HINTS 5 Cycle 1 History Document

November 2017

Data Editing

The following variables were identified to contain invalid or unusual values. Those values were replaced with negative value of -4, "Unreadable or Non-conforming numeric response", negative value of -9, "Missing data (Not Ascertained)" or reasonable regular values.

Height_Inches: G4. About how tall are you without shoes? Inches:

One hundred and forty-three respondents had inches of -9, "Missing data (Not Ascertained)", or -4, "Unreadable or Non-conforming numeric response", which were replaced by 0. The variable Height_Feet is recommended for determining validity of full height.

R_HHAdults: Reconciled number of adults in household

One respondents had missing values (.), which were replaced by -9.

Totalhousehold: O13. Including yourself, how many people live in your household?

Sixty respondents had Totalhousehold of 0, which were replaced by 1.

HowLongModerateExerciseHr: I2. How long are you typically doing these activities? Hours: & HowLongModerateExerciseMn: I2. How long are you typically doing these activities? Minutes:

One respondent had HowLongModerateExerciseHr of 0 and HowLongModerateExerciseMn of 0, both of which were replaced by -9. One respondent had HowLongModerateExerciseHr of 0 and HowLongModerateExerciseMn of -1, where HowLongModerateExerciseHr was replaced with -1. Thirteen respondents had HowLongModerateExerciseHr of 1 and HowLongModerateExerciseMn of 60, where HowLongModerateExerciseMn was replaced with 0.

The following variables were identified to be in a different format in the HINTS 5 Cycle 1 data compared to the cycles of HINTS 4 data. These variables were updated to be formatted consistently with HINTS 4.

HHID: Household ID

The HHID variable was converted from a string variable with a length of 8 to a numeric variable.

Standard Recode

Standard recode/derived variables are listed below.

SEC_RUCA_2010_DESCRIPT: USDA 2010 Secondary Rural-Urban Commuting Area Codes Description (variable found in Stata dataset, only)

The SEC_RUCA_2010 variable was recoded into the string variable SEC_RUCA_2010_DESCRIPT containing the secondary urban-rural commuting area code description. This recoded variable is present on the Stata dataset only, as Stata does not allow format/labels to be applied to the non-integer values that are present in SEC_RUCA_2010.

AgeGrpA: AgeGrpA. 4 Level Age Categories Version A (Derived from Age)

The Age variable was re-coded into 4 categories: 18-34; 35-39; 40-44; 45+. The original negative values were carried over.

AgeGrpB: AgeGrpB. 5 Level Age Categories Version B (Derived from Age)

The Age variable was re-coded into 5 categories: 18-34; 35-49; 50-64; 65-74; 75+. The original negative values were carried over.

EducA: EducA. What is the highest level of school you completed? 4 Levels (Derived from Education)

The Education variable was re-coded into 4 categories: Less than High School; High School Graduate; Some College; College Graduate or More. The original negative values were carried over.

EducB: EducB. What is the highest level of school you completed? 5 Levels (Derived from Education)

The Education variable was re-coded into 5 categories: Less than High School; High School Graduate; Some College; Bachelor's Degree; Post-Baccalaureate Degree. The original negative values were carried over.

RaceEthn: Race/Ethnicity. 7 Levels (Derived from Hisp_Cat and Race_Cat2)

The RaceEthn was created with Hisp_Cat and Race_Cat2 variables. The RaceEthn has 7 categories: Hispanic; Non-Hispanic White; Non-Hispanic Black or African American; Non-Hispanic American Indian or Alaska Native; Non-Hispanic Asian; Non-Hispanic Native Hawaiian or other Pacific Islander; Non-Hispanic Multiple Races Mentioned. If Hisp_Cat had value of 10, "Not Hispanic", and Race_Cat2 had value of -9, "Missing data (Not Ascertained)", the RaceEthn was assigned with value of -9. The RaceEthn was assigned with value of -9 if Hispanic=-9.

RaceEthn5: Race/Ethnicity. 5 Levels (Derived from Hisp_Cat and Race_Cat2)

The RaceEthn5 was created with Hisp_Cat and Race_Cat2 variables. The RaceEthn5 has 5 categories: Non-Hispanic White; Non-Hispanic Black or African American; Hispanic; Non-Hispanic Asian; Non-Hispanic Other. If Hisp_Cat had value of 10, "Not Hispanic", and Race_Cat2 had value of -9, "Missing data (Not Ascertained)", the RaceEthn5 was assigned with value of -9. The RaceEthn5 was assigned with value of -9 if Hispanic=-9.

