

---

# BRFSS Weighting

---

Michelle L. Cook, MPH

Rebecca A. Wood, MSPH

October 14, 2009



---

# Why we weight

- To compensate for unequal probabilities of selection
  - To adjust for non-response and telephone non-coverage
  - To ensure that results are consistent with population data
  - To make populations estimates
-

---

# Current BRFSS Weighting Formula

$$\text{FINALWT} = \text{STRWT} * 1/\text{NPH} * \text{NAD} * \text{POSTSTRAT}$$

---

---

# STRAT

- Accounts for differences in the basic probability of selection among strata (subset of area code and prefix combinations). It is the inverse of the sampling fraction of each stratum.
  - For each geographic stratum (and usually by file month)
  - `denstr=1` (1)
  - `denstr=2` (1.5)

$(\text{nrecstr2}/\text{nrecsel2}) / (\text{nrecstr1}/\text{nrecsel1})$

`FINALWT = STRWT * 1/NPH * NAD * POSTSTRAT`

---

---

## 1/NPH

- The inverse of the number of residential telephone numbers in the respondent's household.
- The more phones in the household, the greater the probability of the selection of a respondent.

$$\text{FINALWT} = \text{STRWT} * 1/\text{NPH} * \text{NAD} * \text{POSTSTRAT}$$

---

---

# NAD

- NAD is the number of adults in the respondent's household.
- Because only one person per household is interviewed, respondents in larger households have a smaller probability of selection than respondents in smaller households.

$$\text{FINALWT} = \text{STRWT} * 1/\text{NPH} * \text{NAD} * \text{POSTSTRAT}$$

---

---

# POSTSTRAT

- Adjusts for non-coverage and non-response and forces the sum of the weighted frequencies to equal population estimates for the region or state.
- Texas only post-stratifies by gender and age – not race/ethnicity.

$$\text{FINALWT} = \text{STRWT} * 1/\text{NPH} * \text{NAD} * \text{POSTSTRAT}$$

---

---

# Look for missing ages

- GET  
FILE='H:\Brfss\_data\brfss08\state\state\_08\_working\_test.sav'.

- MEANS

- age BY sex

- /CELLS = MEAN MEDIAN COUNT.

- 

- What is your age?

- Sex of respondent   Mean   Median N

- Male               54.73   55.00   3725

- Female             54.99   55.00   6864

- Total              54.90   55.00   10589

---

# Recode missing ages and make age groups

- IF (missing(age) AND sex = 1) impage = 55.
  - IF (missing(age) AND sex = 2) impage = 55.
  
  - RECODE
    - impage
    - (18 thru 24 = 1) (25 thru 34 = 2) (35 thru 44 = 3) (45 thru 54 = 4) (55 thru 64 = 5)
    - (65 thru 120 = 6) INTO ageg\_6.
  
  - VARIABLE label ageg\_6 'Age groups for weighting'.
  
  - VALUE label ageg\_6
    - 1 '18-24'
    - 2 '25-34'
    - 3 '35-44'
    - 4 '45-54'
    - 5 '55-64'
    - 6 '65+'.
-



---

# Compute number of phones in household

- **12.17** Do you have more than one telephone number in your household? Do not include cell phones or numbers that are only used by a computer or fax machine.

|   |                       |                       |
|---|-----------------------|-----------------------|
| 1 | Yes                   |                       |
| 2 | No                    | <b>[Go to Q12.19]</b> |
| 7 | Don't know / Not sure | <b>[Go to Q12.19]</b> |
| 9 | Refused               | <b>[Go to Q12.19]</b> |

- **12.18** How many of these telephone numbers are residential numbers?

|   |                               |                        |
|---|-------------------------------|------------------------|
| — | Residential telephone numbers | <b>[6 = 6 or more]</b> |
| 7 | Don't know / Not sure         |                        |
| 9 | Refused                       |                        |

