

## METHODOLOGY

The survey is conducted by the method known as Computer Assisted Telephone Interviewing (CATI) 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 and survey protocols, 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. The interviews take approximately 15-20 minutes. In addition to behavioral risk factors and certain health conditions, 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- to three-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., Saturdays from 10:00 a.m. to 7:00 p.m., and Sundays from 2:00 p.m. to 6:00 p.m.

### **SAMPLE SELECTION**

The sample was selected by random digit dialing (RDD). Telephone directories are not relied upon since they do not include unlisted or new numbers. From 1984 through 1998, sampling was conducted in a multistage cluster design based on the Mitofsky-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 higher-density stratum is sampled at a higher rate than the lower-density stratum. In 2006, the higher-density stratum consisted of banks of listed numbers while the lower-density stratum consisted of banks of unlisted numbers that contained at least one residential number. 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 afternoons, evenings, and weekends). 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 3,675 interviews in 2006.

**Table M.1: Disposition of telephone numbers in the sample: WVBRFSS, 2006**

Disposition	Number	Percent
Completed interview.....	3,675	23.69
Partially completed interview.....	119	0.77
Terminated within questionnaire <50% finished.....	97	0.63
Refusal after respondent selection.....	781	5.04
Selected respondent never reached or was reached but did not begin interview during interviewing period.....	211	1.36
Selected respondent away from residence during the entire interviewing period.....	112	0.72
Language problem after respondent selection.....	7	0.05
Selected respondent physically or mentally unable to complete an interview during the entire interviewing period.....	139	0.90
Hang up or termination after number of adults recorded but before respondent selection, explicit refusal.....	13	0.08
Household contact after number of adults recorded but before respondent selection.....	1	0.01
Household members away from residence during entire interviewing period.....	28	0.18
Hang up or termination, housing unit, unknown if eligible respondent.....	515	3.32
Household contact, eligibility undetermined.....	43	0.28
Language problem before respondent selection.....	10	0.06
Physical or mental impairment before respondent selection.....	15	0.10
Hang up or termination, unknown if private residence.....	1,168	7.53
Contacted, unknown if private residence.....	31	0.20
Telephone answering device, message confirms private residential status.....	221	1.42
Telecommunication technological barrier (such as a call blocking message), message confirms private residence.....	13	0.08
Telephone answering device, not sure if private residence.....	237	1.53
Telecommunication technological barrier, not sure if private residence.....	31	0.20
Telephone number changed status from household or possible household to nonworking during the interviewing period.....	88	0.57
No answer.....	597	3.85
Busy.....	66	0.43
On never-call list.....	1	0.01
Out-of-state.....	2	0.01
Household, no eligible respondent.....	8	0.05
Not a private residence.....	1,753	11.30
Dedicated fax/data/modem line with no human contact.....	405	2.61
Cell phone.....	45	0.29
Fast busy.....	41	0.26
Nonworking/disconnected number.....	5,037	32.48
Total.....	15,510	100.00

## 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 and of various sources of statistical bias can be found in CDC's *Behavioral Risk Factor Surveillance System 2006 Year-to-Date Data Quality Handbook*. While there is no definitive formula for response rate, three primary estimates are most useful for the BRFSS:

**CASRO Rate** uses 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 eligible persons who were 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 monthly CASRO rate ranged from 58.62% to 61.44% in 2006.

**Overall Response Rate** 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 monthly overall response rate ranged from 52.86% to 56.65% in 2006.

**Cooperation Rate** 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 monthly cooperation rate ranged from 78.59% to 79.74% in 2006.

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 also were 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 monthly datasets were submitted electronically to CDC. An annual analysis of the data is provided to the state by CDC.

## DEMOGRAPHIC CHARACTERISTICS OF SAMPLE AND POPULATION

The demographic characteristics of the sample in 2006, both unweighted and weighted to the population, are presented in Table M.2. Data were 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.

---

<sup>1</sup> CASRO rate = 
$$\frac{\text{Completed Interviews}}{\text{Known Eligibles} + [(\text{Known Eligibles} / \{\text{Known Eligibles} \& \text{Ineligibles}\}) \times (\text{Unknowns})]}$$

<sup>2</sup> Overall response rate = 
$$\frac{\text{Completed Interviews}}{\text{Eligible Households}}$$

<sup>3</sup> Cooperation rate = 
$$\frac{\text{Completed Interviews}}{\text{Completed Interviews} + \text{Terminated Before Completion} + \text{Refusals} + \text{Unable to Communicate}}$$

**Table M.2: Demographic summary: WVBRFSS, 2006**

Demographic characteristic	Number of Interviews	Percent of Unweighted Sample	Percent of Weighted Sample <sup>a</sup>
<b>Total</b>	3,794	<b>100.00</b>	<b>100.00</b>
<u>Sex</u>			
Male	1,495	39.40	48.29
Female	2,299	60.60	51.71
<u>Age</u>			
18-24	145	03.82	12.26
25-34	415	10.94	15.47
35-44	601	15.84	17.07
45-54	761	20.06	19.29
55-64	837	22.06	15.88
65+	1,018	26.83	20.04
Unknown	17	00.45	N/A
<u>Education</u>			
<12 Years			
12 Years	620	16.34	15.36
13-15 Years	1,493	39.35	39.87
16+ Years	824	21.72	23.03
Unknown	850	22.40	21.75
	7	00.18	N/A
<u>Household Income</u>			
<\$15,000	583	15.37	13.73
\$15,000-\$24,999	729	19.21	20.36
\$25,000-\$34,999	496	13.07	14.34
\$35,000-\$49,999	524	13.81	17.46
\$50,000-\$74,999	498	13.13	16.64
\$75,000+	509	13.42	17.47
Unknown	455	11.99	N/A

a. Population weight provided by CDC. Weighted to 2006 age and sex postcensus estimates. Not weighted to education or income level. Unknown values for age were replaced by imputed ages for weighting purposes only. The "Percent of Weighted Sample" is the percent with a known value and excludes records with unknown values.

