



## Bureau of Justice Statistics

# Criminal Victimization in the United States, 1994

- **NCVS-measured crimes**
- **Demography of victims**
- **Victims and offenders**
- **Geography**
- **The crime event**
- **Victims and the criminal justice system**
- **Survey instrument**
- **Survey methodology**
- **Glossary**

*A National Crime Victimization Survey Report*

U.S. Department of Justice  
Office of Justice Programs  
Bureau of Justice Statistics



---

# Criminal Victimization in the United States, 1994

*A National Crime Victimization Survey Report*

May 1997, NCJ-162126

## Appendixes

---

I. Survey instrument	104
II. Survey methodology	139
Data collection	139
Sample design and size	139
Estimation procedure	140
Series victimizations	141
Accuracy of estimates	142
Computation and application of standard errors	144
III. Glossary	147

## Appendix I

### Survey instrument

---

A screen questionnaire (form NCVS-1) and a crime incident report (form NCVS-2) are used to obtain information about households, individuals, and the relevant crimes they have experienced. The first form, NCVS-1, is designed to obtain demographic characteristics and to screen for any crime incidents. Each household member age 12 or older is interviewed individually, unless a proxy is used. Proxy interviews are used for children age 12 or 13 when the parents object to an individual interview, as well as for persons who are absent during the entire interviewing period and persons who are otherwise incapable of answering for themselves.

After the first form is completed, the interviewer fills out a second form, the NCVS-2 form, for each reported incident. Along with general questions about the incident, the NCVS-2 form includes questions about the extent of physical injury, economic loss, offender characteristics, and notification of police.

The basic screen questionnaire and incident report reproduced on the following pages are the revised questionnaires being used as a result of the redesign program.

The survey results contained in this report are based on data gathered from residents living throughout the United States, including persons living in group quarters, such as dormitories, rooming houses, and religious group dwellings. Crew members of merchant vessels, Armed Forces personnel living in military barracks, and institutionalized persons, such as correctional facility inmates, were not included in the scope of this survey. Similarly, U.S. citizens residing abroad and foreign visitors to this country were excluded. With these exceptions, individuals age 12 or older living in units selected for the sample were eligible to be interviewed.

#### **Data collection**

Each housing unit selected for the National Crime Victimization Survey (NCVS) remains in the sample for 3 years, with each of seven interviews taking place at 6 month intervals. An NCVS interviewer's first contact with a housing unit selected for the survey is in person. The interviewer may then conduct subsequent visits, except for the fifth, by telephone.

To elicit more accurate reporting of incidents, NCVS uses the self-respondent method which calls for the direct interviewing of each person 12 years or older in the household. An exception is made to use proxy interviewing instead of direct interviewing for the following three cases: 12- and 13- year old persons when a knowledgeable household member insists they not be interviewed directly, incapacitated persons, and individuals absent from the household during the entire field-interviewing period. In the case of temporarily absent household members and persons who are physically or mentally incapable of granting interviews, interviewers may accept other household members as proxy respondents, and in certain situations

non-household members may provide information for incapacitated persons.

As noted in the sample design section, about 30% of the interviews in the 1994 sample were conducted using Computer-Assisted Telephone Interviewing (CATI), a data collection mode which involves interviewing from centralized facilities and using a computerized instrument. In the CATI-eligible part of the sample, all interviews are done by telephone whenever possible, except for the first and fifth interviews, which are still primarily conducted in person. The telephone interviews are conducted by the CATI facilities (Hagerstown, Maryland, and Tucson, Arizona).

#### **Sample design and size**

Survey estimates are derived from a stratified, multi-stage cluster sample. The primary sampling units (PSU's) composing the first stage of the sample were counties, groups of counties, or large metropolitan areas. Large PSU's were included in the sample automatically and are considered to be self-representing (SR) since all of them were selected. The remaining PSU's, called non-self-representing (NSR) because only a subset of them was selected, were combined into strata by grouping PSU's with similar geographic and demographic characteristics, as determined by the 1980 census.

The 1994 NCVS sample households were drawn from the 1980-based sample design. The 1980 design consists of 84 SR PSU's and 153 NSR strata, with one PSU per stratum selected with probability proportionate to population size. The NCVS sample design has been revised to take advantage of the availability of data from the 1990 census. However, the 1990-based sample will not start con-

tributing to the NCVS estimates until 1995.

The two remaining stages of sampling were designed to ensure a self-weighting<sup>1</sup> probability sample of housing units and group-quarter dwellings within each of the selected areas. This involved a systematic selection of enumeration districts (geographic areas used for the 1980 census), with a probability of selection proportionate to their 1980 population size, followed by the selection of segments (clusters of approximately four housing units each) from within each enumeration district. To account for units built within each of the sample areas after the 1980 Census, a sample was drawn of permits issued for the construction of residential housing. Jurisdictions that do not issue building permits were sampled using small land-area segments. These supplementary procedures, though yielding a relatively small portion of the total sample, enabled persons living in housing units built after 1980 to be properly represented in the survey.

Approximately 58,060 housing units and other living quarters were designated for the sample. In order to conduct field interviews, the sample is divided into six groups, or rotations, and each group of households is interviewed once every 6 months over a period of 3 years. The initial interview is used to bound the interviews (bounding establishes a timeframe to avoid duplication of crimes on subsequent interviews), but is not used to compute the annual estimates. Each rotation group is further divided into six panels. A different panel of households, corresponding to one sixth of each rotation group, is interviewed each month during the

<sup>1</sup>Self-weighting means that, prior to any weighting adjustments, each sample housing unit had the same overall probability of being selected.

## Appendix II

### Survey methodology

---

6-month period. Because the survey is continuous, newly constructed housing units are selected as described, and assigned to rotation groups and panels for subsequent incorporation into the sample. A new rotation group enters the sample every 6 months, replacing a group phased out after being in the sample for 3 years.