HHInc: HHInc. What is your {combined} annual household income? 5 Levels (Derived from IncomeRanges)

The IncomeRanges variable was re-coded into 5 categories: Less than $20,000; $20,000 to < $35,000; $35,000 to < $50,000; $50,000 to < $75,000; $75,000 or more. The original negative values were carried over.

BMI: BMI. Body Mass Index (Weight*703)/(Height in inches**2)

The BMI variable was created with weight in pounds and height in inches. If feet of height or weight had value of -9, “Missing data (Not Ascertained)” but neither had value of -4, “Unreadable or Non-conforming numeric response”, the BMI was assigned to -9. If feet of height or weight had value of -4, “Unreadable or Non-conforming numeric response”, the BMI was assigned to -4.

AgeDX: AgeDX. At what age were you diagnosed with cancer? (Derived from WhenDiagnosedCancer)

The variable AgeDX is a copy of variable WhenDiagnosedCancer.

TimeSinceDX: TimeSinceDX. How long ago were you diagnosed with cancer? (Derived from WhenDiagnosedCancer and Age)

The variable TimeSinceDX was created with EverHadCancer, WhenDiagnosedCancer and Age variables. The variable TimeSinceDX has 4 categories: Less than 1 Year since DX; 2-5 Years since DX; 6-10 Years since DX; 11+ Years since DX. If the variable EverHadCancer had value of 1 and either Age or WhenDiagnosedCancer had value of -9, “Missing data (Not Ascertained)”, the TimeSinceDX was assigned to -9. If the variable EverHadCancer had value of 1 and WhenDiagnosedCancer is greater than Age, the TimeSinceDX was assigned to -4. If the variable EverHadCancer had value of -9, the TimeSinceDX was assigned to -6. If the variable EverHadCancer had value of 2 and WhenDiagnosedCancer had value of -2, the TimeSinceDX was assigned to -4. If the variable EverHadCancer had value of 2 and WhenDiagnosedCancer had value of -1, the TimeSinceDX was assigned to -1. .

smokeStat: SmokeStat. Smoking Status (Derived from Smoke100 and SmokeNow)

The variable smokeStat was created with Smoke100 and SmokeNow variables. The variable smokeStat has 3 categories: Current; Former; Never. If Smoke100 had value of 1 and SmokeNow had value of -5, “Multiple responses selected in error”, the smokeStat was assigned to -4. If Smoke100 had value of 1 and SmokeNow had value of -9, “Missing data (Not Ascertained)”, the smokeStat was assigned to -9. If Smoke100 had value of -9, “Missing data (Not Ascertained)”, the smokeStat was assigned to -6.

PHQ4: PHQ4. PHQ-4 total score (Derived composite from LittleInterest, Hopeless, Nervous, and Worrying)

The variable PHQ4 was created with LittleInterest, Hopeless, Nervous and Worrying variables. We created total score as continuous variable: 1) Rescore variables 0-3 and then reverse coding such that 'Not at all=0', 'Several Days'=1, 'More than half the days'=2, 'Nearly every day'=3 2) Compute total score by summing across 4 items 3) Total score range will be 0-12. If one of LittleInterest, Hopeless, Nervous and Worrying variables had value of -5, “Multiple response selected in error”, the PHQ was assigned to -5. If one of LittleInterest, Hopeless, Nervous and Worrying variables had value of -9, “Missing data (Not Ascertained)”, the PHQ was assigned to -9.

WeeklyMinutesModerateExercise: WeeklyMinutesModerateExercise. Minutes per week of at least moderate intensity exercise (Derived from TimesModerateExercise, HowLongModerateExerciseHr, and HowLongModerateExerciseMn)

The variable WeeklyMinutesModerateExercise was created with TimesModerateExercise, HowLongModerateExerciseMn and HowLongModerateExerciseHr variables. If TimesModerateExercise is less than 0 then WeeklyMinutesModerateExercise was assigned to -9.

eCigUse:--> eCigUse. Electronic Cigarette Use (UsedECigEver and UseECigNow Recode)

The variable eCigUse was created with ElectCigLessHarm, UsedECigEver and UseECigNow variables. The variable eCigUse has 3 categories: Current; Former; Never. If ElectCigLessHarm = 6 ("I've never heard of electronic cigarettes"), then eCigUse had a value of 3 "Never". If ElectCigLessHarm = -9, but substantive data existed in UsedECigEver and UseECigNow to make an electronic cigarette user determination, then eCigUse was set to either Current, Former, or Never, rather than missing. If UsedECigEver had value of 1 and UseECigNow had value of -5, "Multiple responses selected in error", the eCigUse was assigned to -4. If UsedECigEver had value of 1 and UseECigNow had value of -9, "Missing data (Not Ascertained)", the eCigUse was assigned to -9. If UsedECigEver had value of -9, "Missing data (Not Ascertained)", the eCigUse was assigned to -6.

