

Documentation for all Data Sets

HEALTH ABC DATA ANALYSIS FILE.....	2
PARTICIPANT HISTORY FILE (PH).....	3
YEAR 12 ANNUAL TELEPHONE CONTACT DATA (Y12Phone).....	6
YEAR 12 PROXY CONTACT DATA (Y12Proxy).....	8
138-MONTH SEMI-ANNUAL TELEPHONE CONTACT DATA (SA138Mo)	9
138-MONTH PROXY CONTACT DATA (SA138Prox)	10
MISSED FOLLOW-UP CONTACT DATA (MissVis).....	11
SPECIAL MISSING VALUE CODES	13
DROPPED VARIABLES.....	15
LISTINGS	15
DATA DICTIONARY	15
Appendix I DROPPED VARIABLES AND ALTERNATES	17
Appendix II 138-MONTH CALCULATED VARIABLES	22

HEALTH ABC DATA ANALYSIS FILE

To use the data, please contact the PI at your site.

Contents:

SAS Datasets

Y12Phone	Year 12 Annual Telephone Interview data
Y12Proxy	Year 12 Proxy Interview data
SA138Mo	138-month Semi-Annual Telephone Contact Data
SA138Prox	138-month Proxy Interview data
MissVis	Missed Follow-up Contact data
Y12Calc	Year 12 calculated (derived) variables

In addition, the following files, not specific to any year but updated each time data are released, can be found at the top of the Current Datasets listing on the Health ABC website under the [Datasets & Documentation](#) link:

PH	Participant History File
Formats	SAS Format Library

PARTICIPANT HISTORY FILE (PH)

1. General description

The PH file contains general information about the participants enrolled in the study. Variables included are:

HABCID	Health ABC Enrollment ID# without the 2-letter prefix
HCFAID	HCFA Screening ID (as assigned by the Coordinating Center)
DOB	Date of Birth
DOD	Date of Death
GENDER	Gender (1=Male; 2=Female)
RACE	Race (1=White; 2=Black)
RACEGEN	Race-gender group (1=White male, 2=White female, 3=Black male, 4=Black female)
SITE	Clinic Site (1=Memphis, 2=Pittsburgh)
CV1AGE	Age at Year 1 Clinic Visit
CV1DATE	Year 1 Clinic Visit Date
SV06AGE	Age at 6-Month Contact
SV06DATE	6-Month Contact Date
VITAL06M	Vital Status at time of 6-month contact
VTYPE06	Visit Type at 6-month contact
CV2AGE	Age at Year 2 Clinic Visit
CV2DATE	Year 2 Clinic Visit Date
VITAL12M	Vital Status at time of Year 2 (12-month) contact
VTYPE12	Visit Type at Year 2 (12-month) contact
SV18AGE	Age at 18-Month Contact
SV18DATE	18-Month Contact Date
VITAL18M	Vital Status at time of 18-month contact
VTYPE18	Visit Type at 18-month contact
CV3AGE	Age at Year 3 Clinic Visit
CV3DATE	Year 3 Clinic Visit Date
VITAL24M	Vital Status at time of Year 3 (24-month) contact
VTYPE24	Visit Type at Year 3 (24-month) contact
SV30AGE	Age at 30-Month Contact
SV30DATE	30-Month Contact Date
VITAL30M	Vital Status at time of 30-month contact
VTYPE30	Visit Type at 30-month contact
CV4AGE	Age at Year 4 Clinic Visit
CV4DATE	Year 4 Clinic Visit Date
VITAL36M	Vital Status at time of Year 4 (36-month) contact
VTYPE36	Visit Type at Year 4 (36-month) contact
SV42AGE	Age at 42-Month Contact
SV42DATE	42-Month Contact Date
VITAL42M	Vital Status at time of 42-month contact
VTYPE42	Visit Type at 42-month contact
CV5AGE	Age at Year 5 Clinic Visit
CV5DATE	Year 5 Clinic Visit Date

VITAL48M	Vital Status at time of Year 5 (48-month) contact
VTYPER48	Visit Type at Year 5 (48-month) contact
SV54AGE	Age at 54-Month Contact
SV54DATE	54-Month Contact Date
VITAL54M	Vital Status at time of 54-month contact
VTYPER54	Visit Type at 54-month contact
CV6AGE	Age at Year 6 Clinic Visit
CV6DATE	Year 6 Clinic Visit Date
VITAL60M	Vital Status at time of Year 6 (60-month) contact
VTYPER60	Visit Type at 60-month contact
SV66AGE	Age at 66-Month Contact
SV66DATE	66-Month Contact Date
VITAL66M	Vital Status at time of 66-month contact
VTYPER66	Visit Type at 66-month contact
CV7AGE	Age at Year 7 Clinic Visit
CV7DATE	Year 7 Clinic Visit Date
VITAL72M	Vital Status at time of Year 7 (72-month) contact
VTYPER72	Visit Type at 72-month contact
SV78AGE	Age at 78-Month Contact
SV78DATE	78-Month Contact Date
VITAL78M	Vital Status at time of 78-month contact
VTYPER78	Visit Type at 78-month contact
CV8AGE	Age at Year 8 Clinic Visit
CV8DATE	Year 8 Clinic Visit Date
VITAL84M	Vital Status at time of Year 8 (84-month) contact
VTYPER84	Visit Type at 84-month contact
SV90AGE	Age at 90-Month Contact
SV90DATE	90-Month Contact Date
VITAL90M	Vital Status at time of 90-month contact
VTYPER90	Visit Type at 90-month contact
CV9AGE	Age at Year 9 Clinic Visit
CV9DATE	Year 9 Clinic Visit Date
VITAL96M	Vital Status at time of Year 9 (96-month) contact
VTYPER96	Visit Type at 96-month contact
SV102AGE	Age at 102-Month Contact
SV102DATE	102-Month Contact Date
VITAL102M	Vital Status at time of 102-month contact
VTYPER102	Visit Type at 102-month contact
CV10AGE	Age at YEAR 10 Clinic Visit
CV10DATE	YEAR 10 Clinic Visit Date
VITAL108M	Vital Status at time of YEAR 10 (108-month) contact
SV114AGE	Age at 114-month Contact
SV114DATE	114-month Contact Date
VITAL114M	Vital Status at time of 114-month contact
CV11AGE	Age at YEAR 11 Clinic Visit
CV11DATE	YEAR 11 Clinic Visit Date
VITAL120M	Vital Status at time of YEAR 11 (120-month) contact
SV126AGE	Age at 126-month Contact
SV126DATE	126-month Contact Date

VITAL126M	Vital Status at time of 126-month contact
CV12AGE	Age at YEAR 12 Clinic Visit
CV12DATE	YEAR 12 Clinic Visit Date
VITAL132M	Vital Status at time of YEAR 12 (132-month) contact
SV138AGE	Age at 138-month Contact
SV138DATE	138-month Contact Date
VITAL138M	Vital Status at time of 138-month contact
VERSIONPH	Participant History File Release Date
VStatus	Vital Status (1=Alive, 2=Dead) as of PH file release date

The birthdate, race, and gender data come from the edited HCFA data. The Year 1 clinic visit date was taken from the corrected final participant list provided by the each clinic (with later corrections as necessary). All other dates were taken from the corresponding Clinic Visit workbook, Home Visit workbook, Annual Telephone interview, Proxy Interview, Proxy Contact Home Visit Workbook or Semi-Annual Telephone Contact form; participants who missed a visit have no corresponding date (set to .A). Age at each contact is a calculated variable based on birthdate and that contact date.

There are 3075 observations in the PH file. The demographic breakdown of participants in this dataset is as follows:

African-American Female	729
African-American Male	552
White Female	855
White Male	939
Memphis Participants	1548
Pittsburgh Participants	1527

2. Cross reference of dataset names with exact source

A complete list of variable names can be found under the “Proc Contents for All Datasets” link (search under PH).

3. Dataset structure and contents

The PH file contains a single observation per participant.

Key variables:

HABCID	HABC Enrollment ID without the 2-letter prefix
HCFAID	HCFA ID (as assigned by the Coordinating Center)

4. Condition of data

a. Known data errors: None at this time. The data have been edited. Editing will, however, be ongoing (iterative), so use of the most recent dataset is always advised.

b. Strength and weaknesses of dataset items: A new variable RACEGEN has been added to PH to ease analyses by race-gender group (1=White male, 2=White female, 3=Black male, 4=Black female)

If a death has been reported on a Missed Visit Form, an Event Form, or the Report of Death, the participant is listed as deceased in the vital status variable (VStatus) variable. Note: VStatus is vital status **as of the release date of the participant history file**, not as of any particular visit. This variable should only be used to determine the most up-to-date snapshot of vital status in Health ABC as of the file date. Vital status variables (VITALxxM) have been created for each Health ABC contact as follows: If a participant had a particular contact, or if they missed a contact but they were determined to still be alive at the time their contact was due (participant refused the contact, was too ill, etc), then VITALxxM is alive. If they missed a contact and were later discovered to have died before the end of their contact window, then VITALxxM is dead. If they missed a contact and were later discovered to have died after the end of their contact window, the VITALxxM is alive for that contact, but dead for the next. Finally, if they missed a contact without a determination of their vital status (participant could not be located, withdrew, no data for a visit for which they are past the visit window, etc) and no further contact with vital status determination has been made since then, then VITALxxM is missing.

The date of death (DOD) variable represents the best available information about the date of death for deceased participants as of the creation date of participant history file (PH).^{*} If a Report of Death form has been entered for the participant, the date of death from that adjudication form is used. If there is no Report of Death form yet, this information is taken from the Event Form dataset, and is therefore an un-confirmed, un-adjudicated date of death.

c. Missing Value Conventions: See Special Missing Value Codes below for special missing value codes applied

5. Dataset index formulation and key variable mapping

The PH file is sorted by HABCID, which is a unique identifier for each participant.

6. General strategies for manipulating and merging the data

Because the Health ABC datasets are sorted by Health ABC Enrollment ID, the HABCID variable is most useful for merging with other datasets.

YEAR 12 ANNUAL TELEPHONE CONTACT DATA (Y12Phone)

1. General description

The Y12Phone file contains information about the participants enrolled in the study gathered from the Year 12 Annual Telephone interview. If a participant did not have a Year 12 Annual Telephone Interview or a Year 12 Proxy Visit, they should have a Missed Follow-up Contact form that explains why. In addition, a variable VISITYPE has been appended to allow the analyst to account for all participants, whether or not they had a Year 12 Annual Telephone Contact. If VISITYPE= 13, the

^{*} Run proc contents in SAS to see creation date of the PH file.

participant's Year 12 data can be found in Y12PHONE. If VISITYPE= 8, the participant's Year 12 data can be found in Y12PROXY. If VISITYPE=3, 4 or 5, the participant did not have a Year 12 contact and their Missed Follow-up Contact data can be found in MISSVIS.

