Guide for Data Users:

The Family Life, Activity, Sun, Health, and Eating Study (FLASHE)

Version 1, Updated 5/31/2016

RECOMMENDED CITATION

Appropriate citation must accompany any publication or presentation of any analysis using variables constructed using code from this document in any form. The following is recommended:

National Cancer Institute (NCI). 2014 Family Life, Activity, Sun, Health and Eating (FLASHE) Survey Data. Rockville, MD: U.S. Department of Health and Human Services, National Cancer Institute. [Year] http://cancercontrol.cancer.gov/brp/hbrb/flashe.html

Acknowledgements

TABLE OF CONTENTS

Section 1: FLASHE overview and purpose of the data user's guide

FLASHE Overview

FLASHE was designed to examine psychosocial, generational (parent-teen), and environmental correlates of cancer preventive behaviors such as diet, activity and sedentary behaviors, sleep, sun-safety, and tobacco use. FLASHE uses a non-probability sample obtained from a survey panel company (Ipsos). The sample is of parent-teen dyads, with the teen being between 12-17 years old. Parents and teens each completed two surveys (one on diet and one on physical activity). A subsample of teens wore an accelerometer. Around 1,800 dyads completed at least one survey. The data files represent the various components of the survey and are household-based, cross-sectional, dyadic (parent-child) data. NCI plans to produce and post public use files (PUF) on their website and require that users sign a data use agreement before they can download the datasets.

Purpose of the FLASHE Data User's Guide

The FLASHE Data User's Guide serves as a complement to other FLASHE resources such as the Methods Summary Report and Codebook, among others. It includes information on understanding how certain measures were conceptualized and constructed including the calculation of weights and how to apply the sampling weights. Other analytic notes are included and will be updated over time as additional FLASHE data become available.

The FLASHE datasets are free of charge and available for download from the website listed above. For additional information about the study methodology including sample size and treatment of unknown/missing values readers are referred to the FLASHE Methods Summary Report.

Requirements for Public-Use Data Dissemination

There are many Federal Laws governing the protection of individual respondents participating in federally-sponsored studies and surveys. Among them include:

Privacy Act of 1974, as amended - "The purpose of this Act is to provide certain safeguards for an individual against invasion of personal privacy by requiring Federal agencies…to collect, maintain, use or disseminate any record of identifiable personal information in a manner that assures that such action is for necessary and lawful purpose, that the information is current and accurate for its

intended use, and that adequate safeguards are provided to prevent misuse of such information." A willful disclosure of <u>individually identifiable data</u> is a misdemeanor, subject to a fine up to $5,000.

E-Government Act of 2002, Title V, Subtitle A, Confidential Information Protection and Statistical Efficiency Act (CIPSEA 2002) - Under this law, all individually identifiable information supplied by individuals or institutions to a federal agency for statistical purposes under the pledge of confidentiality must be kept confidential and may only be used for statistical purposes.   This statute prohibits disclosure or release, for non-statistical purposes, of information collected under a pledge of confidentiality. Under CIPSEA, data may not be released to unauthorized persons. Willful and knowing disclosure of protected data to unauthorized persons is a felony punishable by up to five years imprisonment and up to a $250,000 fine.

US Patriot Act of 2001 - This law permits the Attorney General to petition a court of competent jurisdiction for an ex parte order requiring the Secretary of the Department of Education to provide data relevant to an authorized investigation or prosecution of an offense concerning national or international terrorism. The law states that any data obtained by the Attorney General for these purposes "…may be used consistent with such guidelines as the Attorney General, after consultation with the Secretary, shall issue to protect confidentiality." This law was incorporated into ESRA 2002.

Federal Statistical Confidentiality Order of 1997 - This OMB Order provides a consistent government policy for "…protecting the privacy and confidentiality interests of persons who provide information for Federal statistical programs…" The Order defines relevant terms and provides guidance on the content of confidentiality pledges that Federal statistical programs should use under different conditions. The Order provides language for confidentiality pledges under two conditions- first, when the data may only be used for statistical purposes; second, when the data are collected exclusively for statistical purposes, but the agency is compelled by law to disclose the data.

Section 2: Derived Variable Specifications

Several new variables were derived from the raw survey data for the 2014 Family Life, Activity, Sun, Health, and Eating Study. Additional variables were suppressed, or top-coded / bottom-coded. This section describes the rationale for recoding, suppressing, and top/bottom-coding FLASHE survey variables. It also lists these variables and, in the case of recoded variables, presents the program statements used to develop them.

Some derived variables were created to facilitate data analysis. Other edits were done to protect participant confidentiality. The content below provides further information about the confidentiality assessments and decisions made to protect the confidentiality of FLASHE participants.

Determining Rationale for FLASHE Public-release Data

After reviewing the study design, sampling design and variables collected for the study, it was determined that the identity of the respondents could not be compromised based on matching the FLASHE data against any publicly available data. There are no unique demographic characteristics on the file that could lead towards a disclosure risk in the public domain. However, since the study is dyadic, there is a concern that a parent's / teen's responses could be identified by the respondent and could thus identify the respective dyad's responses. *It was stated during the consent/assent process that the parent's and teen's responses would not be shared with each other.* In order to ensure that the privacy of each party's responses would be maintained, steps need to be made prior to the release of the data. The procedures undertaken to protect privacy must also (1) Not cause any discrepancy with previously released restricted-use data; and (2) ensure the public-use data remains useful and beneficial to researchers.

When evaluating the variables for potential identifying information, one must factor in the dynamic nature of the responses, the time that the respondents participated, and the time the data are made available. That is, what kind of information would the respondent reasonably recall, and how would the passage of time alter some of the data?   For example, an individual's weight changes, so would he/she recall the weight at the time of the survey? It is unlikely that a respondent would recall most of the scaled responses but would most certainly recall specific responses such as diseases, languages and demographics since they are not dynamic. Even weight and height are dynamic though particularly small or large numbers need to be top-coded / bottom-coded, or categorized. All of the risk assessment issues were investigated using both expert data analyses and statistical tools.

Confidentiality Assessment for FLASHE Public Use Data Release

The risk assessment was conducted in two phases using some components of the *WesSDC ToolBox*, which is a proprietary set of SAS macros developed by Westat. In the first phase, summaries of 1-way frequencies leading to proposed recodes and variable suppression were reviewed and discussed with NCI. In addition to frequencies, the reviewed output included a spreadsheet showing the counts of nonmissing values, minimum and maximum values, the number of unique values (categories), and the number of sparse categories. For the Parent and Teen files, only the demographic files had identifying variables to recode or suppress. For all other files, variables with sparse categories (small sample sizes) were reviewed against their sensitive information to determine the risk of releasing the reported data. In all cases, it was determined that the data on the non-demographic files could be released as is, given confidentiality edits on the demographic files. In addition, there is no disclosure risk for public release by non-participants. FLASHE data cannot be used to match with any existing public databases in a meaningful way. In this respect, the objective was to reduce the risk due to identity disclosure, specifically, data masking is necessary to reduce the risk of participants finding their own responses and then knowing what their family member's responses.

Individual identities are at risk by combining indirect identifying variables, such as geography and demographics. In Phase 2, all possible 3-way and 4-way cross-tabulations were processed using the indirect identifying variables (or recodes thereof), and therefore, more recodes and suppressions occurred. The tabulations were re-processed to re-assess the risk levels. Several risk measures of re-identification were used to help guide the data coarsening. Appendix A provides more details about the risk measures used to gauge the reduction in risk.

A high-level summary of the confidentiality edits is given in the sections below. Remedies for eliminating risk included:

- Suppression of variables – this would be necessary if certain variables are easily identifiable by parent/teen;
- Coarsening, collapsing of variables – prepared to maintain reasonable analyses while protecting respondent identity; and
- Use of derived variables previously created for the restricted-use data while suppressing the original responses used in creating the derived variables.

All decisions factored in utility of data for research and thus involved revising collapsing/coarsening ranges. In the case of FLASHE, it should be noted that random perturbation (e.g., data swapping) was not conducted to reduce the risk further since it would create inconsistencies with already released restricted-use data.

Suppression of variables

The following variables were found to be too great of a risk leading to a disclosure of one's identity, and therefore were suppressed from the public use datasets.

- Teen cancer: diagnosis, type and age

- Parent cancer: diagnosis, type and age

- Home and School street address/cross-street

- Teen and parent pregnancy status

- Teen nativity: year arrived in US

Top-and-bottom coding

Continuous variables, or ordinal variables with several categories, were reviewed for outliers. Top-and bottom coding was employed, by using a 1 percent guideline (at least 1 percent of the records below the bottom-code value and 1 percent above the top-code value) to determine the cutpoint. The following table shows the variables that were impacted and their top-and-bottom coding cutpoints:

New Variable			
XPHEIGHTCM_RC	XPHEIGHTCM	149.86	190.5
XPTHEIGHTCM_RC	XPTHEIGHTCM	137.16	187.96
XPWEIGHTKG_RC	XPWEIGHTKG	46.72	142.43
XPTWEIGHTKG_RC	XPTWEIGHTKG	33.57	117.48
XTHEIGHTCM_RC	XTHEIGHTCM	137.16	187.96
XTWEIGHTKG_RC	XTWEIGHTKG	33.57	117.48

Derived Variables

The remainder of this section provides information on all other recodes for FLASHE survey variables. Some derived variables were created to facilitate data analysis. Other edits were made to collapse categories of variables to protect respondent identity.

1.1. Derived Variable Specifications - Parent Demographic Survey

Construct	Parent-reported teen height in centimeters
Derived from Item (No. and Description)	2. What is {TEEN}'s height and weight without shoes? [Feet, Inches]
New variable name and description	XPTHEIGHTCM_RC: Recoded Parent-reported TEEN height in centimeters
Notes for data users	This code converts parent-reported teen height in feet and inches into teen height in centimeters.
Syntax and Annotation	

Create a variable for parent-reported teen height in inches: XPTHEIGHTIN

  If PTHEIGHTFT < 0
     Set XPTHEIGHTIN = -9
  Else if PTHEIGHTIN > 0
     Set XPTHEIGHTIN = ((PTHEIGHTFT x 12) + PTHEIGHTIN)
  Else
     Set XPTHEIGHTIN = (PTHEIGHTFT x 12)
  End if.

Create a variable for parent-reported teen height in centimeters: XPTHEIGHTCM_RC

If XPTHEIGHTIN < 0
    Set XPTHEIGHTCM = -9
Else
    Set XPTHEIGHTCM = (XPTHEIGHTIN x 2.54) round to 2 decimal places
End if.

Note: As described above, this variable was later top/bottom coded. The new variable name is XPTHEIGHTCM_RC

Construct	Parent-reported teen weight in kilograms
Derived from Item (No. and Description)	2. What is {TEEN}'s height and weight without shoes? [Pounds]
New variable name and description	XPTWEIGHTKG_RC: Recoded Parent-reported TEEN weight in kilograms
Notes for data users	This code converts parent-reported teen weight in pounds into teen weight in kilograms.
Syntax and Annotation	

Create a variable for parent-reported teen height in inches: XPTWEIGHTKG_RC

If PTWEIGHT < 0

      Set XPTWEIGHTKG = -9

Else

      Set XPTWEIGHTKG = (PTWEIGHT x 0.453592) (rounded to 2 decimal places)

End if.

Note: As described above, this variable was later top/bottom coded. The new variable name is XPTWEIGHTKG_RC

Construct	Parent-reported height in centimeters
Derived from Item (No. and Description)	15. What is your height and weight without shoes? [Feet, Inches]
New variable name and description	XPHEIGHTCM_RC: Recoded Parent-reported height in centimeters
Notes for data users	This code converts parent-reported height in feet and inches into centimeters.
Syntax and Annotation	

Create variable for parent height in inches: XPHEIGHTIN

  If PHEIGHTFT < 0

      Set XPHEIGHTIN = -9

  Else if PHEIGHTIN > 0

      Set XPHEIGHTIN = ((PHEIGHTFT x 12) + PHEIGHTIN)

  Else

      Set XPHEIGHTIN = (PHEIGHTFT x 12)

  End if.

Create variable for parent height in centimeters: XPHEIGHTCM_RC

If XPHEIGHTIN < 0

      Set XPHEIGHTCM = -9

Else

      Set XPHEIGHTCM = (XPHEIGHTIN x 2.54) **round to 2 decimal places**

End if.

