

Documentation for all Data Sets

HEALTH ABC DATA ANALYSIS FILE.....	2
PARTICIPANT HISTORY FILE (PH).....	3
YEAR 7 TELEPHONE CONTACT DATA (Y7Phone).....	6
YEAR 7 PROXY CONTACT DATA (Y7Proxy).....	7
78-MONTH SEMI-ANNUAL TELEPHONE CONTACT DATA (SA78mo).....	8
78-MONTH PROXY CONTACT DATA (SA78prox).....	10
MISSED FOLLOW-UP CONTACT DATA (MissVis).....	11
SPECIAL MISSING VALUE CODES.....	13
DROPPED VARIABLES.....	14
LISTINGS	15
DATA DICTIONARY (datadict.xls, datadict.txt)	15
Appendix I DROPPED VARIABLES AND ALTERNATES	17
Appendix II 78-MONTH CALCULATED VARIABLES	21

HEALTH ABC DATA ANALYSIS FILE

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

Contents:

SAS Datasets

Y7Phone	Year 7 Telephone Contact
Y7Proxy	Year Proxy Interview Contact
SA78Mo	78-Month Telephone Contact
SA78Prox	78-Month Proxy Contact
MissVis	Missed Follow-up Contact
Y7Calc	Year 7 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

VTYPE48	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
VTYPE54	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
VTYPE60	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
VTYPE66	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
VTYPE72	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
VTYPE78	Visit Type at 78-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, Core Home Visit workbook, Proxy Interview or 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) or by searching the Datadict file (sort by form or database).

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, 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 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.

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

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 7 TELEPHONE CONTACT DATA (Y7Phone)

1. General description

Year 7 is the first year that the annual interview was done by telephone only. The Y7Phone file contains information about the participants enrolled in the study gathered from the Year 7 Telephone Interview. If a participant did not have a Year 7 Telephone Interview or a Year 7 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 7 telephone visit. If VISITYPE=12, the participant's Year 7 data can be found in Y7Phone. If VISITYPE= 8, the participant's Year 7 data can be found in Y7Proxy. If VISITYPE=3, 4 or 5, the participant did not have a Year 7 contact and their Missed Follow-up Contact data can be found in MissVis.

There are 3075 observations in the Y7Phone 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 Y7Phone) or by searching the Datadict file (sort by form or database). 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 Y7Phone 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, 12=Telephone Visit)
MISSREAS	Reason Y7 Visit Missed

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

4. Condition of data

a. Known data errors:

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 6. To avoid confusion, these variables are listed in Y7Calc.

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

5. Dataset index formulation and key variable mapping

The Y7Phone 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 7 PROXY CONTACT DATA (Y7PROXY)

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. Since the Year 7 contact was by phone, some participants were also not able to answer because of hearing difficulties. In response to this situation, the Proxy Interview was developed to allow another person to answer for them.

In year 7, there are 117 participants who have a Proxy Interview and can be found in Y7Proxy.

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 Y7Proxy) or by searching the Datadict file (sort by form or database). 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 Y7Proxy 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)

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.

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

5. Dataset index formulation and key variable mapping

The Y7Proxy 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.

78-MONTH SEMI-ANNUAL TELEPHONE CONTACT DATA (SA78MO)

1. General description

The 78-month visit file contains information about the participants enrolled in the study gathered from the 78-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 SA78PROX. If a participant did not have a 78-month semi-annual telephone interview or a 78-month proxy, they should have a Missed Follow-up Contact form that explains why. A variable, VISITYPE, has been added to this dataset to allow the analyst to account for all participants, whether or not they had a 78-month follow-up telephone call. If VISITYPE=7, the participant's 78-month data can be found in SA78MO. If VISITYPE= 8, the participant's 78-month data can be found in SA78PROX. If VISITYPE=3, 4 or 5, the participant did not have a 78-month contact and their Missed Follow-up Contact data can be found in MissVis.

There are 3075 observations in the SA78MO file.

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

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 SA78MO) or by searching the Datadict file (sort by form or database). 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 SA78MO 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 SA78MO 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 SA78MO 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.

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

78-MONTH PROXY CONTACT DATA (SA78PROX)

1. General description

In some cases, a participant was unable to complete the 78-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 78-month visit, there are 136 participants who have a Proxy Interview and can be found in SA78PROX.

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 SA78PROX) or by searching the Datadict file (sort by form or database). 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 SA78PROX 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)

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 SA78PROX 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.

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 78-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
Year 2 visit	77
18-month semi-annual contact	137
Year 3 visit	94
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	150
Year 6 visit	144
66-month semi-annual contact	142
Year 7 visit	136
78-month semi-annual contact	167

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) or by searching the Datadict file (sort by form or database). 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)

4. Condition of data

a. Known data errors: None.

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 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). 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. Clinic Visit Workbook, Core Home Visit Workbook, Proxy Interview, Proxy Contact Home Visit Workbook, 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 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.

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

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 that 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; Sub-study 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
BJCONTAC	BJID2
BJVISIT	BJID2
BJACROS	N/A (confidentiality)
BJID	HABCID
BJSTFID	N/A (bookkeeping)
BLACROS	N/A (confidentiality)
BLID	HABCID
BLINK	N/A (bookkeeping)
BLACHES	N/A (confidentiality)
BLADDRES	N/A (confidentiality)
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)
GAACROS	N/A (confidentiality)
GAID	HABCID
GASTFID	N/A (bookkeeping)

