HINTS 5 Cycle 3 History Document

January 2020

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.

TimesSunburned: J3. During the past 12 months, how many times have you had a sunburn (even a small part of your skin turns red or hurts for 12 hours or more) from too much sun exposure?
Two respondents had outlying values of 260 and 367, which were replaced by -4, "Unreadable or Non-conforming numeric response".

HowLongModerateExerciseMinutes: H2. On the days that you do any physical activity or exercise of at least moderate physical intensity, how long do you typically do these activities?
One respondent had an outlying value of 999, which was replaced by -4, "Unreadable or Non-conforming numeric response".

Height_Inches: F7 About how tall are you without shoes? Inches:
Two-hundred and eighteen respondents had inches of -9, "Missing data (Not Ascertained)", -7, "Missing data (Web partial- Question never seen), 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
Fifteen respondents had missing values (.), which were replaced by -9.

Totalhousehold: O8. Including yourself, how many people live in your household?
One-hundred and one respondents had Totalhousehold of 0, which were replaced by 1. One respondent had an outlying value of 60, which was replaced by -4, "Unreadable or Non-conforming numeric response".

DrinksPerDay: G8. During the past 30 days, on the days when you drank, about how many drinks did you drink on average?
Two respondents had outlying values of 61 and 96, which were replaced by -4.

Caregiving_HoursPerWeek2: E3. Think about the individual for whom you are currently providing the most care. About how many hours per week do you spend in an average week providing care?

Thirty-eight respondents had an invalid value of 0 for Caregiving_HoursPerWeek2, which was replaced with -4. Values of 0 for Caregiving_HoursPerWeek2 are invalid, because, based on questionnaire skip patterns, only respondents who indicated they were current caregivers received this question.

AverageTimeSitting: H4. During the past 7 days, how much time did you spend sitting on a typical day at home or at work? This may include time spent sitting at a desk, visiting friends, reading, driving or riding in a car, or sitting or lying down to watch television.

Forty respondents had an outlying value of 21 hours or greater for AverageTimeSitting, which was replaced by -4.

The following variables were determined to be different in construct in HINTS 5 Cycle 3 than in previous cycles. These variables were renamed.

ChanceGetCancer: N1. How likely are you to get cancer in your lifetime?

The ChanceGetCancer variable was renamed to ChanceGetCancerNoDX to indicate that, based on the questionnaire skip pattern, it is applicable to those respondents who indicated they have never been diagnosed with cancer.

FreqWorryCancer: N2. How worried are you about getting cancer?

The FreqWorryCancer variable was renamed to FreqWorryCancerNoDx to indicate that, based on the questionnaire skip pattern, it is applicable to those respondents who indicated they have never been diagnosed with cancer.

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 variable 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 variable was assigned with value of -9. If Hisp_Cat had value of 10, "Not Hispanic", and Race_Cat2 had value of -7, "Missing data (Web partial - Question Never Seen)", the RaceEthn variable was assigned with value of -7. The RaceEthn variable was assigned with value of -7 if Hisp_Cat=-7 and -9 if Hisp_Cat=-9.

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

The RaceEthn5 variable 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 -7, "Missing data (Web partial - Question Never Seen)", RaceEthn5 was assigned with value of -7. 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 -7 if Hisp_Cat=-7, and -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 height in feet or weight had value of -9, "Missing data (Not Ascertained)" but neither had value of -4, "Unreadable or Non-conforming numeric response" or -7, "Missing data (Web partial- Question never seen)", the BMI was assigned to -9. If height in feet or weight had value of -4, "Unreadable or Non-conforming numeric response", the BMI was assigned to -4. If height in feet or weight had value of -7, "Missing data (Web Partial- Question never seen)", the BMI was assigned to -7.

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 -9. If the variable EverHadCancer had value of -7, "Missing data (Web partial- Question never seen)", TimeSinceDX was assigned to -7. 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. If the variable EverHadCancer had value of 2 and WhenDiagnosedCancer had value of -7, "Missing data (Web partial- Question never seen)", TimeSinceDX was assigned to -7.