Note: As described above, this variable was later top/bottom coded. The new variable name is XPHEIGHTCM_RC

Construct	Parent-reported weight
Derived from Item (No. and Description)	15. What is your height and weight without shoes? [Pounds]
New variable name and description	XPWEIGHTKG_RC: Recoded Parent-reported weight in kilograms
Notes for data users	This code converts parent-reported weight in pounds into weight in kilograms.
Syntax and Annotation	

Create variable for parent weight in kilograms: XPWEIGHTKG_RC

If PWEIGHT < 0
      Set XPWEIGHTKG = -9
Else
      Set XPWEIGHTKG = (PWEIGHT x 0.45)
End if.

Note: As described above, this variable was later top/bottom coded. The new variable name is XPWEIGHTKG_RC

Construct	Parent-reported race/ethnicity
Derived from Item (No. and Description)	21. Are you Hispanic, Latino/a or Spanish origin? 22. Which one or more of the following would you say is your race? Please select all that apply.
New variable name and description	PETHRAC_RC: Recoded Parent Race/Ethnicity
Notes for data users	This variable is a 4-level composite measure of parental race/ethnicity.
Syntax and Annotation	

Create variable for parent ethnicity/race: PETHRAC_RC

If PETHNICITY = 1 then PETHRAC_RC = 1   /* Hispanic*/

Else if PETHNICITY = -9 then PETHRAC_RC = -9   /* Not ascertained */

Else if PAFRCNAMER = 1 and (PAMERINALSK, PASIAN, PHAWPAC, PWHITE) not equal to 1, then PETHRAC_RC = 2   /* Non-hispanic Black alone*/

Else if PWHITE = 1 and (PAMERINALSK, PASIAN, PHAWPAC, PAFRCNAMER) not equal to 1, then PETHRAC_RC = 3   /* Non-hispanic White alone*/

Else PETHRAC_RC = 4   /* Other */

Construct	Parent-reported teen health insurance – consistency of coverage
Derived from Item (No. and Description)	6. Does {TEEN} have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicaid? 8. During the past 12 months, was there any time when {TEEN} was not covered by ANY health insurance?
New variable name and description	XPTHLTHINS: Consistency of insurance coverage
Notes for data users	This variable combines information from two variables related to teen health coverages and creates a composite measure to represent consistency in teen health insurance coverage.
Syntax and Annotation	

Create variable for Insurance - consistency of coverage (Teen): XPTHLTHINS

PTCURRCOV = "No" OR PTNOHLTHCOV = "Yes" to represent "Currently uninsured or periods of no coverage during past 12 months" – SET XPTHLTHINS to 1.

PTCURRCOV = "Yes" AND PTNOHLTHCOV = "No" to represent "Consistently insured during past 12 months" – SET XPTHLTHINS to 2.

Other cases are -9: Not Ascertained.

SET XPTHLTHINS = -9

IF PTCURRCOV = 2 OR PTNOHLTHCOV = 1
    SET XPTHLTHINS = 1

ELSE IF PTNOHLTHCOV = 2
    SET XPTHLTHINS = 2

End if.

Construct	Parent-reported teen health insurance – type
Derived from Item (No. and Description)	6. Does {TEEN} have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicaid? 7. Is {TEEN} insured by Medicaid or the State Children's Health Insurance Program, S-CHIP?
New variable name and description	XPTCOVTYPE: Type of health insurance coverage
Notes for data users	This variable combines information from two variables related to teen health coverages and creates a composite measure to represent type of teen health insurance coverage.
Syntax and Annotation	

Create variable for Type of Health Insurance Coverage (Teen): XPTCOVTYPE

PTCURRCOV = "Yes" AND PTMEDCAID = "Yes" to represent "Public Health Insurance (Medicaid or SCHIP)" – SET XPTCOVTYPE to 1.
PTCURRCOV = "Yes" AND PTMEDCAID = "No" to represent "Private Health Insurance" – SET XPTCOVTYPE to 2.
PTCURRCOV = "No" to represent "Currently uninsured" – SET XPTCOVTYPE to 3.
Other cases are -9: Not Ascertained.

SET XPTCOVTYPE = -9

IF PTCURRCOV = 1 AND PTMEDCAID = 1
       SET XPTCOVTYPE = 1

ELSE IF PTCURRCOV = 1 AND PTMEDCAID = 2
       SET XPTCOVTYPE = 2

ELSE IF PTCURRCOV = 2
       SET XPTCOVTYPE = 3

End if.

Construct	Parent address in months
Derived from Item (No. and Description)	25. About how long have you lived at your current address?
New variable name and description	XPTIMEADDRMM_RC: Recoded Time at address [in years]
Notes for data users	This variable is a 4-level composite measure of time lived at respondent's current address. Time at current address was categorized as 0-3 years, 3-10 years, 10-15, and 15+.
Syntax and Annotation	

Create variable for parent "time at address in months": XPTIMEADDRMM_RC

SET XPTIMEADRMM = 0

If PTIMEADDRYY = -9
    Set XPTIMEADDRMM = -9

Else if PTIMEADDRYY in (-1, 0)
    Set XPTIMEADDRMM = PTIMEADDRMM

Else if PTIMEADDRYY > 0 AND PTIMESADDRMM = -1
    Set XPTIMEADDRMM = (PTIMEADDRYY x 12)

Else
  Set XPTIMEADDRMM = ((PTIMEADDRYY x 12) + PTIMEADDRMM)

End if.

If 0 <= XPTIMEADDRMM <= 36 then XPTIMEADDRMM_RC = 1
Else if 37 <= XPTIMEADDRMM <= 120 then XPTIMEADDRMM_RC = 2
Else if 121 <= XPTIMEADDRMM <= 180 then XPTIMEADDRMM_RC = 3
Else if 180 <= XPTIMEADDRMM then XPTIMEADDRMM_RC = 4
Else XPTIMEADDRMM_RC = XPTIMEADDRMM

Construct	
Derived from Item (No. and Description)	35. What languages do you usually speak at home? Please select all that apply.
New variable name and description	XPLANGHOME_RC: Language(s) usually spoken by PARENT at home
Notes for data users	This variable is a 2-level composite to represent English only or Not-English only households.
Syntax and Annotation	

Create variable for parent languages at home: XPLANGHOME_RC

No language identified:
PLENGLISH = -9 and all other language variables = -9 – SET XPLANGHOME = -9.

English only:
PLENGLISH = 1 and all other language variables = 0 and PLOTHSPEC <= 0 – SET
XPLANGHOME = 1.

English and other:
PLENGLISH = 1 and at least one other language variable = 1 or PLOTHSPEC > 0 – SET
XPLANGHOME = 2.

Other than English:
PLENGLISH = 0 and at least one other language variable = 1 or PLOTHSPEC > 0 – SET
XPLANGHOME = 3.

SET XPLANGHOME = 0

If PLENGLISH = -9 AND (PLSPANISH = -9 AND PLCANTONSESE = -9 AND
PLVIETNAMES = -9 AND PLTAGALOG = -9 AND PLMANDARIN = -9 AND
PLKOREAN = -9 AND PLASIANIND = -9 AND PLRUSSIAN = -9 AND PLOTHER = -9
AND PLOTHSPEC = -9)
        SET XPLANGHOME = -9

ELSE If PLENGLISH = 1 AND (PLSPANISH = 0 AND PLCANTONSESE =0 AND
PLVIETNAMES = 0 AND PLTAGALOG = 0 AND PLMANDARIN = 0 AND PLKOREAN
= 0 AND PLASIANIND = 0 AND PLRUSSIAN = 0 AND PLOTHER = 0 AND
PLOTHSPEC <=0)
        SET XPLANGHOME = 1

ELSE IF PLENGLISH = 1 AND (PLSPANISH = 1 OR PLCANTONSESE =1 OR PLVIETNAMES = 1 OR PLTAGALOG = 1 OR PLMANDARIN = 1 OR PLKOREAN = 1 OR PLASIANIND = 1 OR PLRUSSIAN = 1 OR PLOTHER = 1 OR PLOTHSPEC >0)
    SET XPLANGHOME = 2

ELSE IF PLENGLISH = 0 AND (PLSPANISH = 1 OR PLCANTONSESE =1 OR PLVIETNAMES = 1 OR PLTAGALOG = 1 OR PLMANDARIN = 1 OR PLKOREAN = 1 OR PLASIANIND = 1 OR PLRUSSIAN = 1 OR PLOTHER = 1 OR PLOTHSPEC >0)
    SET XPLANGHOME = 3

End if.

If XPLANGHOME in (2,3) then XPLANGHOME_RC = 2;/* English and/or Other */
    Else XPLANGHOME_RC = XPLANGHOME

Construct	Parent BMI
Derived from Item (No. and Description)	Not applicable.
New variable name and description	XPBMI: Computed BMI for parent
Notes for data users	This variable computes the parental body mass index from derived variables: XPWEIGHTKG and XPHEIGHTCM
Syntax and Annotation	

Create variable for parent BMI: XPBMI

SET XPBMI = -9
IF XPWEIGHTKG > 0 AND XPHEIGHTCM > 0
     SET XPBMI = XPWEIGHTKG / [XPHEIGHTCM x .01]2 (round to 2 decimal places)
  End if.

Construct	Parent-reported teen BMI
Derived from Item (No. and Description)	Not applicable.
New variable name and description	XPTBMI: Computed teen BMI based on parent-report
Notes for data users	This variable computes the teen body mass index from derived variables: XPTWEIGHTKG and XPTHEIGHTCM
Syntax and Annotation	

Create variable for parent-reported teen BMI: XPTBMI

SET XPTBMI = -9
IF XPTWEIGHTKG > 0 AND XPTHEIGHTCM > 0
     SET XPTBMI = XPTWEIGHTKG/ [XPTHEIGHTCM x .01]2 (round to 2 decimal
     places)
End if.

Construct	Parent-reported teen BMIz-score
Derived from Item (No. and Description)	Not applicable.
New variable name and description	XPTBMIZ: Calculated teen BMIz-score based on parent report of teen's height, weight and teen reported age and sex.
Notes for data users	This variable uses the CDC SAS Program for computing BMI z-scores (standard deviations) for each teen based on parental report of teen height, weight, and BMI; and teen reported age and sex. Readers are referred to the following website for additional detail: http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm
Syntax and Annotation	

Create variable for parent-reported teen BMI-z: XPTBMIZ

SET XPTBMIZ = -9
IF XTAGE > 0 AND XTSEX > 0 AND XPTBMI > 0
Age (mo.): ((XTAGE+10.5) x 12) (Teen Demographic PID)
Sex: XTSEX (Teen Demographic PID)
BMI: XPTBMI
SET XPTBMIZ = result of IF statement
End if.

Construct	Parent-reported teen BMI percentile
Derived from Item (No. and Description)	Not applicable.
New variable name and description:	XPTBMIPRCNT: Calculated teen BMI percentile based on parent report of teen's height, weight and teen reported age and sex.
Notes for data users:	This variable uses the CDC SAS Program for computing percentiles for each teen based on parental report of teen height, weight, and BMI; and teen reported age and sex. Readers are referred to the following website for additional detail: http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm
Syntax and Annotation	

Create variable for parent-reported teen BMI percentile: XPTBMIPRCNT
 	
SET XPTBMIPRCNT = -9
  IF XTAGE > 0 AND XTSEX > 0 AND XPTMBI > 0
  Sex: XTSEX (Teen Demographic PID)
  BMI: XPTBMI
  SET XPTBMIPRCNT = result of IF statement (round to 2 decimal places)
  End if.	

Construct	Parent weight status
Derived from Item (No. and Description)	Not applicable.
New variable name and description	XPWGHTSTAT: This variable computes parental weight status based on calculated parental BMI.
Notes for data users	Weight status categories are defined using standard definitions provided by the Centers for Disease Control and Prevention, which are available at: http://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html
Syntax and Annotation	

Based on Parent BMI: XPBMI
Create variable for Parent Weight Status: XPWGHTSTAT

SET XPWGHTSTAT = -9

IF XPBMI > 0
IF XPBMI >= 30.0
    SET XPWGHTSTAT = 4 (Parent Weight Status is "Obese")

ELSE IF XPBMI >= 25.0
    SET XPWGHTSTAT = 3 (Parent Weight Status is "Overweight")

ELSE IF XPBMI >= 18.5
    SET XPWGHTSTAT = 2 (Parent Weight Status is "Normal weight")

ELSE IF XPBMI < 18.5
    SET XPWGHTSTAT = 1 (Parent Weight Status is "Underweight")

 End if.

