

HINTS 7 History Document

April 2025

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?:

Six respondents had a value of 0 adults in household, which was replaced by 1.

R_HHAdults: Reconciled number of adults in household

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

Totalhousehold: P12. Including yourself, how many people live in your household?

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

AverageTimeSitting: K4. During the past 7 days, how much time did you spend sitting on a typical day at home or at work?

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

HowLongModerateExerciseMinutes: M2. On the days that you do any physical activity or exercise of at least moderate intensity, how long do you typically do these activities?:

Seven respondents had an outlying value greater than 999 minutes for HowLongModerateExerciseMinutes, which was replaced by -4, “Unreadable or Non-conforming numeric response”.

TimesSunburned: J4. 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 an outlying value of 300 or greater for TimesSunburned, which was replaced by -4, “Unreadable or Non-conforming numeric response”.

SleepWeekdayHr: I8a. During the past 30 days how many hours of sleep did you usually get in a 24-hour period on a weekday?:

Six respondents had an outlying value of 0 and 24 respondents had an outlying value of 24, which was replaced by -4, “Unreadable or Non-conforming numeric response”.

SleepWeekendHr2: I8b. During the past 30 days how many hours of sleep did you usually get in a 24-hour period on a weekend day?:

Seven respondents had an outlying value of 0 and 16 respondents had an outlying value of 24, which was replaced by -4, "Unreadable or Non-conforming numeric response".

DrinksPerDay2: M2. During the past 30 days, on the days when you drank alcohol, about how many alcoholic drinks did you drink on average?

Five respondents had an outlying value of 30 or greater for DrinksPerDay2, which was replaced by -4, "Unreadable or Non-conforming numeric response".

WorkHrsPerWeek: R4. About how many hours do you work per week at all of your jobs and businesses combined?

Three respondents had an outlying value greater than 168 for WorkHrsPerWeek, which was replaced by -4, "Unreadable or Non-conforming numeric response".

Standard Recode

Standard recode/derived variables are listed below.

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: 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: 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: 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)", then RaceEthn 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)”, then RaceEthn was assigned with value of -7. RaceEthn was assigned with value of -9 if Hisp_Cat =-9. RaceEthn was assigned a value of -7 if Hisp_Cat=-7.

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)”, then RaceEthn5 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)”, then RaceEthn5 was assigned with value of -7. RaceEthn5 was assigned with value of -9 if Hisp_Cat =-9. RaceEthn5 was assigned a value of -7 if Hisp_Cat=-7.

HHInc: What is your {combined} annual household income? 6 Levels (Derived from IncomeRanges)

The IncomeRanges variable was re-coded into 6 categories: Less than \$20,000; \$20,000 to < \$35,000; \$35,000 to < \$50,000; \$50,000 to < \$75,000; \$75,000 to <\$100,000; \$100,000 or more. The original negative values were carried over.

BMI: Body Mass Index (Weight*703)/(Height in inches2)**

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”, then BMI was assigned to -9. If height in feet had a value greater than 0 and height in inches had a value of -4, “Unreadable or Non-conforming numeric response”, then 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)”, then BMI was assigned to -7.

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 a 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 a value of -7, “Missing data (Web partial- Question never seen)”, TimeSinceDX was assigned to -7. 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. If the variable EverHadCancer had value of 1 and WhenDiagnosedCancer had value of -7, “Missing data (Web partial- Question never seen)”, TimeSinceDX was assigned to -7.

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. Starting in HINTS 6, SmokeNow was administered regardless of the response to Smoke100. In prior HINTS surveys, SmokeNow was only administered if the response of Smoke100 was Yes. Due to this skip pattern difference, SmokeStat was derived differently than in previous versions. If Smoke100 had a value of 1 or 2 (Yes or No) and SmokeNow had a value of 1 or 2 (Every day or Some days), then SmokeStat was assigned a value of 1 (Current). If Smoke100 had a value of 1 (Yes) and SmokeNow had a value of 3 (Not at all), then SmokeStat was assigned a value of 2 (Former). If Smoke100 had a value of 2 (No) and SmokeNow had a value of 3 (Not at all), then SmokeStat was assigned a value of 3 (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 or 2 and SmokeNow had value of -9, "Missing data (Not Ascertained)", then smokeStat was assigned to -9. If Smoke100 had value of -9, "Missing data (Not Ascertained)", the smokeStat was assigned to -6. If Smoke100 had value of -7, "Missing data (Web partial - Question Never Seen)" and SmokeNow had value of -7, "Missing data (Web partial - Question Never Seen)" then smokeStat was assigned to -7.