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 -7, "Missing data (Web partial- Question never seen)", smokeStat was assigned to -7. 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. 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 -7 "Missing data (Web partial- Question never seen)", the PHQ was assigned to -7. 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 and HowLongModerateExerciseMinutes)

The variable WeeklyMinutesModerateExercise was created with TimesModerateExercise and HowLongModerateExerciseMinutes variables. If TimesModerateExercise is 0 then WeeklyMinutesModerateExercise was assigned to 0. If TimesModerateExercise is less than 0 then WeeklyMinutesModerateExercise was assigned to the value of TimesModerateExercise (i.e., the original

negative values were retained). If HowLongModerateExerciseMinutes is less than 0 then WeeklyMinutesModerateExercise was assigned to the value of HowLongModerateExerciseMinutes (i.e., the original negative values were retained).

eCigUse: eCigUse. Electronic Cigarette Use (Derived from UsedECigEver and UseECigNow Recode)

The variable eCigUse was created with UsedECigEver and UseECigNow variables. The variable eCigUse has 3 categories: Current; Former; Never. 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 -7, "Missing data (Web Partial- Question Never Seen)", eCigUse was assigned to -7. 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.

AvgDrinksPerWeek: AvgDrinksPerWeek. Average number of drinks per week (Derived from DrinkDaysPerWeek and DrinksPerDay)

The variable AvgDrinksPerWeek was created with DrinkDaysPerWeek and DrinksPerDay variables. If DrinkDaysPerWeek is less than 0 then AvgDrinksPerWeek was assigned to the value of DrinkDaysPerWeek (i.e., the original negative values were retained). If DrinksPerDay is less than 0 then AvgDrinksPerWeek was assigned to the value of DrinksPerDay (i.e., the original negative values were retained). Any respondents that had outlying values of 140 or more were recoded to -4.

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, eCigUse, AvgDrinksPerWeek.

Labels Modified for Certain Variables

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

Format Editing

Formats Added for Standard Recode Variables

The formats AgeGrpA, AgeGrpB, EducA, EducB, RaceEthn, RaceEthn5f, HHInc, BMI, AgeDX, TimeSinceDX, smokeStat, phq4f, WeeklyMinutesModerateExercise, ECigStat, AvgDrinksPerWeek, and Treatment_H5C3f were created and assigned to the variables AgeGrpA, AgeGrpB, EducA, EducB, RaceEthn, RaceEthn5, HHInc, BMI, AgeDX, TimeSinceDX, smokeStat, PHQ4, WeeklyMinutesModerateExercise, eCigUse, AvgDrinksPerWeek, and Treatment_H5C3, 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: ACCESSO, ADULTSI, CAOTHER, CAREGII, CAREGIM, CAREGIN, DRINKDA, EVERHAD, EVERHADC, EVERHAF, FREQGOP, GENDERC, GENDERC_, HAVEDEF, HAVEDEG, HEALTHF, HEARDHP, NOTICEC, SEEKHEA, SEENFED, SEXUALF, SMOKE1F, SMOKENO, STRONGF, TIMESMO, TIMESSU, USEDECI, USEINTE, and WEARABL.

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

The format for FormType was updated to more accurately indicate if a survey was completed by paper or Web.

Imputation of Income Variable

The income variable (IncomeRanges) has relatively higher percentage (12% for un-weighted percentage or 10% for weighted percentage) of missing values. This variable was imputed via PROC IMPUTE in SUDAAN. The imputation class variables are: Education (O3), RaceEthn (standard recode), RentOrOwn (O11), and SpeakEnglish (O4). 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

```
* 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 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 TotalHousehold 60 to missing;
if TotalHousehold=60 then TotalHousehold=-4;

*Recode TimesSunburned 260 and 367 to missing;
    if TimesSunburned in (260, 367) then
        TimesSunburned=-4;

*Recode HowLongExerciseMinutes 999 to missing;
    if HowLongModerateExerciseMinutes=999 then
        HowLongModerateExerciseMinutes=-4;

* Recode invalid 0 response to Caregiving_HoursPerWeek2          *;
    if Caregiving_HoursPerWeek2 = 0 then Caregiving_HoursPerWeek2 = -
4;

* Recode outliers of AverageTimeSitting (>20 hours)  *;
    if AverageTimeSitting > 20 then AverageTimeSitting = -4;

* Recode outliers of DrinksPerDay                    *;
    if DrinksPerDay in (61, 96) then DrinksPerDay=-4;


* Rename FreqWorryCancer to FreqWorryCancerNoDx to break trend (now
that it is only asked among
    those who never had a cancer diagnosis)                *;
    rename FreqWorryCancer = FreqWorryCancerNoDx;

*    Rename ChanceGetCancer to ChanceGetCancerNoDX to break trend (now
that it is only asked among those who never had a cancer diagnosis)
    *;
    rename ChanceGetCancer = ChanceGetCancerNoDX;
```

