

# HINTS 4 Cycle 4 History Document

---

## July 2021

### IMPORTANT!

#### July 2021 Data Update

The final weighted distribution included in the April 2021 revision of the HINTS 4 Cycle 4 (2014) data had a disproportionate number of males in the higher education groups. This issue arose because the raking step involving gender and education was carried out differently in HINTS 4 Cycle 4 (2014) than in prior HINTS cycles. Raking involves forcing the distributions for several different demographic breakouts to the national totals. In the interest of maintaining a distribution of education for males that is closer to the national population, new weights were created and published in July 2021 that enforced tighter controls on these two variables.

Several different types of analyses were run to compare results using the prior weights (distributed with the April 2021 dataset) and the revised weights (corrected in July 2021). These analyses involved examining percentages, regression estimates, and trends. The results that differed were those based on small sample sizes (e.g., those age 90 and over) or effects of education that were close to statistical significance at the  $p < 0.05$  level.

HINTS data users who have completed, or are currently conducting, analyses using HINTS 4 Cycle 4 (2014) data should re-run final analyses using the revised weights in the data files that were updated in July 2021. Results are not likely to change for effects not related to education; however, it is best to be sure that results based on small sample sizes or results that were/are close to or marginally significant do not change when the revised weights are used.

**HINTS data users who have completed, or are currently conducting, analyses using HINTS 4 Cycle 4 (2014) data that was published prior to April 2021 should continue reading the next section: April 2021 Data Update.**

#### April 2021 Data Update

An error occurred where 5-year American Community Survey (ACS) estimates were used as the source of the population totals used in the calibration step of the weighting. The correct population should have been the 1-year ACS estimates. The 5-year estimates are based on an average of the ACS for the previous 5 years, while the 1-year estimates are based on the results of the ACS for that particular year. The HINTS estimates affected most by this error are population totals or counts (e.g., the total number of adults who have searched for information about cancer from any source). These totals will be, on average, about 2 percent lower for the 5-year estimates than if the 1-year estimates were used.

Linear and logistic regressions using the incorrect weights will be affected less than population totals because the error is in both the numerator and the denominator, which will tend to cancel the error out. Several different types of analyses, which compare results using the weights with the error and a corrected set of weights, were completed to test for differences. These involved looking at percentages, regression estimates, and trends. None of these analyses were substantively different when using the corrected weights. Virtually all resulted in percentages, regression coefficients, and significance tests that did not differ at the first decimal place.

The advice to users who have completed analyses using HINTS 4 Cycle 4 but not published yet is to rerun the analyses with the correct weights found in this April 2021 data package downloaded from the HINTS website. For results that have already been published, it is not advised to do anything except in two scenarios:

1. If the results rely on reporting population counts or totals.
2. If a small change in the statistical significance of a result would affect your conclusions. For example, if the result is based on a result that is significant close to a 5% level (if that is the criteria used in the analysis).

**In both cases, it is advised to rerun the analysis and decide if the results differ enough to merit reporting an erratum to the journal.**

### **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.

#### **MailHHAdults: 2. Including yourself, how many people age 18 or older live in this household?**

Three respondents had the value of 0, which was replaced by 1.

#### **Height\_Inches: F5. (E4.) About how tall are you without shoes? Inches:**

Two hundred and seven 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.

#### **YearCameToUSA: 08. (L8.) In what year did you come to live in the United States?**

Seventeen respondents had values between 1 and 196, which were replaced by -4.

#### **HHAdultAge2: 013. (L13.) Starting with yourself, please mark the gender, and write in the age and month of birth for each adult 18 years of age or older living at this address. Adult 2: Age**

Twenty five respondents had Adult 2 age between 1 and 17. All their ages were replaced by -4.

#### **HHAdultAge3: 013. (L13.) Starting with yourself, please mark the gender, and write in the age and month of birth for each adult 18 years of age or older living at this address. Adult 3: Age**

One hundred and twenty nine respondents had Adult 3 age between 0 and 17. All their ages were replaced by -4.

**HHAdultAge4: 013. (L13.) Starting with yourself, please mark the gender, and write in the age and month of birth for each adult 18 years of age or older living at this address. Adult 4: Age**

One hundred and eight respondents had Adult 4 age between 0 and 17. All their ages were replaced by -4.