Construct	Parent-reported teen weight status
Derived from Item (No. and Description)	Not applicable.
New variable name and description	XPTWGHTSTAT: This variable computes teen weight status based on the calculated, parent-reported, BMI percentile variable: XPTBMIPRCNT
Notes for data users	Weight status categories are defined using standard definitions provided by the Centers for Disease Control and Prevention, which are available at: http://www.cdc.gov/healthyweight/assessing/bmi/childrens_bmi/about_childrens_bmi.html
Syntax and Annotation	

Based on Teen BMI Percentile: XPTBMIPRCNT
Create variable for Teen Weight Status: XPTWGHTSTAT

SET XPTWGHTSTAT = -9

IF XPTBMIPRCNT > 0
IF XPTBMIPRCNT >= 95.0
    SET XPTWGHTSTAT = 4 *(Teen Weight Status is "Obese")*

ELSE IF XPTBMIPRCNT >= 85.0
    SET XPTWGHTSTAT = 3 *(Teen Weight Status is "Overweight")*

ELSE IF XPTBMIPRCNT >= 5.0
    SET XPTWGHTSTAT = 2 *(Teen Weight Status is "Healthy weight")*

Else IF XPTBMIPRCNT < 5.0
    SET XPTWGHTSTAT = 1 *(Teen Weight Status is "Underweight")*

 End if.

Construct	Parent Cancer Diagnosis
Derived from Item (No. and Description)	19. [If parent ever diagnosed with cancer] What type of cancer did you have?
New variable name and description	XPCANCER_RC: Recoded PARENT CANCER - Multiple or Single Cancer Only
Notes for data users	Parent cancer types were combined into two categories (one cancer and multiple cancers).

Syntax and Annotation

Create variable for **parent cancer**: XPCANCER

SET XPCANCER = 0

If PCBLADDER = -9
    SET XPCANCER = -9

Else if PCBLADDER = -1
    SET XPCANCER = -1

Else If SUM
(PCBLADDER, PCBONE, PCBREAST, PCCERV, PCCOLON, PCENDOM, PCHEADNCK, PCHODGKIN, PCLEUKBLD, PCLIVER, PCLUNG, PCMELAN, PCNONHDGKN, PCORAL, PCOVARIAN, PCPANCREA, PCPHARYNGL, PCPROSTATE, PCRECTAL, PCRENKIDNY, PCSKINNMEL) > 1
    SET XPCANCER = 70

Else If PCBLADDER = 1
    SET XPCANCER = 1

Else if PCBONE = 1
    SET XPCANCER = 2

Else if PCBREAST = 1
    SET XPCANCER =3

Else if PCCERV = 1
    SET XPCANCER = 4

Else if PCCOLON = 1
    SET XPCANCER = 5

```
Else if PCENDOM = 1
       SET XPCANCER = 6

Else if PCHEADNCK = 1
       SET XPCANCER = 7

Else if PCHODGKIN = 1
       SET XPCANCER = 8

Else if PCLEUKBLD = 1
       SET XPCANCER = 9

Else if PCLIVER = 1
       SET XPCANCER = 10

Else if PCLUNG = 1
       SET XPCANCER = 11

Else if PCMELAN = 1
       SET XPCANCER = 12

Else if PCNONHDGKN = 1
       SET XPCANCER = 13

Else if PCORAL = 1
       SET XPCANCER = 14

Else if PCOVARIAN = 1
       SET XPCANCER = 15

Else if PCPANCREA = 1
       SET XPCANCER = 16

Else if PCPHARYNGL = 1
       SET XPCANCER = 17

Else if PCPROSTATE = 1
       SET XPCANCER = 18
```

```
Else if PCRECTAL = 1
        SET XPCANCER = 19

Else if PCRENKIDNY = 1
        SET XPCANCER = 20

Else if PCSKINNMEL = 1
        SET XPCANCER = 21

End if.

If XPCANCER > 0 or < 22 then XPCANCER_RC = 1          /* one cancer */
Else if XPCANCER = 70 then XPCANCER_RC = 2       /* multiple cancer */
```

Construct	Parental Age
Derived from Item (No. and Description)	10. What is your age?
New variable name and description	PAGE_RC: Recoded P_Age
Notes for data users	This code creates a four-category variable for parent age.
Syntax and Annotation	
If 18 <= PAGE <= 34 then PAGE_RC = 1 /* Age 18 – 34 */ Else If 35 <= PAGE <= 44 then PAGE_RC = 2 /* Age 35 – 44 */ Else If 45 <= PAGE <= 59 then PAGE_RC = 3 /* Age 45 – 59 */ Else If 60 <= PAGE then PAGE_RC = 4 /* Age 60+ */	

Construct	Number of children in the household
Derived from Item (No. and Description)	39. How many children under the age of 18 live in your household?
New variable name and description	PKIDSINHOME_RC: Recoded P_KidsHousehold
Notes for data users	If a household had more than 2 children they were collapsed into one category.
Syntax and Annotation	
If PKIDSINHOME >2 then PKIDSINHOME_RC = 3 Else PKIDSINHOME_RC = PKIDSINHOME	

Construct	Parent employment
Derived from Item (No. and Description)	33. About how many hours do you work per week at all of your jobs and businesses combined?
New variable name and description	PWORKHRS_RC: Recoded P_WorkHours
Notes for data users	Parent work hours were recoded as less than 15, 15-30, 31-40, and greater than 40 hours.
Syntax and Annotation	
If 0 <= PWORKHRS <= 14 then PWORKHRS_RC = 1 /* hours 0 - 14 */ Else if 15 <= PWORKHRS <= 30 then PWORKHRS_RC = 2 /* hours 15-30 */ Else if 31 <= PWORKHRS <= 40 then PWORKHRS_RC = 3 /* hours 31-40 */ Else if 41 <= PWORKHRS then PWORKHRS_RC = 4 /* hours > 40 */ Else PWORKHRS_RC = PWORKHRS	

Construct	Household income
Derived from Item (No. and Description)	34. Thinking about members of your family living in your household, what is your combined annual income, meaning the total pre-tax income from all sources earned in the last 12 months?
New variable name and description	PHSEHLDINCM_RC: Recoded P_HouseholdIncome
Notes for data users	Household income was recoded as greater than $100,000 and other.
Syntax and Annotation	
If PHSEHLDINCM in (8,9) then PHSEHLDINCM_RC = 2; Else if PHSEHLDINCM in (1 to 7) then PHSEHLDINCM_RC = 1; Else PHSEHLDINCM_RC = PHSEHLDINCM;	

Construct	Parental nativity
Derived from Item (No. and Description)	24. [if not born in United States] What year did you come to live in the United States?
New variable name and description	PNATIVAGE_RC: Recoded number of years since nativity
Notes for data users	Year since coming to live in the United States was classified as 0 – 15 years and greater than 15 years.
Syntax and Annotation	
If 0 <= PNATIVAGE <= 15 then PNATIVAGE_RC = 2 /* 0 – 15 nativity years */ Else if 15 < PNATIVAGE then PNATIVAGE_RC = 2 /* > 15 nativity years */	

Construct	Parental literacy
Derived from Item (No. and Description)	37. How would you rate your ability to read English?
New variable name and description	PLITERACY_RC: Recoded P_Literacy
Notes for data users	The classification of parent literacy was recoded as very good or not very good.
Syntax and Annotation	
If PLITERACY > 1 then PLITERACY_RC = 2 /* not very good */ Else PLITERACY_RC = PLITERACY	

Construct	Marital status
Derived from Item (No. and Description)	13. What is your marital status?
New variable name and description	PMARITAL_RC: Recoded P_MaritalStatus
Notes for data users	Marital status was recoded to four categories of married; divorced, widowed, or separated; never married; or member of an unmarried couple.
Syntax and Annotation	
If PMARITAL in (1) then PMARITAL_RC = 1 /* Married */ Else if PMARITAL in (2,3,4) then PMARITAL_RC = 2 /* Divorced, widowed, or separated*/ Else if PMARITAL in (5) then PMARITAL_RC = 3 /* Never Married */ Else if PMARITAL = 6 then PMARITAL_RC = 4 /*Member of an unmarried couple */ Else PMARITAL_RC = PMARITAL;	

Construct	Home ownership
Derived from Item (No. and Description)	28. Do you currently rent or own your home?
New variable name and description	PHOMEOWN_RC: Recoded P_OwnHome
Notes for data users	Home ownership was collapsed so that parents renting were grouped with ones not paying any rent.
Syntax and Annotation	

If PHOMEOWN in (2,3) then PHOMEOWN_RC = 2   /* Rent or no payment */
Else PHOMEOWN_RC = PHOMEOWN;

1.1. Derived Variable Specifications - Teen Demographic Survey

Construct	Teen-reported height
Derived from Item (No. and Description)	7. XTHEIGHT: What is your height and weight without shoes? (feet, inches)
New variable name and description	XTHEIGHTCM_RC: Recoded teen height in centimeters
Notes for data users	This code converts teen height in feet and inches to teen height in centimeters, using teen-reported data.
Syntax and Annotation	

Create variable for teen height in inches: XTHEIGHTIN

If THEIGHTFT < 0
      Set XTHEIGHTIN = -9

Else if THEIGHTIN > 0
      Set XTHEIGHTIN = ((THEIGHTFT x 12) + THEIGHTIN)

Else
      Set XTHEIGHTIN = (THEIGHTFT x 12)
End if.

Create variable for teen height in centimeters: XTHEIGHTCM

If XTHEIGHTIN < 0
      Set XTHEIGHTCM = -9

Else
      Set XTHEIGHTCM = (THEIGHTIN x 2.54) (round to 2 decimal places)
End if.

Note: As described above, this variable was later top/bottom coded. The new variable name is XTHEIGHTCM_RC

Construct	Teen-reported weight
Derived from Item (No. and Description)	7. XTWEIGHT: What is your height and weight without shoes? (pounds)
New variable name and description:	XTWEIGHTKG_RC: Recoded teen weight in kilograms
Notes for data users:	This code converts teen weight in pounds to teen weight in kilograms, using teen-reported data.
Syntax and Annotation	

Create variable for teen weight in kilograms: XTWEIGHTKG

If TWEIGHT < 0

     Set XTWEIGHTKG = -9

Else

     Set XTWEIGHTKG = (TWEIGHT x 0.453592) (round to 2 decimal places)

End if.

Note: As described above, this variable was later top/bottom coded. The new variable name is XTWEIGHTKG_RC

Construct	Teen-reported languages at home
Derived from Item (No. and Description)	16. What languages do you usually speak at home?
New variable name and description:	XTLANGHOME_RC: Language(s) usually spoken by TEEN at home
Notes for data users:	This variable is a 2-level composite to represent English only or Not-English only households.
Syntax and Annotation	

Create variable for teen languages at home: XTLANGHOME
No language identified:
TLENGLISH = -9 and all other language variables = -9 – SET XTLANGHOME = -9.

English only:
TLENGLISH = 1 and all other language variables = 0 and TLOTHSPEC <= 0 – SET XTLANGHOME = 1.

English and other:
TLENGLISH = 1 and at least one other language variable = 1 or TLOTHSPEC > 0 – SET XTLANGHOME = 2.

Other than English:
TLENGLISH = 0 and at least one other language variable = 1 or TLOTHSPEC > 0 – SET XTLANGHOME = 3.