Label Editing

Labels Added for Standard Recode Variables

Labels were created for the following recoded variables: SEC_RUCA_2010_DESCRIPT, AgeGrpA, AgeGrpB, EducA, EducB, RaceEthn, RaceEthn5, HHInc, BMI, AgeDX, TimeSinceDX, smokeStat, PHQ4, WeeklyMinutesModerateExercise, and eCigUse.

Labels Modified for Certain Variables

Labels were modified for the following variables: NCHSURCODE2013, PR_RUCA_2010, SEC_RUCA_2010, Vegetables, and AccessFamilyMedRec.

Format Editing

Formats Added for Standard Recode Variables

The formats AgeGrpA, AgeGrpB, EducA, EducB, RaceEthn, RaceEthn5f, HHInc, BMI, AgeDX, TimeSinceDX, smokeStat phq4f, WeeklyMinutesModerateExercise and ECigStat were created and assigned to the variables AgeGrpA, AgeGrpB, EducA, EducB, RaceEthn, RaceEthn5, HHInc, BMI, AgeDX, TimeSinceDX, smokeStat, PHQ4, WeeklyMinutesModerateExercise, and eCigUse respectively.

Formats Modified for Certain Variables

All skip patterns in formats were modified (i.e. all instructions to skip questions were deleted). The modified formats are: ACCESSF, ACCESSO, ACTIVED, ADULTSI, BORNINU, CAOTHER, CAREGIF, CAREGIK,

CAREGIL, ELECTCI, EVERHAF, FAMBETW, FREQGOP, GENDERC, HADTEST, HEALTHF, HEARDDN, HEARDHP, HOWLONH, NOTACCF, NOTACCG, OCCUPAF, SEEKHEA, SEXUALF, SMOKE1F, SMOKENO, STRONGG, TABLETH, TIMESMO, UNDERGO, USEDECI, and USEINTE.

The format SEC_RUC was modified to contain the full USDA 2010 Secondary Rural-Urban Commuting Area descriptions to avoid truncation.

Imputation of Income Variable

The income variable (IncomeRanges) has relatively higher percentage (10% for un-weighted percentage or 9% for weighted percentage) of missing values. This variable was imputed via PROC IMPUTE in SUDAAN. The imputation class variables are: Education (O6), RaceEthn (standard recode), RentOrOwn (O16), BornInUSA (O7) and SpeakEnglish (O9). Since the variable SpeakEnglish was asked for people who were born outside USA (BornInUSA=2), the variable SpeakEnglish was declared after BornInUSA in imputation class statement. The copy variables of the imputation class variables and income variable were created, where the missing values were appropriately coded. The copy variables are used for the imputation. The imputed values were saved in a new variable IncomeRanges_IMP.

SAS Code for Data Editing

```
* Reformat HHID from string to numeric to match historic cycles *;

newhhid = input(hhid,8.);
drop hhid;
label newhhid = "Household ID";
rename newhhid = HHID;


* Recode MailHHAdults of 0 to 1 ;

if MailHHAdults=0 then MailHHAdults = 1;


* Recode Height_Feet for Anyone who reports height as 3 feet or less
or as 8 or more;
if 0<=Height_Feet<=3 or Height_Feet>=8 then Height_Feet = -4;

* Recode Height_Inches of -9 and -4 to 0;

if Height_Inches <0 then    Height_Inches = 0;


* Recode Weight less than or equal to 35 to Unreadable or Non-

conforming numeric response;
if 0<=Weight <=35 then  Weight = -4;
```

```
* Recode YearCameToUSA between 1 and 196 to Unreadable or Non-
conforming numeric response          *;
if 1 <= YearCameToUSA <= 196 then   YearCameToUSA = -4;

* Recode SelfAge between 0 and 17 to Unreadable or Non-conforming
numeric response                *;
if 0 <= SelfAge < 18 then  SelfAge = -4;

* Recode HHAdultAge2-HHAdultAge5 between 0 and 17 to Unreadable or
Non-conforming numeric response*;
array a(*) HHAdultAge2-HHAdultAge5;

do I = 1 to dim(a);
      if 0 <= a(I) < 18 then
      a(I) = -4;
end;

* Recode R_HHAdults of missing (.) to -9  ;
if missing(R_HHAdults) = 1 then R_HHAdults = -9;

* Recode Totalhousehold of 0 to 1;
if Totalhousehold=0 then Totalhousehold=1;

* Recode Hours/Minutes of Exercise                                    *;
*if both 0, change both to -9;
if HowLongModerateExerciseHr = 0 & HowLongModerateExerciseMn = 0 then do;
      HowLongModerateExerciseHr = -9;
      HowLongModerateExerciseMn = -9;
end;
*if one -1 and other 0, change 0 to -1;
if HowLongModerateExerciseHr = 0 & HowLongModerateExerciseMn = -1 then
      HowLongModerateExerciseHr = -1;
if HowLongModerateExerciseHr = -1 & HowLongModerateExerciseMn = 0 then
      HowLongModerateExerciseMn = -1;
*if minutes = hours substantively, keep hours;
if HowLongModerateExerciseHr=HowLongModerateExerciseMn/60 then
      HowLongModerateExerciseMn=0;
```