There are 3075 observations in the Y12Phone file. The types of visits (VISITYPE) are as follows:

Annual Phone Visit	1,506
Proxy Phone Visit	234
Missed/Deceased	1,159
Missed/Withdrew	65
Missed/Other	111

2. Cross reference of dataset names with exact source

A complete list of variable names can be found under the “Proc Contents for All Datasets” link (search under Y12Phone). Variable names can also be found on the annotated forms. Please note that not all variables on the forms are contained in the dataset. All variables not found in the dataset are listed in Dropped Variables and Alternates (Appendix I). Alternate variables to use (if applicable) are also listed.

3. Dataset structure and contents

The Y12Phone file contains a single observation per participant.

Key variables:

HABCID	HABC Enrollment ID without the 2-letter prefix
SITE	HABC Clinic site: 1=Memphis; 2=Pittsburgh [*]
VISITYPE	(3=Missed (other), 4=Deceased, 5=Withdrew, 8=Proxy Phone; 13=Annual Telephone Interview)
MISSREAS	Reason Y12 Annual Telephone Contact Missed

4. Condition of data

a. Known data errors:

Year 12 Annual Telephone Interview: None at this time.

b. Strength and weaknesses of dataset items:

Calculated (derived) variables have been created to complement those created for Years 1 through 12. To avoid confusion, these variables are listed in Y12Calc.

c. Missing Value Conventions: See Special Missing Value Codes on below for special missing value codes applied.

5. Dataset index formulation and key variable mapping

The Y12Phone file is sorted by HABCID, which is a unique identifier for each participant.

^{*} Must link to Health ABC participant history file (PH) to add this variable.

6. General strategies for manipulating and merging the data

Because the Health ABC datasets are sorted by Health ABC Enrollment ID, the HABCID variable is most useful for merging with other datasets.

YEAR 12 PROXY CONTACT DATA (Y12Proxy)

1. General description

As the Health ABC cohort has aged, some participants have begun having cognitive difficulties that prevent their being able to answer interview questions, and in some cases, being able to understand the instructions for some measurements. In response to this situation, a new form was created near the end of Year 3, the Proxy Interview. In year 12, there are 234 participants who have a proxy phone visit and can be found in Y12Proxy file.

Note that a “proxy phone” contact is empirically defined as having interview data only, but no physical measurements. This definition was used for the creation of the VISITYPE variable, but was not always followed by the clinics in assigning a value to YACONTAC. Thus VISITYPE is a more reliable variable to use to determine how many proxy contacts have only interview information, vs how many have physical measurements as well.

2. Cross reference of dataset names with exact source

A complete list of variable names can be found under the “Proc Contents for All Datasets” link (search under Y12Proxy). Variable names can also be found on the annotated forms. Please note that not all variables on the forms are contained in the dataset. All variables not found in the dataset are listed in Dropped Variables and Alternates (Appendix I). Alternate variables to use (if applicable) are also listed.

3. Dataset structure and contents

The Y12Proxy file contains a single observation per participant.

Key variables:

HABCID	HABC Enrollment ID without the 2-letter prefix
SITE	HABC Clinic site: 1=Memphis; 2=Pittsburgh*
VISITYPE	(8=Proxy Phone)

4. Condition of data

a. Known data errors: None at this time.

b. Strength and weaknesses of dataset items: No calculated variables have been created yet for this dataset. The only standard calculated variables that might apply to these data are the self-reported function variables; however, it has been reported that proxies over-report functional

* Must link to Health ABC participant history file (PH) to add this variable.

limitation (Elam, et al. Am J Public Health. 1991; 81:1127), and therefore more consideration needs to be applied to how these variables should be calculated.

c. Missing Value Conventions: See Special Missing Value Codes below for special missing value codes applied

5. Dataset index formulation and key variable mapping

The Y12Proxy file is sorted by HABCID, which is a unique identifier for each participant.

6. General strategies for manipulating and merging the data

Because the Health ABC datasets are sorted by Health ABC Enrollment ID, the HABCID variable is most useful for merging with other datasets.

138-MONTH SEMI-ANNUAL TELEPHONE CONTACT DATA (SA138Mo)

1. General description

The 138-month visit file contains information about the participants enrolled in the study gathered from the 138-Month Semi-Annual Telephone Interview. In some cases, a participant was unable to complete the contact by telephone because of deafness, cognitive impairment, or other reasons, and the information was gathered by proxy. Data for these participants can be found in SA138Prox. If a participant did not have a 138-month semi-annual telephone interview or a 138-month proxy, they should have a Missed Follow-up Contact form that explains why. A variable, VISITYYPE, has been added to this dataset to allow the analyst to account for all participants, whether or not they had a 138-month follow-up telephone call. If VISITYYPE=7, the participant's 138-month data can be found in SA138Mo. If VISITYYPE= 8, the participant's 138-month data can be found in SA138PROX. If VISITYYPE=3, 4 or 5, the participant did not have a 138-month contact and their Missed Follow-up Contact data can be found in MissVis.

There are 3075 observations in the SA138Mo file as follows:

SATC Phone Visit	1,434
Proxy Phone Visit	237
Missed/Deceased	1,230
Missed/Withdrew	65
Missed/Other	109

2. Cross reference of dataset names with exact source

A complete list of variable names can be found under the “Proc Contents for All Datasets” link (search under SA138Mo). Variable names can also be found on the annotated forms. Please note that not all variables on the forms are contained in the dataset. All variables not found in the dataset are listed in Dropped Variables and Alternates (Appendix I). Alternate variables to use (if applicable) are also listed.

3. Dataset structure and contents

The SA138Mo file contains a single observation per participant.

Key variables:

HABCID	HABC Enrollment ID without the 2-letter prefix
SITE	HABC Clinic site: 1=Memphis; 2=Pittsburgh *
VISITYPE	(3=Missed (other), 4=Deceased, 5=Withdrew, 7=SATC Phone, 8=Proxy Phone)
MISSREAS	Reason 138MO Visit Missed

4. Condition of data

a. Known data errors:

None at this time.

b. Strength and weaknesses of dataset items:

Calculated (derived) variables have been added to the dataset (see Appendix II).

c. Missing Value Conventions: See Special Missing Value Codes for special missing value codes applied.

5. Dataset index formulation and key variable mapping

The SA138Mo file is sorted by HABCID, which is a unique identifier for each participant.

6. General strategies for manipulating and merging the data

Because the Health ABC datasets are sorted by Health ABC Enrollment ID, the HABCID variable is most useful for merging with other datasets.

138-MONTH PROXY CONTACT DATA (SA138Prox)

1. General description

In some cases, a participant was unable to complete the 138-month contact by telephone because of deafness, cognitive impairment or other reasons. In response to this situation, the Proxy Interview was developed to allow another person to answer for them.

At the 138-month visit, there are 237 participants who have a Proxy Interview and can be found in SA138PROX.

A complete list of variable names can be found under the “Proc Contents for All Datasets” link (search under SA138PROX). Variable names can also be found on the annotated forms. Please note that not all variables on the forms are contained in the dataset. All variables not found in the dataset are listed in Dropped Variables and Alternates (Appendix I). Alternate variables to use (if applicable) are also listed.

* Must link to Health ABC participant history file (PH) to add this variable.

3. Dataset structure and contents

The SA138Prox file contains a single observation per participant.

Key variables:

HABCID	HABC Enrollment ID without the 2-letter prefix
SITE	HABC Clinic site: 1=Memphis; 2=Pittsburgh [*]
VISITYPE	(8=Proxy Phone)

4. Condition of data

a. Known data errors: None at this time.

b. Strength and weaknesses of dataset items: No calculated variables have been created yet for this dataset. The only standard calculated variables that might apply to these data are the self-reported function variables; however, it has been reported that proxies over-report functional limitation (Elam, et al. Am J Public Health. 1991; 81:1127), and therefore more consideration needs to be applied to how these variables should be calculated.

Not all the proxy interview questions were asked at the semi-annual telephone contact. Only the “starred” questions were asked during this contact.

c. Missing Value Conventions: See Special Missing Value Codes for special missing value codes applied.

5. Dataset index formulation and key variable mapping

The SA138PROX file is sorted by HABCID, which is a unique identifier for each participant.

6. General strategies for manipulating and merging the data

Because the Health ABC datasets are sorted by Health ABC Enrollment ID, the HABCID variable is most useful for merging with other datasets.

MISSED FOLLOW-UP CONTACT DATA (MissVis)

1. General description

The MissVis file contains information about the participants who have missed a follow-up contact (died, refused, lost to follow-up, etc) at any time during the study, up through the 138-month contact. If a participant missed an annual or semi-annual contact, they should have a Missed Follow-up Contact form that explains why.

The number of Missed Follow-up Contact forms related to each of these visits is shown below:

6-month follow-up contact	23
---------------------------	----

^{*} Must link to Health ABC participant history file (PH) to add this variable.

Year 2 visit	77
18-month semi-annual contact	137
Year 3 visit	95
30-month semi-annual contact	199
Year 4 visit	160
42-month semi-annual contact	201
Year 5 visit	125
54-month semi-annual contact	151
Year 6 visit	144
66-month semi-annual contact	144
Year 7 visit	137
78-month semi-annual contact	178
Year 8 visit	204
90-month semi-annual contact	207
Year 9 visit	176
102-month semi-annual contact	175
Year 10 visit	211
Year 10.5 (114-month) semi-annual contact	167
Year 11 visit	235
Year 11.5 (126-month) semi-annual contact	221
Year 12 visit	210
Year 12.5 (138-month) semi-annual contact	181

2. Cross reference of dataset names with exact source

A complete list of variable names can be found under the “Proc Contents for All Datasets” link (search under MissVis). Variable names can also be found on the annotated forms.

3. Dataset structure and contents

The MissVis file contains multiple observations per participant.

Key variables:

HABCID	HABC Enrollment ID without the 2-letter prefix
SITE	HABC Clinic site: 1=Memphis; 2=Pittsburgh*
BJID2	Contact missed (9=6-month, 2=Year 2, 10=18-month, 3=Year 3, 11=30-month, 4=Year 4, 12=42-month, 5=Year 5, 13=54-month, 6=Year 6, 14=66-month, 7=Year 7, 15=78-month, 16=Year 8, 17=90-month, 18=96-month, 19=102-month, 20=Year 10, 21=114-month, 22=Year 11, 23=126-month, 24=132-month, 25=138-month)

4. Condition of data

a. Known data errors: None at this time..

b. Strength and weaknesses of dataset items: If a participant missed a visit due to death or withdrawal from the study, the Missed Follow-up Contact corresponding to the first contact missed

* Must link to Health ABC participant history file (PH) to add this variable.

for this reason is usually the last Missed Follow-up Contact for that participant. That is, field centers were instructed not to continue completing Missed Follow-up Contacts for each subsequent contact missed after the death of a participant or their withdrawal from the study. If a participant could not be located at one contact and therefore had a Missed Follow-up Contact completed for that contact, then subsequently was found to have died before that contact, the death was recorded on a Missed Follow-up Contact form for the subsequent contact. That is, the Missed Follow-up Contact information reflects the status of the participant as known to the field center at the time of the scheduled contact. Missed Follow-up Contact data should not be used to determine approximate date of death, nor even numbers of participants who had died as of a particular follow-up contact. The best information available at the time of the data analysis file regarding date of death can be found in the Participant History file (PH, DOD, see page 3). If a participant was found to have both a Missed Follow-up Contact form for a particular contact and the corresponding contact forms (e.g. Annual Telephone Interview, Proxy Interview, or Semi-Annual Follow-Up Contact form), the Missed Follow-up Contact form data for that participant were deleted from the analysis file.

c. Missing Value Conventions: See Special Missing Value Codes below for special missing value codes applied

5. Dataset index formulation and key variable mapping

The MissVis file is sorted by HABCID, which is a unique identifier for each participant. The combination of HABCID and BJID2 is a unique identifier for a participant/contact record in this dataset.