---

---

## Computing number of phones continued

- COMPUTE phones = c12q18.  
IF (c12q17 = 2) phones = 1.  
IF (missing(phones) and c12q17 = 1) phones = 2.  
IF (missing(phones)) phones = 1.
  - The first question asks the respondent if they have more than one phone – if you answer no then the number of phones is 1. If the respondent reports more than one phone but doesn't specify how many then the number of phones is computed as 2. If the questions are left blank then the numbers of phones is assumed to be 1.
-

---

# Number of adults

- I need to randomly select one adult who lives in your household to be interviewed. How many members of your household, including yourself, are 18 years of age or older?
  - 
  - \_\_\_\_\_ Number of adults
  - **If "1,"**
  - Are you the adult?
  - **If "yes,"**
  - Then you are the person I need to speak with. Enter 1 man or 1 woman below (Ask gender if necessary). **Go to page 5.**
  - **If "no,"**
  - Is the adult a man or a woman? Enter 1 man or 1 woman below. May I speak with **[fill in (him/her) from previous question]**? **Go to "correct respondent" on the next page.**
-

---

## Compute weight for probability of selection

- compute  $wt\_prob = numadult/phones.$
  - do if  $denstr = 1.$
  - COMPUTE  $wt1 = wt\_prob.$
  - else if  $denstr = 2.$
  - COMPUTE  $wt1 = wt\_prob * 1.5.$
  - end if.
  - execute.
-

---

## Continued.....

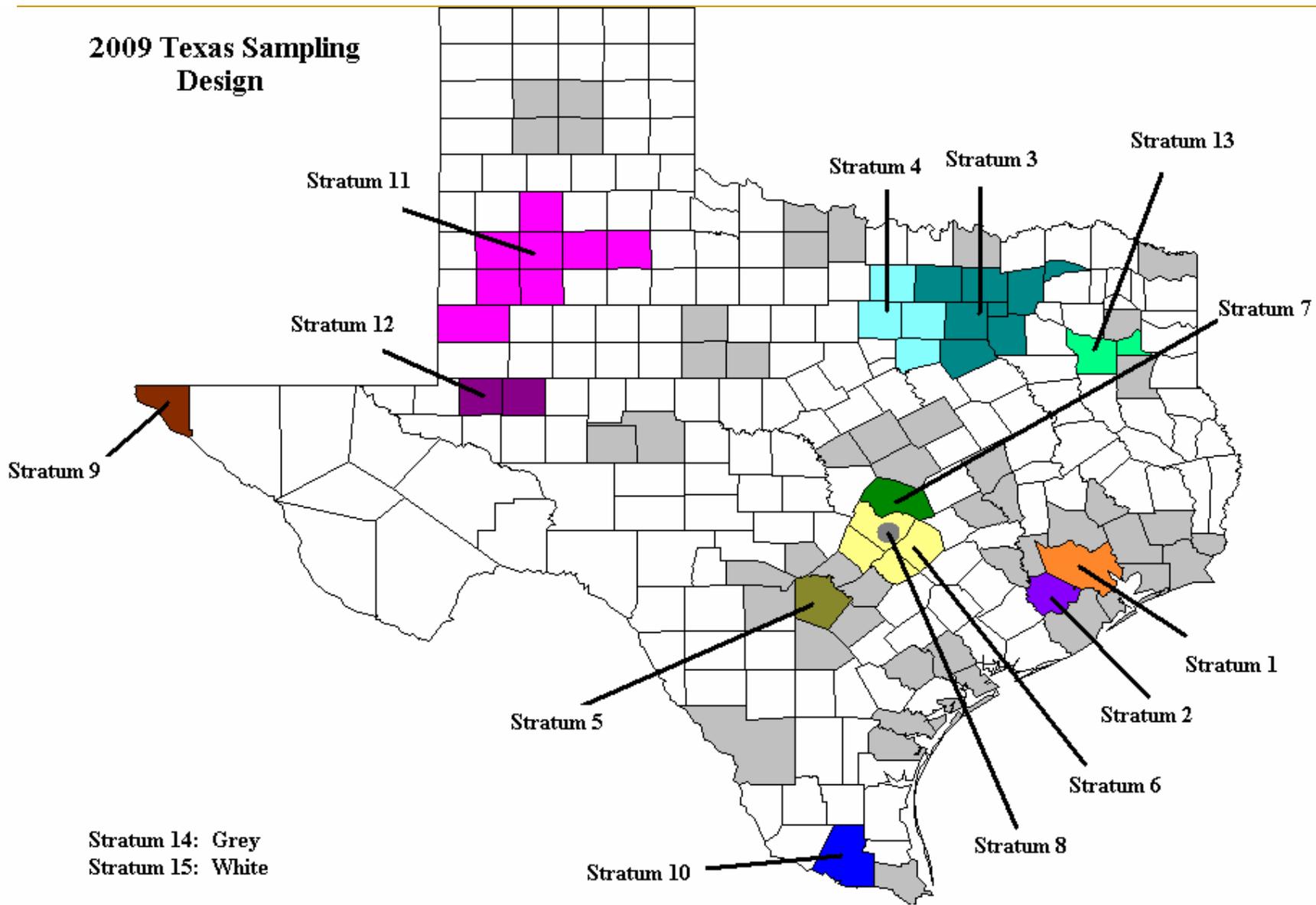
- Sum the weights for the probability of selection and add those values back to the file
  - AGGREGATE
    - /OUTFILE = 'H:\Brfss\_data\brfss08\state\pittagg.sav'
    - /BREAK = ageg\_6 sex
    - /wt1sum = SUM(wt1)
    - /N\_BREAK = N.
  - MATCH FILES /FILE=\*
    - /TABLE='H:\Brfss\_data\brfss08\state\pittagg.sav'
    - /BY ageg\_6 sex.
-