SAS Code for Standard Recode

```
    if sec_ruca_2010 = 1 then sec_ruca_2010_descript = "Metro area
core: primary flow w/in urbanized area (UA), No additional code";
    else if sec_ruca_2010 = 1.1 then sec_ruca_2010_descript = "Metro
area core: primary flow w/in an UA, Secondary flow 30%-%50=50 to
larger UA";
```

```
     else if sec_ruca_2010 = 2 then sec_ruca_2010_descript = "Metro
area high commuting: primary flow 30%+ to a UA, No additional code";
     else if sec_ruca_2010 = 2.1 then sec_ruca_2010_descript = "Metro
area high commuting: primary flow 30%+ to a UA, Secondary flow 30-50%
to larger UA";
     else if sec_ruca_2010 = 3 then sec_ruca_2010_descript = "Metro
area low commuting: primary flow 10-30% to a UA, No additional code";
     else if sec_ruca_2010 = 4 then sec_ruca_2010_descript = "Micropol
area: prim flow w/in Urban Cluster of 10,000-49,999, No additional
code";
     else if sec_ruca_2010 = 4.1 then sec_ruca_2010_descript =
"Micropol area: prim flow w/in Urban Cluster of 10,000-49,999, Sec
flow 30-50% to UA";
     else if sec_ruca_2010 = 5 then sec_ruca_2010_descript = "Micropol
high commuting: prim flow 30%+ to a large UC, No additional code";
     else if sec_ruca_2010 = 5.1 then sec_ruca_2010_descript =
"Micropol high commuting: prim flow 30%+ to large UC, Secondary flow
30-50% to UA";
     else if sec_ruca_2010 = 6 then sec_ruca_2010_descript = "Micropol
low commuting: primary flow 10%-30% to a large UC, No additional
code";
     else if sec_ruca_2010 = 7 then sec_ruca_2010_descript = "Small
town: prim flow w/in Urban Cluster 2,500-9,999(small UC)/No additional
code";
     else if sec_ruca_2010 = 7.1 then sec_ruca_2010_descript = "Small
town: prim flow w/in Urban Cluster 2,500-9,999(small UC)/Sec flow 30-
50% to UA";
     else if sec_ruca_2010 = 7.2 then sec_ruca_2010_descript = "Small
town: prim flow w/in Urban Cluster 2,500-9,999(small UC)/Sec flow 30-
50% to large UC";
     else if sec_ruca_2010 = 8 then sec_ruca_2010_descript = "Small
town/high commuting: prim flow 30%+ to small UC, No additional code";
     else if sec_ruca_2010 = 8.1 then sec_ruca_2010_descript = "Small
town/high commuting: prim flow 30%+ to small UC, Secondary flow 30-50%
to UA";
     else if sec_ruca_2010 = 8.2 then sec_ruca_2010_descript = "Small
town/high commuting: prim flow 30%+ to small UC, Secondary flow 30-50%
to large UC";
     else if sec_ruca_2010 = 9 then sec_ruca_2010_descript = "Small
town/low commuting: prim flow 10-30% to 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 tract outside a UA/UC, Secondary flow 30-50% to
a UA";
     else if sec_ruca_2010 = 10.2 then sec_ruca_2010_descript = "Rural
areas: primary flow to tract outside a UA/UC, Secondary flow 30-50% to
a large UC";
     else if sec_ruca_2010 = 10.3 then sec_ruca_2010_descript = "Rural
areas: primary flow to tract outside a UA/UC, Secondary flow 30-50% to
a small UC";
```

```
      label sec_ruca_2010_descript = "USDA 2010 Secondary Rural-Urban
Commuting Area Codes Description (variable found in Stata dataset,
only)";


*SAS code for standard recode;
      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 (-9,-4) 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;
      if 35<=Age<=49 then AgeGrpB=2;
      if 50<=Age<=64 then AgeGrpB=3;
      if 65<=Age<=74 then AgeGrpB=4;
      if 75<=Age then AgeGrpB=5;
      else if Age in (-9,-4) 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 (-9, -7) then
            EducA = Education;
      label EducA = 'EducA: What is the highest level of school you
completed? Education recoded-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 (-9, -7) then
```

```
        EducB = Education;
    label EducB = 'EducB: What is the highest level of school you
completed? Education recoded-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= -7 then
                RaceEthn = -7;
            else if Race_Cat2= -9 then
                RaceEthn = -9;
        end;
    else if Hisp_Cat= -7 then
        RaceEthn = -7;
    else if Hisp_Cat= -9 then
            RaceEthn = -9;
    label RaceEthn = 'Race/Ethnicity recode (Hisp_Cat and Race_Cat2--
7 Levels)';


    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= (-7) then RaceEthn5 = -7;
    else if Race_Cat2= (-9) then RaceEthn5 = -9;
    end;
    if Hisp_cat= -7 then RaceEthn5=-7;
    if Hisp_Cat = -9 then RaceEthn5 = -9;
    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
```

```sas
            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 (-9, -7) then
            HHInc = IncomeRanges;
        label HHInc = 'HHInc: What is your (combined) annual household
income? (IncomeRanges Recode-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 (-9) and Weight >=-1 ) or (Height_Feet
>=-1 and Weight in (-9)) or (Height_Feet = -9 and Weight = -9) then
            BMI = -9;
        else if Height_Feet = -4 or Weight = -4 then
            BMI = -4;
        else if Height_Feet = -7 or Weight = -7 then
            BMI = -7;
        label BMI = 'BMI.  Body Mass Index (Weight*703)/(Height in
inches**2) (See History Document for more information)';
        if BMI not in (-4, -7, -9) then
            BMI = round(BMI, 0.1);


        AgeDX = WhenDiagnosedCancer;
        label 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 (-7) then