SAS Code for Standard Recode

```
      if sec_ruca_2010 = 1 then sec_ruca_2010_descript = 'Metropolitan area
core: primary flow within an urbanized area (UA), No additional code';
      else if sec_ruca_2010 = 1.1 then sec_ruca_2010_descript = 'Metropolitan
area core: primary flow within an urbanized area (UA), Secondary flow 30% to
50% to a larger UA';
      else if sec_ruca_2010 = 2 then sec_ruca_2010_descript = 'Metropolitan
area high commuting: primary flow 30% or more to a UA, No additional code';
      else if sec_ruca_2010 = 2.1 then sec_ruca_2010_descript = 'Metropolitan
area high commuting: primary flow 30% or more to a UA, Secondary flow 30% to
50% to a larger UA';
      else if sec_ruca_2010 = 3 then sec_ruca_2010_descript = 'Metropolitan
area low commuting: primary flow 10% to 30% to a UA, No additional code';
```

```
      else if sec_ruca_2010 = 4 then sec_ruca_2010_descript = 'Micropolitan
area core: primary flow within an Urban Cluster of 10,000 to 49,999 (large
UC), No additional code';
      else if sec_ruca_2010 = 4.1 then sec_ruca_2010_descript = 'Micropolitan
area core: primary flow within an Urban Cluster of 10,000 to 49,999 (large
UC), Secondary flow 30% to 50% to a UA';
      else if sec_ruca_2010 = 5 then sec_ruca_2010_descript = 'Micropolitan
high commuting: primary flow 30% or more to a large UC, No additional code';
      else if sec_ruca_2010 = 5.1 then sec_ruca_2010_descript = 'Micropolitan
high commuting: primary flow 30% or more to a large UC, Secondary flow 30% to
50% to a UA';
      else if sec_ruca_2010 = 6 then sec_ruca_2010_descript = 'Micropolitan
low commuting: primary flow 10% to 30% to a large UC, No additional code';
      else if sec_ruca_2010 = 7 then sec_ruca_2010_descript = 'Small town
core: primary flow within an Urban Cluster of 2,500 to 9,999 (small UC), No
additional code';
      else if sec_ruca_2010 = 7.1 then sec_ruca_2010_descript = 'Small town
core: primary flow within an Urban Cluster of 2,500 to 9,999 (small UC),
Secondary flow 30% to 50% to a UA';
      else if sec_ruca_2010 = 7.2 then sec_ruca_2010_descript = 'Small town
core: primary flow within an Urban Cluster of 2,500 to 9,999 (small UC),
Secondary flow 30% to 50% to a large UC';
      else if sec_ruca_2010 = 8 then sec_ruca_2010_descript = 'Small town
high commuting: primary flow 30% or more to a small UC, No additional code';
      else if sec_ruca_2010 = 8.1 then sec_ruca_2010_descript = 'Small town
high commuting: primary flow 30% or more to a small UC, Secondary flow 30% to
50% to a UA';
      else if sec_ruca_2010 = 8.2 then sec_ruca_2010_descript = 'Small town
high commuting: primary flow 30% or more to a small UC, Secondary flow 30% to
50% to a large UC';
      else if sec_ruca_2010 = 9 then sec_ruca_2010_descript = 'Small town low
commuting: primary flow 10% to 30% to a small UC, No additional code';
      else if sec_ruca_2010 = 10 then sec_ruca_2010_descript = 'Rural areas:
primary flow to a tract outside a UA or UC, No additional code';
      else if sec_ruca_2010 = 10.1 then sec_ruca_2010_descript = 'Rural
areas: primary flow to a tract outside a UA or UC, Secondary flow 30% to 50%
to a UA';
      else if sec_ruca_2010 = 10.2 then sec_ruca_2010_descript = 'Rural
areas: primary flow to a tract outside a UA or UC, Secondary flow 30% to 50%
to a large UC';
      else if sec_ruca_2010 = 10.3 then sec_ruca_2010_descript = 'Rural
areas: primary flow to a tract outside a UA or UC, Secondary flow 30% to 50%
to a small UC';
      label sec_ruca_2010_descript = "USDA 2010 Secondary Rural-Urban
Commuting Area Codes Description (variable found in Stata dataset, only)";


     if 18 <= Age <= 34 then
           AgeGrpA = 1;
      else if 35 <= Age <= 39 then
           AgeGrpA = 2;
      else if 40 <= Age <= 44 then
           AgeGrpA = 3;
      else if 45 <= Age then
           AgeGrpA = 4;
      else if Age in (-1,-4, -9) then
```

```
            AgeGrpA = Age;
      label AgeGrpA = 'AgeGrpA. 4 Level Age Categories Version A (Derived
from Age; see History Document for more information)';
      if 18 <= Age <= 34 then
            AgeGrpB = 1;
      else if 35 <= Age <= 49 then
            AgeGrpB = 2;
      else if 50 <= Age <= 64 then
            AgeGrpB = 3;
      else if 65 <= Age <= 74 then