Variable omitted	Variable to use
GALINK	N/A (bookkeeping)
GAFSHS	N/A (confidentiality)
GAFSHABC	N/A (confidentiality)
GAFSLN1	N/A (confidentiality)
GAFSLN2	N/A (confidentiality)
GAFSLN3	N/A (confidentiality)
GAFSFN1	N/A (confidentiality)
GAFSFN2	N/A (confidentiality)
GAFSFN3	N/A (confidentiality)
GAFSMO1	N/A (confidentiality)
GAFSMO2	N/A (confidentiality)
GAFSMO3	N/A (confidentiality)
GAFSDY1	N/A (confidentiality)
GAFSDY2	N/A (confidentiality)
GAFSDY3	N/A (confidentiality)
GAFSYR1	N/A (confidentiality)
GAFSYR2	N/A (confidentiality)
GAFSYR3	N/A (confidentiality)
GAHSDY1	N/A (confidentiality)
GAHSDY2	N/A (confidentiality)
GAHSDY3	N/A (confidentiality)
GAHSFN1	N/A (confidentiality)
GAHSFN2	N/A (confidentiality)
GAHSFN3	N/A (confidentiality)
GAHSHABC	N/A (confidentiality)
GAHSLN1	N/A (confidentiality)
GAHSLN2	N/A (confidentiality)
GAHSLN3	N/A (confidentiality)
GAHSMO1	N/A (confidentiality)
GAHSMO2	N/A (confidentiality)
GAHSMO3	N/A (confidentiality)
GAHSYR1	N/A (confidentiality)
GAHSYR2	N/A (confidentiality)
GAHSYR3	N/A (confidentiality)
GAFCDY1	N/A (confidentiality)
GAFCDY2	N/A (confidentiality)
GAFCDY3	N/A (confidentiality)
GAFCFN1	N/A (confidentiality)
GAFCFN2	N/A (confidentiality)
GAFCFN3	N/A (confidentiality)
GAFCHABC	N/A (confidentiality)
GAFCLN1	N/A (confidentiality)

Variable omitted	Variable to use
GAFCLN2	N/A (confidentiality)
GAFCLN3	N/A (confidentiality)
GAFCMO1	N/A (confidentiality)
GAFCMO2	N/A (confidentiality)
GAFCMO3	N/A (confidentiality)
GAFCYR1	N/A (confidentiality)
GAFCYR2	N/A (confidentiality)
GAFCYR3	N/A (confidentiality)
GASPDY1	N/A (confidentiality)
GASPFN1	N/A (confidentiality)
GASPHABC	N/A (confidentiality)
GASPLN1	N/A (confidentiality)
GASPMO1	N/A (confidentiality)
GASPYR1	N/A (confidentiality)
YAACROS	N/A (confidentiality)
YAADDRES	N/A (confidentiality)
YAAPT	N/A (confidentiality)
YACITY	N/A (confidentiality)
YACONTAC	N/A (confidentiality)
YADCITY	N/A (confidentiality)
YADFNAME	N/A (confidentiality)
YADLNAME	N/A (confidentiality)
YADPHONE	N/A (confidentiality)
YADSTATE	N/A (confidentiality)
YADSTRT	N/A (confidentiality)
YADZIP	N/A (confidentiality)
YAFNAME	N/A (confidentiality)
Y Aid	HABCID
YALINK	N/A (bookkeeping)
YALNAME	N/A (confidentiality)
YAMAAPT	N/A (confidentiality)
YAMACITY	N/A (confidentiality)
YAMADATE	N/A (confidentiality)
YAMASTAT	N/A (confidentiality)
YAMASTRT	N/A (confidentiality)
YAMATELE	N/A (confidentiality)
YAMAZIP	N/A (confidentiality)
YAMOVE	N/A (confidentiality)
YAPHONE	N/A (confidentiality)
YAREF11A	N/A (bookkeeping)
YAREF11B	N/A (bookkeeping)
YAREF11C	N/A (bookkeeping)
YAREF12A	N/A (bookkeeping)

Variable omitted	Variable to use
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)
YASTATE	N/A (bookkeeping)
YASTREET	N/A (bookkeeping)
YAVISIT	N/A (bookkeeping)
YAWKPHON	N/A (confidentiality)
YAZIP	N/A (confidentiality)

Appendix II

78-MONTH CALCULATED VARIABLES

Variable	Variable Description	Grouping	Page
CSAINDEX	Climbing stairs ability index	Self-Reported Function Calculated Vars.	24, 27
EASE1F	Ease climbing 1 flight	Self-Reported Function Calculated Vars.	24, 27
EASE1M	Ease walking 1 mile	Self-Reported Function Calculated Vars.	23, 26
EASE2F	Ease climbing 2 flights	Self-Reported Function Calculated Vars.	24, 27
EASEQM	Ease walking 1/4 mile	Self-Reported Function Calculated Vars.	22, 26
WKAINDEX	Walking ability index	Self-Reported Function Calculated Vars.	23, 27