                TimeSinceDX = WhenDiagnosedCancer;
        else if EverHadCancer in (-9) then
                TimeSinceDX = WhenDiagnosedCancer;

        else if EverHadCancer in (2) then
                do;
                    if WhenDiagnosedCancer in (-1) then
                            TimeSinceDX = WhenDiagnosedCancer;
                    if WhenDiagnosedCancer in (-7) 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 (-7) then
                            smokeStat = -7;
                    else if SmokeNow in (-9) then
                            smokeStat = -9;
                end;
        else if Smoke100 in (2) then
                    smokeStat = 3;
        else if Smoke100 in (-7) then
                smokeStat = -7;
        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 if b(1) in (-7) or b(2) in (-7) or b(3) in (-7) or b(4) in  (-7)
then PHQ4 = -7;
else if b(1) in (-9) or b(2) in (-9) or b(3) in (-9) or b(4) in  (-9)
then PHQ4 = -9;
else do;
     do I = 1 to dim(b);
            if PHQ4 not in (-5, -7, -9) and b(I) in (1, 2,3 ,4) then
            PHQ4 = PHQ4 + (4-b(I));
     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=TimesModerateExercise;
else If HowLongModerateExerciseMinutes<0 then
WeeklyMinutesModerateExercise=HowLongModerateExerciseMinutes;
else IF TimesModerateExercise>0 then do;
     if HowLongModerateExerciseMinutes = 0 then
WeeklyMinutesModerateExercise = 0;
     else WeeklyMinutesModerateExercise =
HowLongModerateExerciseMinutes*TimesModerateExercise;
end;
label WeeklyMinutesModerateExercise="WeeklyMinutesModerateExercise.
Minutes per week of at least moderate intensity exercise (Derived from
TimesModerateExercise and HowLongModerateExerciseMinutes; see History
Document for more information)";


if UsedECigEver in (1) then
     do;
            if UseECigNow in (1, 2) then
                 eCigUse = 1;
            else if UseECigNow = 3 then
                 eCigUse = 2;
            else if UseECigNow = (-5) then
                 eCigUse = -4;
            else if UseECigNow in (-7) then
                 eCigUse = -7;
            else if UseECigNow in (-9) then
                 eCigUse = -9;
     end;
else if UsedECigEver in (2) then
            eCigUse = 3;
```

```
else if UsedECigEver in (-7) then
     eCigUse = -7;
else if UsedECigEver in (-9) then
     eCigUse = -6;
label eCigUse = 'eCigUse.  Electronic Cigarette Use (Derived from
UsedECigEver and UseECigNow; see History Document for more
information)';


if DrinkDaysPerWeek=0 then AvgDrinksPerWeek=0;
else if DrinkDaysPerWeek < 0 then AvgDrinksPerWeek=DrinkDaysPerWeek;
else if DrinksPerDay < 0 then AvgDrinksPerWeek=DrinksPerDay;
else if DrinkDaysPerWeek>0 then do;
     if DrinksPerDay=0 then AvgDrinksPerWeek = 0;
     else AvgDrinksPerWeek = DrinkDaysPerWeek*DrinksPerDay;
     if AvgDrinksPerWeek >= 140 then AvgDrinksPerWeek=-4;
end;
label AvgDrinksPerWeek = "AvgDrinksPerWeek. Average number of drinks
per week (Derived from DrinkDaysPerWeek and DrinksPerDay; 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'
                    -7 = "Missing data (Web partial - Question Never
Seen)"
                    -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'
            -7 = "Missing data (Web partial - Question Never
Seen)"
            -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'
            -7 = "Missing data (Web partial - Question Never
Seen)"
            -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"
            -7 = "Missing data (Web partial - Question Never
Seen)"
            -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'
            -7 = "Missing data (Web partial - Question Never
Seen)"
            -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)'
    -4 = 'Unreadable or Non-conforming numeric response'
    -6 = 'Missing Data (Filter Missing)'
    -7 = "Missing data (Web partial - Question Never
Seen)"
    -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'
    -7 = "Missing data (Web partial - Question Never
Seen)"
    -9 = 'Missing Data (Not Ascertained)'
    ;

value smokeStat
    1 = 'Current'
    2 = 'Former'
    3 = 'Never'
    -6 = 'Missing Data (Filter Missing), coded -9 in
Smoke100'
    -7 = "Missing data (Web partial - Question Never
Seen)"
    -9 = 'Missing Data (Not Ascertained)'
    ;

value phq4f
    -5 = 'Multiple Responses Selected in Error'
    -7 = "Missing data (Web partial - Question Never
Seen)"
    -9 = 'Missing Data (Not Ascertained)'
    ;

value eciguse
    1 = 'Current'
    2 = 'Former'
    3 = 'Never'
```

```
                  -6 = 'Missing Data (Filter Missing), coded -9 in
UsedECigEver'
                  -7 = "Missing data (Web partial - Question Never
Seen)"
                  -9 = 'Missing Data (Not Ascertained)';
                  ;

          value AvgDrinksPerWeek
                  -4 = 'Unreadable or Nonconforming Numeric Response'
                  -5 = 'Multiple Responses Selected in Error'
                  -7 = "Missing data (Web partial - Question Never
Seen)"
                  -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    .
                  ECigUse.
                  IncomeRanges_IMP incomer.
                  AvgDrinksPerWeek AvgDrinksPerWeek.
```