            AgeGrpB = 4;
      else if 75 <= Age then
            AgeGrpB = 5;
      else if Age in (-1,-4, -9) then
            AgeGrpB = Age;
      label AgeGrpB = 'AgeGrpB. 5 Level Age Categories Version B (Derived
from Age; see History Document for more information)';

      if Education in (1, 2) then
            EducA = 1;
      else if Education in (3) then
            EducA = 2;
      else if Education in (4, 5) then
            EducA = 3;
      else if Education in (6, 7) then
            EducA = 4;
      else if Education in (-1,-4,-9) then
            EducA = Education;
      label EducA = 'EducA.  What is the highest level of school you
completed? 4 Levels (Derived from Education; see History Document for more
information)';

      if Education in (1, 2) then
            EducB = 1;
      else if Education in (3) then
            EducB = 2;
      else if Education in (4, 5) then
            EducB = 3;
      else if Education in (6) then
            EducB = 4;
      else if Education in (7) then
            EducB = 5;
      else if Education in (-1,-4,-9) then
            EducB = Education;
      label EducB = 'EducB.  What is the highest level of school you
completed? 5 Levels (Derived from Education; see History Document for more
information)';

      if Hisp_Cat in (21, 22, 23, 24, 25) then
            RaceEthn = 1;
      else if Hisp_Cat in (10) then
            do;
```

```
                  if Race_Cat2 in (11) then
                        RaceEthn = 2;
                  else if Race_Cat2 in (12) then
                        RaceEthn = 3;
                  else if Race_Cat2 in (14) then
                        RaceEthn = 4;
                  else if Race_Cat2 in (31, 32, 33, 34, 35, 36, 37) then
                        RaceEthn = 5;
                  else if Race_Cat2 in (51, 52, 53, 54) then
                        RaceEthn = 6;
                  else if Race_Cat2 in (16) then
                        RaceEthn = 7;
                  else if Race_Cat2 in (-1,-4,-9) then
                        RaceEthn = -9;
            end;
      else if Hisp_Cat in (-1,-4,-9) then
            do;
                  RaceEthn = -9;
            end;
      label RaceEthn = 'Race/Ethnicity. 7 Levels (Derived from Hisp_Cat and
Race_Cat2; see History Document for more information)';


      if Hisp_Cat in (21, 22, 23, 24, 25) then RaceEthn5 = 3;
      else if Hisp_Cat in (10) then do;
      if Race_Cat2 in (11) then RaceEthn5 = 1;
      else if Race_Cat2 in (12) then RaceEthn5 = 2;
      else if Race_Cat2 in (31, 32, 33, 34, 35, 36, 37) then RaceEthn5
      = 4;
      else if Race_Cat2 in (51, 52, 53, 54,14,16) then RaceEthn5 = 5;
      else if Race_Cat2 in (-1,-4,-9) then RaceEthn5 = -9;
      end;
      else if Hisp_Cat in (-1,-4,-9) then   do;
      RaceEthn5 = -9;
      end;
      label RaceEthn5 = 'Race/Ethnicity. 5 Levels (Derived from Hisp_Cat and
Race_Cat2; see History Document for more information)';

      if IncomeRanges in (1, 2, 3) then
            HHInc = 1;
      else if IncomeRanges in (4) then
            HHInc = 2;
      else if IncomeRanges in (5) then
            HHInc = 3;
      else if IncomeRanges in (6) then
            HHInc = 4;
      else if IncomeRanges in (7, 8, 9) then
            HHInc = 5;
      else if IncomeRanges in (-1,-4,-9) then
            HHInc = IncomeRanges;
```

```
      label HHInc = 'HHInc.  What is your {combined} annual household income?
5 Levels (Derived from IncomeRanges; see History Document for more
information)';


      if Height_Feet >= 0 and Height_Inches >= 0 and Weight > 0 then
            BMI = (Weight * 703) / ((Height_Feet * 12 +
Height_Inches)**2);
      else if (Height_Feet in (-1, -9) and Weight >=-1 ) or
(Height_Feet >=-1 and Weight in (-1, -9)) or (Height_Feet = -9 and
Weight = -9) then
            BMI = -9;
      else if Height_Feet = -4 or Weight = -4 then
            BMI = -4;
      label BMI = 'BMI.  Body Mass Index (Weight*703)/(Height in inches**2)
(See History Document for more information)';
      if BMI not in (-4, -9) then
            BMI = round(BMI, 0.1);

      AgeDX = WhenDiagnosedCancer;
      AgeDX = 'AgeDX. At what age were you diagnosed with cancer? (Derived
from WhenDiagnosedCancer; see History Document for more information)';