SET XTLANGHOME = 0

If TLENGLISH = -9 AND (TLSPANISH = -9 AND TLCANTONSESE = -9 AND TLVIETNAMES = -9 AND TLTAGALOG = -9 AND TLMANDARIN = -9 AND TLKOREAN = -9 AND TLASIANIND = -9 AND TLRUSSIAN = -9 AND TLOTHER = -9 AND TLOTHSPEC = -9)
        SET XTLANGHOME = -9

ELSE If TLENGLISH = 1 AND (TLSPANISH = 0 AND TLCANTONSESE =0 AND TLVIETNAMES = 0 AND TLTAGALOG = 0 AND TLMANDARIN = 0 AND TLKOREAN = 0 AND TLASIANIND = 0 AND TLRUSSIAN = 0 AND TLOTHER = 0 AND TLOTHSPEC <=0)
        SET XTLANGHOME = 1

ELSE IF TLENGLISH = 1 AND (TLSPANISH = 1 OR TLCANTONSESE =1 OR TLVIETNAMES = 1 OR TLTAGALOG = 1 OR TLMANDARIN = 1 OR TLKOREAN = 1 OR TLASIANIND = 1 OR TLRUSSIAN = 1 OR TLOTHER = 1 OR TLOTHSPEC >0)
    SET XTLANGHOME = 2

ELSE IF TLENGLISH = 0 AND (TLSPANISH = 1 OR TLCANTONSESE =1 OR TLVIETNAMES = 1 OR TLTAGALOG = 1 OR TLMANDARIN = 1 OR TLKOREAN = 1 OR TLASIANIND = 1 OR TLRUSSIAN = 1 OR TLOTHER = 1 OR TLOTHSPEC >0)
    SET XTLANGHOME = 3

If XTLANGHOME in (2,3) then XTLANGHOME_RC = 2;   /* Not English only */
Else XTLANGHOME_RC = XTLANGHOME

Construct	Teen-reported race/ethnicity
Derived from Item (No. and Description)	12. Are you Hispanic, Latino/a or Spanish origin? 13. Which one or more of the following would you say is your race?
New variable name and description:	TETHRAC_RC: Recoded Teen Race/Ethnicity
Notes for data users:	This variable is a 4-level composite measure of teen race/ethnicity.
Syntax and Annotation	

Create variable for teen Hispanic ethnicity/race: TETHRAC_RC

If TETHNICITY = 1 then TETHRAC_RC = 1   /* Hispanic*/

Else if TETHNICITY = -9 then TETHRAC_RC = -9   /* Not ascertained */

Else if TAFRCNAMER = 1 and (TAMERINALSK, TASIAN, THAWPAC, TWHITE) not equal to 1, then TETHRAC_RC = 2   /* Non-hispanic Black alone*/

Else if TWHITE = 1 and (TAMERINALSK, TASIAN, THAWPAC, TAFRCNAMER) not equal to 1, then TETHRAC_RC = 3   /* Non-hispanic White alone*/

Else TETHRAC_RC = 4   /* Other */

Construct	Teen-reported BMI
Derived from Item (No. and Description)	Not applicable.
New variable name and description:	XTBMI: Computed teen BMI based on teen-report
Notes for data users:	This variable computes the teen body mass index from derived variables: XTWEIGHTKG and XTHEIGHTCM
Syntax and Annotation	
Create variable for teen BMI: XTBMI SET XTBMI = -9 IF XTWEIGHTKG > 0 AND XTHEIGHTCM > 0 SET XTBMI = XTWEIGHTKG / [XTHEIGHTCM x .01] 2 (round to 2 decimal places) End if.	

Construct	Teen-reported BMIz-score
Derived from Item (No. and Description)	Not applicable.
New variable name and description:	XTBMIZ: Calculated teen BMIz-score based on teen report of teen's height, weight and teen reported age and sex.
Notes for data users:	This variable uses the CDC SAS Program for computing BMI z-scores (standard deviations) for each teen based on teen report of teen height, weight, BMI, age, and sex. Readers are referred to the following website for additional detail: http://www.cdc.gov/nccdphp/dnpao/ growthcharts/resources/sas.htm
Syntax and Annotation	

Teen BMI-z score and **Teen BMI percentile** can both be calculated using the variables listed below and plugging those values in for each PID into the SAS code found in the link below.

Create variable for teen BMI-z: XTBMIZ
SET XTBMIZ = -9
IF XTAGE > 0 AND XTSEX > 0 AND XTBMI > 0
Age (mo.): ((XTAGE+10.5) x 12)
Sex: XTSEX
BMI: XTBMI
http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm
SET XTBMIZ = result of else statement (round to 2 decimal places)
  End if.

Construct	Teen-reported BMI percentile
Derived from Item (No. and Description)	Not applicable.
New variable name and description:	XTBMIPRCNT: Calculated teen BMI percentile based on teen report of teen's height, weight, age, and sex
Notes for data users:	This variable uses the CDC SAS Program for computing percentiles for each teen based on teen report of teen height, weight, BMI, age, and sex. Readers are referred to the following website for additional detail: http://www.cdc.gov/nccdphp/dnpao/ growthcharts/resources/sas.htm
Syntax and Annotation	

Note: **Teen BMI-z score** and **Teen BMI percentile** can both be calculated using the variables listed below and plugging those values in for each PID into the SAS code found in the link below.

Create variable for teen BMI percentile: XTBMIPRCNT
SET XTBMIPRCNT = -9
IF XTAGE > 0 AND XTSEX > 0 AND XTBMI > 0
  Age (mo.): ((XTAGE+10.5) x 12)
Sex: XTSEX
BMI: XTBMI
http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm
SET XTBMIPRCNT = result of else statement (round to 2 decimal places)
  End if.

Construct	Teen-reported weight status
Derived from Item (No. and Description)	Not applicable.
New variable name and description:	XTWGHTSTAT: This variable computes teen weight status based on the calculated, teen-reported BMI percentile variable: XTBMIPRCNT
Notes for data users:	Weight status categories are defined using standard definitions provided by the Centers for Disease Control and Prevention, which are available at: http://www.cdc.gov/healthyweight/assessing/ bmi/childrens_bmi/about_childrens_bmi.html
Syntax and Annotation	

Based on Teen BMI Percentile: XTBMIPRCNT
Create variable for Teen Weight Status: XTWGHTSTAT

SET XTWGHTSTAT = -9

IF XTBMIPRCNT > 0
IF XTBMIPRCNT >= 95.0
      SET XTWGHTSTAT = 4 *(Teen Weight Status is "Obese")*

ELSE IF XTBMIPRCNT >= 85.0
      SET XTWGHTSTAT = 3 *(Teen Weight Status is "Overweight")*

ELSE IF XTBMIPRCNT >= 5.0
      SET XTWGHTSTAT = 2 *(Teen Weight Status is "Healthy weight")*

ELSE IF XTBMIPRCNT < 5.0
      SET XTWGHTSTAT = 1 *(Teen Weight Status is "Underweight")*

End if.

Construct	Teen work hours
Derived from Item (No. and Description)	18. How many hours a week do you get paid to work?
New variable name and description:	TWORKHRS_RC: Teen work hours (recoded)
Notes for data users:	Teen work hours were recoded so that teens working 20 or more hours were grouped together.
Syntax and Annotation	
If TWORKHRS in (4,5,6,7) then TWORKHRS_RC = 5; /* 20 or more hours per week */ Else TWORKHRS_RC = TWORKHRS;	

1.3 Derived Variable Specifications - Parent Physical Activity Survey

Construct	Parent vigorous activity (minutes per day)
Derived from Item (No. and Description)	6. How much time did you usually spend doing **vigorous** physical activities on one of those days?
New variable name and description:	XPPVIGMINS: Parent vigorous physical activity time in minutes (on one day)
Notes for data users:	This variable combines hours and minutes of daily vigorous activity to arrive at a total number of minutes parents report engaging in vigorous activity in one day.
Syntax and Annotation	

Create variable for parent vigorous physical activity time in minutes: XPPVIGMINS
 	
SET XPPVIGMINS = 0
 	
If PPVIGMINS = -9 SET XPPVIGMINS = -9
 	
Else if PPVIGHRS = -1 AND PPVIGMINS = -1
      SET XPPVIGMINS = -1
 	
Else If PPVIGHRS **<=** 0 AND PPVIGMINS > 0
      SET XPPVIGMINS = PPVIGMINS
 	
Else if PPVIGHRS > 0 AND PPVIGMINS **<=** 0
      SET XPPVIGMINS = (PPVIGHRS x 60)
 	
Else if PPVIGHRS > 0 AND PPVIGMINS > 0
      SET XPPVIGMINS = ((PPVIGHRS x 60) + PPVIGMINS)
 	
 End if.	

Construct	Parent moderate activity (minutes per day)
Derived from Item (No. and Description)	8. How much time did you usually spend doing MODERATE physical activities on one of those days?
New variable name and description:	XPPMODMINS: Parent moderate physical activity time in minutes (on one day)
Notes for data users:	This variable combines hours and minutes of daily moderate activity to arrive at a total number of minutes parents report engaging in moderate activity in one day.
Syntax and Annotation	

Create variable for parent moderate physical activity time in minutes: XPPMODMINS

SET XPPMODMINS = 0

If PPMODMINS = -9
      SET XPPMODMINS = -9

Else if PPMODHRS = -1 AND PPMODMINS = -1
      SET XPPMODMINS = -1

Else If PPMODHRS **<=** 0 AND PPMODMINS > 0
      SET XPPMODMINS = PPMODMINS

Else if PPMODHRS > 0 AND PPMODMINS **<=** 0
      SET XPPMODMINS = (PPMODHRS x 60)

Else if PPMODHRS > 0 AND PPMODMINS > 0
      SET XPPMODMINS = ((PPMODHRS x 60) + PPMODMINS)

End if.

Construct	Parent walking activity (minutes)
Derived from Item (No. and Description)	10. How much time did you usually spend WALKING on one of those days?
New variable name and description:	XPPWLKMINS: Parent walking activity time in minutes (on one day)
Notes for data users:	This variable combines hours and minutes of daily walking activity to arrive at a total number of minutes parents report walking in one day.
Syntax and Annotation	

Create variable for parent walking time in minutes: XPPWLKMINS

SET XPPWLKMINS = 0

If PPWLKMINS = -9
    SET XPPWLKMINS = -9

Else if PPWLKHRS = -1 AND PPWLKMINS = -1
    SET XPPWLKMINS = -1

Else If PPWLKHRS **<=** 0 AND PPWLKMINS > 0
    SET XPPWLKMINS = PPWLKMINS

Else if PPWLKHRS > 0 AND PPWLKMINS **<=** 0
    SET XPPWLKMINS = (PPWLKHRS x 60)

Else if PPWLKHRS > 0 AND PPWLKMINS > 0
    SET XPPWLKMINS = ((PPWLKHRS x 60) + PPWLKMINS)

End if.

Construct	Parent sitting (minutes on a weekday)
Derived from Item (No. and Description)	11. During the LAST 7 DAYS, how much time did you spend SITTING on a WEEKDAY?
New variable name and description:	XPPSITMINS: Parent sitting time in minutes (on a weekday)
Notes for data users:	This variable combines hours and minutes of daily time spend sitting to arrive at a total number of minutes parents report sitting on a weekday.
Syntax and Annotation	

Create variable for parent sitting time in minutes: XPPSITMINS

SET XPPSITMINS = 0

If PPSITMINS = -9
     SET XPPSITMINS = -9

Else if PPSITHRS = -1 AND PPSITMINS = -1
     SET XPPSITMINS = -1

Else If PPSITHRS **<=** 0 AND PPSITMINS > 0
     SET XPPSITMINS = PPSITMINS

Else if PPSITHRS > 0 AND PPSITMINS **<=** 0
     SET XPPSITMINS = (PPSITHRS x 60)

Else if PPSITHRS > 0 AND PPSITMINS > 0
     SET XPPSITMINS = ((PPSITHRS x 60) + PPSITMINS)

End if.

Construct	Parent hours of sleep - weekdays
Derived from Item (No. and Description)	27. What time do you usually go to bed in the evening (turn out the lights in order to go to sleep? Weekday. (Hour, Minute, AM or PM) 28. What time do you usually get out of bed in the morning? Weekday. (Hour, Minute, AM or PM)
New variable name and description:	XPPTBWKDYMT: Parent weekday go to bed (military time) XPPOBWKDYMT: Parent weekday get out of bed (military time) XPSLPWKDYHM: Parent weekday amount of sleep in hours
Notes for data users:	This code converts hours and minutes of times to bed and out of bed on a weekday into military time. This code then calculates elapsed weekday sleep time in hours.
Syntax and Annotation	

Create variable for Parent **To Bed Weekday** in Military Time: XPPTBWKDYMT
Note: For PPTBWKDY, 1 = AM 2 = PM

SET XPPTBWKDYMT = Null

IF PPTBWKDYH > 0
IF PPTBWKDYM = -1 [R made no entry for minute, but hour is present]
      SET XPPTBWKDYMT = (PPTBWKDYH + PPTBWKDY) in Military Time

ELSE
      SET XPPTBWKDYMT = (PPTBWKDYH + PPTBWKDYM + PPTBWKDY) in
      Military Time

End if.
End if.