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 a value of -9, "Missing data (Not Ascertained)", then PHQ was assigned to -9. If one of LittleInterest, Hopeless, Nervous and Worrying variables had a value of -7, "Missing data (Web partial - Question Never Seen)", then PHQ was assigned to -7.

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 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).

E-Cig Use: Electronic Cigarette Use (Derived from EverUsed_ECig and NowUse_ECig)

The variable E-Cig Use was created using the responses to EverUsed_ECig and NowUse_ECig. This variable has four categories: Current, Former, Never, and Missing. Respondents were coded as Current if they selected 1 (Selected) for NowUse_ECig, indicating they currently use them. Respondents were coded as Former if they selected 1 (Selected) for EverUsed_ECig but 2 (Not selected) for NowUse_ECig, indicating they have ever used e-cigarettes but do not currently use them. Respondents were coded as Never if they selected 2 (Not selected) for EverUsed_ECig, indicating they have never used e-cigarettes. Respondents with missing or inapplicable data were coded as -9, Missing.

AvgDrinksPerWeek: Average number of drinks per week (Derived from DrinkDaysPerMonth and DrinksPerDay2)

The variable AvgDrinksPerWeek was created with DrinkDaysPerMonth and DrinksPerDay2 variables. If DrinkDaysPerMonth is less than 0 then AvgDrinksPerWeek was assigned to the value of DrinkDaysPerMonth (i.e., the original negative values were retained). If DrinksPerDay2 is less than 0 then AvgDrinksPerWeek was assigned to the value of DrinksPerDay2 (i.e., the original negative values were retained). If DrinkDaysPerMonth and DrinksPerDay2 are greater than 0 then AvgDrinksPerWeek equals to the product of DrinkDaysPerMonth and DrinksPerDay2 divided by 4. If DrinkDaysPerMonth or DrinksPerDay2 equals to 0 then AvgDrinksPerWeek equals 0.

PCCScale: Patient Centered Communication scale

The PCCScale variable was created with ChanceAskQuestions, FeelingsAddressed, InvolvedDecisions, UnderstoodNextsteps, ExplainedClearly, SpentEnoughTime, and HelpUncertainty. We reverse the values of these 7 variables and take the mean of these 7 variables (if at least half the variables have valid values). And then we transform the mean value linearly to a 0-100 scale. If at least half the input variables did not have valid values and any of the 7 input items were -9, then PCCScale was assigned a value of -9. If at least half the input variables did not have valid values and any of the 7 input items were -7, "Missing data (Web Partial- Question Never Seen)", then PCCScale was assigned a value of -7. If at least half the input variables did not have valid values and the FreqGoProvider variable was 0, "None", then PCCScale was assigned a value of -1. See for more information about the psychometric properties of the PCC scale: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9803193/>

ISEE_Scale: Information Seeking Experience scale

The ISEE_Scale variable was created with CancerLotOfEffort, CancerFrustrated, CancerConcernedQuality and CancerTooHardUnderstand. We take the mean of these 4 variables (if all the variables have valid values). And then we transform the mean value linearly to a 0-100 scale. If any of the 4 input items for ISEE_Scale were -9, then ISEE_Scale was assigned a value of -9. If any of the 4 input items for ISEE_Scale were -7, "Missing data (Web Partial- Question Never Seen)", then ISEE_Scale was assigned a value of -7. If the SeekCancerInfo variable was 2, "No" then ISEE_Scale was assigned a value of -1.