For these 58,060 sample households, complete interviews were obtained for about 47,600 households in 1994, about 95.1% of all eligible housing units. Within the interviewed households some 90,560 persons, or about 92.0%, provided responses; the other individuals for the most part either refused, or were unavailable or unable to answer and no proxy was available. The remaining 10,460 housing units were not interviewed because they were either ineligible — vacant, demolished, or otherwise ineligible — for the survey (about 8,010 units), or the occupants could not be reached or refused to participate (about 2,450 units).

#### *Selection of cases for CATI*

About 30% of the 47,600 households obtained in the 1994 sample were interviewed using the CATI technique. Currently, the NCVS sample PSU's fall into three groups of CATI usage: maximum-CATI PSU's, where all the segments in the PSU are CATI-eligible; half-CATI PSU's, where half of the segments in the PSU are randomly designated to be CATI-eligible; and no-CATI PSU's, where none of the segments are CATI-eligible. The level of CATI usage for each PSU was established with concern toward an optimal workload for the field interviewers. In the half-CATI PSU's, a random sample of about 50% of the segments in each PSU is taken and designated as CATI-eligible. The sample cases in CATI-eligible

segments from the max-CATI and the half-CATI PSU's are interviewed from CATI facilities while the other sample cases are interviewed by the standard NCVS field procedures.

#### **Estimation procedure**

Annual estimates of the levels and rates of victimization are derived by accumulating four quarterly estimates, which in turn are obtained from 17 months of field interviewing, ranging from February of one year through June of the following year. The population and household figures shown on victimization rate tables are based on an average for these 17 months, centering on the ninth month of the data collection period, in this case October 1994.

Sample data from 8 months of field interviewing are required to produce estimates for each quarter. (Quarterly estimates are not published since there may not be sufficient observations to ensure their reliability.) For example, data collected between February and September are required to estimate the first quarter of any given calendar year (see accompanying chart). Each quarterly estimate is composed of equal numbers of field observations from the months during the half-year interval prior to the time of interview. Therefore, incidents occurring in January may be reported in a February interview (1 month between the crime and the interview), in a March interview (2 months), and so on up to 6 months ago for interviews conducted in July. This arrangement minimizes expected biases associated with the tendency of respondents to place victimizations in more recent months of a 6-month reference period rather than the month in which they actually occurred.

The estimation procedure begins with the application of a base weight to the data from each individual interviewed. The base weight is the reciprocal of the probability of each unit's selection for the sample, and provides a rough measure of the population represented by each person in the sample. Next, an adjustment is made to account for households and individuals in occupied units who were selected for the survey but unavailable for interview.

In addition to adjusting for unequal probabilities of selection and observation, the final weight also includes a ratio adjustment to known population totals based on the adjusted counts from the 1990 Decennial Census. Specifically, the final person weight is the product of the values of the following six component weights; the final household weight is the product of all components except the within-household non-interview adjustment component detailed below:

#### *Probabilities of selection*

- Base weight: the inverse of the sampling rate of that unit (person or household) within the stratum.
- Weighting control factor: adjusts for any subsampling due to unexpected events in the field, such as unusually high growth in new construction, area segments larger than anticipated, and other deviations from the overall stratum sampling rate.

#### *Probabilities of observation (Nonresponse)*

- Household noninterview adjustment: adjusts for nonresponse at the household level by inflating the weight assigned to interviewed households so that they represent themselves and noninterviewed households.

- Within-household non-interview adjustment: adjusts for nonresponse at the person level by inflating the weight assigned to the interviewed persons so that they represent themselves and the missed interviews.

*Poststratification ratio adjustment to known population totals*

The distribution of the sample population may differ somewhat from that of the total population in terms of age, race, sex, residence, and other characteristics. Because of this, two stages of ratio estimation are employed to bring the two distributions into closer agreement, thereby reducing the variability of the sample estimates.

- First-stage factor: the first stage of ratio estimation is applied only to non-selfrepresenting PSU's. Its purpose is to reduce sampling error caused by selecting one PSU to represent an entire stratum. It adjusts for race and zone of residence differences between the sample non-self-representing PSU's and the population non-self-representing PSU's. (For self-representing PSU's this factor is set to 1).

- Second-stage factor: the second stage of ratio estimation is applied on an individual basis to bring the distribution of individuals in the sample into closer agreement with independent current estimates of the population according to age, sex and race characteristics.<sup>2</sup> This factor is defined for each person to adjust for the difference between weighted counts of persons (using the above five weight components) and independent estimates of the number of persons, within the defined cells. These independent estimates are projections

<sup>2</sup>Armed Forces personnel who are eligible to be interviewed are not included in the second-stage ratio estimate and receive a factor of 1.

Month of interview by Month of Reference (X's denote months in the 6-month reference period)												
Month of interview	Period of reference within bounded period											
	First Quarter			Second Quarter			Third Quarter			Fourth Quarter		
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
January												
February	X											
March	X	X										
April	X	X	X									
May	X	X	X	X								
June	X	X	X	X	X							
July	X	X	X	X	X	X						
August		X	X	X	X	X	X					
September			X	X	X	X	X	X				
October				X	X	X	X	X	X			
November					X	X	X	X	X	X		
December						X	X	X	X	X	X	
January							X	X	X	X	X	X
February								X	X	X	X	X
March									X	X	X	X
April										X	X	X
May											X	X
June												X
July												

based on the 1990 Census population controls adjusted for the undercount.

For household crimes, the characteristics of the wife in a husband-wife household and the characteristics of the head of household in other types of households are used to determine the ratio adjustment factors. This procedure is considered more precise than simply using the characteristics of the head of household since sample coverage is generally better for females than males.