      if EverHadCancer in (1) then
            do;
                  if Age < 0 then
                        TimeSinceDX = Age;
                  else if WhenDiagnosedCancer <0 and Age >= 18 then
                        TimeSinceDX = WhenDiagnosedCancer;
                  else if WhenDiagnosedCancer >= 0 and Age >= 18 then
                        do;
                              if 0 <= (Age - WhenDiagnosedCancer) <= 1
then
                                    TimeSinceDX = 1;
                              else if 2 <= (Age - WhenDiagnosedCancer) <=
5 then
                                    TimeSinceDX = 2;
                              else if 6 <= (Age - WhenDiagnosedCancer) <=
10 then
                                    TimeSinceDX = 3;
                              else if 11 <= (Age - WhenDiagnosedCancer)
then
                                    TimeSinceDX = 4;
                              else if (Age - WhenDiagnosedCancer) < 0
then
                                    TimeSinceDX = -4;
                        end;
            end;
      else if EverHadCancer in (-9) then
            TimeSinceDX = WhenDiagnosedCancer;
      else if EverHadCancer in (2) then
            do;
```

```
                if WhenDiagnosedCancer in (-1) then
                      TimeSinceDX = WhenDiagnosedCancer;
                else if WhenDiagnosedCancer in (-2) then
                      TimeSinceDX = -4;
          end;
      label TimeSinceDX = 'TimeSinceDX. How long ago were you diagnosed with
cancer? (Derived from WhenDiagnosedCancer and Age; see History Document for
more information)';

      if Smoke100 in (1) then
          do;
                if SmokeNow in (1, 2) then
                      smokeStat = 1;
                else if SmokeNow in (3) then
                      smokeStat = 2;
                else if SmokeNow in (-5) then
                      smokeStat = -4;
                else if SmokeNow in (-9) then
                      smokeStat = -9;
          end;
      else if Smoke100 in (2) then
          do;
                smokeStat = 3;
          end;
      else if Smoke100 in (-9) then
          smokeStat = -6;
      label smokeStat = 'SmokeStat.  Smoking Status (Derived from Smoke100
and SmokeNow; see History Document for more information)';


array b(*) LittleInterest Hopeless Nervous Worrying;
PHQ4= 0;
if b(1) = -5 or b(2) = -5 or b(3) = -5 or b(4) = -5 then PHQ4 =-5;
else do;
if b(1) = -9 or b(2) = -9 or b(3) = -9 or b(4) = -9 then PHQ4 = -9;
else do;
do I = 1 to dim(b);
if PHQ4 not in (-5, -9) and b(I) in (1, 2,3 ,4) then
PHQ4 = PHQ4 + (4-b(I));
end;
end;
end;
label PHQ4 = 'PHQ4. PHQ-4 total score (Derived composite from LittleInterest,
Hopeless, Nervous, and Worrying; see History Document for more information)';
drop I;

If TimesModerateExercise=0 then WeeklyMinutesModerateExercise=0;
Else If TimesModerateExercise<0 then WeeklyMinutesModerateExercise=-9;
else IF TimesModerateExercise>0 then do;
if HowLongModerateExerciseMn < 0 then HowLongModerateExerciseMn=0;
if HowLongModerateExerciseHr < 0 then HowLongModerateExerciseHr=0;
```

```
WeeklyMinutesModerateExercise=HowLongModerateExerciseHr*60*TimesModera
teExercise + HowLongModerateExerciseMn*TimesModerateExercise;
end;
label WeeklyMinutesModerateExercise="WeeklyMinutesModerateExercise. Minutes
per week of at least moderate intensity exercise (Derived from
TimesModerateExercise, HowLongModerateExerciseHr, and
HowLongModerateExerciseMn; see History Document for more information)";

if ElectCigLessHarm = 6 then eCigUse = 3;
else do;
      if UsedECigEver in (1) then
            do;
                  if UseECigNow in (1, 2) then
                        eCigUse = 1;
                  else if UseECigNow in (3) then
                        eCigUse = 2;
                  else if UseECigNow in (-5) then
                        eCigUse = -4;
                  else if UseECigNow in (-9) then
                        eCigUse = -9;
            end;
      else if UsedECigEver in (2) then
            do;
                  eCigUse = 3;
            end;
      else if UsedECigEver in (-9) then
            eCigUse = -6;
end;
label eCigUse = 'eCigUse.  Electronic Cigarette Use (Derived from
ElectCigLessHarm, UsedECigEver, and UseECigNow; see History Document for more
information)';
```