Appendix III
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 GADWQMEZ where 3 is easiest; imputations applied for missing GADWQMEZ	If GADWQMYN in (0,7,8) then EASEQM=7-GADWQMEZ If GADWQMYN=1 then EASEQM=4-GADWQMDF	<p>IF GADWQMYN=9 AND GAMNRS>0 THEN GADWQMYN=1; IF GADWQMEZ=8 and GADWQMYN ne 8 THEN GADWQMEZ=2; IF GADWQMDF=8 and GADWQMYN ne 8 THEN GADWQMDF=2; IF GADWQMYN in (0,7,8) and GADWQMEZ<0 then EASEQM=4; IF (GADWQMYN=0 AND GADWQMEZ<0) THEN EASEQM=7-GADW1MEZ; IF (GADWQMYN=0 AND GADWQMEZ<0) AND GADW1MEZ<0 AND (GADW1MYN=1 OR GADW1MYN=8 OR GADW1MYN<=.z) THEN EASEQM=4; IF (GADWQMYN=0 AND GADWQMEZ<0) AND (GADW1MYN=0 AND GADW1MEZ<=.z) THEN EASEQM=5;</p> <p>IF GADWQMYN=1 AND GADWQMDF<0 THEN EASEQM=7-GADWQMEZ; IF GADWQMYN=1 AND GADWQMDF<0 AND GADWQMEZ<0 THEN EASEQM=2; IF GADWQMYN<0 AND GADWQMDF>0 THEN EASEQM=4-GADWQMDF; IF GADWQMYN<0 AND GADWQMEZ>0 THEN EASEQM=7-GADWQMEZ;</p> <p>if GADWQMYN=8 and GADWQMEZ=8 then EASEQM=4; IF GADWQMYN=9 AND GAMNRS<0 THEN EASEQM=.;</p>	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 BLDW1MEZ where 3 is easiest; imputations applied for missing BLDW1MEZ	If BLDW1MYN=1 then EASE1M=0; if BLDW1MEZ=3 then EASE1M=1; if BLDW1MEZ=2 then EASE1M=2; if BLDW1MEZ=1 then EASE1M=3	<p>IF BLDW1MYN=8 AND (BLDW1MEZ<0 OR BLDW1MEZ=8) THEN EASE1M=1; IF BLDW1MYN<=.z AND BLDW1MEZ<=.z AND EASEQM=4 THEN EASE1M=0; IF BLDW1MYN<=.z AND BLDW1MEZ<=.z AND EASEQM=5 THEN EASE1M=1; IF BLDW1MYN<=.z AND BLDW1MEZ<=.z AND EASEQM=6 THEN EASE1M=2; IF BLDW1MYN=0 AND BLDW1MEZ<=.z AND EASEQM=4 THEN EASE1M=0; IF BLDW1MYN=0 AND BLDW1MEZ<=.z AND EASEQM=5 THEN EASE1M=1; IF BLDW1MYN=0 AND BLDW1MEZ<=.z AND EASEQM=6 THEN EASE1M=2; IF BLDWQMYN=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; if BLDW1MYN=8 and BLDW1MEZ in (1,2,3) then EASE1M=4- BLDW1MEZ; if BLDW1MYN=0 and BLDW1MEZ=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 BLDW10EZ where 3 is easiest; imputations applied for missing BLDW20EZ	EASE1F=7-BLDW10EZ; IF BLDW10YN=1 THEN EASE1F=4-BLDIF;	IF BLDW10YN=9 then EASE1F=.; IF BLDW10EZ=8 THEN BLDW10EZ=2; IF BLDIF=8 THEN BLDIF=2; IF BLDW10EZ<0 THEN EASE1F=7-BLDW20EZ; IF BLDW10YN=1 AND (BLDIF=8 OR BLDIF<0) THEN EASE1F=2; IF BLDW10YN=0 AND BLDW10EZ<0 AND BLDW10EZ<0 THEN EASE1F=4; IF EASE1F=. AND BLDW10YN NE 9 AND BLDW10YN>0 AND (BLDW20YN=1 OR BLDW20YN=8 OR BLDW20YN<=.z) THEN EASE1F=4; IF EASE1F=. AND BLDW20YN=0 AND BLDW20EZ<=.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 BLDW20EZ where 3 is easiest; imputations applied for missing BLDW20EZ	IF BLDW20YN=1 THEN EASE2F=0; ELSE IF BLDW20EZ=3 THEN EASE2F=1; ELSE IF BLDW20EZ=2 THEN EASE2F=2; ELSE IF BLDW20EZ=1 THEN EASE2F=3;	IF BLDW10YN=9 then EASE1F=.; IF (BLDW20YN=8 OR BLDW20YN<0)AND (BLDW20EZ<0 OR BLDW20EZ=8) AND EASE1F>0 THEN EASE2F=1; IF BLDW20YN<=.z AND BLDW20EZ<=.z AND EASE1F=4 THEN EASE2F=0; IF BLDW20YN<=.z AND BLDW20EZ<=.z AND EASE1F=5 THEN EASE2F=1; IF BLDW20YN<=.z AND BLDW20EZ<=.z AND EASE1F=6 THEN EASE2F=2; IF BLDW20YN=0 AND (BLDW20EZ<=.z OR BLDW20EZ=8) AND EASE1F>0 THEN EASE2F=EASE1F-4; IF BLDW10YN=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 \\ilcavallo\HABC\HABC_SAS\Calculated Variables\Programs\SemiAnnual\srfcn_sa78.sas
*
* Lynn Harvey 10/14/05 - Modified year 6 program for use in year 7.5
***** ;
```

```
options ls=132 ps=58 formchar='|---+|---+|=|-\<>*' nocenter pageno=1 nofmterr mprint macrogen;
```

```
DM LOG 'CLEAR' ; DM OUTPUT 'CLEAR' ;
```

```
%include \\ilcavallo\habc\habc_sas\programs\initV8.sas';
libname calc \\ilcavallo\habc\habc_sas\calculated variables\datasets\SemiAnnual\78Month';
```

```
*****
** SELF-REPORT FUNCTION PROGRAM CODE;
** HABC SEMI ANNUAL TELEPHONE CONTACT SELF-REPORT FUNCTION
** CALC VAR CODE
** Creates the following variables for 6mo VISIT:
** 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
** TIREDQM: get tired walking 1/4 mile
** TIRED1F: get tired walking 1 mile
** CSAINDEX: climbing stairs ability index
** WKAINDEX: walking ability index
**
** Adapted from Fran Harris Y2 code: srfcn_y2.sas
** Adapted from Laura Akin Y6 code
**
** Lynn Harvey Akin 10/13/2005
**
***** ;
```