**HHAdultAge5: 013. (L13.) Starting with yourself, please mark the gender, and write in the age and month of birth for each adult 18 years of age or older living at this address. Adult 5: Age**

Forty four respondents had Adult 5 age between 0 and 17. All their ages were replaced by -4.

**R\_HHAdults: Reconciled number of adults in household**

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

**Totalhousehold: 012. Including yourself, how many people live in your household?**

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

## **Standard Recode**

Standard recode/derived variables are listed below.

**AgeGrpA: -->AgeGrpA. (Age Recode -- 4 Levels)**

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. (Age Recode -- 5 Levels)**

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? (Education Recode -- 4 Levels)**

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? (Education Recode -- 5 Levels)**

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. (Hisp\_Cat and Race\_Cat2 Recode -- 7 Levels)**

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.

### **HHInc:-->HHInc. What is your {combined} annual household income? (IncomeRanges Recode -- 5 Levels)**

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?(WhenDiagnosedCancer Recode)**

The variable AgeDX is a copy of variable WhenDiagnosedCancer.

### **TimeSinceDX:-->TimeSinceDX. How long ago were you diagnosed with cancer? (WhenDiagnosedCancer Age Recode)**

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 (Smoke100 and SmokeNow Recode)**

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.

#### **WeeklyMinutesModerateExercise:--> WeeklyMinutesModerateExercise. Minutes per week of at least moderate intensity exercise**

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

#### **PHQ4:→PHQ-4 total score (LittleInterest Hopeless Nervous Worrying Combined)**

The variable PHQ4 was created with LittleInterest, Hopeless, Nervous and Worrying variables. We create 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.

#### **RaceEthn5: →Race/Ethnicity recode (Hisp\_Cat and Race\_Cat2--5 Levels)**

The RaceEthn was created with Hisp\_Cat and Race\_Cat2 variables. The RaceEthn 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 RaceEthn was assigned with value of -9. The RaceEthn was assigned with value of -9 if Hispanic=-9.

## **PCCScale: -- > Patient Centered Communication scale**

The PCCScale variable was created with ChanceAskQuestions, FeelingsAddressed, InvolvedDecisions, UnderstoodNextsteps, HelpUncertainty and DrTakeCareNeeds. We reverse the values of these 6 variables and take the mean of these 6 variables (if at least half the variables have valid values). And then we transform the mean value linearly to a 0-100 scale. If PCCScale value is missing, then we assigned value of -9 to it.

## **Label Editing**

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

## **Format Editing**

### **Formats Added for Standard Recode Variables**

The formats AgeGrpA, AgeGrpB, EducA, EducB, RaceEthn, HHInc, BMI, AgeDX, TimeSinceDX, smokeStat, RaceEthn5f, phq4f, WeeklyMinutesModerateExercise and PccScale were created and assigned to the variables AgeGrpA, AgeGrpB, EducA, EducB, RaceEthn, HHInc, BMI, AgeDX, TimeSinceDX, smokeStat, RaceEthn5, PHQ4, WeeklyMinutesModerateExercise and PCCScale 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: ACTIVED, ADULTSI, BORNINU, CONSID, DRSHOUL, EVERHADC, EVERHAF, FREQGOP, HADTESG, HADTESH, HEARDGE, HEARDHP, INTERNF, LASTMED, OCCUPAF, SEEKHEA, SMOKENO, STRONGF, STRONGG, TIMESMO, USEINTE, WHERESE, and WHERESEF.

## **Imputation of Income Variable**

The income variable (IncomeRanges) has relatively higher percentage (11% 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 (O6), RaceEthn (standard recode), RentOrOwn (O15), 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

```
* 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 99 to Unreadable or Non-
conforming numeric response
*;

    if 1 <= YearCameToUSA <= 99 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 SharedRes_NotHad to -4 if HadTest_Cat is 1,2,3,4,5,6,10,or 91
AND SharedRed_NotHad is 1;
```

```
if HadTest_Cat in (1, 2, 3, 4, 5, 6, 10, 91) and SharedRes_NotHad
=1 then SharedRes_NotHad=-4;
```