SAS Code for Format Editing

SAS Code for Formats Added for Standard Recode Variables

```
value AgeGrpA
1 = '18-34'
2 = '35-39'
3 = '40-44'
4 = '45+'
-4 = 'Unreadable or Nonconforming Numeric Response'
-9 = 'Missing Data (Not Ascertained)'
;

value AgeGrpB
1 = '18-34'
2 = '35-49'
3 = '50-64'
4 = '65-74'
5 = '75+'
-4 = 'Unreadable or Nonconforming Numeric Response'
-9 = 'Missing Data (Not Ascertained)'
```

```
;

value EducA
1 = 'Less than High School'
2 = 'High School Graduate'
3 = 'Some College'
4 = 'College Graduate or More'
-9 = 'Missing Data (Not Ascertained)'
;

value EducB
1 = 'Less than High School'
2 = 'High School Graduate'
3 = 'Some College'
4 = "Bachelor's Degree"
5 = 'Post-Baccalaureate Degree'
-9 = 'Missing Data (Not Ascertained)'
;

value RaceEthn
1 = 'Hispanic'
2 = 'Non-Hispanic White'
3 = 'Non-Hispanic Black or African American'
4 = 'Non-Hispanic American Indian or Alaska Native'
5 = 'Non-Hispanic Asian'
6 = 'Non-Hispanic Native Hawaiian or other Pacific Islander'
7 = 'Non-Hispanic Multiple Races Mentioned'
-4 = 'Unreadable or Nonconforming Numeric Response'
-9 = 'Missing Data (Not Ascertained)'
;

value RaceEthn5f
1 = "Non-Hispanic White"
2= "Non-Hispanic Black or African American"
3="Hispanic"
4="Non-Hispanic Asian"
5="Non-Hispanic Other"
-9="Missing Data--Not Ascertained"
;

value HHInc
1 = 'Less than $20,000'
2 = '$20,000 to < $35,000'
3 = '$35,000 to < $50,000'
4 = '$50,000 to < $75,000'
5 = '$75,000 or More'
-5 = 'Multiple Responses Selected in Error'
-9 = 'Missing Data (Not Ascertained)'
;
```

```
value BMI
-4 = 'Unreadable or Nonconforming Numeric Response'
-9 = 'Missing Data (Not Ascertained)'
;

value AgeDX
-1 = 'Inapplicable, coded 2 in EverHadCancer'
-2 = 'Question Answered in Error (Commission Error)'
-6 = 'Missing Data (Filter Missing)'
-9 = 'Missing Data (Not Ascertained)'
;

value TimeSinceDX
1 = 'Less than 1 Yr Since DX'
2 = '2-5 Yrs Since DX'
3 = '6-10 Yrs Since DX'
4 = '11+ Yrs Since DX'
-1 = 'Inapplicable, coded 2 in EverHadCancer'
-4 = 'Unreadable or Nonconforming Numeric Response'
-6 = 'Missing Data (Filter Missing), coded -9 in EverHadCancer'
-9 = 'Missing Data (Not Ascertained)'
;

value smokeStat
1 = 'Current'
2 = 'Former'
3 = 'Never'
-4 = 'Unreadable or Nonconforming Numeric Response'
-6 = 'Missing Data (Filter Missing), coded -9 in Smoke100'
-9 = 'Missing Data (Not Ascertained)'
;

value phq4f
-5 = 'Multiple Responses Selected in Error'
-9 = 'Missing Data (Not Ascertained)'
;

value WeeklyMinutesModerateExercise
-5 = 'Multiple Responses Selected in Error'
-9 = 'Missing Data (Not Ascertained)'
;

value ecigStat
1 = 'Current'
2 = 'Former'
3 = 'Never'
-4 = 'Unreadable or Nonconforming Numeric Response'
-6 = 'Missing Data (Filter Missing), coded -9 in UsedECigEver'
-9 = 'Missing Data (Not Ascertained)'
;
```

```
      format      AgeGrpA AgeGrpA.
                  AgeGrpB AgeGrpB.
                  EducA EducA.
                  EducB EducB.
                  RaceEthn RaceEthn.
                  RaceEthn5 RaceEthn5f.
                  HHInc HHInc.
                  BMI BMI.
                  AgeDX AgeDX.
                  TimeSinceDX TimeSinceDX.
                  smokeStat smokeStat.
                  phq4 phq4f.
                  WeeklyMinutesModerateExercise
WeeklyMinutesModerateExercise.
                  ECigUse ECigStat.
                 IncomeRanges_IMP  incomer.
```