```
data CALC.SRFCN_78 (keep=HABCID VISIT EASEQM EASE1M EASE1F EASE2F /*tiredqm
tired1*/ CSAINDEX WKAINDEX)
  SRFCN_SA;
  set habc6.BL2(keep=HABCID BLCONTAC BLdwqmyn BLdwqmdf BLMnrs BLMnrs4 BLdwqmez
  BLdwqmt2 BLdw1myn BLdw1mez BLdw10yn BLdif
  BLMnrs2 BLMnrs3 BLdw10ez BLdw10wx BLdw20yn
  BLdw20ez WHERE=(BLCONTAC=7));
  rename blcontac=VISIT;
  *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 BLDW1MYN WAS YES, DONT KNOW, OR MISSING THEN NOT THAT
  EASY WAS ASSIGNED. IF NO
  EASE LEVEL FOR WALKING 1M WAS CODED AND BLDW1MYN WAS NO AND BLDW1MEZ
  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 BLDWQMYN=9 AND (BLMNRS>0 OR BLMNRS4 IN(1,2)) THEN BLDWQMYN=1;
  IF BLDWQMEZ=8 and BLdwqmyn ne 8 THEN BLDWQMEZ=2;
  IF BLDWQMDF=8 and BLdwqmyn ne 8 THEN BLDWQMDF=2;
  IF BLDWQMYN IN (0,7,8) THEN EASEQM=7-BLDWQMEZ;
  IF (BLDWQMYN=0 AND BLDWQMEZ<0) THEN EASEQM=7-BLDW1MEZ;
```

IF (BLDWQMYN=0 AND BLDWQMEZ<0) AND BLDW1MEZ<0 AND (BLDW1MYN=1 OR BLDW1MYN=8 OR BLDW1MYN<=.z) THEN EASEQM=4;
IF (BLDWQMYN=0 AND BLDWQMEZ<0) AND (BLDW1MYN=0 AND BLDW1MEZ<=.z) THEN EASEQM=5;
IF BLDWQMYN=1 THEN EASEQM=4-BLDWQMDF;

IF BLDWQMYN=1 AND BLDWQMDF<0 THEN EASEQM=7-BLDWQMEZ;
IF BLDWQMYN=1 AND BLDWQMDF<0 AND BLDWQMEZ<0 THEN EASEQM=2;
IF BLDWQMYN<0 AND BLDWQMDF>0 THEN EASEQM=4-BLDWQMDF;
IF BLDWQMYN<0 AND BLDWQMEZ>0 THEN EASEQM=7-BLDWQMEZ;

if BLdwqmyn=8 and BLdwqmez=8 then EASEQM=4;
IF BLDWQMYN=9 AND BLMNRS<0 AND (BLMNRS4<0 OR BLMNRS4=8) 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 BLDW1MYN=1 THEN EASE1M=0;
ELSE IF BLDW1MEZ=3 THEN EASE1M=1;
ELSE IF BLDW1MEZ=2 THEN EASE1M=2;
ELSE IF BLDW1MEZ=1 THEN EASE1M=3;

*MISSING VALUE RECODES;
IF BLDW1MYN=8 AND (BLDW1MEZ<0 OR BLDW1MEZ=8) THEN EASE1M=1;
IF BLDW1MYN<=.z AND BLDW1MEZ<=.z AND EASEQM=4 THEN EASE1M=0;
IF BLDW1MYN<=.z AND BLDW1MEZ<=.z AND EASEQM=5 THEN EASE1M=1;
IF BLDW1MYN<=.z AND BLDW1MEZ<=.z AND EASEQM=6 THEN EASE1M=2;
IF BLDW1MYN=0 AND BLDW1MEZ<=.z AND EASEQM=4 THEN EASE1M=0;
IF BLDW1MYN=0 AND BLDW1MEZ<=.z AND EASEQM=5 THEN EASE1M=1;
IF BLDW1MYN=0 AND BLDW1MEZ<=.z AND EASEQM=6 THEN EASE1M=2;
IF BLDWQMYN=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 DOWNGRADES 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 BLdw1myn=8 and BLdw1mez in (1,2,3) then EASE1M=4-BLdw1mez;
if BLdw1myn=0 and BLdw1mez=8 then EASE1M=EASEQM-4;

*CODE TO CREATE A SUMMARY MEASURE OF SELF-REPORTED WALKING ABILITY. THIS SCALE WILL ULTIMATELY RANGE FROM 0 TO 9, BUT AT BASELINE IT RANGES 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 BLDW10YN=9 THEN DO; EASE1F=.; EASE2F=.; END;

IF BLDW10EZ=8 THEN BLDW10EZ=2;

IF BLDIF=8 THEN BLDIF=2;

EASE1F=7-BLDW10EZ;

IF BLDW10EZ<0 THEN EASE1F=7-BLDW20EZ;

IF BLDW10YN=1 THEN EASE1F=4-BLDIF;

IF BLDW10YN=1 AND (BLDIF=8 OR BLDIF<0) THEN EASE1F=2;

IF BLDW10YN=0 AND BLDW10EZ<0 AND BLDW10EZ<0 THEN EASE1F=4;

IF EASE1F=. AND BLDW10YN NE 9 AND BLDW10YN>0 AND
(BLDW20YN=1 OR BLDW20YN=8 OR BLDW20YN<=.z)

THEN EASE1F=4;

IF EASE1F=. AND BLDW20YN=0 AND BLDW20EZ<=.z THEN EASE1F=5;

*EASE2F;

IF BLDW20YN=1 THEN EASE2F=0;

ELSE IF BLDW20EZ=3 THEN EASE2F=1;

ELSE IF BLDW20EZ=2 THEN EASE2F=2;

ELSE IF BLDW20EZ=1 THEN EASE2F=3;

IF (BLDW20YN=8 OR BLDW20YN<0) AND (BLDW20EZ<0 OR BLDW20EZ=8) AND EASE1F>0
THEN EASE2F=1;

IF BLDW20YN<=.z AND BLDW20EZ<=.z AND EASE1F=4 THEN EASE2F=0;

IF BLDW20YN<=.z AND BLDW20EZ<=.z AND EASE1F=5 THEN EASE2F=1;

IF BLDW20YN<=.z AND BLDW20EZ<=.z AND EASE1F=6 THEN EASE2F=2;

IF BLDW20YN=0 AND (BLDW20EZ<=.z OR BLDW20EZ=8) AND EASE1F>0 THEN
EASE2F=EASE1F-4;

IF BLDW10YN=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

RANGE FROM 0 TO 9, BUT AT BASELINE IT RANGES FROM 4-9;

CSAINDEX=EASE1F + EASE2F;

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' ;

FORMAT EASE1F EASE2F EASEQM EASE1M CSAINDEX WKAINDEX SPMISS. ;

RUN;

Documentation for Year 7 Calculated Variable Dataset
(Y7Calc)

YEAR 7 CALCULATED VARIABLE DATASET (Y7calc)..... 29

Appendix I Calculated Variable List

Appendix II Physical Activity Calculated Variables

Appendix III Self-Reported Function Calculated Variables

YEAR 7 CALCULATED VARIABLE DATASET (Y7calc)

1. General description

The Y7calc contains 15 variables derived (calculated) as described below. The documentation is grouped by the type of data from which the variables are calculated (physical activity or self-reported function), and the variables are positioned in the dataset in the same groupings.

There are 3075 observations in the Y7calc 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 Y7calc) 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 Y7calc 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.

5. Dataset index formulation and key variable mapping

The Y7calc 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.

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

Appendix I
CALCULATED VARIABLE LIST
(LINKED)

Variable	Variable Description	Grouping
BKTWTIME	minutes walking briskly/week	Physical Activity Calculated Vars
BRISK180	walks briskly \geq 180 min/week	Physical Activity Calculated Vars
BRISK90	walks briskly \geq 90 min/week	Physical Activity Calculated Vars
GACWKKWK	kcal/kg/week doing child/adult care	Physical Activity Calculated Vars
GAFSKKWK	kcal/kg/week climbing stairs	Physical Activity Calculated Vars
GATWKKWK	Kcal/ kg/ week total walking	Physical Activity Calculated Vars
WALKCAT	minutes walking/week category	Physical Activity Calculated Vars
WALKTIME	minutes walking/week	Physical Activity Calculated Vars
WSKKWK	kcal/kg/week - walking + stairs	Physical Activity Calculated Vars
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.

Appendix II
Physical Activity Calculated 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
GAFSKKWK	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 (Y7 Qaire, page 7, Q14). (Assumes 1 flight up/down takes 30 seconds.)	If GAFS7DAY=1 then GAFSKKWK = 4.0 x GAFSNUM/120 + 1.0 x GAFSLOAD/120 If GAFS7DAY=0 then GAFSKKWK=0;	Correction for outliers: If GAFSNUM>210 then GAFSNUM=210; if GAFSLOAD>210 then GAFSLOAD=210; If GAFS12MO≤.z and GAFS7DAY≤.z then GAFSKKWK=.; if (GAFS12MO in (0,7,8)) and (GAFS7DAY≤.z or GAFS7DAY=8) then GAFSKKWK=0; if GAFS12MO=1 and (GAFS7DAY≤.z or GAFS7DAY=8) then GAFSKKWK=0; if GAFSKKWK<0 then do: if GAFS7DAY=1 and GAFSNUM > 0 and (GAFSLOAD≤.z or GAFSLODK=-1) then GAFSLOAD=0; if GAFS7DAY=1 and (GAFSNUM≤.z or GAFSNUMD=-1) and GAFSLOAD > 0 then GAFSNUM=GAFSLOAD	kkcal/kg/week
GACWKKWK	kcal/kg/week doing child/adult care	Assigns 2.5 kcal/kg/hour doing child/adult care (Y7 Qaire, page 10, Q17)	IF GAVWCURA=1 THEN GACWKKWK=2.5*GA VVAHAW	IF GAVWCURA=0 OR GAVWCURA IN (7,8) THEN GACWKKWK=0; *IMPUTED MISSING CODE; IF GACWKKWK LT 0 THEN DO; IF GAVWCURA=1 AND (GAVVAHAW<=.z OR GAVWDK=-1) THEN GACWKKWK=30; (median value at baseline; used here for consistency) END;	kkcal/kg/week

Physical Activity Calculated Variables

Variable	General description	Detailed Description	How variable is calculated	How to handle missing or special values	Value labels
GATWKKWK	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 (Y7 Qaire, page 8, Q15 all walking)	<p>If GAEW7DAY=1 then GATWKKWK= 4.0 x GAEWTIME* GAEWTIM/60 if GAEWPACE=1, 3.0 x GAEWTIME* GAEWTIM/60 if GAEWPACE=2, 2.0 x GAEWTIME* GAEWTIM/60 if GAEWPACE=3</p> <p>if GAEW7DAY=0 then GATWKKWK=0;</p>	<p>IF GAEW12MO<=.z AND GAEW7DAY<=.z THEN GAEWKKWK=.; else do ; IF (GAEW12MO IN (0,7,8)) AND (GAEW7DAY<=.z) THEN GAEWKKWK=0; IF GAEW12MO=1 AND GAEW7DAY<=.z THEN GAEWKKWK=0; end ;</p> <p>IF GAEWKKWK LT 0 THEN DO; IF GAEWTIME > 0 AND GAEWTIM > 0 AND (GAEWPACE<=.z OR GAEWPACE=8) THEN GAEWMET=3.0; IF GAEWTIME > 0 AND (GAEWTIM<=.z OR GAEWTDK=-1) THEN GAEWTIM=35; (median value at baseline; used here for consistency); IF (GAEWTIME<=.z OR GAEWTMDK=-1) AND GAEWTIM > 0 THEN GAEWTIME=4; (median value at baseline; used here for consistency); IF GAEW7DAY=1 THEN GAEWKKWK=GAEWMET*GAEWTIME*GAE WTIM/60; (median value at baseline; used here for consistency); END;</p> <p>*New variable for total walking*; GATWKKWK=GAEWKKWK;</p>	kcal/kg/week

Physical Activity Calculated Variables

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	GATWKKWK + GAFSKKWK	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=(GAEWTIME x GAEWTIM)	If GATWKKWK=0 then WALKTIME=0;	min
BKTWTIME	minutes walking briskly/week	Sum of minutes brisk exercise walking plus minutes brisk other walking	If GAEWPACE=1 then BKTWTIME= GAEWTIME x GAEWTIM; If GAEWPACE>1 then BKTWTIME=0; If GAEWKKWK=0 then BKTWTIME=0;	SAS SUM function used to sum all non-missing components	min

Physical Activity Calculated Variables

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 \\ilcavallo\HABC\HABC_SAS\Calculated Variables\Programs\Year
7\Phact_y7.sas
*
* Lynn Harvey 10/14/05 - Modified year 6 program for use in year 7
***** ;