For estimates involving *incidents* rather than *victimizations*, further adjustments are made to those cases where an incident involved more than one person. These incidents have more than one chance of being included in the sample so each multiple-victimization is reduced by the number of victims. Thus, if two people are victimized during the same incident, the weight assigned to that incident is the person weight reduced by one-half so that the incident cannot

be counted twice. However, the details of the event's outcome as they related to the victim are reflected in the survey results. No adjustment is necessary in estimating data on household crimes because each separate crime is defined as involving only one household.

**Series victimizations**

A series victimization is defined as six or more similar but separate crimes which the victim is unable to recall individually or describe in detail to an interviewer. These series crimes have been excluded from the tables in this report because the victims were unable to provide details for each event. Data on series crimes are gathered by the calendar quarter(s) of occurrence, making it possible to match the timeframes used in tabulating the data for non-series crimes.

Table I shows the counts of regular and series victimizations for 1994, as well as the results of combining the

## Appendix II

### Survey methodology

---

two, with each series tallied as a single event. A total of 722,740 personal series crimes and 358,770 property series crimes were measured in 1994. Series crimes tended to be crimes of violence.

The effect of combining series and non-series crimes, counting each of the series crimes as a single victimization based on the details of the most recent incident, was included in the initial release of the 1980 data.<sup>3</sup> The report showed that victimization counts and rates were higher in 1979 and 1980 when the series crimes were added. However, rate changes between these 2 years were basically in the same direction and significantly affected the same crimes as those affected when only non-series crimes were analyzed.

#### Accuracy of estimates

The accuracy of an estimate is a measure of its total error, that is, the sum of all the errors affecting the estimate: sampling error as well as non-sampling error.

The sample used for the NCVS is one of a large number of possible samples of equal size that could have been obtained by using the same sample design and selection procedures. Estimates derived from different samples would differ from one another due to sampling variability, or sampling error.

The standard error of a survey estimate is a measure of the variation among that estimate from all possible samples. Therefore, it is a measure of the precision (reliability) with which a particular estimate approximates the average result of all possible samples. The estimate and its associated standard error may be used to construct a

confidence interval. A confidence interval is a range of numbers which has a specified probability that the average of all possible samples, which is the true unknown value of interest in an unbiased design, is contained within the interval. About 68% of the time, the survey estimate will differ from the true average by less than one standard error. Only 10% of the time will the difference be more than 1.6 standard errors, and just 1 time in 100 will it be greater than 2.5 standard errors. A 95% confidence interval is the survey estimate plus or minus twice the standard error, thus there is a 95% chance that the result of a complete census would fall within the confidence interval.

In addition to sampling error, the estimates in this report are subject to nonsampling error. While substantial care is taken in the NCVS to reduce the sources of nonsampling error throughout all the survey operations, by means of a quality assurance program, quality controls, operational controls, and error-correcting procedures, an unquantified amount of nonsampling error remains still.

Major sources of nonsampling error are related to the inability of the respondents to recall in detail the crimes which occurred during the 6 months prior to the interview. Research based on interviews of victims obtained from police files indicates that assault is recalled with the least accuracy of any crime measured by the NCVS. This may be related to the tendency of victims to not report crimes committed by offenders who are not strangers, especially if they are relatives. In addition, among certain groups, crimes which contain elements of assault could be a part of everyday life, and are therefore

forgotten or not considered important enough to mention to a survey interviewer. These recall problems may result in an understatement of the actual rate of assault.

Another source of nonsampling error is the inability of some respondents to recall the exact month a crime occurred, even though it was placed in the correct reference period. This error source is partially offset by interviewing monthly and using the estimation procedure described earlier.

Telescoping is another problem in which incidents that occurred before the reference period are placed within the period. The effect of telescoping is minimized by using the bounding procedure previously described.

The interviewer is provided with a summary of the incidents reported in the preceding interview and, if a similar incident is reported, it can be determined whether or not it is a new one by discussing it with the victim. Events which occurred after the reference period are set aside for inclusion with the data from the following interview.

Other sources of nonsampling error can result from other types of response mistakes, including errors in reporting incidents as crimes, misclassification of crimes, systematic data errors introduced by the interviewer, errors made in coding and processing the data. Quality control and editing procedures were used to minimize the number of errors made by the respondents and the interviewers.

Since field representatives conducting the interviews usually reside in the area in which they interview, the race and ethnicity of the field representatives generally matches that of the local population. Special efforts are

<sup>3</sup>See *Criminal Victimization in the United States; 1979-80 Changes, 1973-80 Trends*, BJS Technical Report, NCJ-80838, July 1982.



