# HINTS FDA History Document

## Iteration One (finalized on April 20, 2016)

### Data Editing

The following variables were identified to contain invalid or unusual values. Those values were replaced with negative value of -2, "Question answered in error", 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? Adultsinhh: 1. Is there more than one person age 18 or older living in this household?

162 respondents had value 1 for both MailHHAdults and Adultsinhh, which was invalid. Adultsinhh value for these 162 records was replaced by 2 and MailHHAdults value was set to -2.

#### R_HHAdults: Reconciled number of adults in household

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

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

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

### Standard Recode

Standard recode/derived variables are listed below.

#### 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.

#### AgeGrpC: -->AgeGrpC. (Age Recode -- 5 Levels)

The Age variable was re-coded into 4 categories: 18-29; 30-49; 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 -- 5 Levels)

The Education variable was re-coded into 5 categories: Less than High School; High School Graduate; Post high school training other than college(vocational or technical; Some College; College Graduate or More. 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.

### 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.

### Smokelessstat:-->Smokelessstat. Smokeless tobacco use status (UsedTobacco20Times and UseTobaccoNow Recode)

The variable SmokelessStat was created with UsedTobacco20Times and UseTobaccoNow variables. The variable SmokelessStat has 3 categories: Current; Former; Never. If UsedTobacco20Times had value of 1 and UseTobaccoNow had value of -5, “Multiple responses selected in error”, the SmokelessStat was assigned to -4. If UsedTobacco20Times had value of 1 and UseTobaccoNow had value of -9, “Missing data (Not Ascertained)”, the SmokelessStat was assigned to -9. If UsedTobacco20Times had value of -9, “Missing data (Not Ascertained)”, the SmokelessStat was assigned to -6.

### CigarStat:--> CigarStat. Cigar Smoking status (NumberCigarsSmoked and SmokeNowCigars Recode)

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

### CigSLTCigar :--> CigSLTCigar. CigSLTCigar Tobacco Status (Indicator of current polytobacco use of Cigarettes, Smokeless, and Cigar only)

The variable CigSLTCigar was created with smokeStat, SmokelessStat, and CigarStat. If one of these variable was equal to 1 (Current users) then CigSLTCigar was set to 1. Else CigSLTCigar was assigned to 0.

### 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.

## Label Editing

Labels were created for the following recoded variables: AgeGrpB, AgeGrpC, EducA, RaceEthn, HHInc, SmokeStat, SmokelesStat, CigarStat, CigSLTCigar and RaceEthn5.

## Format Editing

### Formats Added for Standard Recode Variables

The formats AgeGrpB, AgeGrpC, EducA, RaceEthn, HHInc, SmokeStat, CigSLTCigar and RaceEthn5 were created and assigned to the variables AgeGrpB, AgeGrpC, EducA, RaceEthn, HHInc, SmokeStat, SmokelesStat, CigarStat, CigSLTCigar and RaceEthn5 respectively.

## Imputation of Income Variable

The income variable (IncomeRanges) has relatively higher percentage (11.7% for un-weighted percentage or 8.6% 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 R_HHAdults of missing (.) to -9F *;
if missing(R_HHAdults) = 1 then
R_HHAdults = -9;

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

* Recode adultsinhh and MailhhAdults;
if adultsinhh=1 and MailHHAdults=1 Then do;
adultsinhh=2;
MailHHadults=-2;
end;
```

## SAS Code for Standard Recode

```
*SAS Code for Standard Recode;
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<0 then AgeGrpB = Age;
label AgeGrpB = 'AgeGrpB. (Age Recode -- 5 Levels)';

if 18 <= Age <= 29 then AgeGrpC = 1;
else if 30 <= Age <= 49 then AgeGrpC = 2;
else if 50 <= Age <= 64 then AgeGrpC = 3;
else if 65 <= Age <= 74 then AgeGrpC = 4;
else if 75 <= Age then AgeGrpC = 5;
else if Age<0 then AgeGrpC = Age;
label AgeGrpC = 'AgeGrpC. (Age Recode -- 5 Levels)';