PROMIS_Isolation_t: PROMIS Social Isolation scale T scores

The PROMIS_Isolation_t variable was created with FeelLeftOut, FeelPeopleBarelyKnow, FeelIsolated and FeelPeopleNotWithMe. We take the sum of these 4 variables (if all variables have valid values) and then we transform the sum score to T-score (which has a mean of 50 and standard deviation of 10) using the

PROMIS scoring conversion table. If any of the 4 input variables for PROMIS_Isolation_t were -9, then PROMIS_Isolation_t was assigned a value of -9. If any of the 4 input variables for PROMIS_Isolation_t were -7, “Missing data (Web Partial- Question Never Seen)”, then PROMIS_Isolation_t was assigned a value of -7. See for more information regarding PROMIS measures:

<https://www.healthmeasures.net/explore-measurement-systems/promis>

SexualOrientation_Recode: Sexual Orientation Recode (See History Document)

The variable SexualOrientation_Recode was created using the SexualOrientation2 and SexualOrientation2_OS variables. It has five categories: Lesbian or gay, Straight (that is, not gay or lesbian), Bisexual, Something else, and Don’t know. Responses in SexualOrientation2 coded as Lesbian or gay, Straight, Bisexual, or Don’t know were retained. Responses in SexualOrientation2_OS with terms such as ASEXUAL, DEMISEXUAL/QUEER, HETEROFLEXIBLE, PANSEXUAL, QUEER, or QUEER AND ASEXUAL were recoded as Something else. Any other responses in SexualOrientation2_OS were coded as -9, “Missing Data--Not Ascertained missing”. Additionally, specific HHIDs (72220548 and 72119232) were upcoded to Straight (that is, not gay or lesbian).

Label Editing

Labels Added for Standard Recode Variables

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

Labels Modified for Certain Variables

Labels were modified for the following variables to correct typos and improve clarity:

SleepWeekendHr2, GenderIdentity2, BIRTHSEX_I, MARITALSTATUS_I, EDUCATION_I, RACE_CAT2_I, and SexualOrientation2_OS.

Format Editing

Formats Added for Standard Recode Variables

The formats AgeGrpA, AgeGrpB, EducA, EducB, RaceEthn, RaceEthn5f, HHInc, BMI, TimeSinceDX, smokeStat, phq4f, WeeklyMinutesModerateExercise, ecigStat, AvgDrinksPerWeek, pccscale, ISEscale, , PROMISisolation, and SexualOrientation_Recode were created and assigned to the variables AgeGrpA, AgeGrpB, EducA, EducB, RaceEthn, RaceEthn5, HHInc, BMI, TimeSinceDX, smokeStat, PHQ4, WeeklyMinutesModerateExercise, eCigUse, AvgDrinksPerWeek, PCCScale, ISEE_Scale, PROMIS_Isolation_t, and SexualOrientation_Recode 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 were:

AdultsInHHf, SeekCancerInfof, FreqUseInternetf, WearableDevTrackHealth2f, MisleadingHealthInfof, FreqGoProviderf, TalkedDoctor2f, ManageMultipleProvidersf, ReceiveTelehealthCaref, AccessOnlineRecord3f, LookResultsBeforeHCPf, KnowledgePalliativeCaref, Caregiving2_Nof,

TimesModerateExercisef, TimesSunburnedf, DrinkDaysPerMonthf, SeenFedCourtTobMessagesf, EverUsed_Nonef, Marijuana_NotUsedf, PrefCervCaTestf, EverHadCancerf.

All “&” symbols in formats were removed. The modified formats were: DMAf.