---

# Get populations for post-stratification and calculate final weight

- IF (ageg\_6 = 1 AND c12q20= 1) finalwt = wt1 \* (1326297/wt1sum).
  - IF (ageg\_6 = 1 AND c12q20= 2) finalwt = wt1 \* (1240149/wt1sum).
  
  - IF (ageg\_6 = 2 AND c12q20= 1) finalwt = wt1 \* (1949320/wt1sum).
  - IF (ageg\_6 = 2 AND c12q20= 2) finalwt = wt1 \* (1759646/wt1sum).
  
  - IF (ageg\_6 = 3 AND c12q20= 1) finalwt = wt1 \* (1828577/wt1sum).
  - IF (ageg\_6 = 3 AND c12q20= 2) finalwt = wt1 \* (1780713/wt1sum).
  
  - IF (ageg\_6 = 4 AND c12q20= 1) finalwt = wt1 \* (1663315/wt1sum).
  - IF (ageg\_6 = 4 AND c12q20= 2) finalwt = wt1 \* (1678196/wt1sum).
  
  - IF (ageg\_6 = 5 AND c12q20= 1) finalwt = wt1 \* (1134749/wt1sum).
  - IF (ageg\_6 = 5 AND c12q20= 2) finalwt = wt1 \* (1200580/wt1sum).
  
  - IF (ageg\_6 = 6 AND c12q20= 1) finalwt = wt1 \* (1030440/wt1sum).
  - IF (ageg\_6 = 6 AND c12q20= 2) finalwt = wt1 \* (1351082/wt1sum).
-

# 2009 Texas Sampling Design



---

# Differences in CDC Weights and DSHS Weights

- Different population estimates
    - CDC – Claritas
    - DSHS – State Demographer
  - Different weighting procedure
    - CDC – statewide only (through 2008)
    - DSHS – based on geography
-

---

# Raking

- Raking (also known as iterative proportional fitting or sample balancing) will replace poststratification
  - Raking is preferred when cell counts of the respondents are too small
  - Makes missing key demographic groups like another record
-

# Post-Stratification vs. Raking (2008)

|                         | <b>Post-Stratification</b> | <b>Raking</b> |
|-------------------------|----------------------------|---------------|
| Current Smokers         | 18.5                       | 20.5          |
| No Health Care Coverage | 25.6                       | 27.9          |
| Check-up in Past Year   | 64.9                       | 63.6          |
| No LTPA                 | 28.5                       | 30.2          |

---

Questions????

