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 2014, 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 2014 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 that 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.

The new "raking" weighting method was the primary means of combining the samples of landline telephones and cell phones so that they would yield valid estimates of the entire adult household population. However, changes have been made each year after 2011to deal with the possibility that the same household may appear in both samples. The 2011 survey excluded households from the cell phone sample if they had a landline telephone. The 2012 survey excluded households from the cell sample if fewer than 90 percent of their calls were received on a cell phone. This (2013) survey did the same, but reweighted interviews from the landline sample if the household reported cell phone usage of 90 percent or more, and reweighted interviews from the cell phone sample that also had a landline telephone in use (called "cell-mostly"). The 2014 survey includes all households reached from either cell or landline samples and resolves the effects of possible duplication with weighting.

We do not know what the effects on estimates of these adjustments may be. They are thought to be minor in comparison with the changes accompanying the introduction of the sample of cell phones in 2011.

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 that include one or more unlisted telephone numbers and 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

Reliance on cell phones has increased steadily in Pennsylvania over time. Accordingly, the proportion of cell phones in the Pennsylvania sample has increased.

2011	18%
2012	19%
2013	25%
2014	35% (target)

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 2014 BRFSS survey. All items new to the 2014 survey were field-tested.

Questions of interest to Pennsylvania were added as the state supplement to the core questionnaire. State-added modules and questions in 2014 concerned adult human papilloma virus (HPV), sexual orientation and gender identity, childhood asthma prevalence, health care access, tobacco use, cognitive impairment, sodium or salt-related behavior and adverse childhood experiences.

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 that 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 2014 survey are 37.6 and 34.4, respectively.

Disposition of All Telephone Numbers – Landline Survey
2014 Pennsylvania Behavioral Risk Factor Surveillance System Survey Sample

Disposition Code	Number	Percent
Interview		
Complete	6,976	3.92
Partial complete	691	0.39
Eligible, non-interview		
Household level refusal	1,418	0.80
Refusal after respondent selection	1,484	0.83
Termination within questionnaire	1,001	0.56
Selected respondent away from residence during the entire interviewing period	1,515	0.85
Household answering device	2,548	1.43
Selected respondent physically or mentally unable to complete an interview during the entire interviewing period	471	0.26
Language problem after respondent selection	67	0.04
Unknown eligibility, non-interview		
Unknown if housing unit	17,515	9.83
No answer	5,104	2.86
Telephone answering device, not sure if private residence	8,016	4.50
Telecommunication barrier, not sure if private residence	931	0.52
Contact, unknown if private residence	4,767	2.68
Physical or mental impairment before respondent selection	365	0.20
Language problem before respondent selection	346	0.19
Not eligible		
Out of sample	101	0.06
Dedicated fax/data/modem line with no human contact, no eligible respondent	3,605	2.02
Non-working/disconnected number	11,756	6.60
Special technological circumstances	7,376	4.14
Call forwarding/pager	19	0.01
Cellular phone	288	0.16
Not a private residence	3,638	2.04
Group home	74	0.04
Household, no eligible respondent	22	0.01
Precalls	98,076	55.05
Total	178,170	

Disposition of All Telephone Numbers – Cell Phone Survey
2014 Pennsylvania Behavioral Risk Factor Surveillance System Survey Sample

Disposition Code	Number	Percent
Interview		
Complete	3,016	6.64
Partial complete	257	0.57
Eligible, non-interview		
Refusal after respondent selection	435	0.96
Termination within questionnaire	367	0.81
Selected respondent away from residence during the entire interviewing period	187	0.41
Household answering device	3	0.01
Selected respondent physically or mentally unable to complete an interview during the entire interviewing period	12	0.03
Language problem after respondent selection	14	0.03
Unknown eligibility, non-interview		
Unknown if housing unit	14,825	32.62
No answer	1,002	2.20
Telephone answering device, not sure if private residence	8,147	17.93
Telecommunication barrier, not sure if private residence	594	1.31
Contact, unknown if private residence	13	0.03
Physical or mental impairment before respondent selection	77	0.17
Language problem before respondent selection	266	0.59
Not eligible		
Out of sample	48	0.11
Dedicated fax/data/modem line with no human contact, no eligible respondent	23	0.05
Nonworking number/disconnected	7,893	17.37
Special technological circumstances	6,551	14.41
Call forwarding/pager	3	0.01
Landline phone	274	0.60
Not a private residence	884	1.94
Group home	10	0.02
Household, no eligible respondent	549	1.21
Total	45,450	

Sample Characteristics

The following table compares selected characteristics of the final interview sample for the 2014 BRFSS to the 2013 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, Hispanic origin, marital status, home ownership, type of telephone and education characteristics of the sample.

