

Technical Notes

Survey Management

The BRFSS is a cooperative effort of the Centers for Disease Control and Prevention and participating states. The CDC develops the core questionnaire and provides training, technical assistance, standardized data analyses and funding. The Pennsylvania Department of Health develops supplemental questions (or modules) and responds to requests for data. In 2011, Pennsylvania was divided into eight strata. These strata consisted of eight regional areas: six Pennsylvania health districts and Allegheny and Philadelphia Counties. Sampling and interviewing in 2011 were done by Clearwater Research, Incorporated, which was selected by competitive bid.

Major Changes in 2011

In 2011, the BRFSS added a sample of cell phone numbers to the sample of numbers of landline telephones used in previous years. This was necessary due to the increasing proportion of adults in the United States and in Pennsylvania who live in cell phone-only households. These people tend to be younger, more mobile, and to belong to racial and ethnic minority groups. The loss of this group to the survey population introduced bias to the estimates.

A new weighting methodology called “iterative proportional fitting,” or “raking,” which is explained in the Data Adjustment section, includes the telephone source (landline or cell phone) in the weighting methodology. It also allows the BRFSS to adjust the sample interviews to match the characteristics of the population not only on age, sex and race but also on marital status, education and ownership or rental of the home.

These changes amount to a major re-working of the BRFSS survey and will shift estimates and trend lines in ways which are not related to changes in the actual population. It is advisable to re-benchmark any trend lines and measures, using 2011 as a new starting point.

Further changes in BRFSS methodology are anticipated, but they are thought to be minor compared with the changes in 2011. Cell phone-only respondents made up 18 percent of the Pennsylvania sample in the survey for 2011. This will increase to 20 percent in the 2012 survey and 25 percent in the 2013 survey. In 2012 the cell phone sample will expand to include landline households that use their cell phones for at least 90 percent of their calls. In 2012, respondents living in college group quarters will be added to the sample.

Sample Selection - Landline

Respondents were selected using a two-stage random digit dialing (RDD) sample design. In the first selection stage, a disproportionate stratified sample of telephone numbers was selected from two telephone number strata. One stratum consisted of blocks of numbers containing one or more listed numbers and presumed by the sampling firm to contain a high density of residential telephone numbers. The other stratum consisted of blocks of telephone numbers which include one or more unlisted telephone numbers and which are presumed to contain a medium density of residential numbers. Both strata include only numbers which begin with area codes and exchange prefixes specific to Pennsylvania.

For the sake of efficiency, under the disproportionate stratified sample design, a larger proportion of the sample is selected from the stratum presumed to contain a high density of residential households.

Sampling continued as the selected telephone numbers were called, to determine whether targeted numbers belonged to households with adult residents. Nonresidential telephone numbers were discarded from the sample. Residential numbers were subjected to a second stage of sampling wherein an adult was randomly selected as the respondent from a list of adults residing in the household. The person who answered the telephone generated this list.

Sample Selection – Cell Phone

Cell phone usage differs from that of landline telephones. Accordingly, the management of the cell phone sample has its own characteristics.

Cell phone sample is selected randomly from blocks of numbers dedicated to cell phones. No subsequent within-household selection of a respondent is made if an adult is reached. Numbers selected as part of the landline sample for the survey, that is those ported to a cell-phone, are transferred to the cell phone sample and called. Interviews of respondents selected from cell phone sample for other states but resident in Pennsylvania at the time of the survey, are added to the Pennsylvania sample. These interviews will not contain responses to optional modules or state-added questions. Similarly, interviews of respondents from the Pennsylvania cell phone sample who were found during the interview to have moved to another state are transferred to the BRFSS sample of the other state.

Cell phone respondents are immediately asked whether they are driving or otherwise in an unsafe place. If so, they are re-called at another time, but some interviews are lost when this happens. Fewer follow-back calls are made for cell phones than for landline telephones. Unlike landline calls, a message identifying the survey is left at the first call attempt if there is no answer.

Questionnaire

The survey questionnaire for the statewide survey of Pennsylvania consists of a standardized core, state-selected modules and state-added questions. The CDC developed the core questionnaire with recommendations from all participating jurisdictions. Most of the core questions had been used during the 2010 BRFSS survey. All items new to the 2011 survey were field-tested.

Questions of interest to Pennsylvania were added as the state supplement to the core questionnaire. State-added modules and questions concerned asthma, childhood immunization, childhood asthma prevalence, actions to control high blood pressure, organ donation, injury prevention and gambling.