Imputation of Income Variable

The income variable (IncomeRanges) has relatively higher percentage (12% for un-weighted percentage or 11% for weighted percentage) of missing values. This variable was imputed via PROC IMPUTE in SUDAAN. The imputation class variables are: Education (R8), RaceEthn (standard recode), and FullTimeOcc2_Cat (derived variable from R4 and R5). 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 R_HHAdults of missing (.) to -9 ;
if missing(R_HHAdults) = 1 then R_HHAdults = -9;

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

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

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

*Recode outliers of TimesSunburned 300 and 302 *;
if TimesSunburned in (300,302) then TimesSunburned=-4;

* Recode outliers of DrinksPerDay2 30 or more *;
if DrinksPerDay2 >= 30 then DrinksPerDay2 = -4;

* Recode outliers of SleepWeekdayHr 0 or 24 *;
if SleepWeekdayHr in (0,24) then SleepWeekdayHr = -4;

* Recode outliers of SleepWeekendHr2 0 or 24 *;
if SleepWeekendHr2 in (0,24) then SleepWeekendHr2 = -4;

* Recode outliers of WorkHrsPerWeek more than 168 hours ;
if WorkHrsPerWeek > 168 then WorkHrsPerWeek = -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 (-9,-4) then AgeGrpA=Age;
label AgeGrpA='AgeGrpA: Respondent Age Recode-4 Levels (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: Respondent Age Recode-5 Levels (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;
    end;

```



```

        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) (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= (-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 recode (Hisp_cat and Race_cat2--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) then
        HHInc = 5;
    else if IncomeRanges in (8, 9) then
        HHInc = 6;
    else if IncomeRanges in (-9, -7) then
        HHInc = IncomeRanges;
    label HHInc = 'HHInc: What is your (combined) annual household income?
    (IncomeRanges Recode-6 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) or (Height_Feet >
    0 and Height_Inches = -4) 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);

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) and SmokeNow in (1,2)) or (Smoke100 in (2) and
SmokeNow in (1,2)) then smokeStat = 1;
else if Smoke100 in (1) and SmokeNow in (3) then smokeStat = 2;
else if Smoke100 in (2) and SmokeNow in (3) then smokeStat = 3;
else if Smoke100 in (-9) then smokeStat = -6;
else if SmokeNow in (-5) then smokeStat = -4;
else if SmokeNow in (-7) or Smoke100 in (-7) then smokeStat = -7;
else if SmokeNow in (-9) or Smoke100 in (-9) then smokeStat = -9;
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;
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)";

/* Updated for HINTS7 */
eCigUse = .;
if NowUse_ECig = 1
    then eCigUse = 1; /* Current */
else if EverUsed_ECig = 1 and NowUse_ECig = 2
    then eCigUse = 2; /* Former */
else if EverUsed_ECig = 2 and NowUse_ECig not in (-7,1)
    then eCigUse = 3; /* Never */
else if EverUsed_ECig in (-9, -7) or NowUse_ECig in (-9, -7)
    then eCigUse = -9; /* Missing */
else if NowUse_ECig = -6 then eCigUse = -9; /* Filter Missing */
else if NowUse_ECig = -1 then eCigUse = -9; /* Inapplicable */
label eCigUse = "eCigUse: Electronic Cigarette Use (Derived from
EverUsed_ECig and NowUse_ECig; see History Document for more information)";

/*Updated for HINTS7*/
if DrinkDaysPerMonth > 0 and DrinksPerDay2 > 0
    then AvgDrinksPerWeek = (DrinkDaysPerMonth * DrinksPerDay2) / 4;
else
    AvgDrinksPerWeek = -9; /* Default to missing, to handle
conditions later */

/* Condition 1: DrinkDaysPerMonth = -9, -4, or -7 */
if DrinkDaysPerMonth in (-9, -4, -7) then
    AvgDrinksPerWeek = -9;

/* Condition 2: DrinksPerDay2 = -9, -7, -6, -4, -2, or -1 */
else if DrinksPerDay2 in (-9, -7, -6, -4, -2, -1) then

```

```

    AvgDrinksPerWeek = -9;

    /* Condition 3: DrinkDaysPerMonth = 0 or DrinksPerDay2 = 0 */
    if DrinkDaysPerMonth = 0 or DrinksPerDay2 = 0 then
        AvgDrinksPerWeek = 0;
    if AvgDrinksPerWeek > 75 then AvgDrinksPerWeek=-4;
    label AvgDrinksPerWeek = "AvgDrinksPerWeek: Average number of drinks per week
    (Derived from DrinkDaysPerMonth and DrinksPerDay2; see History Document for
    more information)";

