

Using the BRFSS COPD Data

Data Analysis Considerations

BRFSS uses a complex sampling design. Therefore, to analyze BRFSS data correctly, analysts must use statistical software packages that account for this sampling design. Some appropriate statistical software packages are SUDAAN, SAS, STATA, and SPSS. Learn more about the [Youth Risk Behavior Survey](#), a CDC survey that uses these analysis software packages.

Analyzing the Core COPD Question

Learn more about the [BRFSS 2011 survey data and documentation](#).

Before getting started, make sure to review the documents under “Survey Data Information.” The data files are available in either ASCII or SAS Transport formats under “Data Files.” The files are very large and must be unzipped after downloading. A SAS program (transprt.sas) to convert the XPT (SAS Transport) file into a SAS dataset is available under “SAS Resources.”

For analysis of the core COPD question, use the dataset LLCP2011, which contains the combined landline and cell phone data. The sample design variables are—

<code>_STSTR</code>	Sample Design Stratification Variable
<code>_PSU</code>	Primary Sampling Unit
<code>_LLCPWT</code>	Final Weight: Landline and Cell-Phone Data

Complex sampling design specification in statistical software packages

SUDAAN

The sample design statements for SUDAAN are—

```
PROC ..... design = WR;  
Nest _STSTR _PSU / missunit;  
Weight _LLCPWT;
```

SAS

For appropriate variance estimation, survey procedures must be used (e.g., PROC SURVEYMEANS, PROC SURVEYFREQ).

The appropriate sample design statements for these PROCs are

```
STRATA _STSTR;  
CLUSTER _PSU;  
WEIGHT _LLCPWT;
```

STATA

Survey design can be specified in a SVYSET statement.

```
svyset [pweight=_LLCPWT], strata(_STSTR) psu(_PSU)
```

SPSS

The SPSS Complex Samples add-on module includes sample selection and analysis of complex sample survey data. The sample design for a data set can be specified using the Analysis Preparation Wizard with the following syntax:

```
CSPLAN ANALYSIS  
/PLAN FILE='plan location'  
/PLANVARS ANALYSISWEIGHT=_LLCPWT  
/PRINT PLAN  
/DESIGN STRATA= _STSTR CLUSTER= _PSU  
/ESTIMATOR TYPE=WR.
```

Other important variables included in LLCP2011

_STATE – State FIPS Code

CHCCOPD – (Ever told) you have (COPD) chronic obstructive pulmonary disease, emphysema or chronic bronchitis? This is the core COPD question.

1=Yes

2=No

7=Don't know/Not sure

9=Refused

BLANK=missing

Additional variables are used in the examples below

COPD

If CHCCOPD=1 then COPD=1;

*Respondent reports having been told they have COPD. The numerator for prevalence.;

If CHCCOPD in (2,7) then COPD=2;

*Respondent reports not having been told they have COPD or doesn't know. In addition to affirmative responses, these respondents are included in the denominator for prevalence. A refusal (CHCCOPD=9) or a missing response (CHCCOPD=BLANK) is considered missing and not included in the denominator.;

AGECAT

If 18<=AGE<45 then AGECAT=1;

If 45<=AGE<55 then AGECAT=2;

If 55<=AGE<65 then AGECAT=3;

If 65<=AGE<75 then AGECAT=4;

If 75<=AGE then AGECAT=5;

*These age groups were chosen due to the low prevalence of COPD among the younger population. These age groups were used for age-standardization.;

COPDPREV

If COPD ne . and AGECAT>=1 then COPDPREV=1;

*This variable was used to define the sample population for analyses.;

Examples of analyses of core data using SAS-callable SUDAAN

Reminder: In SUDAAN, data must be sorted prior to performing analyses.

```
PROC SORT data=BRFSS11.COPDCORE;  
By _STSTR_PSU;  
RUN;
```

Example 1. Estimated number of adults with COPD by state

```
PROC CROSSTAB data=BRFSS11.COPDCORE design=WR; *BRFSS11=library. COPDCORE=dataset  
including our calculated variables.;  
Nest _STSTR_PSU / missunit;  
Weight _LLCPWT;  
Subpopn COPDPREV =1;  
Class _STATE COPD/nofreq;  
Tables _STATE*COPD;  
Print nsum="Sample size" /style=nchs;  
Print wsum="Estimated popn." / style=nchs wsumfmt=f12.0;  
Title 'Table. Estimated no of adults with COPD by state';  
RUN;
```