**Table I. Personal and property crimes, 1994**

**Number and percent distribution of series victimizations and of victimizations not in series, by type of crime**

Type of crime	Total victimizations		Series victimizations		Victimizations not in series	
	Number	Percent	Number	Percent	Number	Percent
<b>Personal crimes</b>	<b>12,072,380</b>	<b>100.0%</b>	<b>722,740</b>	<b>6.0%</b>	<b>11,349,640</b>	<b>94.0%</b>
Crimes of violence	11,583,370	100.0	722,740	6.2	10,860,630	93.8
Completed violence	3,379,540	100.0	174,130	5.2	3,205,410	94.8
Attempted/threatened violence	8,203,830	100.0	548,610	6.7	7,655,220	93.3
Rape/Sexual assault	464,970	100.0	32,210	6.9	432,750	93.1
Rape/attempted rape	334,540	100.0	18,370 *	5.5 *	316,160	94.5
Rape	183,690	100.0	16,140 *	8.8 *	167,550	91.2
Attempted rape <sup>1</sup>	150,840	100.0	2,230 *	1.5 *	148,610	98.5
Sexual assault <sup>2</sup>	130,430	100.0	13,840 *	10.6 *	116,590	89.4
Robbery	1,329,140	100.0	30,390	2.3	1,298,750	97.7
Completed/property taken	816,280	100.0	21,150 *	2.6 *	795,130	97.4
With injury	292,130	100.0	4,500 *	1.5 *	287,620	98.5
Without injury	524,150	100.0	16,640 *	3.2 *	507,510	96.8
Attempted to take property	512,860	100.0	9,240 *	1.8 *	503,620	98.2
With injury	126,630	100.0	4,840 *	3.8 *	121,790	96.2
Without injury	386,230	100.0	4,400 *	1.1 *	381,830	98.9
Assault	9,789,260	100.0	660,140	6.7	9,129,120	93.3
Aggravated	2,599,840	100.0	121,690	4.7	2,478,150	95.3
With injury	717,620	100.0	39,040	5.4	678,580	94.6
Threatened with weapon	1,882,220	100.0	82,640	4.4	1,799,570	95.6
Simple	7,189,420	100.0	538,460	7.5	6,650,970	92.5
With minor injury	1,555,000	100.0	88,940	5.7	1,466,060	94.3
Without injury	5,634,430	100.0	449,520	8.0	5,184,900	92.0
Purse snatching/Pocket picking	489,010	100.0	0 *	0.0 *	489,010	100.0
Completed purse snatching	90,160	100.0	0 *	0.0 *	90,160	100.0
Attempted purse snatching	23,160	100.0	0 *	0.0 *	23,160	100.0
Completed pocket picking	375,690	100.0	0 *	0.0 *	375,690	100.0
<b>Property crimes</b>	<b>31,370,970</b>	<b>100.0%</b>	<b>358,770</b>	<b>1.1%</b>	<b>31,012,200</b>	<b>98.9%</b>
Household burglary	5,558,180	100.0	75,460	1.4	5,482,720	98.6
Completed	4,644,100	100.0	71,200	1.5	4,572,900	98.5
Forcible entry	1,745,880	100.0	20,340 *	1.2 *	1,725,540	98.8
Unlawful entry without force	2,898,220	100.0	50,860	1.8	2,847,360	98.2
Attempted forcible entry	914,080	100.0	4,260 *	0.5 *	909,820	99.5
Motor vehicle theft	1,770,570	100.0	6,880 *	0.4 *	1,763,690	99.6
Completed	1,176,980	100.0	4,680 *	0.4 *	1,172,300	99.6
Attempted	593,590	100.0	2,200 *	0.4 *	591,390	99.6
Theft <sup>3</sup>	24,042,220	100.0	276,440	1.1	23,765,790	98.9
Completed	23,012,100	100.0	268,260	1.2	22,743,840	98.8
Less than \$50	9,528,010	100.0	150,870	1.6	9,377,150	98.4
\$50-\$249	7,925,980	100.0	51,750	0.7	7,874,230	99.3
\$250 or more	4,292,050	100.0	40,720	0.9	4,251,340	99.1
Amount not available	1,266,060	100.0	24,930	2.0	1,241,130	98.0
Attempted	1,030,120	100.0	8,170 *	0.8 *	1,021,950	99.2

Note: Detail may not add to total shown because of rounding.

\*Estimate is based on about 10 or fewer sample cases.

<sup>1</sup>Includes verbal threats of rape.

<sup>2</sup>Includes threats.

<sup>3</sup>Thefts includes crimes previously classified as "Personal larceny without contact" and "Household larceny."

## Appendix II

### Survey methodology

made to further match field representatives and the people they interview in areas where English is not commonly spoken. About 90% of all NCVS field representatives are female.

Standard errors measure only those nonsampling errors arising from transient factors affecting individual responses completely at random (simple response variance); they do not reveal any systematic biases in the data. As calculated in the NCVS, the standard errors would partially measure nonsampling error arising from some of the above sources, such as transient memory errors, or accidental errors in recording or coding answers, for example.

#### Computation and application of standard errors

The results presented in this report were tested to determine whether or not the observed differences between groups were statistically significant. Differences were tested for significance at the 90% confidence level, or roughly 1.6 standard errors. Most of the comparisons in this report were significant at the 95% confidence level (about 2.0 standard errors, meaning that the difference between the estimates is greater than twice the standard error of the difference). Comparisons which failed the 90% test were not considered statistically significant. Comparisons qualified by the phrase "some evidence" or "statistically significant" had a significance level between 90% and 95%.

Deriving standard errors which are applicable to a wide variety of items and which can be prepared at a moderate cost requires a number of approximations. Therefore, three generalized variance function (gvf) constant parameters (identified as "a",

"b", and "c" in the following section) were developed for use in calculating standard errors. The parameters provide an indication of the order of magnitude of the standard errors rather than the precise standard error for any specific item.

The gvf represents the curve fitted to the individual standard errors, which were calculated using the Jackknife Repeated Replication technique. The 1994 gvf provided new values for the "a", "b", and "c" parameters calculated from the 1994 data. This updated model also provided new 1993 and 1992 gvf constant parameters.

#### Notation

$x$  = the estimated number (level) of personal or household victimizations or incidents

$y$  = the base; either the total number of persons or households (for victimization rates) or the total of all victimizations (for incident characteristics)

$p$  = the estimated proportion, resulting from dividing the number of victimizations into the base. Also, the percentage or rate expressed in decimal form. The percentage is  $100p$  and the rate per thousand is  $1000p$ .