Create variable for Parent **Out of Bed Weekday** in Military Time: XPPOBWKDYMT
Note: For PPOBWKDY, 1 = AM 2 = PM

SET XPPOBWKDYMT = Null

IF PPOBWKDYH > 0
IF PPOBWKDYM = -1 *[R made no entry for minute, but hour is present]*

SET XPPOBWKDYMT = (PPOBWKDYH + PPOBWKDY) in Military Time

ELSE

SET XPPOBWKDYMT = (PPOBWKDYH + PPOBWKDYM + PPOBWKDY) in Military Time

End if.
End if.

Create NUMERIC variable (rounded to 2 decimal places) for **Parent Weekday Amount of Sleep:** XPSLPWKDYHM

SET XPSLPWKDYHM = -9

IF XPPOBWKDYMT = Null or XPPTBWKDYMT = Null
SET XPSLPTWKDYHM = -9

ELSE
SET XPSLPTWKDYHM = TO *USE SAS UTILITY TO CALCULATE ELAPSED TIME* BETWEEN XPPTBWKDYMT and XPPOBWKDYMT

IF XPSLPTWKDYHM <= 0, SET XPSLPTWKDYHM = -9

End if.

Construct	Parent hours of sleep - weekends
Derived from Item (No. and Description)	27. What time do you usually go to bed in the evening (turn out the lights in order to go to sleep)? Weekend. (Hour, Minute, AM or PM) 28. What time do you usually get out of bed in the morning? Weekend. (Hour, Minute, AM or PM)
New variable name and description:	XPPTBWKNDMT: Parent weekend go to bed (military time) XPPOBWKNDMT: Parent weekend get out of bed (military time) XPSLPWKNDM: Parent weekend amount of sleep in hours
Notes for data users:	This code converts hours and minutes of times to bed and out of bed on a weekend day into military time. This code then calculates elapsed weekend sleep time in hours.
Syntax and Annotation	

Create variable for Parent **To Bed Weekend** in Military Time: XPPTBWKNDMT
Note: For PPTBWKND, 1 = AM 2 = PM

SET XPPTBWKNDMT = Null

If PID <> 6245 *[6245 is a special case that should remain NULL for XPPTBWKNDMT only]*
THEN IF PPTBWKNDH > 0

IF PPTBWKNDM = -1 *[R made no entry for minute, but hour is present]*
       SET XPPTBWKNDMT = (PPTBWKNDH + PPTBWKND) in Military Time

ELSE
       SET XPPTBWKNDMT = (PPTBWKNDH + PPTBWKNDM + PPTBWKND) in
       Military Time

End if.

Create variable for Parent **Out of Bed Weekend** in Military Time: XPPOBWKNDMT
 Note: For PPOBWKND, 1 = AM 2 = PM
 SET XPPOBWKNDMT = Null

IF PPOBWKNDH > 0
IF PPOBWKNDM = -1 [R made no entry for minute, but hour is present]
    SET XPPOBWKNDMT = (PPOBWKNDH + PPOBWKND) in Military Time

ELSE
    SET XPPOBWKNDMT = (PPOBWKNDH + PPOBWKNDM + PPOBWKND) in
    Military Time

End if.
End if.

Create NUMERIC variable (rounded to 2 decimal places) for **Parent Weekend Amount of Sleep**:
XPSLPWKNDHM

SET XPSLPWKNDHM = -9

IF XPPOBWKNDMT = Null or XPPTBWKNDMT = Null
    SET XPSLPWKNDHM = -9

ELSE
SET XPSLPWKNDHM = TO *USE SAS UTILITY TO CALCULATE ELAPSED TIME*
BETWEEN XPPTBWKNDMT and XPPOBWKNDMT

IF XPSLPWKNDHM <= 0, SET XPSLPWKNDHM = -9
End if.

1.4 Derived Variable Specifications – Teen Physical Activity Survey

Construct	Teen School Time Start (in military time)
Derived from Item (No. and Description)	9: What time does your school day typically start? (Hour, Minute, AM or PM)
New variable name and description:	XTPSCHSTRTMT: Teen school start in military time
Notes for data users:	This variable indicates time that the school day typically starts, in military time.
Syntax and Annotation	

Create variable for **Teen School Start** in Military Time: XTPSCHSTRTMT
Note: For TPSCHSTRT, 1 = AM 2 = PM

SET XTPSCHSTRTMT = Null

IF TPSCHSTRTH > 0
IF TPSCHSTRTM = -1 *[R made no entry for minute, but hour is present]*
        SET XTPSCHSTRTMT = (TPSCHSTRTH + TPSCHSTRT) in Military Time

ELSE
        SET XTPSCHSTRTMT = (TPSCHSTRTH + TPSCHSTRTM + TPSCHSTRT) in Military Time

 End if.
 End if.

Construct	Teen School Time End (in military time)
Derived from Item (No. and Description)	10. What time does your school day typically end? (Hour, Minute, AM or PM)
New variable name and description:	XTPSCHENDMT: Teen school end in military time
Notes for data users:	This variable indicates time that the school day typically ends, in military time.
Syntax and Annotation	

Create variable for **Teen School End** in Military Time: XTPSCHENDMT
Note: For TPSCHEND, 1 = AM 2 = PM

SET XTPSCHENDMT = Null

IF TPSCHENDH > 0
IF TPSCHENDM = -1 *[R made no entry for minute, but hour is present]*
        SET XTPSCHENDMT = (TPSCHENDH + TPSCHEND) in Military Time

ELSE
        SET XTPSCHENDMT = (TPSCHENDH + TPSCHENDM + TPSCHEND) in Military Time

End if.
End if.

Construct	Teen sleep - weekday
Derived from Item (No. and Description)	72. What time do you usually go to bed in the evening (turn out the lights in order to go to sleep? Weekday. (Hour, Minute, AM or PM) 73. What time do you usually get out of bed in the morning? Weekday. (Hour, Minute, AM or PM)
New variable name and description:	XTPTBWKDYMT: Teen weekday go to bed (military time) XTPOBWKDYMT: Teen weekday get out of bed (military time) XTSLPWKDYHM: Teen weekday amount of sleep in hours
Notes for data users:	This code converts hours and minutes of times to bed and out of bed on a weekday into military time. This code then calculates elapsed weekday sleep time in hours.
Syntax and Annotation	

Create variable for **Teen to Bed Weekday** in Military Time: XTPTBWKDYMT
Note: For TPTBWKDY, 1 = AM 2 = PM

SET XTPTBWKDYMT = Null
IF TPTBWKDYH > 0
IF TPTBWKDYM = -1 *[R made no entry for minute, but hour is present]*
     SET XTPTBWKDYMT = (TPTBWKDYH + TPTBWKDY) in Military Time

ELSE
     SET XTPTBWKDYMT = (TPTBWKDYH + TPTBWKDYM + TPTBWKDY) in
     Military Time

End if.
End if.

Create variable for **Teen Out of Bed Weekday** in Military Time: XTPOBWKDYMT
Note: For TPOBWKDY, 1 = AM 2 = PM

SET XTPOBWKDYMT = Null
IF TPOBWKDYH > 0

IF TPOBWKDYM = -1 *[R made no entry for minute, but hour is present]*

SET XTPOBWKDYMT = (TPOBWKDYH + TPOBWKDY) in Military Time

ELSE

SET XTPOBWKDYMT = (TPOBWKDYH + TPOBWKDYM + TPOBWKDY) in
Military Time

End if.
End if.

*Use any available SAS utilities to calculate the hours of sleep based on the derived military times.
Create NUMERIC variable (rounded to 2 decimal places) for **Teen Weekday Amount of Sleep**:
XTSLPWKDYHM

SET XTSLPWKDYHM = -9

IF XTPOBWKDYMT = Null or XTPTBWKDYMT = Null

SET XTSLPWKDYHM = -9

ELSE SET XTSLPWKDYHM = TO *USE SAS UTILITY TO CALCULATE ELAPSED
TIME* BETWEEN XTPTBWKDYMT and XTPOBWKDYMT

IF XTSLPWKDYHM <= 0, SET XTSLPWKDYHM = -9

End if.

Construct	Teen sleep - weekend
Derived from Item (No. and Description)	72. What time do you usually go to bed in the evening (turn out the lights in order to go to sleep)? Weekend. (Hour, Minute, AM or PM) 73. What time do you usually get out of bed in the morning? Weekend. (Hour, Minute, AM or PM)
New variable name and description:	XTPTBWKNDMT: Teen weekend go to bed (military time) XTPOBWKNDMT: Teen weekend get out of bed (military time) XTSLPWKNDHM: Teen weekend amount of sleep in hours
Notes for data users:	This code converts hours and minutes of times to bed and out of bed on a weekend day into military time. This code then calculates elapsed weekend sleep time in hours.
Syntax and Annotation	

Create variable for **Teen To Bed Weekend** in Military Time: XTPTBWKNDMT
Note: For TPTBWKND, 1 = AM 2 = PM

SET XTPTBWKNDMT = Null

IF TPTBWKNDH > 0
IF TPTBWKNDM = -1 *[R made no entry for minute, but hour is present]*
     SET XTPTBWKNDMT = (TPTBWKNDH + TPTBWKND) in Military Time

ELSE
     SET XTPTBWKNDMT = (TPTBWKNDH + TPTBWKNDM + TPTBWKND) in
     Military Time

End if.
End if.

Create variable for **Teen Out of Bed Weekend** in Military Time: XTPOBWKNDMT
Note: For TPOBWKND, 1 = AM 2 = PM

SET XTPOBWKNDMT = Null

IF TPOBWKNDH > 0
IF TPOBWKNDM = -1 *[R made no entry for minute, but hour is present]*
      SET XTPOBWKNDMT = (TPOBWKNDH + TPOBWKND) in Military Time

ELSE
      SET XTPOBWKNDMT = (TPOBWKNDH + TPOBWKNDM + TPOBWKND) in
      Military Time

End if.
End if.

*Use any available SAS utilities to calculate the hours of sleep based on the derived military times. Create NUMERIC variable (rounded to 2 decimal places) for **Teen Weekend Amount of Sleep**: XTSLPWKNDHM

SET XTSLPWKNDHM = -9

IF XTPOBWKNDMT = Null or XTPTBWKNDMT = Null
      SET XTSLPWKNDHM = -9

ELSE SET XTSLPWKNDHM = TO *USE SAS UTILITY TO CALCULATE ELAPSED TIME* BETWEEN XTPTBWKNDMT and XTPOBWKNDMT

IF XTSLPWKNDHM <= 0, SET XTSLPWKNDHM = -9

End if.

Section 3: Weighting Procedures

3.1   Weighting Overview

Survey/study weights were created for all FLASHE respondents. The FLASHE survey weights are not traditional sampling weights because FLASHE used a non- probability (e.g. convenience) sample. Given that inferences based on such a convenience sample could be largely different from the general population, "analysis weights" were created with the intention of making the weighted sample more similar to the general population by raking the weights to the target population on key demographics. The report of the AAPOR Task Force on Non-Probability Sampling (page 72) indicates that "in many non-probability samples, post-stratification is the only form of weighting." The report goes on to explain that "Tourangeau and his co-authors (2013) summarized the results across eight studies that attempted to reduce biases in non-probability opt-in panels by using weighting methods when the biases were due to coverage and selection effects…..Overall, the adjustments reduce to some extent, but do not by any means eliminate coverage, nonresponse, and selection biases inherent in opt-in panels."

The analysis weights started from a base weight of one for each respondent and then, within cells defined by demographic variables, were inflated by matching the sum of the weights to population control totals derived from larger surveys such as the American Community Survey (ACS) and the Current Population Survey (CPS). These raked weights are intended to reduce bias on the estimates due to the undercoverage of the consumer panels, the nonprobability sampling strategy, differential sampling rates by quota cells (such as the oversampling of African Americans), and differential response rates by sample characteristics.