Example 2. Prevalence of COPD by sex for a single state (Alabama)

```
PROC DESCRIPT data=BRFSS11.COPDCORE design=WR;  
Nest _STSTR_PSU / missunit;  
Weight _LLCPWT;  
Subpopn COPDPREV=1 and _STATE=1; *The FIPS code for Alabama is 1;  
Var COPD; *The categorical variable we are interested in;  
Catlevel 1; *The value for which we want to calculate percentages. Respondents who report that they HAVE  
been told they have COPD;  
Class SEX/nofreq; *We are interested in COPD prevalence by sex;  
Tables SEX;  
PRINT percent="Pct" lowpct="Pct-low" uppct="Pct-up" sepercent="SE"  
/style=nchs percentfmt=f8.3 lowpctfmt=f8.3 uppctfmt=f8.3 sepercentfmt=f8.3;  
*Percent will provide the percentage of respondents who report COPD by sex. Pct-low and Pct-up will provide  
the lower and upper bounds of the 95% confidence interval. Sepercent will provide the standard error of the  
estimate to be used to calculate the relative standard error;  
Title 'Table. Prevalence of COPD by sex for Alabama';  
RUN;
```

Example 3. Age-adjusted prevalence of COPD by state

```
PROC DESCRIPT data=BRFSS11.COPDCORE design=WR;  
Nest _STSTR_PSU / missunit;  
Weight _LLCPWT;  
Subpopn COPDPREV=1;  
Var COPD; *The categorical variable we are interested in;
```

```

Catlevel 1; *The value for which we want to calculate percentages. Respondents who report that they HAVE
been told they have COPD;
Stdvar AGECAT; *We want the prevalence to be age-adjusted;
Stdwgt 0.530535 0.181652 0.117542 0.088967 0.081304; *Adjustment weights for the age groups were
calculated based on Table 1 data from the report available at http://www.cdc.gov/nchs/data/statnt/statnt20.pdf.
e.g., 53.05% of the population aged 18 years and older is 18–44 years of age;
Class _STATE AGECAT /nofreq; *We are interested in COPD prevalence by state. The standardization
variable AGECAT must also be included in the Class statement;
Tables _STATE;
PRINT percent="Pct" lowpct="Pct-low" uppct="Pct-up" sepercent="SE"
/style=nchs percentfmt=f8.3 lowpctfmt=f8.3 uppctfmt=f8.3 sepercentfmt=f8.3;
*Percent will provide the percentage of respondents who report COPD by state. Pct-low and Pct-up will provide
the lower and upper bounds of the 95% confidence interval. Sepercent will provide the standard error of the
estimate to be used to calculate the relative standard error;
Title 'Table. Prevalence of COPD by State-Age-Adjusted to the 2000 Projected U.S. Population';
RUN;

```

Analyzing the Optional COPD Module Data

Analysis of the optional COPD module data is more complicated. Some states asked the module of all respondents who reported COPD, whether they were landline or cell- phone users. Other states limited the use of the module to landline respondents only. Still other states used a split sample, with up to three possible questionnaire versions. Each of these situations affected the sample weights differently. As a result, there are five data files that contain data from the COPD module: LLCP2011, LAND2011, LAND11V1, LAND11V2, LAND11V3 containing data for a total of 23 states and territories.

Sample SAS code to compile the COPD module data for ALL states that used the module

```

DATA BRFSS11.copdmodA; *Temporary data file for states that used the COPD module for all respondents
with COPD, both landline and cell-phone respondents.;
SET BRFSS11.llcp2011;
if _state in (4,9,17,26,27,30,37,47,54);
finwgt=_llcpwt; *finwgt will be the weight variable in the final data file.;
run;

```

```

DATA BRFSS11.copdmodB; *Temporary data file for states that limited the use of the module to landline
respondents only;
SET BRFSS11.LAND2011; *This data file can be downloaded from
http://www.cdc.gov/brfss/technical\_infodata/surveydata/2011/2011\_landline.htm;
if _state in (11,19,21,25,32,41,72);
finwgt=_landwt;
run;

```

```

DATA BRFSS11.copdmodC; *Temporary data file for states that use of the module on their Version 1
questionnaire;
SET BRFSS11.LAND11V1; *This data (Version 1) can be downloaded from
http://www.cdc.gov/brfss/technical\_infodata/surveydata/2011/2011\_landline\_multiple.htm;
if _state in (6,20,34,39);