6. General strategies for manipulating and merging the data

Because the Health ABC datasets are sorted by Health ABC Enrollment ID, the HABCID variable is most useful for merging with other datasets. The MissVis file must first be subsetted by BJID2 to the contact desired before merging with a contact-specific, one-record-per-participant dataset.

SPECIAL MISSING VALUE CODES

SAS allows for stratification of missing values. The following missing values have been assigned:

- . = 'Missing Form'
- .A = 'A:Not Applicable'
- .E = 'E:Recoded to Missing'
- .F = 'F:Variable Missing from Form'
- .L = 'L:Permanently Lost'
- .M = 'M:Missing'
- .N = 'N:Not Required'
- .T = 'T:Missing Due to Technical Problems' (reading center data only)
- .U = 'Unacceptable'

Description

. : Missing Form

Used when a value is missing because the entire form has not been entered.

A: Not Applicable

Used when a value is missing but the value is not required (due to simple skip pattern logic)

E: Recoded to Missing

Used to flag that a value was entered originally but should not have been (due to a skip pattern logic error) and that the value has been recoded to missing

F: Variable Missing from Form

Used to flag a variable that was not originally on the form (form was revised during the visit year) and therefore there is no value for this participant

L: Permanently Lost

Used to flag a tracking variable when a measurement involving a Reading Center was done, but the data either never made it to the Reading Center, or was lost at the Reading Center. A list of lost measurements, along with a brief description of what happened, can be found in Appendix I of the Reading Center Dataset documentation). Every effort was made to track down these data before they were declared "permanently lost" and the flag assigned.

M:Missing

Used to flag missing values when the value is required (i.e., true missing values).

N:Not Required

Used when a value is missing but the value is not required (not due to simple skip pattern logic). For example, for checkbox variables which are "Check all that apply" each one, individually, is not required. In these cases, a summary calculated variable (not included on the dataset) was used to edit missing responses. Some variables whose skip pattern logic is non-standard (i.e., the skip pattern involves several variables and forms) also have .N flags when missing, whether or not a response was required due to the skip pattern.

T:Missing Due to Technical Problems

Used when a value is missing from the Reading Center dataset due to technical difficulties. An explanation of when this value has been assigned can be found under Strengths and Weaknesses of (Reading Center) dataset items for each Reading Center (Reading Center data documentation; Substudy documentation).

U:Unacceptable

Used with certain Reading Center data when the data exist but cannot be used., for example, DXA data when the whole scan has been reviewed as unacceptable

General Strategies for Using Special Missing Values

In SAS, when using special missing values in logical expressions, the missing value is no longer only equal to '.' To express a value equal to missing, the code should be written: <= .z or alternately: le .z

To express a value not equal to missing, the code should be written >.z or alternately: gt .z
.Z is the greatest value of missing available in SAS.

DROPPED VARIABLES

A number of variables appear on the annotated forms but will not be found in the datasets. These variables are listed in Appendix I. There are several reasons why variables were dropped:

1. Participant confidentiality: identifying information such as participant name, acrostic, address, telephone number, etc. are omitted from the analysis file. All participants are instead identified by the HABC Enrollment ID# (HABCID).
2. Bookkeeping variables: a number of variables were put on the forms merely for bookkeeping purposes and are extremely unlikely to be useful for analysis. If an investigator notices that such a variable has been dropped and believes it should not have been, s/he should feel free to contact the Coordinating Center (HABCHelp@psg.ucsf.edu) and let us know that it should be included in future datasets.
3. Redundant variables: in many cases the same information was collected numerous times. We have made a special effort to clean up one version of each of these; and to avoid analysis using uncleaned data, the uncleaned version is omitted.

Appendix I lists the omitted variables in alphabetical order. There are two columns; the first, entitled "Variable omitted," lists the name of the variable not included in the dataset. The second column, entitled "Variable to use," lists the cleaned variable, for redundant variables, or the reason the variable was not included.

LISTINGS

A PDF listing of the SAS proc contents printout for all SAS datasets can be found under the "Proc Contents for All Datasets" link.

A text file, Formats.lst, showing all formats and value descriptions (e.g.: 1=White, 2=Black) contained in the SAS Format Library can also be downloaded. Click on FormatsList under the List of Current Datasets on the Health ABC website.

The following files are zipped together in a self-extracting document also available on the Health ABC website under the Current Datasets listing

Datadict.xls Microsoft Excel 4.0 spreadsheet containing information about all variables included in the SAS datasets

Datadict.dat Tab-delimited text file containing the same information as above.

DATA DICTIONARY (datadict.xls, datadict.txt)

General description

This is a searchable/sortable file that contains all the variables included in the SAS datasets. The following fields are included:

Variable	Variable Name
Label	Description of the variable
Form	Form or reading center origin of the variable
Page	Page number (not applicable for reading center or PH data)
Database	Database location of the variable
Variable Types	Type of variable (text, categorical (numeric), continuous (numeric), date, time, etc.)
Possible Values	Range of possible values associated with the variable (used for range edits)
SAS Format	SAS format assigned to the variable
Page Order	Variable order on the data collection forms (useful for sorting)

General Strategies for Use

The data dictionary is provided in two formats: Excel 2000 and tab-delimited text. They contain exactly the same information. The tab-delimited file was generated to provide easy access to those who do not have Excel version 2000 or higher.

The file is currently sorted by variable name. It can be used to search for details on a particular variable, to group by database or form, or to find a variable location on a form. It is a little more user friendly than the standard SAS proc contents, as it can be sorted, easily searched, and it provides additional details such as page number and form.

Appendix I

DROPPED VARIABLES AND ALTERNATES

Variable omitted	Variable to use
BJID	HABCID
BJACROS	N/A (confidentiality)
BJCONTAC	BJID2
BJSTFID	N/A (bookkeeping)
BJTYPE	BJID2
BJVISIT	BJID2
BLID	HABCID
BLACROS	N/A (confidentiality)
BLINK	N/A (bookkeeping)
BLMOVE	N/A (bookkeeping)
BLREF13A	N/A (bookkeeping)
BLREF13B	N/A (bookkeeping)
BLREF13C	N/A (bookkeeping)
BLREF14A	N/A (bookkeeping)
BLREF14B	N/A (bookkeeping)
BLREF14C	N/A (bookkeeping)
BLREF15A	N/A (bookkeeping)
BLREF15B	N/A (bookkeeping)
BLREF15C	N/A (bookkeeping)
BLREF16A	N/A (bookkeeping)
BLREF16B	N/A (bookkeeping)
BLREF16C	N/A (bookkeeping)
BLREF17A	N/A (bookkeeping)
BLREF17B	N/A (bookkeeping)
BLREF17C	N/A (bookkeeping)
BLREF18A	N/A (bookkeeping)
BLREF18B	N/A (bookkeeping)
BLREF18C	N/A (bookkeeping)
BLREF19A	N/A (bookkeeping)
BLREF19B	N/A (bookkeeping)
BLREF19C	N/A (bookkeeping)
BLREF19D	N/A (bookkeeping)
BLREF19E	N/A (bookkeeping)
BLREF19F	N/A (bookkeeping)
BLREF20A	N/A (bookkeeping)
R2ID	HABCID
R2ACROS	N/A (confidentiality)
R2CONTAC	VISITYPE

Variable omitted	Variable to use
R3ID	HABCID
R3ACROS	N/A (confidentiality)
R3CONTAC	VISITYPE
R4ID	HABCID
R4ACROS	N/A (confidentiality)
R4CONTAC	VISITYPE
R5ID	HABCID
R5ACROS	N/A (confidentiality)
R5CONTAC	VISITYPE
R6ID	HABCID
R6ACROS	N/A (confidentiality)
R6CONTAC	VISITYPE
R7ID	HABCID
R7ACROS	N/A (confidentiality)
R7CONTAC	VISITYPE
R8ID	HABCID
R8ACROS	N/A (confidentiality)
R8CONTAC	VISITYPE
R9ID	HABCID
R9ACROS	N/A (confidentiality)
R9CONTAC	VISITYPE
RAID	HABCID
RAACROS	N/A (confidentiality)
RACONTAC	VISITYPE
RBID	HABCID
RBACROS	N/A (confidentiality)
RBCONTAC	VISITYPE
RBHCHRS	RBHCTIM
RBHCMINS	RBHCTIM
RCEWHRS	RCEWTIM
RCEWMINS	RCEWTIM
RCID	HABCID
RCACROS	N/A (confidentiality)
RCCONTAC	VISITYPE
RDID	HABCID
RDACROS	N/A (confidentiality)
RDCONTAC	VISITYPE
RDHIA1HR	RDH1TIME
RDHIA1MN	RDH1TIME
REID	HABCID
REACROS	N/A (confidentiality)
RECONTAC	VISITYPE
REMA1HR	REM1TIME

Variable omitted	Variable to use
REMI1MN	REMI1TIME
RFID	HABCID
RFACROS	N/A (confidentiality)
RFCONTAC	VISITYPE
RGID	HABCID
RGACROS	N/A (confidentiality)
RGCONTAC	VISITYPE
RHID	HABCID
RHACROS	N/A (confidentiality)
RHCONTAC	VISITYPE
RIID	HABCID
RIACROS	N/A (confidentiality)
RICONTAC	VISITYPE
RJID	HABCID
RJACROS	N/A (confidentiality)
RJCONTAC	VISITYPE
RKID	HABCID
RKACROS	N/A (confidentiality)
RKCONTAC	VISITYPE
RKREF39A	N/A (bookkeeping)
RKREF39B	N/A (bookkeeping)
RKREF39C	N/A (bookkeeping)
RKREF40A	N/A (bookkeeping)
RKREF40B	N/A (bookkeeping)
RKREF40C	N/A (bookkeeping)
RLID	HABCID
RLACROS	N/A (confidentiality)
RLCONTAC	VISITYPE
RLREF41A	N/A (bookkeeping)
RLREF41B	N/A (bookkeeping)
RLREF41C	N/A (bookkeeping)
RLREF42A	N/A (bookkeeping)
RLREF42B	N/A (bookkeeping)
RLREF42C	N/A (bookkeeping)
RMID	HABCID
RMACROS	N/A (confidentiality)
RMCONTAC	VISITYPE
RMREF43A	N/A (bookkeeping)
RMREF43B	N/A (bookkeeping)
RMREF43C	N/A (bookkeeping)
RMREF44A	N/A (bookkeeping)
RMREF44B	N/A (bookkeeping)
RMREF44C	N/A (bookkeeping)