    /*Creating PCC Scale*/
    if ChanceAskQuestions < 0 then ChanceAskQuestionsRev=.;
    if ChanceAskQuestions = 1 then ChanceAskQuestionsRev=4;
    if ChanceAskQuestions = 2 then ChanceAskQuestionsRev=3;
    if ChanceAskQuestions = 3 then ChanceAskQuestionsRev=2;
    if ChanceAskQuestions = 4 then ChanceAskQuestionsRev=1;
    *****;
    if FeelingsAddressed < 0 then FeelingsAddressedRev=.;
    if FeelingsAddressed = 1 then FeelingsAddressedRev=4;
    if FeelingsAddressed = 2 then FeelingsAddressedRev=3;
    if FeelingsAddressed = 3 then FeelingsAddressedRev=2;
    if FeelingsAddressed = 4 then FeelingsAddressedRev=1;
    *****;
    if InvolvedDecisions < 0 then InvolvedDecisionsRev=.;
    if InvolvedDecisions = 1 then InvolvedDecisionsRev=4;
    if InvolvedDecisions = 2 then InvolvedDecisionsRev=3;
    if InvolvedDecisions = 3 then InvolvedDecisionsRev=2;
    if InvolvedDecisions = 4 then InvolvedDecisionsRev=1;
    *****;
    if UnderstoodNextSteps < 0 then UnderstoodNextStepsRev=.;
    if UnderstoodNextSteps = 1 then UnderstoodNextStepsRev=4;
    if UnderstoodNextSteps = 2 then UnderstoodNextStepsRev=3;
    if UnderstoodNextSteps = 3 then UnderstoodNextStepsRev=2;
    if UnderstoodNextSteps = 4 then UnderstoodNextStepsRev=1;
    *****;
    if ExplainedClearly < 0 then ExplainedClearlyRev=.;
    if ExplainedClearly = 1 then ExplainedClearlyRev=4;
    if ExplainedClearly = 2 then ExplainedClearlyRev=3;
    if ExplainedClearly = 3 then ExplainedClearlyRev=2;
    if ExplainedClearly = 4 then ExplainedClearlyRev=1;
    *****;
    if SpentEnoughTime < 0 then SpentEnoughTimeRev=.;
    if SpentEnoughTime = 1 then SpentEnoughTimeRev=4;
    if SpentEnoughTime = 2 then SpentEnoughTimeRev=3;
    if SpentEnoughTime = 3 then SpentEnoughTimeRev=2;
    if SpentEnoughTime = 4 then SpentEnoughTimeRev=1;
    *****;
    if HelpUncertainty < 0 then HelpUncertaintyRev=.;
    if HelpUncertainty = 1 then HelpUncertaintyRev=4;
    if HelpUncertainty = 2 then HelpUncertaintyRev=3;
    if HelpUncertainty = 3 then HelpUncertaintyRev=2;
    if HelpUncertainty = 4 then HelpUncertaintyRev=1;
    *****;

    ** Create the PCCMean variable;
    If nmiss
    (ChanceAskQuestionsRev,FeelingsAddressedRev,InvolvedDecisionsRev,

```

```

UnderstoodNextStepsRev, ExplainedClearlyRev, SpentEnoughTimeRev,
HelpUncertaintyRev) < 4
then PCCMean =
MEAN(ChanceAskQuestionsRev, FeelingsAddressedRev, InvolvedDecisionsRev,
UnderstoodNextStepsRev, ExplainedClearlyRev, SpentEnoughTimeRev,
HelpUncertaintyRev);