Response Rates

Interviews were conducted in the evenings and on weekends in order to reach people when they were more likely to be at home, as well as during the day. For landline sample, at least 15 calls were placed at different times of the day and night on different days of the week before any sample number was classified as “no answer.”

Interviewers who were experienced in converting refusals to completed interviews re-contacted people who refused to participate in the survey.

For cell phone sample, if a number has not been reached within the first six attempts, the record will receive a final disposition code on the seventh attempt. However, if any contact has been made within the first six attempts, the number will receive additional attempts, up to 12 total attempts.

The final dispositions of the landline and cell phone samples, as recorded by the data collection firm, are shown in the tables on the following pages. These dispositions allow calculation of the CASRO response rates. CASRO response rates may be thought of as the percentage of eligible telephone numbers which yielded an interview, adjusted in a standard way for the large numbers of telephone numbers of unknown eligibility. The CASRO response rates for the landline and cell phone samples for the 2011 survey are 44.6 and 20.0, respectively.

**Disposition of All Telephone Numbers – Landline Survey
2011 Pennsylvania Behavioral Risk Factor Survey Sample**

<u>Disposition Code</u>	<u>Number</u>	<u>Percent</u>
Interview		
Complete	8,631	6.72
Partial Complete	788	0.61
Eligible, Non-Interview		
Termination within questionnaire	1,105	0.86
Refusal after respondent selection	3,597	2.80
Selected respondent never reached or was reached but did not begin interview during interviewing period	1,261	0.98
Selected respondent away from residence during the entire interviewing period	547	0.43
Language problem after respondent selection	110	0.09
Selected respondent physically or mentally unable to complete an interview during the entire interviewing period	573	0.45
Hang-up or termination after number of adults recorded but before respondent selection	38	0.03
Contacted, but terminated after number of adults recorded but before respondent selection	1	0.00
Unknown Eligibility, Non-Interview		
Household members away from residence during entire interviewing period	184	0.14
Hang-up or termination, housing unit, unknown if eligible respondent	312	0.24
Household contact, eligibility undetermined	117	0.09
Language problem before respondent selection	337	0.26
Physical or mental impairment before respondent selection	362	0.28
Hang-up or termination, unknown if private residence	13,779	10.73
Contact, unknown if private residence	2,202	1.72
Telephone answering device, message confirms private residential status	1,308	1.02
Telecommunication technological barrier, message confirms private residential status	63	0.05
Telephone answering device, not sure if private residence	6,696	5.22
Telecommunication technological barrier, not sure if private residence	153	0.12
Telephone number is no longer in service or has been changed	833	0.65
No answer	4,348	3.39
Busy	210	0.16
Not Eligible		
Out-of-state	9	0.01
Household, no eligible respondent	12	0.01
Not a private residence	4,150	3.23
Dedicated fax/data/modem line with no human contact	3,576	2.79
Cellular Phone	372	0.29
Fast busy	2,235	1.74
Non-working/disconnected number	4,943	3.85
Precalls	65,518	51.04
Total	128,370	

**Disposition of All Telephone Numbers – Cell Phone Survey
2011 Pennsylvania Behavioral Risk Factor Survey Sample**

<u>Disposition Code</u>	<u>Number</u>	<u>Percent</u>
Interview		
Complete	1,853	2.78
Partial Complete	225	0.34
Eligible, Non-Interview		
Termination within questionnaire	193	0.29
Refusal after respondent selection	1,657	2.48
Selected respondent never reached or was reached but did not begin interview during interviewing period	566	0.85
Selected respondent away from residence during the entire interviewing period	1	0.00
Language problem after respondent selection	70	0.10
Selected respondent physically or mentally unable to complete an interview during the entire interviewing period	12	0.02
Hang-up or termination after number of adults recorded but before respondent selection	1	0.00
Unknown Eligibility, Non-Interview		
Hang-up or termination, housing unit, unknown if eligible respondent	749	1.12
Cell contact – eligibility undetermined	246	0.37
Contact – eligibility undetermined	1,844	2.76
Language problem before respondent selection	279	0.37
Physical or mental impairment before respondent selection	27	0.04
Hang-up or termination, unknown if private residence	14,065	21.08
Contact, unknown if private residence	1	0.00
Telephone answering device, message confirms private residential status	4,488	6.73
Telecommunication technological barrier, message confirms private residential status	2,741	4.11
Telephone answering device, not sure if private residence	9,574	14.35
Telecommunication technological barrier, not sure if private residence	1,367	2.05
Telephone number is no longer in service or has been changed	1,940	2.91
No answer	769	1.15
Busy	45	0.07
Not Eligible		
Out-of-state	55	0.08
Cell phone – not an adult	1,616	2.42
Not a private residence	583	0.87
Cell phone – business only	1,715	2.57
Landline phone	403	0.60
Dedicated fax/data/modem line with no human contact	20	0.03
Cellular Phone with landline in household	6,237	9.35
Fast busy	1,276	1.91
Non-working/disconnected number	11,567	17.34
Cell phone wrong number	535	0.80
Total	66,720	