$s(p)$  = the estimated standard error of  $p$

It follows that:

$$s(\text{percentage}) = s(100p) = 100 s(p)$$

$$s(\text{rate}) = s(1000p) = 1000 s(p)$$

a,b,c = the generalized variance function parameters (see chart)

#### GVF parameters from 1994-92 data-year estimates

Generalized variance functions	a	b	c
1994 parameter set			
Estimates			
Overall person crime (1)	-.00004144	2008	1.612
Personal crime domain (2)	-.00006269	2278	1.804
Overall property crime (3)	-.00008894	1501	1.276
Property crime domain (4)	-.00005292	2185	1.153
1993 revised parameter set			
Estimates			
Overall person crime (1)	-.00005221	2530	2.031
Personal crime domain (2)	-.00007899	2870	2.273
Overall property crime (3)	-.00011206	1891	1.608
Property crime domain (4)	-.00006668	2753	1.453
1992 revised parameter set			
Estimates			
Overall person crime (1)	-.00009951	4822	3.871
Personal crime domain (2)	-.00015053	5470	4.332
Overall property crime (3)	-.00021356	3604	3.064
Property crime domain (4)	-.00012707	5247	2.769

Parameter set #1 is used for the overall person crime estimates (Table 1). These are the person crime estimates by crime category for the whole population, not disaggregated by any victim, offender, or incident characteristics, nor any variable related to reporting to police.

Parameter set #2 is used for the person crime domain estimates. These are the person crime estimates disaggregated by victim, offender, or incident characteristics, or any variable related to reporting to police.

Parameter set #3 is used for the property crime estimates for the whole population (Table 1). These are the property crime estimates by crime category for the whole population, not disaggregated by any household characteristics, nor any variable related to reporting to police.

Parameter set #4 is used for the property crime domain estimates. These are the property crime estimates disaggregated by household characteristics, or any variable related to reporting to police.

For the statistic from Table 1 that corresponds to the crime category "all crimes" (person and property crimes together), parameter set #3 should be used. When the person and property estimates are combined (all crimes) and disaggregated by victim, household, incident characteristics, as well as any variable related to reporting to police, parameter set #4 should be used for the best estimate of the corresponding variance.

**Formula 1. Levels:** Standard errors for the estimated *number of victimizations or incidents* may be calculated by using the following formula:

$$s(x) = \sqrt{ax^2 + bx + cx^{3/2}}$$

The following example illustrates the proper use of this Formula 1. Table 1 (page 6) shows 795,130 completed robberies in 1994; this estimate and the appropriate parameters are substituted in the formula as follows:

$$s(x) = \sqrt{-0.00004144(795,130)^2 + (2008)(795,130) + (1.612)(795,130)^{3/2}} = 52,090$$

Therefore, the 95% confidence interval around the estimated number of robbery victimizations is about equal to 795,130 plus or minus twice (1.96) the standard error, or 102,096: an interval of 693,034 to 897,226.

**Formula 2. Proportions, Percentages, or Rates:** Standard errors for the estimated *victimization rates or percentages* are calculated using the following formula:

$$s(p) = \sqrt{\frac{bp(1.0-p)}{y} + \frac{cp(\sqrt{p}-p)}{\sqrt{y}}}$$

The following example demonstrates the use of Formula 2. Table 3 (page 8) shows an estimated robbery rate of 11.3 per 1,000 persons between the ages of 20 and 24, based on a total of 18,304,850 persons in this age range. Substituting the appropriate values into the formula yields:

$$s(p) = \sqrt{\frac{2278(.0113)(1.0-.0113)}{18,304,850} + \frac{1.804(.0113)(\sqrt{.0113}-.0113)}{\sqrt{18,304,850}}} = 0.0014 \text{ or } 1.4 \text{ per thousand}$$

Thus, the 95% confidence interval is 11.3 per 1,000 plus or minus 2.7: an interval of 8.6 to 14.0 per 1,000.

**Formula 3. Differences in rates or percentages with different bases:** The standard error of a *difference between two rates or percentages having different bases* is calculated using the formula:

$$s(p_1 - p_2) = \sqrt{\text{var}(p_1) + \text{var}(p_2) - 2ps(p_1)s(p_2)}$$

where:

$p$  is the year-to-year correlation between  $p_1$  and  $p_2$  (see chart on next page); and  $\text{var}(p_1)$  and  $\text{var}(p_2)$  are the square of the standard error of  $p$  using Formula 2 for each rate and substituting:

$p_1$  = first percent or rate (expressed as a proportion in decimal form)

$y_1$  = base from which first percent or rate was derived

$p_2$  = second percent or rate (expressed as a proportion in decimal form)

$y_2$  = base from which second percent or rate was derived

If estimates are uncorrelated,  $p = 0$ . Hence, omitting the term containing  $p$  in the formula will provide an accurate standard error for the difference between uncorrelated estimates. On the other hand, if the two estimates have a strong positive correlation, omitting the last term will cause overestimation of the true standard error. If the numbers have a strong negative correlation, this will cause underestimation of the actual standard error.

The following example illustrates the use of Formula 3. Table 4 (page 9) lists the victimization rate for aggravated assault for males as 15.3 per 1,000 and the rate for females as 8.1 per 1,000. The total number of males in the population is 103,369,260 and the total of females, 110,378,010. Noting that  $p = 0$  because the two estimates are for the same year and placing the appropriate values in the formulas yields:

$$\begin{aligned} \text{var}(p_1) &= \frac{2278(.0153)(1.0-.0153)}{103,369,260} + \frac{1.804(.0153)(\sqrt{.0153}-.0153)}{\sqrt{103,369,260}} = 0.000000626 \\ \text{var}(p_2) &= \frac{2278(.0081)(1.0-.0081)}{110,378,010} + \frac{1.804(.0081)(\sqrt{.0081}-.0081)}{\sqrt{110,378,010}} = 0.000000280 \end{aligned}$$

Standard error of the difference =

$$\sqrt{0.000000626 + 0.000000280} = .00095 \text{ or } .95 \text{ per thousand}$$

The 95% confidence interval around the difference of 7.2 per thousand is approximately the difference plus or minus 1.9 per thousand (a difference between 5.3 and 9.1 per thousand).

