# Methodology

The survey is conducted by telephone and represents a collaborative effort between the WVBPH and CDC. The Bureau provides telephones, office space, interviewers, and supervision of the data collection. Financial assistance, a standardized set of core questions, computer-assisted telephone interviewing software, data processing services, and analytic consultation are provided by CDC.

A prepared introductory statement and the core questions were developed and tested in the field by CDC. Interviews require approximately 15-20 minutes to complete. In addition to behavioral risk factors, they cover standard demographic characteristics and selected preventive health practices. A very limited number of questions of topical interest may be added by individual states to the survey.

Phone calls and interviews are conducted by the WVBPH for approximately a two-week period each month. The monthly interview schedule reduces the possibility of bias because of seasonal variations in certain lifestyles. To assure maximum response rates, calls are made weekdays from noon to 9:00 p.m. and on Saturdays from 10:00 a.m. to 7:00 p.m.

## SAMPLE SELECTION

According to figures from the 2000 U.S. Census, 95.3% of West Virginia households have telephones, compared to 97.6% of households in the United States. The sample was selected by random digit dialing (RDD). Telephone directories are not used since they do not include unlisted or new numbers. From 1984 to 1998, sampling was conducted in a multistage cluster design based on the Waksberg Sampling Method for Random Digit Dialing. Since 1999, the sampling method known as Disproportionate Stratified Sampling (DSS) has been used. Both methods eliminate many unassigned and business phone numbers from the selection process.

CDC provides banks of telephone numbers that are presumed to contain either more household numbers (higher-density stratum) or fewer household numbers (lower-density stratum). The higherdensity stratum is sampled at a higher rate than the lower-density stratum. In 2003 the higher-density stratum consisted of banks of numbers that contained listed residential numbers while the lower-density stratum consisted of banks of numbers that contained unlisted residential numbers. The higher-density stratum was sampled at a rate of 1.5 to 1 compared to the lower-density stratum. The data ultimately were weighted to account for differences in selection probability. Calls were made until each number resulted in a completed interview or a refusal or was disqualified. A number was disqualified if it was nonresidential or nonworking, if there was no eligible respondent available during the survey, if the selected respondent was unable to communicate, or if the number had been called at least 15 times without success (encompassing a minimum of three attempts each during afternoon, evening, and weekend). Within each household, the actual respondent was chosen randomly to avoid possible biases related to the time of day and household telephone answering preferences. Since the number of adult residents and the number of telephone lines may differ from household to household, resulting in different probabilities of being selected, data were weighted to compensate for this bias. Table M.1 on the following page shows the results for all the telephone numbers attempted in obtaining a total of 3,349 interviews during 2003.

Disposition	Number	Percent
Completed interview	3 310	30.65
Partially completed interview	39	0.36
Terminated within questionnaire <50% finished	72	0.67
Refusal after respondent selection.	588	5.44
Selected respondent never reached or was reached but did not begin		
interview during interviewing period	194	1.80
Selected respondent away from residence during the entire interviewing		
period	152	1.41
Selected respondent physically or mentally unable to complete an		
interview during the entire interviewing period	93	0.86
Hang up or termination after number of adults recorded but before		
respondent selection, explicit refusal	16	0.15
Household members away from residence during entire interviewing		
period	33	0.31
Hang up or termination, housing unit, unknown if eligible		
respondent	337	3.12
Household contact, eligibility undetermined	55	0.51
Physical or mental impairment before respondent selection	9	0.08
Hang up or termination, unknown if private residence	745	6.90
Contacted, unknown if private residence	50	0.46
Telephone answering device, message confirms private residential		
status	100	0.93
Telecommunication technological barrier (such as a call blocking		
message), message confirms private residence	10	0.09
Telephone answering device, not sure if private residence	164	1.52
Telecommunication technological barrier, not sure if private	24	0.10
residence	21	0.19
Telephone number changed status from household or possible	02	0.77
household to nonworking during the interviewing period	83	0.77
No answer	547	5.06
Busy	53	0.49
Un never-call list.	1	0.01
Not a private residence	12	0.11
Dedicated for/data/modem line with no human contact	1,078	9.98
Foot busy	23	2.94
Nonworking/disconnected number	25	24.07
	2,097	24.77
Total	10,800	100.00