Sample Characteristics

The following table compares the final interview sample for the 2011 BRFSS to the 2010 Population estimates for the adult population of Pennsylvania. The observations used to calculate the estimates presented in the main report were weighted to account for differences between the population and the distribution of age, sex, race and Hispanic origin characteristics in the sample.

**Distribution of 2011 Pennsylvania BRFSS Survey Sample and
2010 Pennsylvania Adult Population Estimates for Selected Characteristics**

		2011 BRFSS Survey Sample		2010 Population Estimates	
		Number	Percent	Number	Percent
All Adults		11,509	100.00	9,910,224	100.00
Sex	Male	4,610	40.06	4,760,912	48.04
	Female	6,899	59.94	5,149,312	51.96
Race	White	9,981	86.72	8,314,232	83.90
	Black	942	8.18	992,798	10.02
	Other	368	3.20	603,194	6.09
	Unknown/Refused	218	1.89	N/A	N/A
Hispanic Origin	Yes	291	2.53	459,421	4.64
	No	11,134	96.74	9,450,803	95.36
	Unknown/Refused	84	0.73	N/A	N/A
Age	18-24	628	5.46	1,261,381	12.73
	25-34	1,218	10.58	1,511,119	15.25
	35-44	1,441	12.52	1,615,669	16.30
	45-54	2,024	17.59	1,940,404	19.58
	55-64	2,474	21.50	1,622,344	16.37
	65-74	1,854	16.11	979,538	9.88
	75+	1,743	15.14	979,769	9.89
	Unknown/Refused	127	1.10	N/A	N/A

Note¹: Race data include Hispanics.

Note²: Population estimates allocate unknowns, so they are included in demographic categories. This is further indicated by the use of “N/A” or not applicable for the 2010 population estimate “Unknown/Refused” entries.

Determining Accuracy of the Estimates and Significance Using Confidence Intervals

Tables included in this report show the 95 percent confidence intervals associated with all reported percentages. They appear in the table columns labeled (CI).

Confidence intervals are a way to measure sampling error and define the range of values where percentages estimated by multiple samples of the same population would be found (95 percent of the time). The size of the confidence interval is directly related to the probability of selection and characteristics of the people surveyed within the universe being sampled. Percentages for two different subgroups of the population are significantly different if their confidence intervals or ranges do not overlap.

Confidence intervals were calculated using SUDAAN, a software package developed by the Research Triangle Institute that properly estimates sample variances for complex sample designs.

Percentages were not calculated and shown for subgroups of the population when their sample size was less than 50. The method used to determine the reliability of percentages calculated from sample sizes of 50 or more consisted of a comparison of the relative standard error of the calculated percentage with the relative standard error of the same percentage outcome for a simple random sample. If the relative standard error for the percentage being tested was smaller than the relative error of the same percentage outcome for the simple random sample, then the calculated percentage was considered reliable.

Data Adjustment

Before 2011, BRFSS weights were based on a number of design factors such as the number of adults and the number of telephones in the household, as well as differential sampling of households which are more and less easily reached. After design factors were taken care of, an additional factor was applied which compensated for shortfalls due to non-response and under-coverage.

This was called the post-stratification adjustment. Usually the Pennsylvania survey has been post-stratified by two sexes and at least six age groups, involving six age-by-sex categories, for a total of twelve cells.

Beginning with the 2011 sample, the weighting process became much more complex. After weights for design factors were applied the sample was adjusted, using the raking method to agree proportionally with more external references (called margins).

The adjustments are applied to the sample sequentially, with the objective of bringing the total of each cell of each margin into the same proportion as the target population. The proportion for the cells of the first margin (age x gender) are processed and the sample weighted, then another margin (race/ethnicity) is processed and the sample weighted again. When the sample has been weighted for all the margins once, the process repeats, reweighting the sample. This continues until no cell in any margin differs from the target proportions by more than 0.025% (i.e., convergence) or until the margins have been processed 75 times.

For states that use regional weighting, as Pennsylvania does, there are 12 raking control margins as shown below.