The ratio of a difference to the standard error of the difference is the "z score," which is associated with a given statistical level of significance. For example, a ratio with an absolute value of 2.0 (1.96, to be exact) or greater indicates that the difference is significant at the 95% confidence level (or greater); a ratio with an absolute value between 1.6 and 2.0 indicates the difference is significant at a confidence level between 90% and 95%;

## Appendix II

### Survey methodology

a ratio with an absolute value less than 1.6 denotes a confidence level less than 90%. In the previous example, the ratio of the difference (.0072) to the standard error (.00095) is 7.58. Thus the aggravated assault rate for males and females was significantly different at a confidence level exceeding 95%.

**Formula 4.** Differences in rates or percentages with the same base: The standard error of *the difference between two rates or percentages derived from the same base* is calculated using the formula:

$$s(p_1 - p_2) = \sqrt{\text{var}(p_1) + \text{var}(p_2) - 2ps(p_1)s(p_2)}$$

where  $q = 1 - p$

$$p = -\sqrt{\frac{p_1 p_2}{q_1 q_2}}; \text{ and}$$

$q = 1 - p$ ; and all other terms are as defined in Formula 3, except that  $y_1$  and  $y_2$  are the same common base,  $y$ .

The following example, which uses Table 43 (page 42), illustrates the use of Formula 4. The proportion of single-offender violent crimes victimizations involving relatives was 11.2% and the proportion involving acquaintances (well-known or casual) was 41.0%, out of a total of 8,169,830 single-offender violent crime victimizations. Substituting the appropriate values into the formula gives:

Year-to-Year Correlation between Estimates			
Because of the year-to-year overlap in the sample, the same households and persons contribute to annual estimates for different years. This year-to-year correlation between estimates is measured by $p$ . In general:			
$p = 0$ when estimates are for the same year			
$p \neq 0$ for year-to-year comparisons			
When comparing estimates that are 1 year apart, use $p$ as shown below.			
When comparing estimates that are 2 years apart, multiply $p$ by 1/2.			
When comparing estimates that are more than 2 years apart, assume $p=0$ .			
Following are NCVS year-to-year correlation values for major crime categories for 1992-94.			
Type of crime	1993-94 correlation	1992-94 correlation	Revised 1992-94 correlation
Total crimes	0.41	0.14	0.29
Total personal crimes	0.30	0.11	0.29
Crimes of violence	0.31	0.11	0.22
Rape/Sexual assault	0.04	0.02	0.03
Robbery	0.04	0.01	0.03
Assault	0.30	0.10	0.21
Purse snatching/Pocket picking	0.03	0.01	0.02
Total property crimes	0.38	0.13	0.27
Burglary	0.21	0.07	0.15
Motor vehicle theft	0.08	0.03	0.06
Theft	0.34	0.12	0.24

$$\text{var}(p_1) = \frac{2278(.112)(1.0-.112)}{8,169,830} + \frac{1.804(.112)(\sqrt{.112} - .112)}{\sqrt{8,169,830}} = 0.000043$$

$$\text{var}(p_2) = \frac{2278(.410)(1.0-.410)}{8,169,830} + \frac{1.804(.410)(\sqrt{.410} - .410)}{\sqrt{8,169,830}} = 0.000127$$

Standard error of the difference =

$$\sqrt{0.000043 + 0.000127} + 2\sqrt{0.0876} \\ \frac{(0.006557)(0.011269)}{0.011269} \\ = 0.0146 \text{ or } 1.5 \text{ percent}$$

The confidence interval around the difference at *one* standard error is from 31.3% to -28.3% (29.8% plus or minus 1.5%). The ratio of the difference (-0.298) to its standard error (0.015) is 19.87. Since 19.87 is greater than 2.0, the difference between these two percentages is statistically significant at a confidence level exceeding 95%.

## Appendix III

### Glossary

**Age** — The appropriate age category is determined by the respondent's age on the last day of the month before the interview.

**Annual household income** — The total income of the household head and all members of the household for the 12 months preceding the interview. Includes wages, salaries, net income from businesses or farms, pensions, interest, dividends, rent, and any other form of monetary income.

**Aggravated assault** — Attack or attempted attack with a weapon, regardless of whether or not an injury occurred and attack without a weapon when serious injury results.

*With injury* — An attack without a weapon when serious injury results or an attack with a weapon involving any injury. Serious injury includes broken bones, lost teeth, internal injuries, loss of consciousness, and any unspecified injury requiring two or more days of hospitalization.

*Threatened with a weapon* — Threat or attempted attack by an offender armed with a gun, knife, or other object used as a weapon, not resulting in victim injury.

**Assault** — An unlawful physical attack or threat of attack. Assaults may be classified as aggravated or simple. Rape, attempted rape, and sexual assaults are excluded from this category, as well as robbery and attempted robbery. The severity of assaults ranges from minor threat to incidents which are nearly fatal.

**Household burglary** — Unlawful or forcible entry or attempted entry of a residence. This crime usually, but not always, involves theft. The illegal entry may be by force, such as breaking a window or slashing a screen, or

may be without force by entering through an unlocked door or an open window. As long as the person entering has no legal right to be present in the structure a burglary has occurred. Furthermore, the structure need not be the house itself for a burglary to take place; illegal entry of a garage, shed, or any other structure on the premises also constitutes household burglary. If breaking and entering occurs in a hotel or vacation residence, it is still classified as a burglary for the household whose member or members were staying there at the time the entry occurred.