## LIMITATIONS

The target population consists of civilian, noninstitutionalized persons 18 years of age and older who reside in households with telephones. Some questions in the questionnaire also pertain to children who live in such households. State residents who do not fit the target population are not represented in prevalence estimates.

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 fewer than 50 are considered statistically unreliable.

## **ESTIMATES, CONFIDENCE INTERVALS, SIGNIFICANCE, AND RELIABILITY**

The prevalence rates presented in this report are derived from surveying a sample of adults rather than all adults in the population; therefore, the rates are estimates of the true values. For this reason, estimates are presented together with their associated confidence intervals. A confidence interval is a range of values around an estimate, which reflects sampling error and represents the uncertainty of the estimate. This report presents 95% confidence intervals (95% CI)<sup>4</sup>. Therefore, we can be 95% confident that the confidence interval contains the true value that we are estimating.

Significant is the term used in this report to describe prevalence estimates that have been tested and found to be significantly different. Statistically significant differences between estimates are traditionally determined using statistical tests such as a t-test or chi-squared test. However, when comparing estimates from surveys with a large number of respondents, such as the BRFSS, these statistical tests can indicate statistically significant differences even when there are only small variations in prevalence. This method would label most of the estimate comparisons in this report as significantly different. Therefore, this report uses the following more conservative method for determining significance. Two prevalence estimates are said to be “significantly” different when the 95% confidence intervals (CIs) associated with each of the estimates do not overlap. In other words, it can be stated with 95% certainty that the difference found between the two prevalence estimates is not a random occurrence. Although this is not the “classical” statistical test of differences, it is a better method of highlighting the BRFSS results important to the design of effective and efficient health promotion interventions. Identifying differences as significant by this method targets the characteristics most strongly associated with a particular health condition or risk behavior, and directs attention to the largest changes in prevalence over time. Adjectives such as slight, minor, and little are used in this report to describe notable differences that are not considered significant because the confidence intervals do overlap.

Reliability refers to the precision of an estimate. If an estimate is termed reliable, there is confidence that the same, or a very similar, estimate would be obtained if the survey were to be repeated within the same time period. Estimates that are determined to be unreliable may not reflect the true prevalence; therefore, they should be reported and interpreted with caution. Throughout this report, unreliable estimates are noted with this message: “Use caution when interpreting and reporting this estimate. See discussion of unstable estimates on page 9.” Based on CDC recommendations, estimates in this report were termed unreliable if any of the three following conditions were met:

- 1) The estimate is based on responses from fewer than 50 respondents.
- 2) The 95% confidence interval of the estimate has a width or range greater than 20 (e.g., 95% CI = 10.0-30.5).
- 3) The estimate has a relative standard error (RSE) of 30.0% or higher. The RSE is obtained by dividing the standard error of the estimate by the estimate itself. It is calculated by the SAS software.

---

<sup>4</sup> Confidence intervals were derived from the surveyfreq procedure in SAS, a commonly used 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 data were reweighted to be representative of 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. The 12 groups of counties plus the 24 stand-alone counties resulted in 36 geographical entities (see Appendix L).

In some prior reports, the county prevalence estimates were compared to a middle-year United States prevalence estimate. County maps were included that classified counties according to the degree of difference from the United States prevalence: significantly higher, higher, lower, or significantly lower.<sup>5</sup> In this report, county estimates were compared to the total West Virginia estimate for the same time period. This method better identifies disparities between counties. It also clearly identifies counties in need of health promotion interventions. *The county maps included in this report classify counties according to the degree of difference from the West Virginia prevalence, not the United States prevalence.* County estimates, as well as county classifications compared to both West Virginia and the United States, can be found in [Appendix M](#). Extensive county data also can be found in the WVBPB publication *West Virginia County Health Profiles, 2004* available online at <http://www.wvdhhr.org/bph/oehp/hsc/profiles2004/default.htm>.

Unlike some previous reports, this report does not include county prevalence estimates of heavy drinking. Based on the reliability standards discussed above, a majority of the county estimates were determined to be unreliable, primarily attributable to the low statewide prevalence of this risk factor. Use caution when interpreting county estimates of heavy drinking published in earlier reports. It is likely that many of the estimates are unreliable.

## PRESENTATION

In the sections that follow, the prevalence data are presented in a variety of ways, including by state rank, yearly state and national prevalence, and demographic variables. It should be stressed that the risk factor prevalence estimates 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 estimate (%), and the 95% confidence interval for the prevalence (95% CI).

Prevalence estimates are calculated by excluding unknown responses from the denominators. Consequently, estimates may be slightly higher than would have been the case had the unknown responses been included. In editions of this report before 2003, many estimates representing the years 1984 through 1996 were calculated by including unknown responses. In the present report, all such rates have been re-calculated to exclude unknown responses. Therefore, discrepancies may exist between the time trends and appendixes in this report and those in older editions.

The risk factor sections include West Virginia's rank among the US States and DC. For example, if hypertension-related questions were administered by all 50 US States and DC, ranking 1<sup>st</sup> in

---

<sup>5</sup> Significance can be affected by both prevalence level and county sample size.

hypertension would mean having the highest prevalence of hypertension while ranking 51<sup>st</sup> would mean having the lowest prevalence. Some questions are not asked by all US States and DC. 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 prevalence estimates and rankings for many risk factors are presented in Appendixes A and M.