# Table M.1: Disposition of telephone numbers in the sample: WVBRFSS, 2003

# **QUALITY CONTROL**

The degree to which completed interviews are obtained from among the telephone numbers selected for the sample can be shown numerically by response rates. A higher response rate indicates a lower potential for bias in the data. A discussion of response rates as well as various sources of statistical bias can be found in CDC's *Behavioral Risk Factor Surveillance System 2003 Year-to-Date Data Quality Handbook.* While there is no definitive formula for response rate, three primary estimates are most useful for BRFSS:

<u>CASRO</u> is a response rate formula<sup>1</sup> developed by the Council of American Survey Research Organizations (CASRO). The resulting estimate reflects telephone sampling efficiency and the degree of cooperation among eligibles contacted. The formula assumes that numbers that are never contacted contain the same percentage of eligible households as the records whose eligibility status is known. Quality control guidelines by CDC suggest a minimum acceptable value of 40%. West Virginia's CASRO rate for the year 2003 was 60%.

<u>**Overall Response Rate**</u> is a conservative response rate<sup>2</sup> that includes a higher percentage of all households in the denominator. Quality control guidelines by CDC suggest a minimum acceptable value of 30%. West Virginia's overall response rate for the year 2003 was 58%.

<u>Cooperation Rate</u> is a calculation<sup>3</sup> that is not affected by differences in telephone sampling efficiency. It is the proportion of all cases interviewed of all eligible units that were actually contacted. Non-contacts are excluded from the denominator. This rate is based on contacts with households containing an eligible respondent. The denominator of the rate includes completed interviews plus the number of non-interviews that involve the identification of and contact with an eligible respondent. Quality control guidelines by CDC suggest a minimum acceptable value of 65%. West Virginia's cooperation rate for the year 2003 was 82%.

The survey results were edited daily to assure proper completion. For verification, call backs were completed randomly to confirm that interviews had been conducted as indicated. After all phone numbers received a final disposition each month, the data were edited to check for entries that were invalid or inconsistent with other entries. Data were also checked for answers that were outside the expected range of values, such as extreme values for height, weight, exercise times, or alcohol consumption. Once all of the data were corrected or verified as correct, the results were sent to CDC via electronic mail. An annual analysis of the data is provided to the state by CDC.

1	CASRO rate =	Completed Interviews
2	Overall response rate	Known Eligibles + [(Known Eligibles/{Known Eligibles & Ineligibles}) x (Unknowns)] =Completed Interviews
3	Cooperation rate -	Eligible Households
		Completed Interviews + Terminated Before Completion + Refusals + Unable to Communicate

# DEMOGRAPHIC CHARACTERISTICS OF SAMPLE AND POPULATION

The demographic characteristics of the 2003 sample, both unweighted and weighted to the population, are presented below.

# Table M.2: Demographic Summary: WVBRFSS, 2003

Demographic characteristic	# Interviews	% Unweighted Sample	% Weighted Sample <sup>a</sup>
Total	3,349	100.0	100.0
Sex			
Male	1,323	39.5	47.9
Female	2,026	60.5	52.1
Age			
18-24	203	6.1	12.4
25-34	455	13.6	15.6
35-44	557	16.6	17.7
45-54	672	20.1	19.3
55-64	647	19.3	14.3
65+	803	24.0	20.5
Unknown	12	0.4	0.2
Education			
<12 Years	636	19.0	18.7
12 Years	1,321	39.4	40.0
13-15 Years	741	22.1	22.9
16+ Years	647	19.3	18.3
Unknown	4	0.1	0.2
Household Income			
<\$15.000	526	15.7	13.1
\$15.000-\$24.999	724	21.6	20.5
\$25,000-\$34,999	463	13.8	13.7
\$35,000-\$49,999	478	14.3	14.8
\$50,000-\$74,999	393	11.7	12.7
\$75,000+	329	9.8	11.1
Unknown	436	13.0	14.2

a. Population weight provided by CDC. Weighted to 2003 age and sex postcensual estimates. Not weighted to education or income level.

Compared to the 2003 census estimates, male respondents and persons aged 18 to 24 were underrepresented in the sample, while females and the elderly (65 and older) were overrepresented, a frequent result of telephone surveys. Survey responses were therefore weighted by the census age and sex distribution in order to more accurately estimate the actual prevalence of behavioral risk factors in the adult population of West Virginia.

## LIMITATIONS

Self-reported behavior obtained by telephone must be interpreted with caution. The validity of survey results depends on the accuracy of the responses given by the persons interviewed. This may be affected by the ability to recall past behavior. For example, individuals may not accurately recall blood pressure or cholesterol levels. In addition, respondents may have a tendency to understate behaviors known to be unhealthy, socially unacceptable, or illegal. These biases may vary depending on the specific risk factor.