/*Create composite score - linear variable on a scale of 100*/
PCCScale = round(((pccmean-1)*100)/3, 0.1);
if PCCScale < 0 then PCCScale = .;
if ChanceAskQuestions = -7 or FeelingsAddressed = -7 or InvolvedDecisions = -7
or UnderstoodNextSteps = -7 or ExplainedClearly = -7 or SpentEnoughTime = -7
or HelpUncertainty = -7 then PCCScale = -7;
if PCCScale = . then PCCScale = -9;
if FreqGoProvider = 0 then PCCScale = -1;
label PCCScale = 'PCCScale: Patient Centered Communication scale';

/*Create ISEE_scale */
** create the ISEEMean variable;
if CancerLotOfEffort >= 0 and CancerFrustrated >= 0 and
CancerConcernedQuality >= 0 and CancerTooHardUnderstand >= 0 Then
ISEE_Mean =
MEAN(CancerLotOfEffort, CancerFrustrated, CancerConcernedQuality,
CancerTooHardUnderstand);

/*Create composite score - linear variable on a scale of 100*/
ISEE_Scale = round(((ISEE_mean-1)*100)/3, 0.1);

if ISEE_Scale < 0 then ISEE_Scale = .;
if CancerLotOfEffort = -7 or CancerFrustrated = -7 or
CancerConcernedQuality = -7 or CancerTooHardUnderstand = -7 then
ISEE_Scale = -7;
if ISEE_Scale = . then ISEE_Scale = -9;
if SeekCancerInfo = 2 then ISEE_Scale = -1;
label ISEE_Scale = 'ISEE_Scale: Information Seeking Experience scale';

/*PROMIS Social Isolation*/
if feelleftout < 0 or feelpeoplebarelyknow < 0 or feelisolated < 0 or
feelpeoplenotwithme < 0 then PROMIS_Isolation_t = .;
else do;
    SIsSum = sum(feelleftout, feelpeoplebarelyknow, feelisolated,
feelpeoplenotwithme);
    if SIsSum = 4 then PROMIS_Isolation_t = 34.8;
    else if SIsSum = 5 then PROMIS_Isolation_t = 40.4;
    else if SIsSum = 6 then PROMIS_Isolation_t = 43.3;
    else if SIsSum = 7 then PROMIS_Isolation_t = 45.7;
    else if SIsSum = 8 then PROMIS_Isolation_t = 47.8;
    else if SIsSum = 9 then PROMIS_Isolation_t = 49.8;
    else if SIsSum = 10 then PROMIS_Isolation_t = 51.8;
    else if SIsSum = 11 then PROMIS_Isolation_t = 53.9;
    else if SIsSum = 12 then PROMIS_Isolation_t = 56.1;
    else if SIsSum = 13 then PROMIS_Isolation_t = 58.1;
    else if SIsSum = 14 then PROMIS_Isolation_t = 60.1;
    else if SIsSum = 15 then PROMIS_Isolation_t = 62.0;
    else if SIsSum = 16 then PROMIS_Isolation_t = 63.8;
    else if SIsSum = 17 then PROMIS_Isolation_t = 65.5;

```

```

        else if SIsun=18 then PROMIS_Isolation_t=67.5;
        else if SIsun=19 then PROMIS_Isolation_t=69.9;
        else if SIsun=20 then PROMIS_Isolation_t=74.2;
    end;
    if feelleftout=-7 or feelpeoplebarelyknow=-7 or feelisolated=-7 or
    feelpeoplenotwithme=-7 then PROMIS_Isolation_t=-7;
    if PROMIS_Isolation_t=. then PROMIS_Isolation_t=-9;
    label PROMIS_Isolation_t="PROMIS Social Isolation Scale T Scores
    (Derived from FeelLeftOut, FeelPeopleBarelyKnow, FeelIsolated,
    FeelPeopleNotWithMe; see History Document for more information)";