## 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 (-1, -4, -9) then
    AgeGrpA = Age;
label AgeGrpA = '-->AgeGrpA. (Age Recode -- 4 Levels)';

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. (Age Recode -- 5 Levels)';

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? (Education Recode -- 4 Levels)';

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? (Education Recode -- 5 Levels)';

    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. (Hisp_Cat and Race_Cat2
Recode -- 7 Levels)';

    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? (IncomeRanges Recode -- 5 Levels)';

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)';
if BMI not in (-4, -9) then
    BMI = round(BMI, 0.1);

AgeDX = WhenDiagnosedCancer;
label AgeDX = '-->AgeDX. At what age were you diagnosed with
cancer? (WhenDiagnosedCancer Recode)';

if EverHadCancer in (1) then do;
    if Age in (-9) then TimeSinceDX = Age;
    else if WhenDiagnosedCancer in (-9, -4) 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;
    end;

label TimeSinceDX = '-->TimeSinceDX. How long ago were you
diagnosed with cancer? (WhenDiagnosedCancer Age Recode)';

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 (Smoke100 and
SmokeNow Recode)';

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 (LittleInterest Hopeless Nervous
Worrying Combined)';
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*TimesModerat
eExercise + HowLongModerateExerciseMn*TimesModerateExercise;
end;
label WeeklyMinutesModerateExercise="WeeklyMinutesModerateExercise.
Minutes per week of at least moderate intensity exercise";

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 recode (Hisp_Cat and Race_Cat2--5
Levels)';

if ChanceAskQuestions=1 then QuestionsR=4;
if ChanceAskQuestions=2 then QuestionsR=3;
if ChanceAskQuestions=3 then QuestionsR=2;
if ChanceAskQuestions=4 then QuestionsR=1;

if FeelingsAddressed=1 then FeelingsR=4;
if FeelingsAddressed=2 then FeelingsR=3;
if FeelingsAddressed=3 then FeelingsR=2;
if FeelingsAddressed=4 then FeelingsR=1;

if InvolvedDecisions=1 then DecisionsR=4;
if InvolvedDecisions=2 then DecisionsR=3;
if InvolvedDecisions=3 then DecisionsR=2;
if InvolvedDecisions=4 then DecisionsR=1;

if UnderstoodNextSteps=1 then NextStepsR=4;
if UnderstoodNextSteps=2 then NextStepsR=3;
if UnderstoodNextSteps=3 then NextStepsR=2;
if UnderstoodNextSteps=4 then NextStepsR=1;

if HelpUncertainty=1 then UncertaintyR=4;
if HelpUncertainty=2 then UncertaintyR=3;
if HelpUncertainty=3 then UncertaintyR=2;
if HelpUncertainty=4 then UncertaintyR=1;

if DrTakeCareNeeds=1 then NeedsR=4;
if DrTakeCareNeeds=2 then NeedsR=3;
if DrTakeCareNeeds=3 then NeedsR=2;
if DrTakeCareNeeds=4 then NeedsR=1;

/*Take mean of 6 variables if at least half the variables have valid
values*/
if nmiss(QuestionsR, FeelingsR, DecisionsR, NextStepsR, UncertaintyR,
NeedsR)<4
then pccmean=mean(QuestionsR, FeelingsR, DecisionsR, NextStepsR,
UncertaintyR, NeedsR);
/*Creat composite score - linear variable on a scale of 100*/
PCCScale=round(( (pccmean-1)*100)/3,0.1);
if PCCScale=. then PCCScale=-9;
label PCCScale='PCCScale. Patient Centered Communication scale';

```

## SAS Code for Label Editing