```

```

data CALC.Y7PHACT (KEEP=HABCID GACWKKWK GAFSKKWK GATWKKWK WALKTIME WALKCAT
BKTWTIME BRISK180

```

```

BRISK90 WSKKWK );

```

```

merge habc7.ga ;
by habcid;

```

```

Label GAFSKKWK='KCAL/KG/WEEK CLIMBING STAIRS'
      GATWKKWK='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'
      GACWKKWK='KCAL/KG/WEEK DOING CHILD/ADULT CARE'

```

```

;

```

```

*CLIMBING STAIRS;
*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 GAFSNUM>210 then GAFSNUM=210;
IF GAFSLOAD>210 THEN GAFSLOAD=210;

```

```

IF GAFS12MO<=.z AND GAFS7DAY<=.z THEN GAFSKKWK=.;
IF (GAFS12MO=0 OR GAFS12MO=8 OR GAFS12MO=7) AND (GAFs7day=8 OR
GAFs7day<=.z) THEN GAFSKKWK=0;
IF GAFs7day=0 THEN GAFSKKWK=0;
IF GAFS12MO=1 AND (GAFs7day<=.z OR GAFs7day=8) THEN GAFSKKWK=0;
IF GAFs7day=1 THEN GAFSKKWK=(4.0*GAFSNUM/120) + (1.0*GAFSLOAD/120);