```

```
if _state in (20,34,39) then finwgt=_Indwtv1; *These states included the COPD module as part of their Version 1 questionnaire only. Therefore, the entire weight is kept.;
else if _state=6 then finwgt=_Indwtv1/2; *This state included the COPD module as part of their Version 1 and Version 2 questionnaires, which were evenly split. Therefore, this sample weight is divided in half.;
run;
```

```
DATA BRFSS11.copdmodD; *Temporary data file for states that use of the module on their Version 2 questionnaire;
SET BRFSS11.LAND11V2; *This data (Version 2) can be downloaded from
http://www.cdc.gov/brfss/technical\_infodata/surveydata/2011/2011\_landline\_multiple.htm;
if _state in (6,23,31,49);
if _state in (23,49) then finwgt=_Indwtv2 ;
else if _state =6 then finwgt=_Indwtv2/2;
else if _state =31 then finwgt=_Indwtv2/3; *This state included the COPD module as part of their Version 2 and Version 3 questionnaires, but the split was not even. Their sample using Version 3 was about twice the size of that using Version 2. Therefore, the sample weight for Version 2 is divided by 3.;
run;
```

```
DATA BRFSS11.copdmodE; *Temporary data file for states that use of the module on their Version 3 questionnaire;
SET BRFSS11.LAND11V3; *This data (Version 3) can be downloaded from
http://www.cdc.gov/brfss/technical\_infodata/surveydata/2011/2011\_landline\_multiple.htm;
if _state =31;
finwgt=_Indwtv3*2/3 ; *This state's sample using Version 3 was about twice the size of that using Version 2. Therefore, the sample weight for Version 3 is multiplied by 2/3.;
run;
```

```
DATA BRFSS11.COPDMODULE; *Concatenated data file for all states that used the COPD module;
SET BRFSS11.copdmodA BRFSS11.copdmodB BRFSS11.copdmodC BRFSS11.copdmodD
BRFSS11.copdmodE;
run;
```

For analysis of the optional COPD module data for all states, use the merged data file. The sample design variables for this data file are—

_STSTR	Sample Design Stratification Variable (Unchanged)
_PSU	Primary Sampling Unit (Unchanged)
FINWGT	Final Weight for Concatenated Data File

For analysis of the COPD module data for a single state

State	FIPS code	Data set	Weight variable
Arizona	4	LLCP2011	_LLCPWT
California*	6	LAND11V1 LAND11V2	_LNDWTV1/2 _LNDWTV2/2
Connecticut	9	LLCP2011	_LLCPWT
District of Columbia	11	LAND2011	_LANDWT
Illinois	17	LLCP2011	_LLCPWT
Iowa	19	LAND2011	_LANDWT
Kansas	20	LAND11V1	_LNDWTV1
Kentucky	21	LAND2011	_LANDWT
Maine	23	LAND11V2	_LNDWTV2
Massachusetts	25	LAND2011	_LANDWT
Michigan	26	LLCP2011	_LLCPWT
Minnesota	27	LLCP2011	_LLCPWT
Montana	30	LLCP2011	_LLCPWT
Nebraska*	31	LAND11V2 LAND11V3	_LNDWTV2/3 _LNDWTV3*2/3
Nevada	32	LAND2011	_LANDWT
New Jersey	34	LAND11V1	_LNDWTV1
North Carolina	37	LLCP2011	_LLCPWT
Ohio	39	LAND11V1	_LNDWTV1
Oregon	41	LAND2011	_LANDWT
Tennessee	47	LLCP2011	_LLCPWT
Utah	49	LAND11V2	_LNDWTV2
West Virginia	54	LLCP2011	_LLCPWT
Puerto Rico	72	LAND2011	_LANDWT

*For the states with split samples, create temporary files from each data set with newly calculated weight variables. Concatenate the temporary files. Use the concatenated data set and the new weight variables for analysis.

Sample SAS code to compile the COPD module data for a state (CA) using a split sample

```

DATA BRFSS11.cacopdmod1; *Temporary data file for CA Version 1 questionnaire;
SET BRFSS11.LAND11V1;
if _state=6;
finwgt=_Lndwtv1/2;
run;
DATA BRFSS11.cacopdmod2; *Temporary data file for CA Version 2 questionnaire;
SET BRFSS11.LAND11V2;
if _state=6;
finwgt=_Lndwtv2/2;
run;

DATA BRFSS11.COPDMODCA; *Concatenated data file for CA COPD module;
SET BRFSS11.cacopdmod1 BRFSS11.cacopdmod2;
run;

```

For analysis of the optional COPD module data for CA, use the concatenated data file. The sample design variables for this data file are

<u>_STSTR</u>	Sample Design Stratification Variable (Unchanged)
<u>_PSU</u>	Primary Sampling Unit (Unchanged)
FINWGT	Final weight for Concatenated data file.