785	912	96	54	5,126	2,023	360	13
	Percent Change	1.6%	2.1%	1.1%	1.9%	-0.6%	-0.1%	-0.3%	0%
1995	*Published	2,844	999	83	50	5,289	2,121	290	12
	**Revised	2,914	1,024	84	52	5,335	2,149	289	13
	Percent Change	2.5%	2.5%	1.2%	4.0%	0.9%	1.3%	-0.3%	8.3%

\* Source: GES; Traffic Safety Facts 1995, p.17.

\*\* Based on adjustments to the sampling weights.

Exhibit 5 Persons Injured by Person Type and Vehicle Type, 1993-1995 Published and Revised Estimates(In Thousands)												
Year	Occupants by Vehicle Type							Nonmotorists				Total
	Passenger Cars	Light Trucks	Large Trucks	Motorcycles	Buses	Other	Total	Pedestrian	Pedalcyclist	Other	Total	
1993 *Published	2,257	590	32	58	17	4	2,958	93	65	9	166	3,125
** Revised	2,265	601	32	59	17	4	2,978	94	68	9	171	3,149
Percent Change	0.4%	1.9%	0%	1.7%	0%	0%	0.7%	1.1%	4.6%	0%	3.0%	0.8%
1994 *Published	2,332	619	32	56	15	3	3,056	90	60	9	159	3,215
** Revised	2,364	631	30	57	15	4	3,102	92	62	9	161	3,265
Percent Change	1.4%	1.9%	-6.3%	1.8%	0%	33.3%	1.5%	2.2%	3.3%	0%	1.3%	1.6%
1995 *Published	2,416	709	30	55	18	4	3,232	84	61	9	154	3,386
** Revised	2,469	722	30	57	19	4	3,303	86	67	10	163	3,465
Percent Change	2.2%	1.8%	0%	3.6%	5.5%	0%	2.2%	2.4%	9.8%	11.1%	5.8%	2.3%

\* Source: GES; *Traffic Safety Facts 1995* p.18.

\*\* Based on adjustments to the sampling weights

For additional copies of this research note, please call (202) 366-4198 or fax your request to (202) 366-7078. For questions regarding the data reported in this research, contact Nancy Bondy [202-366-5353] or Barbara Rhea

[202-366-2714] of the National Center for Statistics and Analysis. This research note and other general information on highway traffic safety may be accessed by Internet users at <http://www.nhtsa.dot.gov/people/nca>.

U.S. Department  
of Transportation  
**National Highway  
Traffic Safety  
Administration**  
400 Seventh Street, S.W. , NRD-31  
Washington, D.C. 20590