    /*SexualOrientation_Recode*/
    if upcase(SexualOrientation2_OS) in (
        'ASEXUAL',
        'DEMISEXUAL/QUEER',
        'HETEROFLEXIBLE',
        'PANSEXUAL',
        'QUEER',
        'QUEER AND ASEXUAL'
    ) then SexualOrientation_Recode = 4;
    else if notdigit(trim(SexualOrientation2_OS)) > 0 then
    SexualOrientation_Recode = -9;

    if HHID in ("72220548","72119232") then SexualOrientation_Recode = 2;
    if SexualOrientation2 in (1,2,3,5) then
    SexualOrientation_Recode=SexualOrientation2;
    label SexualOrientation_Recode="Sexual Orientation Recode (See History
    Document)";

```

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'
-7 = 'Missing data (Web partial - Question Never Seen)'
-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'
-7 = 'Missing data (Web partial - Question Never Seen)'
-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'
-4 = 'Unreadable or Nonconforming Numeric Response'
-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 to < $100,000"
6 = "$100,000 or greater"
-5 = "Multiple Responses Selected in Error"
-7 = "Missing data (Web partial - Question Never Seen)"
-9 = "Missing data (Not Ascertained) "
;

value BMI
-4 = 'Unreadable or Nonconforming Numeric Response'
-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'
-4 = 'Unreadable or Nonconforming Numeric Response'
-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 WeeklyMinutesModerateExercise
-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)'
;

value ecigStat
1 = 'Current'
2 = 'Former'
3 = 'Never'
-4 = 'Unreadable or Nonconforming Numeric Response'
-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)'
;

value pccscale
-1 = 'Inapplicable, coded 0 in FreqGoProvider'
-7 = 'Missing data (Web partial - Question Never Seen)'
-9 = 'Missing data (Not Ascertained)'

```



```

;

value ISEEScale
-1 = 'Inapplicable, coded 2 in SeekCancerInfo'
-7 = 'Missing data (Web partial - Question Never Seen)'
-9 = 'Missing data (Not Ascertained)'
;

value PROMISisolation
-7 = 'Missing data (Web partial - Question Never Seen)'
-9 = 'Missing data (Not Ascertained)'
;

value SexualOrientation_Recode
1 = "Lesbian or gay"
2 = "Straight, that is, not gay or lesbian"
3 = "Bisexual"
4 = "Something else"
5 = "Don't Know"
-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 RaceE5f.
            HHInc HHInc.
            BMI BMI.
            TimeSinceDX TimeSinceDX.
            smokeStat smokeStat.
            phq4 phq4f.
            WeeklyMinutesModerateExercise WeeklyMinutesModerateExercise.
            ECigUse ecigStat.
            IncomeRanges_IMP IncomeRangesf.
            AvgDrinksPerWeek AvgDrinksPerWeek.
            PCCScale pccscale.
            ISEE_Scale ISEEScale.
            PROMIS_Isolation_t PROMISisolation.
            SexualOrientation_Recode SexualOrientation_Recode.

;

```

SAS Code for Imputation of Income Variable

```

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

    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
        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 = .;

    COPY_FullTimeOcc2_Cat = FullTimeOcc2_Cat;
    if FullTimeOcc2_Cat in (-4, -9, -7) then
        COPY_FullTimeOcc2_Cat = .;

    ID = _N_;

run;

proc freq data=hints;
    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;
    tables COPY_FullTimeOcc2_Cat*FullTimeOcc2_Cat / list missing;
run;

proc impute data=hints method=wshd notsorted;
    weight PERSON_FINWT0;
    impvar COPY_IncomeRanges;
    impby COPY_Education COPY_RaceEthn COPY_FullTimeOcc2_Cat;
    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=hints;
    by ID;
run;

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

data hints;
    merge hints (in=A) imputel (in=B);

```

```

        by ID;

        if A = 1 and B = 1;
run;

data _null_;
    set hints;

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

data hints;
    set hints;

    if missing(IncomeRanges_IMP) = 1 then
        IncomeRanges_IMP = IncomeRanges;
    label IncomeRanges_IMP = '-->IncomeRanges_IMP.  Imputed IncomeRanges
variable via PROC HOTDECK in SUDAAN (see History Document for more
information)';

    drop COPY_Education COPY_RaceEthn /*COPY_RentOrOwn COPY_SpeakEnglish*/
    COPY_FullTimeOcc2_Cat
    ID
    COPY_IncomeRanges;
run;

```