1. Age group by gender
2. Detailed race/ethnicity
3. Education
4. Marital status
5. Tenure (rent or own home)
6. Gender by race/ethnicity
7. Age group by race/ethnicity
8. Phone usage groups
9. Region
10. Region by age group
11. Region by gender
12. Region by race/ethnicity

The sample design in Pennsylvania includes the six districts and Allegheny and Philadelphia counties, making eight regions. For those states that do not use regional weighting, only the first eight control variables are used in the raking. The population control totals for age group, gender, race/ethnicity, and region and the control percentages for education, marital status and tenure are obtained from the Nielsen Company, Inc. and the American Community Survey. The estimate that 16.5 percent of the 2011 Pennsylvania adult population are cell phone only users is derived from the National Health Interview Survey (NHIS) [See Blumberg, et. Al.].

All of the percentages reported here were calculated with weighted data and should be representative of the adult population of Pennsylvania. It should be noted that the percentages might not add to 100, due to rounding. When calculating the percentages of prevalence for each health topic in this report, responses of “Don’t know/Not sure” and “Refused” were removed from the denominators.

Comparison of estimates

This report presents estimates for a single year, 2011. The changes which occurred in the BRFSS survey between 2010 and 2011, specifically the addition of the cell phone sample and the employment of a different weighting technique, make direct comparisons of estimates of health conditions between 2011 and previous years invalid. This situation is reflected in the line graphs presented in this report for each topic. No line connects the 2011 estimate with trend lines from earlier years.

Other Department of Health reports are based on three years of data. The three-year report in the Department's interactive web tool Epidemiological Querying and Mapping System (EpiQMS) and certain burden reports are of this nature. These reports will continue to be updated with data from the landline sample only, weighted with the same post-stratification method used in earlier years. When sample data for 2011, 2012 and 2013 are available, the three year reports will be prepared with combination landline and cell phone data, weighted by raking.

Reports based on BRFSS data can come from a wide variety of sources. The record level data from every state can be downloaded from the BRFSS web site (<http://www.cdc.gov/brfss/>), with certain fields suppressed to protect the confidentiality of respondents. Other sources may include or exclude data from the denominators used to calculate percentages. For example, the exclusion or inclusion of "Don't Know/Not Sure" or "Refused" responses could potentially affect the final response percentage calculation.

If you have any questions about these differences, please contact the Bureau of Health Statistics and Research by phone at 717-783-2548 or by mail at 555 Walnut St., 6th Floor, Harrisburg, PA 17101-1914.

Synthetic Estimation Process for Local Data

The BRFSS is an ongoing telephone survey consisting of interviews conducted each month. In 2011, the sample dataset includes 11,509 surveys divided into eight different Pennsylvania health regions (Allegheny and Philadelphia are separate).

On the state level, data from the BRFSS serve several purposes. BRFSS data help to identify subgroups, which should be targeted for health promotion and disease prevention programs due to elevated risks. Multiple years of BRFSS data are useful for tracking Pennsylvania's progress in achieving selected Healthy People 2020 National Health Objectives. Data from Pennsylvania, when compared to similar data from other states, identifies the need for increased health promotion and disease prevention program efforts. In 2011, comparable data were available from all 50 states, the District of Columbia, Puerto Rico, the U.S. Virgin Islands and Guam.

On the local level, BRFSS data may also be used to estimate the prevalence of risks in local areas such as counties, if the data are combined for several years or the counties or county groups of interest are over-sampled. However, for most counties, the number of respondents in the BRFSS sample data set is insufficient to produce reliable estimates.

In cases where local data on behavioral risk are not available, synthetic estimates can be computed based on either national data or statewide data from the BRFSS. Synthetic estimates are calculated using population estimates for subgroups of interest and the state or national risk factor prevalence rates for those groups. Below is an example of how one can compute synthetic estimates for a local area:

Step 1

Obtain the population estimates for the local geographic area of interest. Sum the population estimates into a table with the same breakdown as a table listing the national or state estimates (see the table below).

Step 2

To estimate the number of persons who have the behavioral risk in each subgroup, multiply the subgroup-specific rates by the population estimates for each group. For example, multiply the 2010 (latest available) Dauphin County census population of 40,985 for ages 18-29 by the 2011 fair or poor health prevalence of 9 percent (0.09) for that age group at the state level. The 2011 synthetic estimate for those in fair or poor health ages 18-29 in Dauphin County is 3,689.

Step 3

To obtain the total number of persons who indicated fair or poor health, repeat Step 2 for all subgroups and then sum the subgroup estimates to get a total estimate.