```

```

*IMPUTED MISSING CODE;
IF GAFSKKWK LT 0 THEN DO;
IF GAFs7day=1 AND GAFSNUM GT 0 AND (GAFSLOAD<=.z OR GAFSLODK=-1) THEN
GAFSLOAD=0;
IF GAFs7day=1 AND (GAFSNUM<=.z OR GAFSNUMD=-1) AND GAFSLOAD GT 0 THEN
GAFSNUM=GAFSLOAD;
IF GAFs7day=1 THEN GAFSKKWK=(4.0*GAFSNUM/120) + (1.0*GAFSLOAD/120); END;

```

```

*WALKING FOR EXERCISE;

```

```

IF GAEWPACE=1 THEN GAEWMET=4.0;
IF GAEWPACE=2 THEN GAEWMET=3.0;
IF GAEWPACE=3 THEN GAEWMET=2.0;

```

```

IF GAEW12MO<=.z AND GAEW7DAY<=.z THEN GAEWKKWK=. ;
else do ;
    IF (GAEW12MO IN (0,7,8)) AND (GAEW7DAY<=.z) THEN GAEWKKWK=0 ;
    IF GAEW7DAY=0 THEN GAEWKKWK=0 ;
    IF GAEW12MO=1 AND GAEW7DAY<=.z THEN GAEWKKWK=0 ;
    IF GAEW7DAY=1 THEN GAEWKKWK=GAEWMET*GAEWTIME*GAEWTIM/60 ;
end ;

*IMPUTED MISSING CODE ;
IF GAEWKKWK LT 0 THEN DO ;
    IF GAEWTIME > 0 AND GAEWTIM > 0 AND (GAEWPACE<=.z OR GAEWPACE=8)
THEN GAEWMET=3.0 ;
    IF GAEWTIME > 0 AND (GAEWTIM<=.z OR GAEWTDK=-1) THEN GAEWTIM=35 ;
    IF (GAEWTIME<=.z OR GAEWTMDK=-1) AND GAEWTIM > 0 THEN GAEWTIME=4 ;
    IF GAEW7DAY=1 THEN GAEWKKWK=GAEWMET*GAEWTIME*GAEWTIM/60 ;
END ;

*New variable for total walking* ;
GATWKKWK=GAEWKKWK ;

IF GATWKKWK=0 THEN WALKTIME=0 ;
ELSE IF GATWKKWK>0 THEN WALKTIME=GAEWTIME*GAEWTIM ;

IF WALKTIME=0 THEN WALKCAT=0 ;
ELSE IF 0<WALKTIME<150 THEN WALKCAT=1 ;
ELSE IF WALKTIME GE 150 THEN WALKCAT=2 ;

*updated below to fix calculation of BKEWTIME EK 9/25/03 ;
IF GAEWPACE=1 THEN BKEWTIME=GAEWTIME*GAEWTIM ;
ELSE IF GAEWPACE>1 THEN BKEWTIME=0 ;
ELSE IF GATWKKWK=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 ;

*WALKING AND STAIRS ;

WSKKWK=SUM(OF GAFSKKWK GATWKKWK) ;

*CAREGIVING* ;

IF GAVWCURA=0 OR GAVWCURA IN (7,8) THEN GACWKKWK=0 ;
IF GAVWCURA=1 THEN GACWKKWK=2.5*GAVWAHAW ;
*IMPUTED MISSING CODE ;
IF GACWKKWK LT 0 THEN DO ;
IF GAVWCURA=1 AND (GAVWAHAW<=.z OR GAVWDK=-1) THEN GACWKKWK=30 ; END ;

format brisk180 brisk90 yndk. walkcat walk9x. ;
run ;

```

```

%MACRO SKIP ;
PROC CONTENTS DATA=CALC.Y7PHACT ;
RUN ;

PROC FREQ DATA=CALC.Y7PHACT ;
    TABLES GACWKKWK GAWTIM GAFSKKWK GATWKKWK WALKTIME WALKCAT BKTWTIME BRISK180
BRISK90
    WSKKWK ;
RUN ;

PROC MEANS DATA=CURRENT.Y6CALC ;
    VAR WALKTIME BKTWTIME ;
RUN ;

PROC MEANS DATA=CALC.Y7PHACT ;
    VAR WALKTIME BKTWTIME ;
RUN ;

PROC MEANS DATA=CALC.Y8PHACT ;
    VAR WALKTIME BKTWTIME ;
RUN ;

PROC FREQ DATA=CURRENT.Y6CALC ;
    TABLES WALKCAT BRISK90 BRISK180 ;
RUN ;

PROC FREQ DATA=CALC.Y7PHACT ;
    TABLES WALKCAT BRISK90 BRISK180 ;
RUN ;

PROC FREQ DATA=CALC.Y8PHACT ;
    TABLES WALKCAT BRISK90 BRISK180 ;
RUN ;
%MEND SKIP ;

%macro skip ;

*COMPOSITE MEASURES. SINCE THE SUM FUNCTION ASSIGNS ZERO TO MISSING VALUES,
THE IMPUTED VERSION OF THE COMPONENT VARIABLES WERE USED;
*HOUSEHOLD DUTIES;
*HHKKWK=SUM(OF FAPAKKWK FAHCKKWK FALWKKWK FAGSKKWK FALDKKWK);
*EXERCISE AND RECREATION;
*EXKKWK=SUM(OF FAACKKWK FATRKKWK FAHIKKWK FAMIKKWK);
*WORK, VOLUNTEER, CAREGIVING;
*GRAND TOTAL;

*PROGRAM CODE TO CREATE SUMMARY MEASURES OF PHYSICAL ACTIVITY
BASED ON CALORIC EXPENDITURE. 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;

*STEP 1: CALCULATION OF KCALORIES PER KILOGRAM OF BODY WEIGHT PER WEEK
EXPENDED IN EACH ACTIVITY. ASSIGNED MET VALUES AND ASSUMPTIONS FOLLOW.
FOR EASE OF COMPUTATION WHILE THE DATA IS STILL DIRTY, KCAL/KG/WK FOR