Raking procedures were implemented to create the final analysis weights. Raking uses an iterative procedure to benchmark the sum of the weights to the marginal population total for one dimension (a single variable or the cross-classification of several variables) at a time. The process is iterated until the control totals for all dimensions are simultaneously satisfied (at least within a specified tolerance). It can be thought of as a multidimensional post-stratification procedure. Final raked weights were created using Westat's proprietary SAS Macro.

It should be noted that by using weights, the variance of the estimates will be larger due to the variation of the weight compared to the no weighting approach. Because of the imbalance of the FLASHE sample, a large variation of weights can be expected between:

- Male and female parents;

- Hispanics and Non-Hispanics for both parents and adolescents;

- Some census divisions; and

- Some income categories.

Although weights were created for this quota sample, caution should be taken in the interpretation of the weighted estimates and any statistical tests. If weighted estimates from this survey were made for the raked variables, the resulting estimates would be equal to the weighted estimate from the larger probability sample (either ACS or CPS). However, the expected value of weighted estimates for other variables is not necessarily equal to the corresponding population value and is thus possibly subject to bias. Even with this possible bias, the weighted estimates should be closer to the underlying population values than the weighted estimates. At this point, variance estimation for weighted quota samples remains a challenging issue for the field of survey research.

3.2   Imputation of Missing Raking Variables

Because raking was to be used in the weighting process and the variables to be used for raking were missing for some survey respondents, the missing values were imputed. First, these values were filled in using the screener data, when available. If screener information was not available, the variable was imputed using Westat's proprietary SAS Macro AutoImpute (Judkins et al, 2007). This macro uses the predictive mean matching method (PMM), where a linear model is fitted to predict the variables to be imputed and the imputed value is taken from the record with a similar predicted value. Variables are imputed sequentially and in iterations. The demographic variables in the parent and adolescent survey data were used as predictors in imputing each variable. These imputed variables were used in raking and the same variables in ACS and CPS were used to construct population control totals. The missing rate of demographic variables for each survey dataset was low, ranging from 0 to 3.11 percent.

3.3 Types of Weights Developed

Due to the complexity of the study design of FLASHE and to facilitate planned analyses, several sets of weights were developed. Creating weights for each individual survey dataset in addition to the combined datasets allows data analysts to use the maximum number of respondents based on the analytical need. For example, if the data from both the Physical Activity and Diet parent surveys

are being analyzed, the weights developed for respondents who completed both surveys needs to be used. On the other hand, if only parent Physical Activity data is being analyzed, the weights developed for the parent respondents to the Physical Activity survey should be used rather than the weights developed for parents who completed both surveys. Three sets of weights were developed for the parent survey data:

- Weights for the parent Physical Activity survey respondents (WT_P_PA);

- Weights for the parent Diet survey respondents (WT_P_Diet); and

- Weights for parents that completed both the Physical Activity and Diet surveys (WT_P_BOTH).

Four sets of weights were developed for the adolescent survey data:

- Weights for the adolescent Physical Activity survey respondents(WT_T_PA);

- Weights for the adolescent Diet survey respondents(WT_T_Diet);

- Weights for adolescents that completed both the Physical Activity and Diet surveys(WT_T_BOTH); and

- Weights for adolescents who wore an accelerometer as part of the Motion Study (WT_MOTION).

The fourth set of weights for the adolescent data is important because the adolescents included in the accelerometer motion study were a random subsample of the full adolescent sample. A separate set of weights is needed for this subgroup in order for the sum of the final weights for this subgroup to equal the target adolescent population. Those who returned one or both of the surveys but had no accelerometer data (either because they were not in the Motion Study or due to failure to wear the device) did not receive the motion study weight.

Care should be taken in selecting which weights to use in data analysis. For any analysis involving the adolescent accelerometer data, the motion study weight (WT_MOTION) should be used. For any analysis involving the surveys only, the other parent or adolescent weights should be used. For example, if the analysis variable is from the Parent Physical Activity survey, WT_T_PA should be used. If variables from both the parent Physical Activity survey and the parent Diet survey are used in one analysis, WT_P_BOTH is the best weight to use.

Note that no weight was produced for the specific use in dyadic analysis, which uses data from both the parent surveys and the adolescent surveys, because the population control totals cannot be easily defined. As a common practice, dyadic analyses are unweighted. However, if a weight is preferred during dyadic analysis and one major outcome variable is the focus of the analysis, the weight for the outcome variable can be used as the best alternative. For example, if the outcome variable is from the parent Physical Activity survey, the weight for the parent Physical Activity respondents can be used. Although some respondents to the parent Physical Activity survey will drop out of the analysis due to not responding to the adolescent survey, the weighted distribution of the outcome variable may be close to that of all respondents to the parent Physical Activity survey.

3.4　　　Weighting of Parent Data

Three parent weights were created to facilitate the analysis of parent data for survey respondents defined in three different ways: respondents to the parent Physical Activity survey, respondents to the parent Diet survey, and respondents to both surveys. The same set of raking dimensions was applied in creating these weights. Hence, one set of population control totals was developed and used for all parent weighting.

3.4.1　　　Creation of Parent Population Control Totals

To create the population control total file for raking the parent weights, the target population of the FLASHE survey was defined as: Parent, step-parent or legal guardian of a 12-17 year old that lives in the same household as the adolescent at least 50 percent of the time during the week.

The two most commonly used datasets for constructing control totals for the US population were considered: the American Community Survey (ACS) and the Current Population Survey (CPS), Annual Social and Economic Supplement (ASEC). [6] As neither of the two datasets was controlled to the Census for the special target population specified above, ACS was selected due to its large sample size. The ACS 1-year Public Use Microdata Sample (PUMS) provides data from 3 million addresses per year, while the CPS ASEC annual sample size is about 100,000 addresses.

One difficulty in constructing the parent control totals was that legal guardians cannot be easily identified based on the relationship within household reported in ACS. Assumptions had to be made in identifying legal guardians other than parents or grandparents. This accounted for only a small portion of the population. Issues considered include the relationship to householder[7] of the household, whether grandparents are responsible for the adolescents, and the subfamily[8]

relationship. Also, the amount of time staying in the household during the week cannot be identified in ACS or CPS ASEC.

The following steps were taken to construct the population control totals for the parents using the 2013 ACS 1-year PUMS file:

1. Adolescents 12-17 years old were first extracted from the ACS PUMS file. Adults and adolescents in group quarters were excluded from the construction of the parent control totals.

2. If parents in the same household/family or grandparents who take care of the adolescents extracted were available in the ACS PUMS file based on the relationship variables, they were extracted as parents/legal guardians. This accounts for 94.42 percent of all identified parents.

3. If no parents/legal guardians were found in the household in step 2, adults aged 18+ in the same household with certain relationships to the householder were assumed to be the parents. If more than two such adults were available in the household, the first two based on the person sequential number created during data collection were counted as parents.

---

6 The Current Population Survey (CPS Annual Social and Economic Supplement (ASEC) will be referred to as CPS ASEC in this document.

7 One person in each household is designated as the householder. In most cases, this is the person or one of the people in whose name the home is owned, being bought, or rented and who is listed on line one of the survey questionnaire. If there is no such person in the household, any adult household member 15 years old and over could be designated as the householder.

8 Subfamily – A subfamily is a married couple (husband and wife interviewed as members of the same household) with or without never-married children under 18 years old, or one parent with one or more never-married children under 18 years old. A subfamily does not maintain its own household, but lives in a household where the householder or householder's spouse is a relative. The number of subfamilies is not included in the count of families, since subfamily members are counted as part of the householder's family. Subfamilies are defined during processing of data. Same-sex married couples are only shown as the householder and spouse, and are not included in subfamilies.

Given the relationship variables available within households or subfamilies, it was not possible to identify roommates, roomers, or unmarried partners of the householder. Also, for some non-relative adolescents of the householder, no parents were identified. In total, 2,225 adolescents did not have parents/legal guardians identified. The adolescents with no parent/legal guardian assigned accounted for about 0.9% of the 236,864 adolescents who were aged 12-17 and who were not in group quarters in the 2013 ACS sample.

3.4.2　　　Raking Variables for Parent Data

In constructing raking dimensions, all variables that are available in both the FLASHE survey and the ACS data were considered. Eight socio-demographic variables expected to be related to the outcome variables in the survey were selected. Categorization of the selected socio-demographic variables was created to avoid small sample sizes. Each raking dimension was constructed based on one socio-demographic variable. The eight raking dimensions are:

1. **Age:** 20-39/40-44/46-49/50 or more;

2. **Gender:** Male/Female;

3. **Race/Ethnicity:** Non-Hispanic Black/Non-Hispanic White/Other;

4. **Census Region:** Northeast/Midwest/South/West;

5. **Income:** $0-$34,999/$35,000-$74,999/$75,000 or more;

6. **Marital Status:** Married/Other;

7. **Homeownership:** Own/Other; and

8. **Work Status:** Employed/Not Employed.

In creating the marginal population control totals, ACS weighted counts were computed for each raking dimension. In total, eight control total files were created. Table 2-1 shows the ACS sample size and population totals of identified parents/legal guardians by the raking variables.

Table 2-1.   ACS sample size and population totals of identified parents/legal guardians by
the raking variables

	ACS sample size	ACS population control
Overall	291,786	29,797,729
Age		
20-39	72,335	8,319,077
40-44	72,014	7,500,021
45-49	71,960	7,042,940
50 and more	75,477	6,935,691
Gender		
Male	129,690	13,046,623
Female	162,096	16,751,106
Race/Ethnici		
Non-Hispanic White only	190,557	17,882,995
Non-Hispanic Black only	27,372	3,470,738
Other	73,857	8,443,996
Census Region		
Northeast	50,579	5,093,615
Midwest	64,029	6,418,179
South	107,072	11,097,773
West	70,106	7,188,162
Income		
$0-$34,999	56,880	6,395,791
$35,000-$74,999	85,388	8,897,674
$75,000 or more	149,518	14,504,264
Marital Status		
Married	236,409	23,251,555
Other	55,377	6,546,174
Homeownership		
Own	219,858	20,920,259
Other	71,928	8,877,470
Work Status		
Employed	224,785	22,901,087
Other	67,001	6,896,642

3.4.3 Raking of Parent Data

Each of the three sets of FLASHE parent weights was raked to the parent population control totals using Westat's SAS Macro FSRake. In this process, the weighted estimates by the first raking dimension were adjusted to agree with the first set of controls (e.g., gender), then the weighted estimates by the second raking dimension were adjusted to agree with the second set of controls (e.g., age group), etc. This process was carried out through all dimensions and then repeated until the sum of FLASHE weights simultaneously equaled the corresponding control totals. Note that the sum of the raked weights equals the population totals marginally for each dimension, but not for the cross-classification of the dimensions.

The three raked weights for the parent surveys are summarized in Tables 5-2 – 5-4, respectively. Summary statistics of the weights including sample size (n), mean, standard deviation (STD), Coefficient of Variance (CV), sum, maximum value (MAX) and minimum value (MIN) are presented in the tables, by the raking variables and for all respondents to the corresponding survey(s). The sample size shows that there were no raking categories with a small sample size. The sum of the weights is equal to the population control total as a result of raking. STD, MAX and MIN show the spread and range of the weights.