Variable omitted	Variable to use
RNID	HABCID
RNACROS	N/A (confidentiality)
RNCONTAC	VISITYPE
RNREF45A	N/A (bookkeeping)
RNREF45B	N/A (bookkeeping)
RNREF45C	N/A (bookkeeping)
RNREF45D	N/A (bookkeeping)
RNREF45E	N/A (bookkeeping)
RNREF45F	N/A (bookkeeping)
RNREF46A	N/A (bookkeeping)
ROID	HABCID
ROACROS	N/A (confidentiality)
ROCONTAC	VISITYPE
RPID	HABCID
RPACROS	N/A (confidentiality)
RPCONTAC	VISITYPE
RUID	HABCID
RUACROS	N/A (confidentiality)
RUCONTAC	VISITYPE
RVID	HABCID
RVACROS	N/A (confidentiality)
RVCONTAC	VISITYPE
RWID	HABCID
RWACROS	N/A (confidentiality)
RWCONTAC	VISITYPE
RXID	HABCID
RXACROS	N/A (confidentiality)
RXCONTAC	VISITYPE
RXADDYN	N/A (bookkeeping)
RXSSESPY	N/A (bookkeeping)
RYID	HABCID
RYACROS	N/A (confidentiality)
RYCONTAC	VISITYPE
RYCIYN	N/A (bookkeeping)
RYKNOK	N/A (bookkeeping)
RYKYN	N/A (bookkeeping)
RZID	HABCID
RZACROS	N/A (confidentiality)
RZCONTAC	VISITYPE
RZPPOA	N/A (bookkeeping)
RZPAYN	N/A (bookkeeping)
RZP2YN	N/A (bookkeeping)
SIID	HABCID

Variable omitted	Variable to use
S1ACROS	N/A (confidentiality)
S1CONTAC	VISITYPE
S1C1YN	N/A (bookkeeping)
T4ID	HABCID
T4ACROS	N/A (confidentiality)
T4CONTAC	VISITYPE
T4FNM	N/A (confidentiality)
T4LNM	N/A (confidentiality)
T4REASON	N/A (bookkeeping)
YAID	HABCID
YAACROS	N/A (confidentiality)
YACONTAC	VISITYPE
YAVISIT	N/A (bookkeeping)
YALINK	N/A (bookkeeping)
YAMOVE	N/A (confidentiality)
YAREF11A	N/A (bookkeeping)
YAREF11B	N/A (bookkeeping)
YAREF11C	N/A (bookkeeping)
YAREF12A	N/A (bookkeeping)
YAREF12B	N/A (bookkeeping)
YAREF12C	N/A (bookkeeping)
YAREF13A	N/A (bookkeeping)
YAREF13B	N/A (bookkeeping)
YAREF13C	N/A (bookkeeping)
YAREF14A	N/A (bookkeeping)
YAREF14B	N/A (bookkeeping)
YAREF14C	N/A (bookkeeping)
YAREF15A	N/A (bookkeeping)
YAREF15B	N/A (bookkeeping)
YAREF15C	N/A (bookkeeping)
YAREF16A	N/A (bookkeeping)
YAREF16B	N/A (bookkeeping)
YAREF16C	N/A (bookkeeping)
YAREF17A	N/A (bookkeeping)
YAREF17B	N/A (bookkeeping)
YAREF17C	N/A (bookkeeping)
YAREF17D	N/A (bookkeeping)
YAREF17E	N/A (bookkeeping)
YAREF17F	N/A (bookkeeping)
YAREF18A	N/A (bookkeeping)

Appendix II

138-MONTH CALCULATED VARIABLES

Variable	Variable Description	Grouping
CSAINDEX	Climbing stairs ability index	Self-Reported Function Calculated Vars.
EASE1F	Ease climbing 1 flight	Self-Reported Function Calculated Vars.
EASE1M	Ease walking 1 mile	Self-Reported Function Calculated Vars.
EASE2F	Ease climbing 2 flights	Self-Reported Function Calculated Vars.
EASEQM	Ease walking 1/4 mile	Self-Reported Function Calculated Vars.
WKAINDEX	Walking ability index	Self-Reported Function Calculated Vars.

Documentation for Year 12 Calculated Variable Dataset

YEAR 12 CALCULATED VARIABLE DATASET (Y12Calc).....24

- Appendix I Calculated Variable List
- Appendix II Self-Reported Function
- Appendix III Smoking Habits
- Appendix IV Physical Activity and Exercise

YEAR 12 CALCULATED VARIABLE DATASET (Y12Calc)

1. General description

The Y12Calc contains 26 variables derived (calculated) as described below. The documentation is grouped by the type of data from which the variables are calculated (self-reported function, smoking habits, etc.), and the variables are positioned in the dataset in the same groupings.

There are 3075 observations in the Y12Calc file.

2. Cross reference of dataset names with exact source

A complete list of variable names can be found under the “Proc Contents for All Datasets” link (search under Y12Calc) and in Appendix I. NOTE: SOME CALCULATED VARIABLES USE THE SAME VARIABLE NAMES EVERY YEAR. THE ANALYST IS REMINDED TO RENAME VARIABLES BEFORE MERGING WITH OTHER YEARS AS NECESSARY.

3. Dataset structure and contents

The Y12Calc file contains a single observation per participant.

Key variables:

HABCID	Health ABC Enrollment ID# without the 2-letter prefix
GENDER	Gender (1=Male; 2=Female)*
RACE	Race (1=White; 2=Black)*
RACEGEN	Race-gender group (1=White male, 2=White female, 3=Black male, 4=Black female)*
SITE	Clinic Site (1=Memphis, 2=Pittsburgh)*

4. Condition of data

a. Strengths and weaknesses of dataset items: When a calculated variable includes imputations for missing variables, the method of imputation is **bolded** in the documentation. There were no major logic changes used to derive calculated variables for year 12 as compared with previous years.

5. Dataset index formulation and key variable mapping

The Y12Calc file is sorted by HABCID, which is a unique identifier for each participant.

* Must link to Health ABC participant history file (PH) to add this variable.

6. General strategies for manipulating and merging the data

Because the Health ABC datasets are sorted by Health ABC Enrollment ID, the HABCID variable is most useful for merging with other datasets.

Appendix I
CALCULATED VARIABLE LIST

Variable	Variable Description	Grouping
EASE1F	Ease climbing 1 flight	Self-reported function
EASE2F	Ease climbing 2 flights	Self-reported function
CSAINDEX	Climbing stairs ability index	Self-reported function
EASE10P	Ease lift/carry 10 lbs	Self-reported function
EASE20P	Ease lift/carry 20 lbs	Self-reported function
LCAINDEX	Lift/carry ability index	Self-reported function
EASEUP	Ease standing from chair without using arms	Self-reported function
SMK12	Smoking status at Year 12 visit	Smoking Habits
Y12MCKKWK	Kcal/kg/week doing major chores	Physical Activity
Y12FSKKWK	Kcal/kg/week climbing stairs	Physical Activity
Y12TWKKWK	Kcal/kg/week total walking	Physical Activity
WALKTIME	Minutes walking/week	Physical Activity
WALKCAT	Minutes walking/week category	Physical Activity
BKTWTIME	Minutes walking briskly/week	Physical Activity
BRISK90	Walks briskly \geq 90 min/week	Physical Activity
BRISK180	Walks briskly \geq 180 min/week	Physical Activity
WSKKWK	Kcal/kg/week - walking + stairs	Physical Activity

Appendix II
Self-Reported Function Calculated Variables

Investigator Name: Eleanor Simonsick

Variable	General Description	Detailed Description	How variable is calculated	How to handle missing or special values	Value labels
EASEQM	Ease walking 1/4 mile	Recode of Y12DWQMEZ where 3 is easiest; imputations applied for missing Y12DWQMEZ	If Y12DWQMYN in (0,7,8) then EASEQM=7-Y12DWQMEZ If Y12DWQMYN=1 then EASEQM=4-Y12DWQMDF	IF Y12DWQMYN=9 AND (Y12MNRS>0 OR Y12MNRS4 IN(1,2)) THEN Y12DWQMYN=1; IF Y12DWQMEZ=8 and Y12DWQMYN ne 8 THEN Y12DWQMEZ=2; IF Y12DWQMDF=8 and Y12DWQMYN ne 8 THEN Y12DWQMDF=2; IF (Y12DWQMYN=0 AND Y12DWQMEZ<0) THEN EASEQM=7-Y12DW1MEZ; IF (Y12DWQMYN=0 AND Y12DWQMEZ<0) AND Y12DW1MEZ<0 AND (Y12DW1MYN=1 OR Y12DW1MYN=8 OR Y12DW1MYN<=.z) THEN EASEQM=4; IF (Y12DWQMYN=0 AND Y12DWQMEZ<0) AND (Y12DW1MYN=0 AND Y12DW1MEZ<=.z) THEN EASEQM=5; IF Y12DWQMYN=1 AND Y12DWQMDF<0 THEN EASEQM=7-Y12DWQMEZ; IF Y12DWQMYN=1 AND Y12DWQMDF<0 AND Y12DWQMEZ<0 THEN EASEQM=2; IF Y12DWQMYN<0 AND Y12DWQMDF>0 THEN EASEQM=4-Y12DWQMDF; IF Y12DWQMYN<0 AND Y12DWQMEZ>0 THEN EASEQM=7-Y12DWQMEZ; IF Y12DWQMYN=8 and Y12DWQMEZ=8 then EASEQM=4; IF Y12DWQMYN=9 AND Y12MNRS<0 AND (Y12MNRS4<0 OR Y12MNRS4=8) THEN EASEQM=.	6=very easy 5=somewhat easy 4=not that easy 3=a little difficult 2=somewhat difficult 1=very difficult 0=unable to do

Self-Reported Function Calculated Variables

Variable	General Description	Detailed Description	How variable is calculated	How to handle missing or special values	Value labels
EASE1M	Ease walking 1 mile	Recode of Y12DW1MEZ where 3 is easiest; imputations applied for missing Y12DW1MEZ	If Y12DW1MYN=1 then EASE1M=0; if Y12DW1MEZ=3 then EASE1M=1; if Y12DW1MEZ=2 then EASE1M=2; if Y12DW1MEZ=1 then EASE1M=3	<p>IF Y12DW1MYN=8 AND (Y12DW1MEZ<0 OR Y12DW1MEZ=8) THEN EASE1M=1; IF Y12DW1MYN<=.z AND Y12DW1MEZ<=.z AND EASEQM=4 THEN EASE1M=0; IF Y12DW1MYN<=.z AND Y12DW1MEZ<=.z AND EASEQM=5 THEN EASE1M=1; IF Y12DW1MYN<=.z AND Y12DW1MEZ<=.z AND EASEQM=6 THEN EASE1M=2; IF Y12DW1MYN=0 AND Y12DW1MEZ<=.z AND EASEQM=4 THEN EASE1M=0; IF Y12DW1MYN=0 AND Y12DW1MEZ<=.z AND EASEQM=5 THEN EASE1M=1; IF Y12DW1MYN=0 AND Y12DW1MEZ<=.z AND EASEQM=6 THEN EASE1M=2; IF Y12DWQMYN=1 THEN EASE1M=0;</p> <p>IF EASEQM GT 0 AND ((EASEQM-3) LT EASE1M) THEN DO; IF EASEQM LE 3 THEN EASE1M=0; ELSE EASE1M=EASEQM-3; END;</p> <p>if Y12DW1MYN=8 and Y12DW1MEZ in (1,2,3) then EASE1M=4-Y12DW1MEZ; if Y12DW1MYN=0 and Y12DW1MEZ=8 then EASE1M=EASEQM-4;</p>	3=very easy 2=somewhat easy 1=not that easy 0=difficult
WKAINDEX	Walking ability index	Summary measure of self-reported walking ability.	WKAINDEX=EASEQM + EASE1M	Imputed version of component variables used	This scale ranges from 0 (unable) to 9 (very easy)