Distribution of 2014 Pennsylvania BRFSS Survey Sample and 2013 Pennsylvania Adult Population Estimates for Selected Characteristics

		2014 BRFSS Survey Sample		2013 Population Estimates	
		Number	Percent	Number	Percent
All adults		11,003	100.00	10,057,548	100.00
Sex	Male	4,510	40.99	4,853,526	48.26
	Female	6,493	59.01	5,204,022	51.74
Race	White	9,600	87.25	8,398,335	83.50
	Black	882	8.02	1,038,345	10.32
	Other	337	3.06	620,868	6.18
	Unknown/refused	184	1.67	N/A	N/A
Hispanic Origin	Yes	224	2.04	524,179	5.21
	No	10,689	97.15	9,533,369	94.79
	Unknown/refused	90	0.82	N/A	N/A
Age	18-24	501	4.55	1,248,038	12.41
	25-34	892	8.11	1,600,329	15.91
	35-44	1,192	10.83	1,536,491	15.28
	45-54	1,791	16.28	1,840,649	18.30
	55-64	2,469	22.44	1,742,069	17.32
	65-74	2,198	19.98	1,109,434	11.03
	75+	1,787	16.24	980,538	9.75
	Unknown/refused	173	1.57	N/A	N/A

Note[1]: Race data include Hispanics.

Note[2]: 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 2013 population estimate "Unknown/refused" entries.

Determining Accuracy of the Estimates and Significance Using Confidence Intervals

Tables included in this report show the 95% 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 that are more and less easily reached. After design factors were taken care of, an additional factor was applied that 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 percent (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

Beginning with the 2013 survey, CDC added four additional geographic margins. The geographies (Allegheny, Philadelphia, and rest of state) are similar to stratification that Pennsylvania already does and may be more important in other states.

13. County
14. County by race/ethnicity
15. County by age group
16. County by gender

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 44.1 percent of the 2014 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.

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 website (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 Informatics and Information Technology by phone at 717-782-2448 or by mail at 2150 Herr St., Harrisburg, PA 17103-1625.

Synthetic Estimation Process for Local Data

The BRFSS is an ongoing telephone survey consisting of interviews conducted each month. In 2014, the sample dataset includes 11,003 surveys divided into eight different Pennsylvania health regions: Northwest, Southwest, Northcentral, Southcentral, Northeast, Southeast, Allegheny and Philadelphia.

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 2014, 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 calculated 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 calculate 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 2013 (latest available) Dauphin County census population of 41,480 for ages 18-29 by the 2014 fair or poor health prevalence of 10 percent (0.10) for that age group at the state level. The 2014 synthetic estimate for those in fair or poor health ages 18-29 in Dauphin County is 4,148.

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	2013 Dauphin County Census Population		Fair or Poor Health from 2014 Pa. BRFSS		Estimate of Dauphin County Adults Indicating Fair or Poor Health, 2014
18-29	41,480	x	10 %	=	4,148
30-44	50,729	x	11 %	=	5,580
45-64	78,399	x	21 %	=	16,464
65+	39,267	x	24 %	=	9,424
				Total	35,617

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	=	**35,617**
Total Population Age 18+ in Dauphin County	=	**209,875**

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 percentage.

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

Synthetically Estimated Percentage With Fair or Poor Health in Dauphin County = (35,617/ 209,875) X 100

Synthetically Estimated Percentage With Fair or Poor Health in Dauphin County = **17.0 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 calculate the local-area synthetic estimates should be included in every report of the estimates.

References

Blumberg, Stephen J., et. al.. "Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, July-December 2014", National Health Statistics Reports, Number 70, December 18, 2013.

Bureau of Health Statistics and Research; Behavioral Health Risks of Pennsylvania Adults 2014 Harrisburg, Pa., Pennsylvania Department of Health; #303.300P, and preceding issues.

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

Fruits and Vegetables. Digital image. *Why Follow a Green Diet*. Shannon Crawford, 25 Feb. 2015. Web. 20 Oct. 2015. <http://www.weight-n-watch.com/diet-nutrition/why-follow-a-green-diet.html>.

Group Workout. Digital image. *Coaches Corner: Should I Do More Exercise?* Ryan Mitchell, n.d. Web. 20 Oct. 2015. <http://evolutionfitnesscentre.com.au/coaches-corner-should-i-do-more-exercise.html>.

Outdoor Fitness Class. Digital image. *The Adventure Collection Blog*. Candice Gaukel Andrews, 15 June 2015. Web. 20 Oct. 2015. <http://www.adventurecollection.com/natural-areas-exercise-gyms-and-adventure-sports-arenas/>.

Map of Pennsylvania. Digital image. *First-time Freshman Undergraduate Admissions Counselors for Pennsylvania (By County, On-campus Programs Only)*. N.p., n.d. Web. 20 Oct. 2015. <http://www.calu.edu/prospective/undergraduate-admissions/contact-an-admissions-counselor/meet-our-counselors/pennsylvania/index.htm>.

Pennsylvania State Flag. Digital image. *Pennsylvania State Flag*. N.p., n.d. Web. 20 Oct. 2015. <http://www.50states.com/flag/paflag.htm#.VikirZiFN9A>.