Age Group	2010 Dauphin County Census Population		Fair or Poor Health from 2011 Pa. BRFSS		Estimate of Dauphin County Adults Indicating Fair or Poor Health, 2011
18-29	40,985	x	9 %	=	3,689
30-44	51,044	x	13 %	=	6,636
45-64	77,015	x	18 %	=	13,863
65+	36,841	x	27 %	=	9,947
					Total
					34,134

Caution: Synthetic estimates can be useful for planning purposes. However, these estimates should not be used if there is reason to believe that local rates for subgroups of interest would diverge widely from the state or national rates. The prevalence of most health-related conditions varies considerably with age and often with other factors, such as sex, race and income. A more precise estimate may be obtained using age, sex and race-specific prevalence rates. The BRFSS is not a reliable source of prevalence rates specific to age-sex-race categories; national data would be a more reliable basis for synthetic estimates.

It is important to qualify estimates whenever they are used. A clear citation of the sources of the data used to compute the local area synthetic estimates should be included in every report of the estimates.

Step 4

To calculate the synthetic estimated percentage of Dauphin County adults with fair or poor health, pull the “Total Estimated Number of Adults” and the “Total Population Age 18+” in Dauphin County from “Step 3.”

Total Synthetically Estimated Number of Adults With Fair or Poor Health in Dauphin County = **34,134**

Total Population Age 18+ in Dauphin County = **205,885**

Divide the synthetically estimated number of adults with fair or poor health by the adult population. Then multiply by 100 so that the result will be expressed as a percent.

$$\begin{array}{l} \text{Synthetically Estimated Percentage} \\ \text{With Fair or Poor Health in Dauphin County} \end{array} = \frac{\text{Total Synthetically Estimated Number of Adults With Fair or Poor Health in Dauphin County}}{\text{Total Population Age 18+ in Dauphin County}} \times 100$$

$$\begin{array}{l} \text{Synthetically Estimated Percentage} \\ \text{With Fair or Poor Health in Dauphin County} \end{array} = (34,134 / 205,885) \times 100$$

$$\begin{array}{l} \text{Synthetically Estimated Percentage} \\ \text{With Fair or Poor Health in Dauphin County} \end{array} = \mathbf{16.6 \text{ Percent}}$$

This step gives you a synthetically estimated percentage of adults.

Caution: Synthetic estimates can be useful for planning purposes. However, these estimates should not be used if there is reason to believe that local rates for subgroups of interest would diverge widely from the state or national rates. The prevalence of most health-related conditions varies considerably with age and often with other factors, such as sex, race and income. A more precise estimate may be obtained using age, sex and race-specific prevalence rates. The BRFSS is not a reliable source of prevalence rates specific to age-sex-race categories; national data would be a more reliable basis for synthetic estimates.

It is important to qualify estimates whenever they are used. A clear citation of the sources of the data used to compute the local-area synthetic estimates should be included in every report of the estimates.

References

Blumberg, Stephen J., et. al. “Wireless Substitution: State-level Estimates From the National Health Interview Survey, January 2007-June 2010”, National Health Statistics Reports, Number 39, April 20, 2011.

Bureau of Health Statistics and Research; Behavioral Health Risks of Pennsylvania Adults 2010 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

Bureau of Health Statistics and Research; Behavioral Health Risks of Pennsylvania Adults 2009 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

Bureau of Health Statistics and Research; Behavioral Health Risks of Pennsylvania Adults 2008 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

Bureau of Health Statistics and Research; Behavioral Health Risks of Pennsylvania Adults 2007 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

Bureau of Health Statistics and Research; Behavioral Health Risks of Pennsylvania Adults 2006 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

Bureau of Health Statistics and Research; Behavioral Health Risks of Pennsylvania Adults 2005 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

Bureau of Health Statistics and Research; Behavioral Health Risks of Pennsylvania Adults 2004 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

Bureau of Health Statistics and Research; Behavioral Health Risks of Pennsylvania Adults 2003 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

Bureau of Health Statistics and Research; Behavioral Health Risks of Pennsylvania Adults 2002 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

Bureau of Health Statistics and Research; Behavioral Health Risks of Pennsylvania Adults 2001 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

Bureau of Health Statistics; Behavioral Health Risks of Pennsylvania Adults 2000 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

Bureau of Health Statistics; Behavioral Health Risks of Pennsylvania Adults 1999 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

Division of Health Statistics; Behavioral Health Risks of Pennsylvania Adults 1998 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P.

National Center for Health Statistics, Office of Disease Prevention and Health Promotion, HealthyPeople.gov (<http://www.healthypeople.gov/2020/default.aspx>).