if Education in (1, 2) then EducA = 1;*less than high school;
else if Education in (3) then EducA = 2;*High school completed;
else if Education in (4) then EducA = 3; *Post high school training other
than college(vocational or technical);
else if Education in (5) then EducA = 4; *Some college;
else if Education in (6, 7) then EducA = 5; *College Graduate or more;
else if Education<0 then EducA = Education;
label EducA = 'EducA.  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 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)';


if UsedTobacco20Times in (1) then do;
if UseTobaccoNow in (1, 2) then     smokelessstat = 1;
else if UseTobaccoNow in (3) then smokelessstat = 2;
else if UseTobaccoNow in (-5) then smokelessstat = -4;
else if UseTobaccoNow in (-9) then smokelessstat = -9;
end;
else if UsedTobacco20Times in (2) then do;
smokelessstat = 3;
end;
else if UsedTobacco20Times in (-9) then   smokelessstat = -6;
label smokelessstat = 'smokelessstat.  Smokeless tobacco use status
(UsedTobacco20Times and UseTobaccoNow Recode)';


if NumberCigarsSmoked in (4,5) then do;
if SmokeNowCigars in (1, 2) then CigarStat = 1;
else if SmokeNowCigars in (3) then CigarStat = 2;
else if SmokeNowCigars in (-5) then CigarStat = -4;
else if SmokeNowCigars in (-9) then CigarStat = -9;
end;
else if NumberCigarsSmoked in (0,1,2,3) then do;
CigarStat = 3;
end;
else if NumberCigarsSmoked in (-9) then CigarStat = -6;
else if NumberCigarsSmoked in (-5) then CigarStat=-4;
label CigarStat = 'CigarStat.  Cigar Smoking status (NumberCigarsSmoked and
SmokeNowCigars Recode)';


Smoke1=smokestat;
if smoke1>1 or smoke1<0 then smoke1=0;
smoke2=smokelessstat;
if smoke2>1 or smoke2<0 then smoke2=0;
smoke3=cigarstat;
if smoke3>1 or smoke3<0 then smoke3=0;
smokeall=sum(smoke1, smoke2, smoke3);
if smokeall>1 then CigSLTCigar=1;
else CigSLTCigar=0;
drop smoke1 smoke2 smoke3 smokeall;
label CigSLTCigar='CigSLTCigar. CigSLTCigar Tobacco Status (Indicator of
current polytobacco use of Cigarettes, Smokeless, and Cigar only)';
```

```
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)';
```

## SAS Code for Format Editing

### SAS Code for Formats Added for Standard Recode Variables

```
value AgeGrpC
1 = '18-29'
2 = '30-49'
3 = '50-64'
4 = '65-74'
5 = '75+'
-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 = 'vocational or technical'
4 = 'Some College'
5 = 'College Graduate or More'
-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 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 CigSLTCigar
1="Current CigSLTCigar Tobacco User"
0="Not Current CigSLTCigar Tobacco User";

value RaceEthnN
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"
;

format
AgeGrpC AgeGrpC.
AgeGrpB AgeGrpB.
EducA EducA.
RaceEthn RaceEthn.
HHInc HHInc.
SmokeStat smokeStat.
smokelessstat smokeStat.
CigarStat smokeStat.
CigSLTCigar CigSLTCigar.
RaceEthn5 RaceEthnN.
IncomeRanges_imp IncomeR.;
```

## SAS Code for Imputation of Income Variable

```
* Impute IncomeRanges via PROC HOTDECK ;

data HINTS_FDA;
      set t.hints_fda_r;

      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=HINTS_FDA;
      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=HINTS_FDA 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=impute1 replace;
run;

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

proc contents data=impute1;
run;

proc sort data=HINTS_FDA;
      by ID;
```

```
run;

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

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

      if A = 1 and B = 1;
run;

data check2;
set HINTS_FDA;
if IncomeRanges not in (-9) and COPY_IncomeRanges ^= IncomeRanges_IMP;
run;

data T.HINTS_FDA;
      set HINTS_FDA;

      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;
```