Self-Reported Function Calculated Variables

Variable	General Description	Detailed Description	How variable is calculated	How to handle missing or special values	Value labels
EASE1F	Ease Climbing 1 flight	Recode of Y12DW10EZ where 3 is easiest; imputations applied for missing Y12DW20EZ	EASE1F=7-Y12DW10EZ; IF Y12DW10YN=1 THEN EASE1F=4-Y12DIF;	IF Y12DW10YN=9 then EASE1F=.; IF Y12DW10EZ=8 THEN Y12DW10EZ=2; IF Y12DIF=8 THEN Y12DIF=2; IF Y12DW10EZ<0 THEN EASE1F=7-Y12DW20EZ; IF Y12DW10YN=1 AND (Y12DIF=8 OR Y12DIF<0) THEN EASE1F=2; IF Y12DW10YN=0 AND Y12DW10EZ<0 AND Y12DW10EZ<0 THEN EASE1F=4; IF EASE1F=. AND Y12DW10YN NE 9 AND Y12DW10YN>0 AND (Y12DW20YN=1 OR Y12DW20YN=8 OR Y12DW20YN<=.z) THEN EASE1F=4; IF EASE1F=. AND Y12DW20YN=0 AND Y12DW20EZ<=.z THEN EASE1F=5;	6=very easy 5=somewhat easy 4=not that easy 3=a little difficult 2=somewhat difficult 1=very difficult 0=unable to do
EASE2F	Ease Climbing 2 flights	Recode of Y12DW20EZ where 3 is easiest; imputations applied for missing Y12DW20EZ	IF Y12DW20YN=1 THEN EASE2F=0; ELSE IF Y12DW20EZ=3 THEN EASE2F=1; ELSE IF Y12DW20EZ=2 THEN EASE2F=2; ELSE IF Y12DW20EZ=1 THEN EASE2F=3;	IF Y12DW10YN=9 then EASE1F=.; IF (Y12DW20YN=8 OR Y12DW20YN<0)AND (Y12DW20EZ<0 OR Y12DW20EZ=8) AND EASE1F>0 THEN EASE2F=1; IF Y12DW20YN<=.z AND Y12DW20EZ<=.z AND EASE1F=4 THEN EASE2F=0; IF Y12DW20YN<=.z AND Y12DW20EZ<=.z AND EASE1F=5 THEN EASE2F=1; IF Y12DW20YN<=.z AND Y12DW20EZ<=.z AND EASE1F=6 THEN EASE2F=2; IF Y12DW20YN=0 AND (Y12DW20EZ<=.z OR Y12DW20EZ=8) AND EASE1F>0 THEN EASE2F=EASE1F-4; IF Y12DW10YN=1 THEN EASE2F=0; IF .z<EASE1F<=3 THEN EASE2F=0; ELSE IF (EASE1F-3) LT EASE2F THEN EASE2F=EASE1F-3;	3=very easy 2=somewhat easy 1=not that easy 0=difficult
CSAINDEX	Climbing stairs ability index	Summary measure of self-reported walking ability.	CSAINDEX=EASE1F + EASE2F	Imputed version of component variables used	This scale ranges from 0 (unable) to 9 (very easy)

```

*****
***
*   Saved as:
*   \\Fu-hsing-c\HABC\HABC_SAS\Calculated Variables\Programs\Year
12\srfcn_Y12.sas
*
*   Todd Glasser 01/06/2011 - Modified year 11 program for use in Year
12
*****
* ;

*****
* ;
**  HABC Year 12 SELF-REPORT PHYSICAL FUNCTION CALC VAR CODE
** ;
**
** ;
**  Creates the following variables:
** ;
**      EASEQM:   ease of walking 1/4 mile
** ;
**      EASE1M:   ease of walking 1 mile
** ;
**      EASE1F:   ease of climbing 1 flight
** ;
**      EASE2F:   ease of climbing 2 flights
** ;
**      EASE10P:  ease of lifting/carrying 10 lbs
** ;
**      EASE20P:  ease of lifting/carrying 20 lbs
** ;
**      EASEUP:   ease of standing up from chair w/o using arms
** ;
**      CSAINDEX: climbing stairs ability index
** ;
**      WKAINDEX: walking ability index
** ;
**      LCAINDEX: lift/carry ability index
** ;
**
** ;
** Adapted from FY12n Harris Y2 code: srfcn_y2.sas
** ;
*****
* ;

*DM LOG 'CLEAR' ; *DM OUTPUT 'CLEAR' ;

%include '\\Fu-hsing-c\HABC\HABC_SAS\Programs\initV8_ase.sas';
libname calc '\\Fu-hsing-c\habc\habc_sas\calculated
variables\datasets\Year 12';

*options ls=90 ps=56 formchar='|----|+|----+=|-\<>*' nodate nofmterr;

```

```

data calc.Y12srfcn (keep=HABCID EASE1F EASE2F EASEQM EASE1M EASE10P
EASE20P EASEUP CSAINDEX WKAINDEX
                    LCAINDEX);
    ***Y12 data***;
/*
    merge habcrts.r2 (keep=habcid r2contac where=(r2contac=20))
                    habcrts.Y12 (keep=habcid Y12contac Y12dwqmyn
Y12dwqmdfY124mnrs wherY12(r4contac=20))
                    habcrts.Y12 (keep=habcid Y12contac Y12dwqmez
Y12dwlmyY125dwlmez wherY12(r5contac=20))
                    habcrts.Y12 (keep=habcid Y12contac Y12dwl0yn
Y12dif where=Y126contac=20))
                    habcrts.Y12 (keep=habcid Y12contac Y12dwl0ez
Y12dw20ynY127dw20ez wherY12(r7contac=20))
                    habcrts.Y12 (keep=habcid Y12contac Y12difsta
Y12ezstaY128dstamt wherY12(r8contac=20))
                    habcrts.Y12 (keep=habcid Y12contac Y12dif10
Y12d10amtY12aez10lb rad201Y12 raez201b whY12e=(racontac=20)) ;
*/
    set daf.y12visit (keep=habcid Y12dwqmyn Y12dwqmdf Y12mnrs
                    Y12dwqmez Y12dwlmyY12dwlmez
                    Y12dwl0yn Y12dif
                    Y12dwl0ez Y12dw20yn Y12dw20ez
                    Y12difsta Y12ezsta Y12dstamt
                    Y12dif10 Y12d10amt Y12ez10lb Y12d20lbs
Y12ez20lb) ;
    by habcid ;
    LABEL EASEQM = 'EASE WALKING 1/4 MILE, 6=VERY EASY'
EASE1M = 'EASE WALKING 1 MILE, 3=VERY EASY'
WKAINDEX = 'WALKING ABILITY INDEX, 9=BEST'
EASE1F = 'EASE CLIMBING 1 FLIGHT, 6=VERY EASY'
EASE2F = 'EASE CLIMBING 2 FLIGHTS, 3=VERY
EASY'
CSAINDEX = 'CLIMBING STAIRS ABILITY INDEX,
9=BEST'
EASE10P = 'EASE LIFT/CARRY 10 LBS, 6=VERY EASY'
EASE20P = 'EASE LIFT/CARRY 20 LBS, 3=VERY EASY'
LCAINDEX = 'LIFT/CARRY ABILITY INDEX, 9=BEST'
EASEUP = 'EASE STANDING FROM CHAIR WITHOUT
USING ARMS, 6=VERY EASY' ;

    *EASEQM: CODE FOR EASE OF WALKING 1/4 MILE
    IF EASE LEVEL MISSING, EASE LEVEL WALKING 1 MILE WAS ASSIGNED.
    IF NO EASE LEVEL FOR WALKING
        1M WAS CODED AND Y12dwlmyY12dwlmez WAS YES, DONT KNOW, OR MISSING THEN
        NOT THAT EASY WAS ASSIGNED. IF NO
            EASE LEVEL FOR WALKING 1M WAS CODED AND Y12dwlmyY12dwlmez WAS NO AND
            Y12dwlmez WAS MISSING THEN SOMEWHAT
            EASY WAS ASSIGNED. THE 4 CASES WHO REPORTED DIFFICULTY WERE
            RECODED TO NOT THAT EASY FOR QM
            AND DIFFICULTY FOR 1M;

    IF Y12dwqmyn=9 AND (Y12mnrs>0 and Y12mnrs ne 22) THEN
Y12dwqmyn=1;
    IF Y12dwqmez=8 and Y12dwqmyn ne 8 THEN Y12dwqmez=2;
    IF Y12dwqmdf=8 and Y12dwqmyn ne 8 THEN Y12dwqmdf=2;
    IF Y12dwqmyn IN (0,7,8) AND Y12dwqmez<0 THEN EASEQM=4;

```

```

ELSE IF Y12dwqmyn IN (0,7,8) THEN EASEQM=7-Y12dwqmez;
IF (Y12dwqmyn=0 AND Y12dwqmez<0) THEN EASEQM=7-Y12dwlmez;
IF (Y12dwqmyn=0 AND Y12dwqmez<0) & Y12dwlmez<0 & (Y12dwlmyn
in(1,8) OR
      Y12dwlmyn<=.z) THEN EASEQM=4;
IF (Y12dwqmyn=0 & Y12dwqmez<0) & (Y12dwlmyn=0 & Y12dwlmez<=.z)
THEN EASEQM=5;
IF Y12dwqmyn=1 THEN EASEQM=4-Y12dwqmdf;

IF Y12dwqmyn=1 AND Y12dwqmdf<0 THEN EASEQM=7-Y12dwqmez;
IF Y12dwqmyn=1 AND Y12dwqmdf<0 AND Y12dwqmez<0 THEN EASEQM=2;
IF Y12dwqmyn<0 AND Y12dwqmdf>0 THEN EASEQM=4-Y12dwqmdf;
IF Y12dwqmyn<0 AND Y12dwqmez>0 THEN EASEQM=7-Y12dwqmez;

if Y12dwqmyn=8 and Y12dwqmez=8 then EASEQM=4;
IF Y12dwqmyn=9 AND (Y12mnrs<0 or Y12mnrs=22) THEN EASEQM=.;

```

```

*EASE1M: CODE FOR EASE OF WALKING 1 MILE. AS SELF-REPORTED
WALKING ABILITY IS A KEY DEPENDENT
MEASURE IN HEALTH ABC, MISSING VALUES WERE RECODED TO NON-
MISSING WHENEVER A REASON
GUESS COULD BE MADE BASED ON OTHER AVAILABLE INFORMATION. SEE
BELOW FOR CODE AND
RECODING RULES FOR MISSING VALUES;