SAS Code for Imputation of Income Variable

```
* Impute IncomeRanges via PROC HOTDECK
*;
data HINTS5CYCLE3;
     set HINTS5CYCLE3;

     COPY_Education = Education;
     if COPY_Education in (-9, -7) then
          COPY_Education = .;

     COPY_RaceEthn = RaceEthn;
     if COPY_RaceEthn in (-9, -7) then
          COPY_RaceEthn = .;

     COPY_RentOrOwn = RentOrOwn;
     if COPY_RentOrOwn in (-5, -7, -9) then
```

```sas
            COPY_RentOrOwn = .;

        COPY_SpeakEnglish = SpeakEnglish;
        if COPY_SpeakEnglish in (-5, -7, -9) then
            COPY_SpeakEnglish = .;

        COPY_IncomeRanges = IncomeRanges;
        if COPY_IncomeRanges in (-7, -9) then
            COPY_IncomeRanges = .;

        ID = _N_;

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

proc freq data= HINTS5CYCLE3;
    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_IncomeRanges*IncomeRanges / list missing;
run;

proc impute data= HINTS5CYCLE3 method=wshd notsorted;
    weight TG_all_FINWT0;
    impvar COPY_IncomeRanges;
    impby COPY_Education COPY_RaceEthn COPY_RentOrOwn
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= HINTS5CYCLE3;
    by ID;
run;

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

data HINTS5CYCLE3;
    merge HINTS5CYCLE3 (in=A) impute1 (in=B);
    by ID;
```

```sas
      if A = 1 and B = 1;
run;

data _null_;
      set HINTS5CYCLE3;

      if IncomeRanges not in (-7, -9) and COPY_IncomeRanges ^=
IncomeRanges_IMP then
            put ID IncomeRanges COPY_IncomeRanges IncomeRanges_IMP;
run;

data HINTS5CYCLE3;
      set HINTS5CYCLE3;

      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;
run;
```