*Completed burglary* — A form of burglary in which a person who has no legal right to be present in the structure successfully gains entry to a residence, by use of force, or without force.

*Forcible entry* — A form of completed burglary in which force is used to gain entry to a residence. Some examples include breaking a window or slashing a screen.

*Unlawful entry without force* — A form of completed burglary committed by someone having no legal right to be on the premises, even though no force is used.

*Attempted forcible entry* — A form of burglary in which force is used in an attempt to gain entry.

**Commercial crimes** — Crimes against commercial establishments of any type are not included in the survey. Commercial establishments include stores, restaurants, businesses, service stations, medical offices or hospitals, or other similar establishments. For victimizations occurring in commercial establishments, the crime is included or not included depending upon whether the survey respondent was threatened or harmed in some way or personal property was taken.

**Crime classification** — Victimizations and incidents are classified based upon detailed characteristics of the event provided by the respondent. Neither victims nor interviewers classify crimes at the time of interview. During data processing, a computer program classifies each event into one type of crime, based upon the entries on a number of items on the survey questionnaire. This ensures that similar events will be classified using a standard procedure. The glossary definition for each crime indicates the major characteristics required to be so classified. If an event can be classified as more than one type of crime, a hierarchy is used which classifies the crime according to the most serious event that occurred. The hierarchy is: rape, sexual assault, robbery, assault, burglary, motor vehicle theft, theft.

**Ethnicity** — A classification based on Hispanic culture and origin, regardless of race.

**Head of household** — A classification which defines one and only one person in each housing unit as the head. Head of household implies that the person rents or owns (or is in the process of buying), the housing unit. The head of household must be at least 18, unless all members of the household are under 18, or the head is married to someone 18 or older.

**Hispanic** — A person who describes himself as Mexican-American, Chicano, Mexican, Mexicano, Puerto Rican, Cuban, Central American, South American, or from some other Spanish culture or origin, regardless of race.

**Household** — A person or group of people meeting either of the following criteria. (1) people whose usual place of residence is the same housing unit, even if they are temporarily absent.

## Appendix III

### Glossary

---

(2) people staying in a housing unit who have no usual place of residence elsewhere.

**Incident** — A specific criminal act involving one or more victims and offenders. For example, if two people are robbed at the same time and place, this is classified as two robbery victimizations but only one robbery incident.

**Marital status** — Every person is assigned to one of the following classifications: (1) married, which includes persons in common-law unions and those who are currently living apart for reasons other than marital discord (employment, military service, etc.); (2) separated or divorced, which includes married persons who are legally separated and those who are not living together because of marital discord; (3) widowed; and (4) never married, which includes persons whose marriages have been annulled and those who are living together and not in a common-law union.

**Metropolitan area** — See "Metropolitan Statistical Area."

**Metropolitan Statistical Area (MSA)** — The Office of Management and Budget (OMB) defines this as a population nucleus of 50,000 or more, generally consisting of a city and its immediate suburbs, along with adjacent communities having a high degree of economic and social integration with the nucleus. MSA's are designated by counties, the smallest geographic units for which a wide range of statistical data can be attained. However, in New England, MSA's are designated by cities and towns since these subcounty units are of great local significance and considerable data is available for them. Currently, an area is defined as an MSA if it meets one of two standards:

(1) a city has a population of at least 50,000; (2) the Census Bureau defines an urbanized area of at least 50,000 people with a total metropolitan population of at least 100,000 (or 75,000 in New England). The Census Bureau's definition of urbanized areas, data on commuting to work, and the strength of the economic and social ties between the surrounding counties and the central city determine which counties not containing a main city are included in an MSA. For New England, MSA's are determined by a core area and related cities and towns, not counties. A metropolitan statistical area may contain more than one city of 50,000 and may cross State lines.

**Motor vehicle** — An automobile, truck, motorcycle, or any other motorized vehicle legally allowed on public roads and highways.

**Motor vehicle theft** — Stealing or unauthorized taking of a motor vehicle, including attempted thefts.

*Completed motor vehicle theft* — The successful taking of a vehicle by an unauthorized person.

*Attempted motor vehicle theft* — The unsuccessful attempt by an unauthorized person to take a vehicle.

**Multiple offenders** — Two or more persons inflicting some direct harm to a victim. The *victim-offender relationship* is determined by the offender with the closest relationship to the victim. The following list ranks the different relationships from closest to most distant: spouse, ex-spouse, parent, child, other relative, nonrelative well-known person, casual acquaintance, or stranger. (see *Nonstranger* and *Stranger*)

**Non-Hispanic** — Persons who report their culture or origin as something other than "Hispanic" as defined above. This distinction is made regardless of race.

**Nonstranger** — A classification of a crime victim's relationship to the offender. An offender who is either related to, well known to, or casually acquainted with the victim is a nonstranger. For crimes with more than one offender, if any of the offenders are nonstrangers, then the group of offenders as a whole is classified as nonstranger. This category only applies to crimes which involve contact between victim and the offender; the distinction is not made for crimes of theft since victims of this offense rarely see the offenders.

**Offender** — The perpetrator of a crime; this term usually applies to crimes involving contact between the victim and the offender.

**Offense** — A crime. When referring to personal crimes, the term can be used to refer to both victimizations and incidents.

**Personal crimes** — Rape, sexual assault, personal robbery, assault, purse snatching and pocket picking. This category includes both attempted and completed crimes.

**Place of occurrence of crime** — The location at which a crime occurred, as specified by the victim. Survey measures of crimes occurring in commercial establishments, restaurants, nightclubs, public transportation and other similar places include only those crimes involving NCVS measured crimes against persons, not the establishments. Crimes against commercial establishments and other places are not measured by the survey.

---

**Property crimes** — Property crimes including burglary, motor vehicle theft, or theft. This category includes both attempted and completed crimes.