```

```

IF Y12dwlmyn=1 THEN EASE1M=0;
ELSE IF Y12dwlmez=3 THEN EASE1M=1;
ELSE IF Y12dwlmez=2 THEN EASE1M=2;
ELSE IF Y12dwlmez=1 THEN EASE1M=3;

```

```

*MISSING VALUE RECODES;

```

```

IF Y12dwlmyn=8 AND (Y12dwlmez<0 OR Y12dwlmez=8) THEN EASE1M=1;
IF Y12dwlmyn<=.z AND Y12dwlmez<=.z AND EASEQM=4 THEN EASE1M=0;
IF Y12dwlmyn<=.z AND Y12dwlmez<=.z AND EASEQM=5 THEN EASE1M=1;
IF Y12dwlmyn<=.z AND Y12dwlmez<=.z AND EASEQM=6 THEN EASE1M=2;
IF Y12dwlmyn=0 AND Y12dwlmez<=.z AND EASEQM=4 THEN EASE1M=0;
IF Y12dwlmyn=0 AND Y12dwlmez<=.z AND EASEQM=5 THEN EASE1M=1;
IF Y12dwlmyn=0 AND Y12dwlmez<=.z AND EASEQM=6 THEN EASE1M=2;
IF Y12dwqmyn=1 THEN EASE1M=0;

```

```

*IN SOME CASES THE AMOUNT OF EASE REPORTED FOR WALKING 1 MILE
WAS GREATER THAN
THE AMOUNT OF EASE REPORTED FOR 1/4 MILE. THE CODE BELOW
DOWNGY12DES THE EASE
LEVEL REPORTED FOR 1M TO THAT REPORTED FOR 1/4M;

```

```

IF EASEQM GT 0 AND ((EASEQM-3) LT EASE1M) THEN DO;
  IF EASEQM LE 3 THEN EASE1M=0; ELSE
  EASE1M=EASEQM-3;
END;

```

```

if Y12dwlmyn=8 and Y12dwlmez in (1,2,3) then EASE1M=4-
Y12dwlmez;
if Y12dwlmyn=0 and Y12dwlmez=8 then EASE1M=EASEQM-4;

```

```

*CODE TO CREATE A SUMMARY MEASURE OF SELF-REPORTED WALKING
ABILITY. THIS
SCALE WILL ULTIMATELY Y12NGE FROM 0 TO 9, BUT AT BASELINE IT
Y12NGES FROM
4-9, AS NO ONE HAS DIFFICULTY. THIS SCALE CAN BE CREATED IN TWO
WAYS: 1) KEYING
OFF THE RESPONSE TO EASE1M AND ONLY LOOKING AT EASEQM WHEN
DIFFICULTY IS
REPORTED FOR 1M OR 2) SUMMING THE RESPONSES TO THE TWO
QUESTIONS. I
DID IT BOTH WAYS AND FOUND THE LATTER APPROACH SUPERIOR;

```

```

WKAINDEX=EASEQM + EASE1M;

```

```

*CODE FOR EASE OF CLIMBING STAIRS;

```

```

*EASE1F;

```

```

IF Y12dw10yn=9 THEN DO; EASE1F=.; EASE2F=.; END;
IF Y12dw10ez=8 THEN Y12dw10ez=2;
IF Y12dif=8 THEN Y12dif=2;
EASE1F=7-Y12dw10ez;
IF Y12dw10ez<0 THEN EASE1F=7-Y12dw20ez;
IF Y12dw10yn=1 THEN EASE1F=4-Y12dif;
IF Y12dw10yn=1 AND (Y12dif=8 OR Y12dif<0) THEN EASE1F=2;
IF Y12dw10yn=0 AND Y12dw10ez<0 AND Y12dw10ez<0 THEN EASE1F=4;

IF EASE1F=. & Y12dw10yn NE 9 & Y12dw10yn>0 & (Y12dw20yn=1 OR
Y12dw20yn=8 OR Y12dw20yn<=.z ) THEN EASE1F=4;
IF EASE1F=. AND Y12dw20yn=0 AND Y12dw20ez<=.z THEN EASE1F=5;

```

```

*EASE2F;

```

```

IF Y12dw20yn=1 THEN EASE2F=0;
ELSE IF Y12dw20ez=3 THEN EASE2F=1;
ELSE IF Y12dw20ez=2 THEN EASE2F=2;
ELSE IF Y12dw20ez=1 THEN EASE2F=3;
IF (Y12dw20yn=8 OR Y12dw20yn<0) AND (Y12dw20ez<0 OR Y12dw20ez=8)
AND EASE1F>0 THEN EASE2F=1;
IF Y12dw20yn<=.z AND Y12dw20ez<=.z AND EASE1F=4 THEN EASE2F=0;
IF Y12dw20yn<=.z AND Y12dw20ez<=.z AND EASE1F=5 THEN EASE2F=1;
IF Y12dw20yn<=.z AND Y12dw20ez<=.z AND EASE1F=6 THEN EASE2F=2;
IF Y12dw20yn=0 AND (Y12dw20ez<=.z OR Y12dw20ez=8) AND EASE1F>0
THEN EASE2F=EASE1F-4;
IF Y12dw10yn=1 THEN EASE2F=0;

IF .z<EASE1F<=3 THEN EASE2F=0; ELSE
IF (EASE1F-3) LT EASE2F THEN EASE2F=EASE1F-3;

```

```

*CODE TO CREATE A SUMMARY MEASURE OF SELF-REPORTED STAIR
CLIMBING ABILITY. THIS SCALE WILL
Y12NGE FROM 0 TO 9, BUT AT BASELINE IT Y12NGES FROM 4-9;

```

```

CSAINDEX=EASE1F + EASE2F;

*CODE FOR LIFTING/CARRYING;

IF Y12dif10=0 AND (Y12ez10lb=8 OR Y12ez10lb<0) THEN EASE10P=5;
ELSE IF Y12dif10 IN (0,7,8) and Y12ez10lb ne 8 THEN EASE10P=7-
Y12ez10lb;
else if Y12dif10 in (0,7,8) and Y12ez10lb=8 then EASE10P=4;
IF Y12dif10=1 AND (Y12d10amt=8 OR Y12d10amt<0) THEN EASE10P=.;
ELSE IF Y12dif10=1 THEN EASE10P=4 - Y12d10amt;
IF Y12dif10<=.z & Y12d10amt>0 & Y12ez10lb<0 THEN EASE10P=4 -
Y12d10amt;
IF Y12dif10<=.z & Y12d10amt<0 & Y12ez10lb>0 THEN EASE10P=7 -
Y12ez10lb;
IF EASE10P=. & Y12d20lbs NE 1 & Y12d20lbs NE 8 THEN EASE10P=7-
Y12ez20lb;
IF EASE10P=. & Y12d20lbs=0 & (Y12ez20lb LT 0 OR Y12d20lbs=8)
THEN EASE10P=6;
IF EASE10P=. AND Y12dif10=0 AND Y12d20lbs=1 THEN EASE10P=4;
IF EASE10P=. AND Y12dif10=8 AND Y12d20lbs=1 THEN EASE10P=4;
IF EASE10P=. & Y12dif10 IN (0,8) & (Y12d20lbs<0 OR Y12d20lbs=8)
&
(Y12ez20lb<0 OR Y12ez20lb=8) THEN EASE10P=4;
IF EASE10P=. AND Y12dif10=0 AND Y12ez10lb<=.z AND Y12d10amt>0
THEN
EASE10P=4 - Y12d10amt;

*IF Y12ez20lb=8 and Y12d20lbs ne 8 THEN Y12ez20lb=2;
EASE20P=4-Y12ez20lb;
if Y12d20lbs=8 and (Y12ez20lb=8 or Y12ez20lb<0) then EASE20P=1;
if Y12d20lbs=0 and Y12ez20lb=8 then EASE20P=EASE10P-4;

IF Y12d20lbs=1 THEN EASE20P=0;
IF (0 LE EASE10P LE 3) THEN EASE20P=0;

IF (Y12d20lbs<=.z OR Y12d20lbs=0) and (Y12ez20lb<=.z OR
Y12ez20lb=8)
AND EASE10P=4 THEN EASE20P=0;
IF (Y12d20lbs=0 OR Y12d20lbs<=.z) AND (Y12ez20lb<=.z OR
Y12ez20lb=8)
AND EASE10P=6 THEN EASE20P=2;
IF (Y12d20lbs=0 OR Y12d20lbs<=.z) AND (Y12ez20lb<=.z OR
Y12ez20lb=8)
AND EASE10P=5 THEN EASE20P=1;

IF (EASE10P-3) LT EASE20P THEN DO;
EASE20P=EASE10P-3;
IF EASE20P<0 THEN EASE20P=0;
END;

*CODE FOR SUMMARY INDEX;

LCAINDEX=EASE10P + EASE20P;

```

```

*OTHER FUNCTION MEASURES;

IF Y12ezsta=8 and Y12difsta ne 8 THEN Y12ezsta=2;
IF Y12dstamt=8 and Y12difsta ne 8 THEN Y12dstamt=2;
IF Y12difsta IN (0,7,8) THEN EASEUP=7 - Y12ezsta;
IF Y12difsta=1 THEN EASEUP=4 - Y12dstamt;
if Y12difsta=8 and Y12ezsta=8 then EASEUP=4;
IF Y12difsta<=.z AND Y12dstamt>0 AND Y12ezsta<0 THEN EASEUP=4 -
Y12dstamt;
IF Y12difsta=8 AND Y12ezsta<0 AND Y12dstamt<0 THEN EASEUP=4;
IF Y12difsta=0 AND Y12ezsta<=.z THEN EASEUP=5;
IF Y12difsta=0 & Y12ezsta<=.z & Y12dstamt >0 THEN EASEUP=4 -
Y12dstamt;

*Formats;
FORMAT EASE1F EASE2F EASEQM EASE1M EASE10P EASE20P EASEUP
CSAINDEX WKAINDEX
LCAINDEX SPMISS. /*TIRED1F TIREDQM YNDK.*;/

run ;

%macro skip ;

proc freq data=calc.Y12srfcn ;
tables EASE1F EASE2F EASEQM EASE1M EASE10P EASE20P EASEUP
CSAINDEX WKAINDEX
LCAINDEX ;
* tables ease1m * Y12dwlmyrn * Y12dwqmyn * Y12dwlmez / list
missing nocum ;
* tables Y12dwqmyn Y12dwqmdf Y12mnrs ;
* tables Y12dwqmyn * (Y12dwqmdf Y12mnrs) / list missing ;
* tables (Y12dwqmez Y12dwqmdf) * Y12dwqmyn / list missing ;
* tables Y12dwqmyn * Y12dwqmez * Y12dwlmez / list missing ;
* tables Y12dwqmyn * Y12dwqmdf * Y12dwqmez / list missing nocum ;
* tables easeqm ;
* tables
EASEQM*Y12dwqmyn*Y12dwqmez*Y12dwqmdf*Y12dwlmyrn*Y12dwlmez*Y12mnrs / list
missing nocum ;
* tables Y12dwlmyrn*Y12dwlmez*EASEQM*Y12dwqmyn / list missing
nocum ;
* tables EASE1M*Y12dwlmyrn*Y12dwlmez*EASEQM*Y12dwqmyn / list
missing nocum ;
* tables WKAINDEX*EASEQM*EASE1M / list missing ;
* tables Y12dw10yn*Y12dw10ez*Y12dw20yn*Y12dw20ez*Y12dif / list
missing ;
* tables EASE1F*Y12dw10yn*Y12dw10ez*Y12dw20yn*Y12dw20ez*Y12dif /
list missing ;
* tables EASE2F*Y12dw20yn*Y12dw20ez*EASE1F / list missing ;
* tables CSAINDEX*EASE1F*EASE2F / list missing ;
* tables TIRED1F*FADW10WX ;
* tables EASE10P*Y12dif10*Y12ez10lb*Y12d10amt*Y12d20lbs*Y12ez20lb
/ list missing ;
* tables EASE20P*Y12ez20lb*Y12d20lbs*EASE10P / list missing ;
* tables LCAINDEX*EASE10P*EASE20P / list missing ;
* tables EASEUP*Y12difsta*Y12ezsta*Y12dstamt / list missing ;