```
label
```

```

ActiveDutyArmedForces = "03. (L3.) Have you ever served on
active duty in the U.S. Armed Forces, military Reserves or National
Guard?"

```

AverageDailyTVGames = "H6. (F8.) Over the past 30 days, how many hours per day, on average, did you sit and watch TV or movies, surf the web, or play computer games?"

CancerCheckups = "M8. Have you ever received instructions from a doctor or other health care professional about where you should return...after completing your cancer treatment?"

CancerConcernedQuality = "A6c. Based on the results of your most recent search for information about cancer how much do you agree or disagree with each of the following statements? Concerned about Quality."

CancerFrustrated = "A6b. Based on the results of your most recent search for information about cancer, how much do you agree or disagree with each of the following statements? Frustrated."

CancerLotOfEffort = "A6a. Based on the results of your most recent search for information about cancer, how much do you agree or disagree with each of the following statements? Took effort."

CancerTooHardUnderstand = "A6d. Based on the results of your most recent search for information about cancer, how much do you agree or disagree with each of the following statements? Hard to understand."

ChanceAskQuestions = "C5a. In the past 12 months, how often did your health professional: Give you the chance to ask all the health-related questions you had? "

ElectCigLessHarm = "J10. Compared to smoking cigarettes, would you say that electronic cigarettes are..."

EverHadPSATest = "L8. (I4.) Have you ever had PSA test?"

ExplainedClearly = "C5e. In the past 12 months, how often did your health professional: Explain things in a way you could understand?"

FeelingsAddressed = "C5b. In the past 12 months, how often did your health professional: Give the attention you needed to your feelings and emotions?"

Genetics\_Diabetes = "E4a. How much do you think genetics determine whether or not a person will develop each of the following conditions? Diabetes or high blood sugar?"

Genetics\_HighBP = "E4d. How much do you think genetics determine whether or not a person will develop each of the following conditions? High blood pressure or hypertension?"

HeardGeneticTest = "E2. (A11.) Genetic tests that analyze your DNA, diet and lifestyle are currently being marketed by companies directly to consumers. Have you ever heard or read about these tests?"

HelpUncertainty = "C5g. In the past 12 months, how often did your health professional: Help you deal with feelings of uncertainty about your health or health care?"

InvolvedDecisions = "C5c. In the past 12 months, how often did your health professional: Involve you in decisions about your health care as much as you wanted?"

RegExercise\_Enjoyment = "H5d. People start or continue exercising regularly for lots of reasons. How much does getting enjoyment reflect why you would start or continue?"

```
RegExercise_Guilt = "H5c. People start or continue  
exercising regularly for lots of reasons. How much does feeling guilty  
when you skip reflect why you would start or continue?"
```

```
TimesModerateExercise = "H1. (F5.) In a typical week, how  
many days do you do any physical activity of at least moderate  
intensity?"
```

```
TimesStrengthTraining = "H3. (F7.) In a typical week, how  
many days do you do leisure-time physical activities specifically  
designed to strengthen your muscles?"
```

```
UnderstoodNextSteps = "C5d. In the past 12 months how often  
did your health professional: Make sure you understood the things you  
needed to do to take care of your health?"
```

```
EverTalkedHPVShot = "L1. Has a doctor or other health care  
professional ever talked with you about the HPV shot or vaccine?"
```

```
HeardOfClinicalTrial = "E1. Clinical trials are research  
studies that involve people. Have you ever heard of a clinical trial?"
```

```
SpentEnoughTime = "C5f. In the past 12 months, how often  
did your health professional: Spend enough time with you?"
```

```
;
```

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

format    AgeGrpA AgeGrpA.
          AgeGrpB AgeGrpB.
          EducA EducA.
          EducB EducB.
          RaceEthn RaceEthn.
          HHInc HHInc.
          BMI BMI.
          AgeDX AgeDX.
          TimeSinceDX TimeSinceDX.
          smokeStat smokeStat.
          RaceEthn5 RaceEthn5f.
          phq4 phq4f.
          WeeklyMinutesModerateExercise
WeeklyMinutesModerateExercise.
          IncomeRanges_IMP incomer.
          PccScale pccscale.

;

```

## SAS Code for Imputation of Income Variable

```

* Impute IncomeRanges via PROC HOTDECK
*;
data HINTS4CYCLE4;
    set HINTS4CYCLE4;

    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=HINTS4CYCLE4;
    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=HINTS4CYCLE4 method=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=imputel replace;
run;

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

proc contents data=imputel;
run;

```

```

proc sort data=HINTS4CYCLE4;
    by ID;
run;

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

data HINTS4CYCLE4;
    merge HINTS4CYCLE4 (in=A) impute1 (in=B);
    by ID;

    if A = 1 and B = 1;
run;

data _null_;
    set HINTS4CYCLE4;

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

data HINTS4CYCLE4;
    set HINTS4CYCLE4;

    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;

```