**Purse snatching/Pocket picking** — Theft or attempted theft of property or cash directly from the victim by stealth, without force or threat of force.

**Race** — Racial categories for this survey are white, black, and other. The "other" category is composed mainly of Asian Pacific Islanders, and American Indian, Aleut, and Eskimo. The race of the head of household is used in determining the race of the household for computing household crime demographics.

**Rape** — Forced sexual intercourse including both psychological coercion as well as physical force. Forced sexual intercourse means vaginal, anal or oral penetration by the offender(s). This category also includes incidents where the penetration is from a foreign object such as a bottle. Includes attempted rapes, male as well as female victims, and both heterosexual and homosexual rape. Attempted rape includes verbal threats of rape.

**Rate of victimization** — see "Victimization rate".

**Region** — The States have been divided into four groups or census regions:

*Midwest* — Includes the 12 States of Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin.

*Northeast* — Includes the 9 states of Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont.

*South* — Includes the District of Columbia and the 16 States of Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia.

*West* — Includes the 13 states of Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming.

**Robbery** — Completed or attempted theft, directly from a person, of property or cash by force or threat of force, with or without a weapon, and with or without injury.

*Completed/property taken* — The successful taking of property from a person by force or threat of force, with or without a weapon, and with or without injury.

*Completed with injury* — The successful taking of property from a person, accompanied by an attack, either with or without a weapon, resulting in injury.

*Completed without injury* — The successful taking of property from a person by force or the threat of force, either with or without a weapon, but not resulting in injury.

*Attempted to take property* — The attempt to take property from a person by force or threat of force without success, with or without a weapon, and with or without injury.

*Attempted without injury* — The attempt to take property from a person by force or the threat of force without success, either with or without a weapon, but not resulting in injury.

*Attempted with injury* — The attempt to take property from a person without success, accompanied by an attack,

either with or without a weapon, resulting in injury.

**Rural area** — A place not located inside the Metropolitan Statistical Area. This category includes a variety of localities, ranging from sparsely populated rural areas to cities with populations less than 50,000.

**Sample** — The set of housing units selected by the U. S. Census Bureau to be interviewed for the survey. All occupants of the household age 12 or older are interviewed. See Appendix IV, p. 158 for sample inclusions and exclusions.

**Series** — Six or more similar but separate events, which the respondent is unable to describe separately in detail to an interviewer.

**Sexual assault** — A wide range of victimizations, separate from rape or attempted rape. These crimes include attacks or attempted attacks generally involving unwanted sexual contact between victim and offender. Sexual assaults may or may not involve force and include such things as grabbing or fondling. Sexual assault also includes verbal threats.

**Simple assault** — Attack without a weapon resulting either in no injury, minor injury (for example, bruises, black eyes, cuts, scratches or swelling) or in undetermined injury requiring less than 2 days of hospitalization. Also includes attempted assault without a weapon.

*With minor injury* — An attack without a weapon resulting in such injuries as bruises, black eyes, cuts or in undetermined injury requiring less than 2 days of hospitalization.

*Without injury* — An attempted assault without a weapon not resulting in injury.

## Appendix III

### Glossary

---

**Stranger** — A classification of the victim's relationship to the offender for crimes involving direct contact between the two. Incidents are classified as involving strangers if the victim identifies the offender as a stranger, did not see or recognize the offender, or knew the offender only by sight. Crimes involving multiple offenders are classified as involving nonstrangers if any of the offenders was a nonstranger. Since victims of theft without contact rarely see the offender, no distinction is made between strangers and nonstrangers for this crime.

**Suburban areas** — A county or counties containing a central city, plus any contiguous counties that are linked socially and economically to the central city. On data tables, suburban areas are categorized as those portions of metropolitan areas situated "outside central cities."

**Tenure** — The NCVS recognizes two forms of household tenancy: (1) owned, which includes dwellings that are mortgaged, and (2) rented, which includes rent-free quarters belonging to a party other than the occupants, and situations where rental payments are in kind or services.

**Theft** — Completed or attempted theft of property or cash without personal contact. Incidents involving theft of property from within the sample household would classify as theft if the offender has a legal right to be in the house (such as a maid, delivery person, or guest). If the offender has no legal right to be in the house, the incident would classify as a burglary.

*Completed* — To successfully take without permission property or cash without personal contact between the victim and offender.

*Attempted* — To unsuccessfully attempt to take property or cash without personal contact.

**Urban areas** — The largest city (or grouping of cities) in a Metropolitan Statistical Area (see definition of Metropolitan Statistical Area).

**Victim** — The recipient of a criminal act, usually used in relation to personal crimes, but also applicable to households.

**Victimization** — A crime as it affects one individual person or household. For personal crimes, the number of victimizations is equal to the number of victims involved. The number of victimizations may be greater than the number of incidents because more than one person may be victimized during an incident. Each crime against a household is assumed to involve a single victim, the affected household.

**Victimization rate** — A measure of the occurrence of victimizations among a specified population group. For personal crimes, this is based on the number of victimizations per 1,000 residents age 12 or older. For household crimes, the victimization rates are calculated using the number of incidents per 1,000 households.

**Victimize** — To commit a crime against a person or household.

**Violence, crimes of** — Rape, sexual assault, personal robbery or assault. This category includes both attempted and completed crimes. It does not include purse snatching and pocket picking. Murder is not measured by the NCVS because of an inability to question the victim.

*Completed violence* — The sum of all completed rapes, sexual assaults, robberies, and assaults. See individual crime types for definition of completed crimes.

*Attempted/threatened violence* — The unsuccessful attempt of rape, sexual assault, personal robbery or assault. Includes attempted attacks or sexual assaults by means of verbal threats. See individual crime types for definition of attempted crimes.