Table 5-2.   Parent weight for the respondents to the physical activity survey

	N	MEA	STD	CV*100	SUM	MAX	MIN
Overall	802	6,536	11,556	70	29,797,729	80,468	3,870
Age							
20-39	537	15,492	11,285	73	8,319,077	80,468	3,870
40-44	448	16,741	11,159	67	7,500,021	79,811	4,058
45-49	402	17,520	12,724	73	7,042,940	75,878	4,471
50 and more	415	16,713	11,060	66	6,935,691	75,520	4,781
Gender							
Male	476	27,409	14,527	53	13,046,623	80,468	7,467
Female	1,326	12,633	6,937	55	16,751,106	52,981	3,870
Race/ethnicity							
Non-Hispanic White	1,247	14,341	7,466	52	17,882,995	49,666	3,870
Non-Hispanic Black	312	11,124	4,880	44	3,470,738	38,789	4,058
Other	243	34,749	16,748	48	8,443,996	80,468	10,230
Census Region							
Northeast	330	15,435	10,792	70	5,093,615	74,894	4,058
Midwest	515	12,462	7,347	59	6,418,179	54,205	3,870
South	651	17,047	11,604	68	11,097,773	72,229	5,156
West	306	23,491	14,371	61	7,188,162	80,468	5,267
Income							
$0-$34,999	421	15,192	10,003	66	6,395,791	80,468	5,244
$35,000-$74,999	719	12,375	8,214	66	8,897,674	62,427	3,870
$75,000 or more	662	21,910	13,344	61	14,504,264	79,811	6,820
Marital Status							
Married	1,295	17,955	12,290	68	23,251,555	80,468	4,463
Other	507	12,912	8,406	65	6,546,174	64,794	3,870
Homeownership							
Own	1,290	16,217	11,569	71	20,920,259	79,811	3,870
Other	512	17,339	11,493	66	8,877,470	80,468	4,842
Work status							
Employed	1,191	19,228	12,615	66	22,901,087	80,468	5,514
Other	611	11,287	6,483	57	6,896,642	54,580	3,870

Table 5-3.　　Parent weight for the respondents to diet survey

	N	MEA	STD	CV*100	SUM	MAX	MIN
Overall	1,754	16,988	12,065	71	29,797,729	87,237	3,986
Age							
20-39	514	16,185	12,028	74	8,319,077	87,237	3,986
40-44	447	16,779	11,346	68	7,500,021	82,319	3,996
45-49	388	18,152	13,449	74	7,042,940	81,068	4,522
50 and more	405	17,125	11,418	67	6,935,691	80,112	5,035
Gender							
Male	458	28,486	15,369	54	13,046,623	87,237	7,853
Female	1,296	12,925	7,100	55	16,751,106	56,497	3,986
Race/ethnicity							
Non-Hispanic White	1,223	14,622	7,716	53	17,882,995	51,907	3,986
Non-Hispanic Black	297	11,686	5,254	45	3,470,738	40,180	3,996
Other	234	36,085	17,787	49	8,443,996	87,237	10,118
Census Region							
Northeast	322	15,819	11,303	71	5,093,615	79,862	3,996
Midwest	499	12,862	7,692	60	6,418,179	55,640	3,986
South	637	17,422	11,968	69	11,097,773	73,227	5,246
West	296	24,284	15,271	63	7,188,162	87,237	5,485
Income							
$0-$34,999	413	15,486	10,300	67	6,395,791	87,237	5,130
$35,000-$74,999	694	12,821	8,661	68	8,897,674	70,059	3,986
$75,000 or more	647	22,418	14,045	63	14,504,264	82,319	6,849
Marital status							
Married	1,257	18,498	12,912	70	23,251,555	87,237	4,711
Other	497	13,171	8,480	64	6,546,174	69,467	3,986
Homeownership							
Own	1,259	16,617	12,083	73	20,920,259	82,319	3,986
Other	495	17,934	11,981	67	8,877,470	87,237	5,055
Work status							
Employed	1,159	19,759	13,208	67	22,901,087	87,237	5,640
Other	595	11,591	6,734	58	6,896,642	58,558	3,986

Table 5-4.   Parent weight for the respondents to both pa and diet surveys

	N	MEA	STD	CV*100	SUM	MAX	MIN
Overall	1,708	17,446	12,200	70	29,797,729	87,262	4,150
Age							
20-39	502	16,572	12,083	73	8,319,077	87,262	4,150
40-44	432	17,361	11,622	67	7,500,021	82,695	4,291
45-49	381	18,485	13,501	73	7,042,940	80,472	4,714
50 and more	393	17,648	11,587	66	6,935,691	80,296	5,279
Gender							
Male	450	28,992	15,403	53	13,046,623	87,262	8,137
Female	1,258	13,316	7,251	54	16,751,106	56,493	4,150
Race/Ethnicity							
Non-Hispanic White	1,197	14,940	7,745	52	17,882,995	51,758	4,150
Non-Hispanic Black	283	12,264	5,363	44	3,470,738	41,083	4,291
Other	228	37,035	17,761	48	8,443,996	87,262	10,676
Census Region							
Northeast	314	16,222	11,424	70	5,093,615	79,496	4,291
Midwest	486	13,206	7,808	59	6,418,179	56,442	4,150
South	618	17,958	12,122	68	11,097,773	74,099	5,449
West	290	24,787	15,339	62	7,188,162	87,262	5,818
Income							
$0-$34,999	397	16,110	10,543	65	6,395,791	87,262	5,550
$35,000-$74,999	676	13,162	8,789	67	8,897,674	69,601	4,150
$75,000 or more	635	22,841	14,115	62	14,504,264	82,695	7,149
Marital status							
Married	1,235	18,827	13,011	69	23,251,555	87,262	4,887
Other	473	13,840	8,813	64	6,546,174	71,060	4,150
Homeownership							
Own	1,229	17,022	12,219	72	20,920,259	82,695	4,150
Other	479	18,533	12,097	65	8,877,470	87,262	5,321
Work Status							
Employed	1,125	20,357	13,333	65	22,901,087	87,262	5,984
Other	583	11,830	6,726	57	6,896,642	58,710	4,150

The differences among the mean weights by the raking variables illustrate the variation of the mean weights, which was due to an imbalanced outgoing sample, lack of sample in the panel(s) for males and for Blacks, oversampling Black, and unequal response rates in the screener and in the survey(s). For example, the mean weight of males is much higher than that of females. This is because the Ipsos panel and partner panels did not have enough panel members who are male parents of adolescents. The mean weight of Blacks is lower than that of other race categories due to the oversampling of Blacks. As described in Chapter 2, the target sample size of Blacks was supposed to account for 25 percent of the overall sample. However, due to lack of Black parents of adolescents in the Ipsos panel, the black respondents were oversampled at a

lower degree and accounted for about 17 percent of respondents. Other differences in mean weights were due to the lack of balancing of the outgoing sample and unequal response rates by socio-demographics in the screener and the survey(s).

3.5   Weighting of Adolescent Data

Four adolescent weights were created to facilitate the analysis of adolescent data for study respondents defined in four different ways:

- Weights for respondents to the adolescent Physical Activity survey (WT_T_PA);

- Weights for respondents to the adolescent Diet survey (WT_T_Diet);

- Weights for respondents to both surveys (WT_T_BOTH); and

- Weights for respondents to the motion study (WT_MOTION).

Note that the motion study weights were created for any adolescents who wore the accelerometer, no matter whether a survey was completed or not. The same set of raking dimensions was applied in creating these weights. Hence, one set of population control totals were developed and used for adolescent weighting. If an adolescent responded to the Physical Activity survey, the Diet survey and the motion study, this adolescent has positive values for all four weights. See Section 5.3 for which weight to use in analyses.

3.5.1   Creation of Population Control Totals for Adolescent Data

The adolescent target population in the FLASHE study is adolescents 12-17 years old with a parent who lives with the adolescent at least 50 percent of the time during the week. Unlike the parent control totals, the CPS ASEC was used to develop adolescent population control totals. The CPS ASEC data offer more detailed age categories than the ACS data. Thus, the population control totals could be created using the age range (12-17 years old) that was also used in the FLASHE study. Although the CPS ASEC sample size is much smaller than ACS, the CPS ASEC data was post-stratified to the US Census data by the same variables in the dimensions in raking FLASHE adolescent weights (race/ethnicity, sex and age). In addition, the age categories (12-13, 14, 15, 16-17) for FLASHE raking were also used in post-stratifying CPS ASEC data within Black and White respectively.

To construct the adolescent population control totals using the 2014 CPS ASEC, adolescents aged 12-17 years old who lived with an adult in the same household were extracted from the 2014 CPS ASEC file. Adolescents in group quarters were excluded from the construction of the parent control totals.

3.5.2   Raking Variables for Adolescent Data

In constructing raking dimensions, all variables that are available in both the FLASHE survey and the CPS data were considered. Four socio-demographic variables that were expected to be related to the outcome variables in the survey were selected. Categorization of the selected socio-demographic variables was conducted to avoid small categories. Each raking dimension was constructed based on one variable. The four raking dimensions are listed as below.

1. **Age:** 12-13/14-15/16-17;

2. **Gender:** Male/Female;

3. **Race/Ethnicity:** Non-Hispanic Black/Non-Hispanic White/Other; and

4. **Census Region:** Northeast/Midwest/South/West.

In creating the marginal population control totals, CPS ASEC weighted counts were computed as the population control totals by each raking dimension. In total, four population control total files were created. Table 5-5 shows the CPS ASEC sample size and population control totals for adolescents by the raking variables.

Table 5-5.   CPS ASEC sample size and population control totals for adolescents by the raking variables

	CPS sample size	Population control totals
Overall	13,620	25,494,345
Age		
12-13	4,472	8,279,172
14-15	4,593	8,321,924
16-17	4,555	8,893,249
Gender		
Male	6,994	12,909,782
Female	6,626	12,584,563
Race/Ethnicity		
Non-Hispanic White only	7,657	13,978,158
Non-Hispanic Black only	1,434	3,496,106
Other	4,529	8,020,082
Census Region		
Northeast	2,393	4,319,930
Midwest	3,083	5,584,939
South	4,317	9,549,595
West	3,827	6,039,881

3.5.3   Raking of Adolescent Data

Each of the four sets of FLASHE adolescent weights was raked to the adolescent population control totals using Westat's SAS Macro FSRake. As described previously, after raking, the sum of FLASHE weights equaled the population totals marginally for each dimension

The four raked weights for the adolescent surveys are summarized in Tables 5-6 – 5-9, respectively. Summary statistics of weights including sample size (n), mean, standard deviation (STD), Coefficient of Variance (CV), sum, maximum value (MAX) and minimum value (MIN) are presented in the tables, by the raking variables and for all respondents to the corresponding survey(s)/study. The sample size shows that there were no small raking categories. The sum of the weights is equal to the control total as a result of raking. Note that 6 adolescent respondents reported age 11 and 4 adolescent respondents reported age 18. Age 11 was treated as in the category 12-13 and age 18 was treated as in the category 16-17.

The overall CV for the adolescent weights is around 0.35, which is much smaller than the CV (around 0.7) for the parent weights. The main reason is that the weight variation by gender was minimal for adolescents, in comparison to the large difference in mean weights between male parents and female parents. In addition, there are four raking dimensions for adolescents in comparison to the eight raking dimensions for parents. Additional raking dimensions usually increase weight variation so the lower number of dimensions also contributes to the lower variation.

Table 5-9 shows the differences among the mean weights by race and region, which are due to lack of balancing in the outgoing sample and unequal response rates by demographics in the screener and the survey(s).