Other sources of bias may result from greater difficulty in contacting some persons, from higher refusal rates, or from lower telephone coverage. Given the possibility that persons not interviewed for these reasons may behave differently from the general population, estimates for the population based on the survey sample may be biased. Weighting the data by age and sex distribution is done in order to correct for over or underrepresentation of these groups.

Finally, breaking down the data into smaller categories decreases the sample size of the individual strata, thereby decreasing the power to determine statistically significant differences. Prevalence rates based on denominators of less than 50 are considered statistically unreliable.

## ESTIMATES AND CONFIDENCE INTERVALS

Because the prevalence rates shown in tables throughout the report are derived from surveying a sample of people rather than all adults in the population, the resultant rates are estimates. For this reason, the estimated rates are presented together with their associated confidence intervals. Confidence intervals reflect sampling error and represent the range of values among which the true value would be found. The prevalence tables show 95% confidence intervals, meaning the true value would be within the given interval 95% of the time. When confidence ranges do not overlap, the estimates they are based upon may be termed significantly different. Confidence intervals were derived from the surveymeans procedure in SAS, a common statistical software package. This procedure estimates sample variances (which are used to calculate confidence intervals) for complex sample designs.

## **COUNTY-LEVEL DATA**

County prevalence rates were calculated by using multiple years of aggregated BRFSS data. The weighting procedures were the same as those for state-level data, with the exception that the 2003 age and sex population distribution for the state was replaced by the 2000 age and sex population distribution by county. Aggregated sample sizes were large enough for 24 of the 55 counties to stand alone, that is, to yield individual county prevalence calculations. The data from the remaining 31 counties that had sample sizes too small to stand alone were combined into 12 groupings of counties. The aim was to arrive at as many groups of contiguous counties as possible, provided that the groups' sample sizes were sufficiently large for statistical analysis. Similarity in poverty level was an additional factor in deciding which counties to group together. Whenever a risk factor prevalence was calculated for a group of counties, each county within the group was considered to have the same prevalence. The 12 groups of counties plus the 24 stand-alone counties resulted in 36 geographical entities (see Appendix I). The county prevalence rates were then compared to the U.S. prevalence for 2001. Counties were classified according to the degree of difference from the U.S. prevalence: significantly higher, higher, lower, and

significantly lower.<sup>4</sup> Risk factor rates by county are shown in Appendix J. Extensive county data also can be found in the WVBPH publication *West Virginia County Health Profiles, 2004*.

#### PRESENTATION

In the sections that follow, the prevalence data are presented in a variety of ways, including by state rank, yearly state prevalence, and demographic variables. It should be stressed that the risk factor prevalence rates for the demographic variables (age, sex, education, and income) show the percentages of persons within the group – not in the total survey sample – who report the behavior being examined. This method of presenting risk factor prevalence facilitates identification of at-risk populations for health promotion efforts. Each table shows the number of respondents (# Resp.) who were asked the question, the weighted prevalence rate (%), and the 95% confidence interval for the prevalence (95% CI).

Prevalence rates are calculated by excluding unknown responses from the denominators. Consequently, rates may be slightly higher than would have been the case had unknown responses been included. In prior publications, rates representing the years 1984 through 1996 were often calculated by including the unknown responses. In this report, all rates have been re-calculated with the unknown responses excluded. Therefore, discrepancies may exist between the time trends and appendixes in this report and prior publications.

The risk factor sections include West Virginia's rank among the 54 BRFSS participants, with 1<sup>st</sup> as highest in prevalence and 54<sup>th</sup> as lowest. For example, ranking 1<sup>st</sup> in hypertension would mean having the highest prevalence of hypertension of all BRFSS participants; conversely, ranking 54<sup>th</sup> would mean having the lowest prevalence. Some questions are not asked by all BRFSS participants. In these cases, the rankings should be interpreted with caution, as they may be different if information were available from all participants. In addition, readers should note that differences between states often are less than one percentage point and that statistical significance was not tested when determining rankings. The rates and rankings were calculated by Health Statistics Center staff. State and county prevalences and rankings for many risk factors are presented in Appendixes H and J.

<sup>&</sup>lt;sup>4</sup> Statistical significance can be affected by both prevalence level and county sample size.