```

```

*      tables EASEHHW*Y12difHW*FAEZHW*FADHWAMT / list missing ;
*      tables Y12dwqmez Y12dwlmyN Y12dwlmez ;
*      tables Y12dw10yn Y12dif ;
*      tables Y12dw10ez Y12dw20yn Y12dw20ez ;
*      tables Y12difsta ;
*      tables Y12dif10 Y12d10amt Y12ez10lb Y12d20lbs Y12ez20lb ;
*      format EASEQM Y12dwqmyn Y12dwqmez Y12dwqmdf Y12dwlmyN Y12dwlmez
Y12mnrs
      EASE1M WKAINDEX /*FADWQMT2*/
      EASE1F Y12dw10yn Y12dw10ez Y12dw20yn Y12dw20ez Y12dif
      EASE2F CSAINDEX /*TIRED1F FADW10WX */
      EASE10P Y12dif10 Y12ez10lb Y12d10amt Y12d20lbs Y12ez20lb
      EASE20P LCAINDEX
      /*EASEHHW Y12difHW FAEZHW FADHWAMT*/;
      FORMAT EASE1F EASE2F EASEQM EASE1M EASE10P EASE20P EASEUP
CSAINDEX WKAINDEX
      LCAINDEX SPMISS. /*TIRED1F TIREDQM YNDK.*//;
run ;

proc contents data=calc.Y12srfcn ;
  title 'Year 12 - Self Reported Physical Functioning' ;
run ;
%mend skip ;
%skip;

***check for duplicates***;
data dupes;
  set calc.Y12srfcn;
  by habcid;
  if not(first.habcid and last.habcid);
run;

proc print;
title4 'Duplicates in SRFCN Y12';
run;

/*
libname calc_y10 '\\Fu-hsing-c\habc\habc_sas\calculated
variables\datasets\Year 10';
proc freq data=calc_y10.Y10srfcn ;
  tables EASE1F EASE2F EASEQM EASE1M EASE10P EASE20P EASEUP
CSAINDEX WKAINDEX
      LCAINDEX ;
  title 'Year 10 - Self Reported Physical Functioning' ;
  title2 "For comparison";
run ;
proc contents data=calc_y10.Y10srfcn ;
run ;
*/

```

Appendix III
Smoking Calculated Variables

Using Clinic Visit, Home Visit and Annual Telephone Interview Variables

Investigator Name: Stephen Kritchevsky

Variable	Descriptive Title	Detailed Description	How variable is calculated	How to handle missing or special values	Value labels
SMK12	Smoking Status at Year 12	0 is never smoker; 1 is former smoker; 2 is current smoker	If Y12smoke=1 then SMK12=1; Else if Y12smoke=0 and smk1=0 and (smk3=0 or smk3 le .z) and (smk5=0 or smk5 le .z) and (smk8=0 or smk8 le .z) and (smk9=0 or smk9 le .z) (smk10=0 or smk10 le .z) (smk11=0 or smk11 le .z) then SMK12=0; Else if Y12smoke=0 and (smk1 in (1,2) or smk3 in (1,2) or smk5 in(1,2) or smk8 in (1,2) or smk9 in (1,2) or smk10 in (1,2) or smk11 in (1,2) then SMK12=2; Else if Y12smoke le .z or Y12smoke in (7,8) then SMK12=.A; Else SMK12=.M;	If Y12SMOKE ≤ .z or Y12SMOKE in (7,8) then SMK12=.A	0=Never 1=Current 2=Former

```

*****
***
* Create calculated variables requested by SKritchevsky for:
* SMK12      Smoking status at Year 12
*
*
* Todd Glasser 01/06/2011 - Modified year 11 program for use in Year
12
*****
* ;

*DM LOG 'CLEAR' ; *DM OUTPUT 'CLEAR' ;

%include '\\Fu-hsing-c\habc\habc_sas\programs\initV8_ase.sas';
libname calc '\\Fu-hsing-c\habc\habc_sas\calculated
variables\datasets\Year 12';

*options ls=141 ps=46 formchar='|----|+|---+=|-/\<>*' nodate nofmterr
pageno=1 nocenter;

***** Smoking status at year 12;
data Y12smoke ;
  set daf.y12visit (keep=habcid y12smoke);
run;

data bl; set current.y1calc (keep=habcid smk1);run;
data y3; set current.y3calc (keep=habcid smk3);run;
data cm; set current.y5calc (keep=habcid smk5);run;
data y8; set current.y8calc (keep=habcid smk8);run;
data y9; set current.y9calc (keep=habcid smk9);run;
data Y10; set current.Y10calc (keep=habcid smk10);run;
data Y11; set current.Y11calc (keep=habcid smk11);run;

data smoke;
  merge bl y3 cm y8 y9 y10 y11 y12smoke;
  by habcid;
  label SMK12 = 'Smoking status at Year 12 visit';
  format SMK12 ncfmt. ;
  if Y12smoke le .z or Y12smoke in (7,8) then SMK12=.A;
  else if Y12smoke=1 then SMK12=1;
  *** Use smoking status at prior years to determine never/former
status ***;
  else if Y12smoke=0 and smk1=0 and (smk3=0 or smk3 le .z) and (smk5=0
or smk5 le .z)
    and (smk8=0 or smk8 le .z) and (smk9=0 or smk9 le .z) or (smk10=0
or smk10 le .z) or (smk11=0 or smk11 le .z) then SMK12=0;
  else if Y12smoke=0 and (smk1 in (1,2) or smk3 in (1,2) or smk5
in(1,2) or smk8 in (1,2) or
    smk9 in (1,2) or smk10 in (1,2) or smk11 in (1,2)) then SMK12=2;
  else SMK12=.M;
  ** 0=never, 1=current, 2=former **;
run;

data calc.Y12smoke;
  set smoke (keep=habcid SMK12);
run;

```

```

%macro skip ;

proc contents data=calc.Y12smoke;
run;
proc freq data=calc.Y12smoke;
  tables SMK12;
run;
proc freq data=smoke;
  table smk12*Y12smoke / list missing;
  tables smk12*y12smoke*SMK10*smk9*smk8*smk5*smk3*smk1 / list missing
nopercent;
  format _all_;
run;

***check for duplicates***;
data dupes;
  set calc.Y12smoke;
  by habcid;
  if not(first.habcid and last.habcid);
run;
proc print data=dupes;
title4 'Duplicates in Y12Smoke';
run ;

%mend skip ;

```

Appendix IV
Physical Activity Calculated Variables
Using Clinic Visit, Home Visit and Annual Telephone Interview Variables

Investigator Name: Eleanor Simonsick

Step 1: assign met estimates to each activity

Met estimates derive from Ainsworth BE, Haskell WL, Leon AS, Jacobs DR, Montoye HJ, Sallis JF, Paffenbarger RS. Compendium of physical activities: classification of energy costs of human physical activities. Med Sci Sports Exerc 25(1):71-80.

Variable	General description	Detailed Description	How variable is calculated	How to handle missing or special values	Value labels
Y12FSKKWK	kcal/kg/week climbing stairs	Assigns 4.0 kcal/kg/hour of stair climbing plus an additional 1.0 kcal/kg/hour carrying a load (Y10 Qaire, page 12, Q19). (Assumes 1 flight up/down takes 30 seconds.)	If Y12FS7DAY=1 then Y12FSKKWK = 4.0 x Y12FSNUM/120 +1.0 x Y12FSLOAD/120; If Y12FS7DAY=0 then Y12FSKKWK=0;	Correction for outliers: If Y12FSNUM>210 then Y12FSNUM=210; if Y12FSLOAD>210 then Y12FSLOAD=210; If Y12FS12MO≤.z and Y12FS7DAY≤.z then Y12FSKKWK=.; if (Y12FS12MO in (0,7,8)) and (Y12FS7DAY≤.z or Y12FS7DAY=8) then Y12FSKKWK=0; if Y12FS12MO=1 and (Y12FS7DAY≤.z or Y12FS7DAY=8) then Y12FSKKWK=0; if Y12FSKKWK<0 then do: if Y12FS7DAY=1 and Y12FSNUM > 0 and (Y12FSLOAD≤.z or Y12FSLODK=-1) then Y12FSLOAD=0; if Y12FS7DAY=1 and (Y12FSNUM≤.z or Y12FSNUMD=-1) and Y12FSLOAD > 0 then Y12FSNUM=Y12FSLOAD	kkcal/kg/week

Variable	General description	Detailed Description	How variable is calculated	How to handle missing or special values	Value labels
Y12MCKKWK	kcal/kg/week doing major chores	Assigns 3.5 kcal/kg/hour doing major chores (Y10 Qaire, page 11, Q17). Note: this question combines activities listed separately at baseline. Y12MCKKWK is therefore equivalent to the sum of baseline variables FPPAKKWK and FPHCKKWK.	3.5 x Y12HCTIM/60 if Y12HC7DAY=0 then Y12MCKKWK=0;	If Y12HC12MO≤.z and Y12HC7FAY≤.z then Y12MCKKWK=.; if (Y12HC12MO=0 or Y12HC12MO in (7,8) and (Y12HC7DAY≤.z or Y12HC7DAY=0) then Y12MCKKWK=0; if Y12HC12MO=1 and Y12HC7DAY≤.z then Y12MCKKWK=0; if Y12MCKKWK<0 then do: if Y12HC12MO=1 and Y12HC7DAY=1 and (Y12HCTIM≤.z or Y12HCDK=-1) then Y12HCTIM=120 (median value for non-missings at baseline; used here for consistency)	kkcal/kg/week
Y12TWKKWK	kcal/kg/week total walking	Assigns 4.0 kcal/kg/hour walking briskly, 3.0 to walking at moderate pace, and 2.0 to strolling (Y10 Qaire, page 12, Q18 - all walking)	If Y12EW7DAY=1 then Y12TWKKWK= 4.0 x Y12EWTIME* Y12EWTIM/60 if Y12EWPACE=1, 3.0 x Y12EWTIME* Y12EWTIM/60 if Y12EWPACE=2, 2.0 x Y12EWTIME* Y12EWTIM/60 if Y12EWPACE=3 if Y12EW7DAY=0 then Y12TWKKWK=0;	If Y12EW12MO≤.z and Y12EW7RCY≤.z then Y12TWKKWK=.; if (Y12EW12MO in (0,7,8) and Y12EW7RCY≤.z then Y12TWKKWK=0; if Y12EW12MO=1 and Y12EW7RCY≤.z then Y12TWKKWK=0; if Y12TWKKWK<0 then do: if Y12EWTIME > 0 and Y12EWTIM > 0 and (Y12EWPACE≤.z or Y12EWPACE=8) then Y12TWKKWK=3.0 x Y12EWTIME*Y12EWTIM/60 (median value at baseline; used here for consistency); if Y12EWTIME > 0 and (Y12EWTIM≤.z or Y12EWTDK=-1) then Y12EWTIM=35 (median value at baseline; used here for consistency); if (Y12EWTIME≤.z or Y12EWTMDK=-1) and Y12EWTIM > 0 then Y12EWTIME=4 (median value at baseline; used here for consistency)	kkcal/kg/week