Table 5-6.   Summary of the adolescent weight for the respondents to physical activity survey

	n	MEA	STD	CV*100	SUM	MAX	MIN
Overall	1,670	15,266	5,261	34	25,494,345	32,326	9,628
Age							
12-13	559	14,811	5,142	35	8,279,172	27,840	9,696
14-15	578	14,398	4,884	34	8,321,924	27,646	9,628
16-17	533	16,685	5,494	33	8,893,249	32,326	11,258
Gender							
Male	829	15,573	5,336	34	12,909,782	32,326	9,730
Female	841	14,964	5,171	35	12,584,563	31,985	9,628
Race/Ethnicity							
Non-Hispanic White	1,068	13,088	2,310	18	13,978,158	18,448	10,070
Non-Hispanic Black	273	12,806	1,892	15	3,496,106	17,638	9,628
Other	329	24,377	4,095	17	8,020,082	32,326	17,645
Census Region							
Northeast	303	14,257	3,892	27	4,319,930	24,319	11,229
Midwest	479	11,660	2,824	24	5,584,939	20,852	9,628
South	599	15,943	4,337	27	9,549,595	27,728	12,803
West	289	20,899	6,060	29	6,039,881	32,326	14,926

Table 5-7.   Summary of the adolescent weights for the respondents to the diet survey

	n	MEA	STD	CV*100	SUM	MAX	MIN
Overall	1,667	15.294	5.450	36	25,494.345	33.244	9.603
Age							
12-13	555	14.917	5.397	36	8,279.172	29.456	9.603
14-15	569	14.626	5.199	36	8,321.924	29.497	9.617
16-17	543	16.378	5.606	34	8,893.249	33.244	10.838
Gender							
Male	824	15.667	5.583	36	12,909.782	33.244	9.755
Female	843	14.928	5.295	35	12,584.563	32.727	9.603
Race/Ethnicity							
Non-Hispanic White	1,066	13.113	2.409	18	13,978.158	18.799	10.157
Non-Hispanic Black	277	12.621	1.841	15	3,496.106	17.775	9.603
Other	324	24.753	4.428	18	8,020.081	33.244	17.961
Census Region							
Northeast	307	14.071	3.879	28	4,319.930	23.751	11.077
Midwest	480	11.635	2.842	24	5,584.939	20.591	9.603
South	601	15.890	4.372	28	9,549.595	27.301	12.733
West	279	21.648	6.344	29	6,039.881	33.244	15.505

Table 5-8.   Summary of the adolescent weight for the respondents to both pa and diet surveys

	n	MEA	STD	CV*10	SUM	MAX	MIN
Overall	1,590	16,034	5,674	35	25,494,345	34,448	10,192
Age							
12-13	535	15,475	5,534	36	8,279,172	29,867	10,192
14-15	541	15,382	5,401	35	8,321,924	30,295	10,338
16-17	514	17,302	5,895	34	8,893,249	34,448	11,755
Gender							
Male	786	16,425	5,819	35	12,909,782	34,448	10,357
Female	804	15,652	5,505	35	12,584,563	33,900	10,192
Race/Ethnicity							
Non-Hispanic White	1,025	13,637	2,411	18	13,978,158	19,235	10,441
Non-Hispanic Black	257	13,604	1,976	15	3,496,106	18,775	10,192
Other	308	26,039	4,407	17	8,020,082	34,448	18,700
Census Region							
Northeast	289	14,948	4,173	28	4,319,930	25,735	11,967
Midwest	459	12,168	3,004	25	5,584,939	21,917	10,192
South	570	16,754	4,633	28	9,549,595	29,218	13,587
West	272	22,205	6,636	30	6,039,881	34,448	16,019

Table 5-9.   Summary of the adolescent weight for the respondents to motion study

	n	MEA	STD	CV*10	SUM	MAX	MIN
Overall	509	50,087	16,087	32	25,494,345	101,877	29,280
Age							
12-13	166	49,875	15,746	32	8,279,172	92,304	29,987
14-15	177	47,017	15,215	32	8,321,924	90,129	29,280
16-17	166	53,574	16,719	31	8,893,249	101,877	33,096
Gender							
Male	259	49,845	15,327	31	12,909,781	94,727	29,280
Female	250	50,338	16,865	34	12,584,564	101,877	31,490
Race/ethnicity							
Non-Hispanic White	318	43,956	11,492	26	13,978,159	72,783	29,280
Non-Hispanic Black	73	47,892	6,782	14	3,496,105	68,950	30,614
Other	118	67,967	17,605	26	8,020,082	101,877	40,984
Census Region							
Northeast	82	52,682	8,252	16	4,319,930	75,621	44,441
Midwest	162	34,475	5,403	16	5,584,939	49,823	29,280
South	186	51,342	7,763	15	9,549,595	73,740	43,335
West	79	76,454	13,797	18	6,039,881	101,877	59,871

3.6　Limitations of the FLASHE Weights and the Use of Quota Cell Variables

As noted earlier, the FLASHE sample which was selected via the Ipsos Consumer Panel is a quota sample, not a probability sample. Although weighing has been conducted, caution should be taken in the interpretation of the weighted estimates and any statistical tests. As indicated on page 21 of the Report of the AAPOR Task Force on Non-Probability Sampling, "population subject to recruitment is likely to be a small and unrepresentative portion of the target population of interest…..With non- probability samples, it may be better to call this problem 'exclusion bias' rather than 'coverage bias,' since the vast majority of the target population is likely to have no chance of inclusion in the sample." On page 33, the report goes on to say that "The selection bias in most non-probability methods creates the substantial risk that the distribution of important covariates in the sample will differ significantly from their distribution in the target population and to such an extent that inferences are misleading if not simply wrong". To create a "pseudo design-based weight" (See page 67), the Task Force reports that "a common theme is to estimate the probability by computing the ratio of the sample size to the estimated population total within some categories (e.g., post-stratification). Clearly, the creation or estimation of a pseudo weight requires strong assumptions for some non-probability sampling recruitment methods."

The one-step raking (similar to post-stratification) procedure in FLASHE weighting was used to manage this risk of large biases. The effectiveness of this adjustment depends on the correlation between the raking variables and the analysis variable. As the result of raking, the weighted estimates for the raking variables are unbiased, i.e., the weighted estimates equal to the population estimates from the larger probability samples, e.g., ACS and CPS ASEC. The weighted estimates for variable other than the raking variable are subject to bias in making inference to the United States population. Level of bias may depend on the selection bias within the raking dimensions and the correlation between the raking variables and the analysis variables.

In terms of variance estimation, the Report of the AAPOR Task Force on Non-Probability Sampling provides general guidance (See pages 74-75): "An approach often taken is to assume a SRS design and estimate a standard deviation under this assumption. This approach sometimes is also used for probability samples, but the literature is very clear that ignoring the sample design produces biased estimates of precision. The same is undoubtedly true for most non-probability

samples."

The quota cell variables in FLASHE are race (non-Hispanic Black / other) for adults and gender and age for adolescents. Given that there is no probability sampling within the quota cell it is unclear whether it is appropriate to use the quota cell as the stratification variable in analytical procedures for survey data such PROC SURVEY. Currently, there is no theoretical support for a preferred variance estimation method for a finite population based on non-probability sampling. Sergey Dorofeev, Peter Grant (2006) suggested that overestimating variance is probably better than underestimating, as over-reporting significant finding will be minimized. Under this scheme, it is advised not to use the quota cell as the stratification variable in analysis, which usually reduces the estimated variance. One alternative is to include the quota cell as a covariate in the modeling practice to account for the preset quota boundary.

Again, there is no theoretical rationale for this.

A common practice in analyzing quota samples is to only take the unequal weights into account in the analysis without considering the quota cell variable. For the FLASHE study, the unequal weights by the quota cell should not contribute much to the overall variance. African American adults accounted for about 17 percent of the adult respondents in FLASHE, while African Americans made up 13.5 percent of the US population. Thus, the oversampling was moderate. The adolescent respondents by gender and age groups were approximately proportional to the population counts. Thus, ignoring the quota cell variable in the analysis should have a minimum effect. Again, the analysis of survey data based on non-probably samples is still being explored by researchers and at this time there is no valid theoretical method for estimation.

To communicate these limitations in journal articles, it should be declared that the data is a quota sample from a consumer panel and that survey weights were constructed through raking procedures on key socio-demographic variables as an attempt to reduce potential bias from population values. After raking, any statistical inference based on variables other than the raking variables are subject to selection bias. The results and discussion sections can interpret the analysis results as in any other analysis, but should avoid generalizing to the U.S. general population. For example, odds ratios can be interpreted as "category A is X times more likely than category B for a certain event".

References

Reg Baker, J. Michael Brick, Nancy A. Bates, Mike Battaglia, Mick P. Couper, Jill A. Dever, Krista J. Gile, Roger Tourangeau. Report of the AAPOR Task Force on Non-Probability Sampling. June 2013.

Sergey Dorofeev and Peter Grant. Statistics for Real-Life Sample Surveys: Non-Simple-Random Samples and Weighted Data. Cambridge University Press, 2006.

Tourangeau, Roger, Frederick G. Conrad, and Mick P. Couper. (2013). *The Science of Web Surveys.*

New York: Oxford University Press.

Section 4: Data Note

There were 10 adolescent records whose age was outside the range of the age requirement of 12 to 17. It was decided to drop these records from the public use file rather than top-code the 18 year olds to 17 and bottom-code the 11 year olds to 12. Because the adolescent records were dropped, the corresponding parent records were also dropped.

Section 5: Diet Survey Outcome Variables

A later release of the FLASHE diet survey datasets will include derived diet variables, including daily frequency of intake of various food groups. This section will be updated to accompany the release of those data and will include information on how those variables were computed.

Section 6: Teen Physical Activity Survey Outcome Variables

When the adolescent physical activity survey dataset is released (planned for release in 2016), it will include derived variables for adolescent physical activity behavior using the Youth Activity Profile. This section will be updated to accompany the release of those data and will include information on how those variables were computed.

Section 7: Parent Physical Activity Survey: Imputation

When the parent physical activity survey dataset is released (planned for release in 2016), it will be accompanied by a dataset that includes imputed values for a small subset of variables that have a high percent of missing data in the parent physical activity survey. These include questions on parents' goals for their adolescents and health conditions that may limit physical activity. This section will be updated to accompany the release of the imputed data and will include information on the imputation procedures.

Section 8: Motion Study Variables

A subset of adolescents participated in a motion study, during which they wore accelerometers for a period seven days. A dataset containing variables that have been derived from accelerometer data is planned for release by the end of 2016. This section will be updated to accompany the release of those data.

Appendix A. Description of risk measures

The *SDCNway* macro is one in a series of SAS macros, referred to as the *WesSDC ToolBox*, developed for statistical disclosure control (SDC). SDC techniques can be defined as the set of methods to reduce the risk of disclosing information on individuals, businesses or other organizations. Such methods are only related to the dissemination step and are usually based on restricting the amount of information disclosure or modifying the data released. The *SDCNway* software is designed to help the user analyze the disclosure risk elements in the data set. The results from the risk analysis can be used to guide the disclosure control treatments such as data suppression, recoding, swapping (directed or random), or other data perturbation techniques. Li and Krenzke (2013) review the use of *SDCNway* and compare it with other software and disclosure risk measures in an evaluation study[1].

The methodology for the *SDCNway* software is simple and extensive. It is simple because it is runs cross-tabulations as the basic approach. It is extensive since it processes an exhaustive number of tabulations and captures key information along the way. The basic idea of *SDCNway* is to identify the attributes or combinations of attributes that make a record different from the others in survey data. The unique or rare cases are associated with high disclosure risk. If a sample case can be uniquely identified by a small number of less detailed attributes, it is even more risky because it is highly likely to be a population unique. Disclosure risk may arise if an intruder intends to identify individuals and disclose their identities or attributes through the matching of known information to external sources.

For forming the tabulations, it is recommended to use factual identifiers such as demographic and geographical variables, etc. Variables with subjective responses, such as cognitive items, are not visible or identifiable by data intruders and would typically be excluded from the extensive multi-way analysis.

Within *SDCNway*, a limited recoding utility is offered, if requested by the user. The user can recode some non-missing values of a variable to be missing so that these values will not be involved in the subsequent risk analysis. That is, the values will not be included in any combination of attributes to determine sample uniqueness or rareness. After initial recodes are implemented, the next phase of the risk analysis is to process all possible tables of certain dimensions for a specified number of variables. Violations are flagged when table cell counts are less than a given threshold rule – three for example (i.e., Rule of 3). For each category of each variable, the proportion of table cells with violations is computed among all cells in which a variable is involved. The algorithm counts the number of violations in which a record is involved for the set of tables generated.

---

[1] Li and Krenzke (2013). Comparing Approaches That Are Used to Identify High Risk Values in Microdata, Census Statistical Disclosure Risk Research report prepared for the U.S. Census Bureau.

Other re-identification metrics were used in the FLASHE disclosure risk assessment, including those introduced in the Appendix of El Emam (2011)[2]. The metrics are listed as follows:

1. pR_α is the average number of cases in cells with cell count less than $\frac{1}{\tau}$, where τ is a threshold value (set equal to 0.2 for FLASHE).

$$pR_\alpha = \frac{1}{n} \sum_{k \in K} f_k \times I\left(\frac{1}{f_k} > \tau\right)$$

where n = the number of respondents.

2. jR_α is the average number of cases in the cells with $\frac{1}{F_k}$ less than τ.

$$jR_\alpha = \frac{1}{n} \sum_{k \in K} f_k \times I\left(\frac{1}{F_k} > \tau\right)$$

3. jR_c is the average of re-identification risk $\frac{1}{F_k}$.

$$jR_c = \frac{1}{n} \sum_{k \in K} \frac{f_k}{F_k}$$

F_k was estimated by the sum of weights.

[2] El Emam, K. (2011). Methods for the de-identification of electronic health records for genomic research. Genome Medicine, 3, 1-9. (appendix available at: http://genomemedicine.com/content/supplementary/gm239-s1.pdf.