---

---

# BRFSS Data Processing

---



---

# Process Monthly Files

- Get monthly file from contractor
  - Run disposition codes for Texas and all geographic strata
  - Rename variables to match questionnaire (ex. C01Q01)
  - Add variable labels, add value labels and set missing values
  - Save monthly file in SPSS format with variables in same order as asked in the questionnaire
-

---

## Combine Monthly Files and Compute Calculated Variables

- Combine monthly files to make year to date or yearly file
  - Select completed or partially completed interviews (disposition codes 110 or 120)
  - Create calculated variables
  - Save year to date or annual file with variable in order of questionnaire and with calculated variables placed after the original questions
-

---

# Calculate Preliminary DSHS Weight

- Calculate a preliminary weight on year to date or final annual file
-

---

# Add CDC Final Weight

- Sort year to date or annual DSHS file and add variable cdcfinwt from quarterly or annual CDC file
  - Compare frequencies of original variables and calculated variables between DSHS and CDC files
-

---

# Make Public Use Data File

- Blank out county information on individual data record for counties that have less than 50 cases in the annual data file
  - Strip off variables that may link that record to an individual
  - Have users of data sign confidentiality agreement
-

---

# Calculated Variables

---



---

# Simple Recode

- Recode a variable down to a few groups
  - Example – age in years to age group
  
  - RECODE c12q01
  - (18 thru 24 = 1) (25 thru 34 = 2) (35 thru 44 = 3) (45 thru 54 = 4) (55 thru 64 = 5) (65 thru 99 = 6) INTO agegr6.
  - VARIABLE LABEL agegr6 'Age Group'.
  - VALUE LABEL agegr6
  - 1 '18 to 24'
  - 2 '25 to 34'
  - 3 '35 to 44'
  - 4 '45 to 54'
  - 5 '55 to 64'
  - 6 '65+'.
  - FORMATS agegr6 (F1.0).
-

---

# Summary Variable from Two Questions

- Use data from two or more survey questions to make a new variable
- Example – compute Race/Ethnicity from Hispanic question and Race questions

## 12.2 Are you Hispanic or Latino?

|   |                       |
|---|-----------------------|
| 1 | Yes                   |
| 2 | No                    |
| 7 | Don't know / Not sure |
| 9 | Refused               |

---

- **12.3** Which one or more of the following would you say is your race?

**(Check all that apply)**

**Please read:**

- 1 White
- 2 Black or African American
- 3 Asian
- 4 Native Hawaiian or Other Pacific Islander
- 5 American Indian or Alaska Native

**Or**

- 6 Other [specify]\_\_\_\_\_

**Do not read:**

- 8 No additional choices
- 7 Don't know / Not sure
- 9 Refused

- **12.4** Which one of these groups would you say best represents your race?

- 1 White
- 2 Black or African American
- 3 Asian
- 4 Native Hawaiian or Other Pacific Islander
- 5 American Indian or Alaska Native
- 6 Other (specify)\_\_\_\_\_

**Do not read:**

- 7 Don't know / Not sure
- 9 Refused

---

## Summary Variable from Questions with Skip Patterns

- Sometimes need to compute a variable from a set of questions when the answer to the first question determines if the respondent is asked the second question
  - Example – Mammogram Questions
-

---

**18.1** A mammogram is an x-ray of each breast to look for breast cancer. you ever had a mammogram?

- 1 Yes
- 2 No [Go to Q18.3]
- 7 Don't know / Not sure [Go to Q18.3]
- 9 Refused [Go to Q18.3]

**18.2** How long has it been since you had your last mammogram?  
**Read only if necessary:**

- 1 Within the past year (anytime less than 12 months ago)
- 2 Within the past 2 years (1 year but less than 2 years ago)
- 3 Within the past 3 years (2 years but less than 3 years ago)
- 4 Within the past 5 years (3 years but less than 5 years ago)
- 5 5 or more years ago

**Do not read:**

- 7 Don't know / Not sure
  - 9 Refused
-

---

Questions????

---