Step 2: create composite variables

Variable	General description	Detailed Description	How variable is calculated	How to handle missing or special values	Value labels
WSKKWK	kcal/kg/week - walking + stairs	Sum of exercise walking, other walking, and stair climbing variables	Y12TWKKWK + Y12FSKKWK	Imputed version of component variables used; SAS SUM function used to sum all non-missing components	kkcal/kg/week
WALKTIME	minutes walking/week	Adds minutes exercise walking and other walking in past week	WALKTIME=(Y12EWTIME x Y12EWTIM)	If Y12TWKKWK=0 then WALKTIME=0;	min
BKTWTIME	minutes walking briskly/week	Sum of minutes brisk exercise walking plus minutes brisk other walking	If Y12EWPACE=1 then BKTWTIME= Y12EWTIME x Y12EWTIM; If Y12EWPACE>1 then BKTWTIME=0; If Y12EWKKWK=0 then BKTWTIME=0;		min

Step 3: create categorical variables

Variable	General description	Detailed Description	How variable is calculated	How to handle missing or special values	Value labels
WALKCAT	minutes walking/week category	Categories of time spent on any walking per week	If WALKTIME=0 then WALKCAT=0; if 0<WALKTIME<150 then WALKCAT=1; if WALKTIME ≥ 150 then WALKCAT=2	Imputed version of component variables used	0= 0 min/wk 1= between 0 and 150 min/wk 2= more than 150 min/wk
BRISK90	walks briskly ≥ 90 min/week	Dichotomous variable for walking briskly ≥90 min/week vs <90 min/week	If BKTWTIME=0 then BRISK90=0; if 0<BKTWTIME<90 then BRISK90=0; if BKTWTIME ≥ 90 then BRISK90=1	Imputed version of component variables used	0=No 1=Yes
BRISK180	walks briskly ≥ 180 min/week	Dichotomous variable for walking briskly ≥180 min/week vs <180 min/week	If BKTWTIME=0 then BRISK180=0; if 0<BKTWTIME<180 then BRISK180=0; if BKTWTIME ≥ 180 then BRISK180=1	Imputed version of component variables used	0=No 1=Yes

```

*****
***
*   Saved as   \\Fu-hsing-c\HABC\HABC_SAS\Calculated
Variables\Programs\Year 12\Phact_y12.sas
*
*   Todd Glasser 01-06-2011 - Modified year 11 program for use in Year
12
*****
* ;

%include '\\Fu-hsing-c\habc\habc_sas\programs\initV8.sas';
libname daf 'M:\VSS_HABC\ToddFiles\Year12.12.5';
libname calc '\\Fu-hsing-c\habc\habc_sas\calculated
variables\datasets\Year 12';

data calc.y12phact (KEEP=HABCID Y12MCKKWK Y12FSKKWK Y12TWKKWK WALKTIME
WALKCAT BKTWTIME
                                BRISK180 BRISK90 WSKKWK );

    set daf.y12visit;
    by habcid;

    Label Y12MCKKWK='KCAL/KG/WEEK DOING MAJOR CHORES'
          Y12FSKKWK='KCAL/KG/WEEK CLIMBING STAIRS'
          Y12TWKKWK='KCAL/KG/WEEK TOTAL WALKING'
          WALKTIME='MINUTES WALKING/WEEK'
          WALKCAT='MINUTES WALKING/WEEK CATEGORY'
          BKTWTIME='MINUTES WALKING BRISKLY/WEEK'
          BRISK90='WALKS BRISKLY >= 90 MIN/WEEK'
          BRISK180='WALKS BRISKLY >= 180 MIN/WEEK'
          WSKKWK='KCAL/KG/WEEK - WALKING + STAIRS';

    *****----- MAJOR CHORES (17);
    IF Y12HC12MO<=.z AND Y12HC7DAY<=.z THEN Y12MCKKWK=. ;
    IF (Y12HC12MO=0 OR Y12HC12MO IN (7,8)) AND (Y12HC7DAY<=.z OR
Y12HC7DAY=0) THEN Y12MCKKWK=0;
    IF Y12HC7DAY=0 THEN Y12MCKKWK=0;
    IF Y12HC12MO=1 AND Y12HC7DAY<=.z THEN Y12MCKKWK=0;
    IF Y12HC7DAY=1 THEN Y12MCKKWK=3.5*Y12hctim/60;
    *IMPUTED MISSING CODE;
    IF Y12MCKKWK LT 0 THEN DO;
    IF Y12HC12MO=1 AND Y12HC7DAY=1 AND (Y12hctim<=.z OR Y12HCDK=-1) THEN
Y12hctim=120;
    IF Y12HC7DAY=1 THEN Y12MCKKWK=3.5*Y12hctim/60;
    END;

    *****----- WALKING FOR EXERCISE (18);
    IF Y12EWPACE=1 THEN Y12EWMET=4.0;
    IF Y12EWPACE=2 THEN Y12EWMET=3.0;
    IF Y12EWPACE=3 THEN Y12EWMET=2.0;
    IF Y12EW12MO<=.z AND Y12EW7DAY<=.z THEN Y12EWKWK=. ;
    else do ;
        IF (Y12EW12MO IN (0,7,8)) AND (Y12EW7DAY<=.z) THEN
Y12EWKWK=0;
        IF Y12EW7DAY=0 THEN Y12EWKWK=0;
        IF Y12EW12MO=1 AND Y12EW7DAY<=.z THEN Y12EWKWK=0;
        IF Y12EW7DAY=1 THEN
Y12EWKWK=Y12EWMET*Y12EWTIME*Y12EWTIM/60;

```

```

end ;
*IMPUTED MISSING CODE;
IF Y12EWKWK LT 0 THEN DO;
    IF Y12EWTIME > 0 AND Y12EWTIM > 0 AND (Y12EWPACE<=.z OR
Y12EWPACE=8) THEN Y12EWMET=3.0;
    IF Y12EWTIME > 0 AND (Y12EWTIM<=.z OR Y12EWTDK=-1) THEN
Y12EWTIM=35;
    IF (Y12EWTIME<=.z OR Y12EWTMDK=-1) AND Y12EWTIM > 0 THEN
Y12EWTIME=4;
    IF Y12EW7DAY=1 THEN
Y12EWKWK=Y12EWMET*Y12EWTIME*Y12EWTIM/60;
    END;

****----- Fix calculation of BKEWTIME EK 9/25/03;
IF Y12EWPACE=1 THEN BKEWTIME=Y12EWTIME*Y12EWTIM;
    ELSE IF Y12EWPACE>1 THEN BKEWTIME=0;
    ELSE IF Y12TWKWK=0 THEN BKEWTIME=0;
BKTWTIME=BKEWTIME;
IF BKTWTIME=0 THEN BRISK90=0;
    ELSE IF 0<BKTWTIME<90 THEN BRISK90=0;
    ELSE IF BKTWTIME GE 90 THEN BRISK90=1;
IF BKTWTIME=0 THEN BRISK180=0;
    ELSE IF 0<BKTWTIME<180 THEN BRISK180=0;
    ELSE IF BKTWTIME GE 180 THEN BRISK180=1;

****----- New variable for total walking;
Y12TWKWK=Y12EWKWK;
IF Y12TWKWK=0 THEN WALKTIME=0;
    ELSE IF Y12TWKWK>0 THEN WALKTIME=Y12EWTIME*Y12EWTIM;
IF WALKTIME=0 THEN WALKCAT=0;
    ELSE IF 0<WALKTIME<150 THEN WALKCAT=1;
    ELSE IF WALKTIME GE 150 THEN WALKCAT=2;

****----- CLIMBING STAIRS (19);
* ASSUME UP/DOWN 1 FLIGHT TAKES 30 SECONDS AND ONE ADDITIONAL MET
FOR CARRYING A LOAD;
* Cutoff for flights of stairs - only for purpose of calculated
variables ***;
*** See email from MLE dated 4/6/01
***;
IF Y12FSNUM>210 then Y12FSNUM=210;
IF Y12FSLOAD>210 THEN Y12FSLOAD=210;
IF Y12FS12MO<=.z AND Y12FS7DAY<=.z THEN Y12FSKWK=.;
IF (Y12FS12MO=0 OR Y12FS12MO=8 OR Y12FS12MO=7) AND (Y12fs7day=8 OR
Y12fs7day<=.z) THEN Y12FSKWK=0;
IF Y12fs7day=0 THEN Y12FSKWK=0;
IF Y12FS12MO=1 AND (Y12fs7day<=.z OR Y12fs7day=8) THEN Y12FSKWK=0;
IF Y12fs7day=1 THEN Y12FSKWK=(4.0*Y12FSNUM/120) +
(1.0*Y12FSLOAD/120);
*IMPUTED MISSING CODE;
IF Y12FSKWK LT 0 THEN DO;
    IF Y12fs7day=1 AND Y12FSNUM GT 0 AND (Y12FSLOAD<=.z OR Y12FSLODK=-1)
THEN Y12FSLOAD=0;
    IF Y12fs7day=1 AND (Y12FSNUM<=.z OR Y12FSNUMD=-1) AND Y12FSLOAD GT 0
THEN Y12FSNUM=Y12FSLOAD;
    IF Y12fs7day=1 THEN Y12FSKWK=(4.0*Y12FSNUM/120) +
(1.0*Y12FSLOAD/120); END;

```

```

*****----- WALKING AND STAIRS;
WSKKWK=SUM(OF Y12FSKKWK Y12TWKKWK);

format brisk180 brisk90 yndk. walkcat walk9x.;
run;

%MACRO SKIP ;
proc contents data=CALC.y12phact varnum;
title4 'Final Dataset';
run;
proc freq data=CALC.y12phact;
tables brisk90 brisk180 walkcat/missing;
run;
proc univariate data=CALC.y12phact;
var bktwtime Y12mckkwk Y12fskkwk Y12twkkwk walktime wskkwk;
run;
%MEND SKIP ;

***check for duplicates***;
data dupes;
set CALC.y12phact;
by habcid;
if not(first.habcid and last.habcid);
run;

proc print;
title4 'Duplicates in phact y12';
run;

```