```

EACH ACTIVITY WILL BE CALCULATED DIRECTLY FROM MINUTES ACTIVITY PERFORMED (OR LOADS WASHED, BAGS OF GROCERIES, FLIGHTS OF STAIRS CLIMBED). IT WILL BE ASSUMED THAT A MISSING VALUE IN THESE FIELDS INDICATES NON-PARTICIPATION AND A VALUE OF ZERO WILL BE ASSIGNED;

*MET ASSIGNMENTS AND OTHER DECISION RULES:

Q16 - GARDENING/YARDWORK = 3.5 (THIS HAS BEEN DOWNGRADED FROM 4.5)
Q17 - MAJOR CHORES = 3.5 (THIS HAS BEEN DOWNGRADED FROM 4.5)
SINCE THESE ARE COMBINED IN LATER ROUNDS THE SAME MET VALUE WAS ASSIGNED
Q18 - LIGHT HOUSE WORK = 2.5
Q19 - GROCERY SHOPPING = 3.5 FOR 4 BAGS PLUS 1.25 FOR THOSE CARRIED AND 1.25 FOR THOSE UNPACKED
Q20 - LAUNDRY - ASSUMING 30 MINUTES OF WORK PER LOAD THEN 1.0 PER LOAD WASHED PLUS .5 PER LOAD PUT AWAY
Q21 - STAIRS = 4.0 WITH THE ASSUMPTION 1 FLIGHT UP/DOWN TAKES 30 SECONDS PLUS 1.0 MET FOR EVERY FLIGHT CLIMBED WITH A LOAD
Q22 - WALKING FOR EXERCISE - 4.0 IF BRISK, 3.0 IF MODERATE, AND 2.0 IF STROLL (DOWNGRADED TO 3.0 AND 2.0)
Q23 - WALKING FOR OTHER REASONS - 4.0 IF BRISK, 3.0 IF MODERATE, AND 2.0 IF STROLL (DOWNGRADED TO 3.0 AND 2.0)
Q24 - AEROBICS = 5.0
Q25 - WEIGHT TRAINING = 6.0 (DOWNGRADED TO 6)
Q26 - HIGH INTENSITY EXERCISE - 4.0 IF LIGHT, 6.0 IF MODERATE, 8.0 IF VIGOROUS (DOWNGRADED FROM 7.0 AND 10.0)
Q27 - MODERATE INTENSITY EXERCISE = 3.0 FOR ALL ACTIVITIES AND INTENSITY LEVELS (DOWNGRADED FROM 4.0)
Q28 - PAID WORK = 1.5 FOR LEVEL 1, 2.0 FOR LEVEL 2, 2.5 FOR LEVEL 3, AND 3.0 FOR LEVEL 4
Q29 - VOLUNTEER WORK = 1.5 FOR LEVEL 1, 2.0 FOR LEVEL 2, 2.5 FOR LEVEL 3, AND 3.0 FOR LEVEL 4
Q30 - CAREGIVING = 2.5;

*HERES THE CODE;

*CODE TO EXAMINE KCALS EXPENDED IN WALKING AND VIGOROUS EXERCISE, SIMILAR TO PAFFENBARGER;

```
/******  
WSKCAL=P2WTK*WSKKWK;  
EXKCAL=P2WTK*EXKKWK;  
HAKCAL=SUM(OF WSKCAL EXKCAL);  
*CATEGORICAL HARVARD ALUM VARIABLE;  
IF HAKCAL=0 THEN HACAT=0;  
IF 0 < HAKCAL < 500 THEN HACAT=1;  
IF 500 LE HAKCAL < 1000 THEN HACAT=2;  
IF 1000 LE HAKCAL < 1500 THEN HACAT=3;  
IF 1500 LE HAKCAL < 2000 THEN HACAT=4;  
IF HAKCAL GE 2000 THEN HACAT=5;  
*****/
```

*CODE TO CALCULATE AMOUNT OF WALKING IN MINUTES, AMOUNT OF BRISK WALKING IN MINUTES, AND ENGAGEMENT IN INTENSE EXERCISE;

```

*CODE FOR MINUTES OF HIGH INTENSITY EXERCISE. INCLUDES AEROBICS
AND WEIGHT/CIRCUIT TRAINING AND ANY HIGH INTENSITY EXERCISE;

/*****
HIGHXMIN=SUM(OF FAACTIME FATRTIME FAH1TIME FAH2TIME FAH3TIME FAH4TIME);
IF FAACKKWK=0 AND FATRKKWK=0 AND FAH1KKWK=0 THEN HIGHXMIN=0;
IF HIGHXMIN=0 THEN HIGHX90=0;
IF 0<HIGHXMIN<90 THEN HIGHX90=0;
IF HIGHXMIN GE 90 THEN HIGHX90=1;

IF FAHIKKWK=0 THEN HIGHEX=0;
IF FAHIKKWK>0 THEN HIGHEX=1;
*****/
%mend skip ;

options ls=159 ps=51 formchar='|----|+|----+=|-\<>*' nodate nofmterr;

%MACRO SKIP ;
proc contents data=CALC.y7phact;
  title4 'Final Dataset';
run;
%MEND SKIP ;

***check for duplicates***;
data dupes;
  set CALC.y7phact;
  by habcid;
  if not(first.habcid and last.habcid);
run;

proc print;
title4 'Duplicates in phact y7';
run;