SAS Code for Imputation of Income Variable

```
* Impute IncomeRanges via PROC HOTDECK
*;
data HINTS5CYCLE1;
      set HINTS5CYCLE1;

      COPY_Education = Education;
      if COPY_Education in (-9) then
            COPY_Education = .;

      COPY_RaceEthn = RaceEthn;
      if COPY_RaceEthn in (-9) then
            COPY_RaceEthn = .;

      COPY_RentOrOwn = RentOrOwn;
      if COPY_RentOrOwn in (-5, -9) then
            COPY_RentOrOwn = .;

      COPY_SpeakEnglish = SpeakEnglish;
      if COPY_SpeakEnglish in (-1, -2, -5, -6, -9) then
            COPY_SpeakEnglish = .;

      COPY_BornInUSA = BornInUSA;
      if COPY_BornInUSA in (-9) then
            COPY_BornInUSA = .;

      COPY_IncomeRanges = IncomeRanges;
      if COPY_IncomeRanges in (-9) then
            COPY_IncomeRanges = .;
```

```
      ID = _N_;

      format COPY_Education Educati. COPY_RaceEthn RaceEthn.
COPY_RentOrOwn RentOrO.
                  COPY_SpeakEnglish SpeakEn. COPY_BornInUSA BornInU.;
run;

proc freq data= HINTS5CYCLE1;
      tables COPY_Education*Education / list missing;
      tables COPY_RaceEthn*RaceEthn / list missing;
      tables COPY_RentOrOwn*RentOrOwn / list missing;
      tables COPY_SpeakEnglish*SpeakEnglish / list missing;
      tables COPY_BornInUSA*BornInUSA / list missing;
      tables COPY_IncomeRanges*IncomeRanges / list missing;
run;

proc impute data= HINTS5CYCLE1method=wshd notsorted;
      weight person_finwt0;
      impvar COPY_IncomeRanges;
      impby COPY_Education COPY_RaceEthn COPY_RentOrOwn COPY_BornInUSA
COPY_SpeakEnglish;
      impname COPY_IncomeRanges="IncomeRanges_IMP";
      impid ID;
      output IMPID IMPBY IMPUTEVAL / filename=impute1 replace;
run;

proc freq data=impute1;
      tables IncomeRanges_IMP / missing;
run;

proc contents data=impute1;
run;

proc sort data= HINTS5CYCLE1;
      by ID;
run;

proc sort data=impute1 (keep=ID IncomeRanges_IMP);
      by ID;
run;

data HINTS5CYCLE1;
      merge HINTS5CYCLE1 (in=A) impute1 (in=B);
      by ID;

      if A = 1 and B = 1;
run;

data _null_;
      set HINTS5CYCLE1;
```

```
      if IncomeRanges not in (-9) and COPY_IncomeRanges ^=
IncomeRanges_IMP then
            put ID IncomeRanges COPY_IncomeRanges IncomeRanges_IMP;
run;

data HINTS5CYCLE1;
      set HINTS5CYCLE1;

      if missing(IncomeRanges_IMP) = 1 then
            IncomeRanges_IMP = IncomeRanges;
      label IncomeRanges_IMP = '-->IncomeRanges_IMP.  Imputed
IncomeRanges variable via PROC HOTDECK in SUDAAN';
      format IncomeRanges_IMP IncomeR.;

      drop COPY_Education COPY_RaceEthn COPY_RentOrOwn
COPY_SpeakEnglish COPY_BornInUSA
            ID
            COPY_IncomeRanges;

      format MultiOcc;
run;
```