```

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 BLDWQMEZ where 3 is easiest; imputations applied for missing BLDWQMEZ	If BLDWQMYN in (0,7,8) then EASEQM=7-BLDWQMEZ If BLDWQMYN=1 then EASEQM=4-BLDWQMDF	<p>IF BLDWQMYN=9 AND (BLMNRS>0 OR BLMNRS4 IN(1,2)) THEN BLDWQMYN=1; IF BLDWQMEZ=8 and BLDWQMYN ne 8 THEN BLDWQMEZ=2; IF BLDWQMDF=8 and BLDWQMYN ne 8 THEN BLDWQMDF=2; IF (BLDWQMYN=0 AND BLDWQMEZ<0) THEN EASEQM=7-BLDW1MEZ; IF (BLDWQMYN=0 AND BLDWQMEZ<0) AND BLDW1MEZ<0 AND (BLDW1MYN=1 OR BLDW1MYN=8 OR BLDW1MYN<=.z) THEN EASEQM=4; IF (BLDWQMYN=0 AND BLDWQMEZ<0) AND (BLDW1MYN=0 AND BLDW1MEZ<=.z) THEN EASEQM=5;</p> <p>IF BLDWQMYN=1 AND BLDWQMDF<0 THEN EASEQM=7-BLDWQMEZ; IF BLDWQMYN=1 AND BLDWQMDF<0 AND BLDWQMEZ<0 THEN EASEQM=2; IF BLDWQMYN<0 AND BLDWQMDF>0 THEN EASEQM=4-BLDWQMDF; IF BLDWQMYN<0 AND BLDWQMEZ>0 THEN EASEQM=7-BLDWQMEZ;</p> <p>IF BLDWQMYN=8 and BLDWQMEZ=8 then EASEQM=4; IF BLDWQMYN=9 AND BLMNRS<0 AND (BLMNRS4<0 OR BLMNRS4=8) THEN EASEQM=.;</p>	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 GADW1MEZ where 3 is easiest; imputations applied for missing GADW1MEZ	If GADW1MYN=1 then EASE1M=0; if GADW1MEZ=3 then EASE1M=1; if GADW1MEZ=2 then EASE1M=2; if GADW1MEZ=1 then EASE1M=3	<p>IF GADW1MYN=8 AND (GADW1MEZ<0 OR GADW1MEZ=8) THEN EASE1M=1; IF GADW1MYN<=.z AND GADW1MEZ<=.z AND EASEQM=4 THEN EASE1M=0; IF GADW1MYN<=.z AND GADW1MEZ<=.z AND EASEQM=5 THEN EASE1M=1; IF GADW1MYN<=.z AND GADW1MEZ<=.z AND EASEQM=6 THEN EASE1M=2; IF GADW1MYN=0 AND GADW1MEZ<=.z AND EASEQM=4 THEN EASE1M=0; IF GADW1MYN=0 AND GADW1MEZ<=.z AND EASEQM=5 THEN EASE1M=1; IF GADW1MYN=0 AND GADW1MEZ<=.z AND EASEQM=6 THEN EASE1M=2; IF GADWQMYN=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; if GADW1MYN=8 and GADW1MEZ in (1,2,3) then EASE1M=4-GADW1MEZ; if GADW1MYN=0 and GADW1MEZ=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 GADW10EZ where 3 is easiest; imputations applied for missing GADW10EZ	EASE1F=7-GADW10EZ; IF GADW10YN=1 THEN EASE1F=4-GADIF;	IF GADW10YN=9 then EASE1F=.; IF GADW10EZ=8 THEN GADW10EZ=2; IF GADIF=8 THEN GADIF=2; IF GADW10YN=1 AND (GADIF=8 OR GADIF<0) THEN EASE1F=2; IF GADW10YN=0 AND GADW10EZ<0 AND GADW10EZ<0 THEN EASE1F=4; IF EASE1F=., AND GADW10YN NE 9 AND GADW10YN>0 AND (GADW20YN=1 OR GADW20YN=8 OR GADW20YN<=.z) THEN EASE1F=4; IF EASE1F=., AND GADW20YN=0 AND GADW20EZ<=.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 GADW20EZ where 3 is easiest; imputations applied for missing GADW20EZ	IF GADW20YN=1 THEN EASE2F=0; ELSE IF GADW20EZ=3 THEN EASE2F=1; ELSE IF GADW20EZ=2 THEN EASE2F=2; ELSE IF GADW20EZ=1 THEN EASE2F=3;	IF GADW10YN=9 then EASE1F=.; IF (GADW20YN=8 OR GADW20YN<0)AND (GADW20EZ<0 OR GADW20EZ=8) AND EASE1F>0 THEN EASE2F=1; IF GADW20YN<=.z AND GADW20EZ<=.z AND EASE1F=4 THEN EASE2F=0; IF GADW20YN<=.z AND GADW20EZ<=.z AND EASE1F=5 THEN EASE2F=1; IF GADW20YN<=.z AND GADW20EZ<=.z AND EASE1F=6 THEN EASE2F=2; IF GADW20YN=0 AND (GADW20EZ<=.z OR GADW20EZ=8) AND EASE1F>0 THEN EASE2F=EASE1F-4; IF GADW10YN=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   \\ilcavallo\HABC\HABC_SAS\Calculated Variables\Programs\Year
7\srfcn_y7.sas
*
*   Lynn Harvey 10/14/05 - Modified year 6 program for use in year 7
***** ;

*options ls=132 ps=58 formchar='|----|+|----+=|-\<>*' nocenter pageno=1 nofmterr
mprint macrogen;

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

%include '\\ilcavallo\habc\habc_sas\programs\initV8.sas';
*libname calc '\\ilcavallo\habc\habc_sas\calculated variables\datasets\Year 7';

*****;
** SELF-REPORT FUNCTION PROGRAM CODE;
**                                     **;

** CALC VAR CODE                                                                **;
**                                                                              **;
** Creates the following variables for 6mo VISIT:                             **;
**   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                                     **;
**   TIREDQM:  get tired walking 1/4 mile                                       **;
**   TIRED1F:  get tired walking 1 mile                                         **;
**   CSAINDEX: climbing stairs ability index                                   **;
**   WKAINDEX: walking ability index                                           **;
**                                                                              **;
** Adapted from Fran Harris Y2 code: srfcn_y2.sas                             **;
** Adapted from Laura Akin Y6 code                                           **;
**                                                                              **;
** Lynn Harvey Akin      10/13/2005                                           **;
**                                                                              **;
*****;
data CALC.Y7SRFCN (keep=HABCID EASEQM EASE1M EASE1F EASE2F /*tiredqm tired1f*/
CSAINDEX WKAINDEX);
  set habc7.ga(keep=HABCID GADWQMYN GADWQMDF GAMNRS GADWQMEZ
              GADW1MYN GADW1MEZ GADW10YN GADIF
              GAMNRS2 GADW10EZ GADW20YN
              GADW20EZ ) ;

*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 GADW1MYN WAS YES, DONT KNOW, OR MISSING THEN NOT THAT EASY WAS
ASSIGNED. IF NO
EASE LEVEL FOR WALKING 1M WAS CODED AND GADW1MYN WAS NO AND GADW1MEZ 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;

*****
* Modified code below because in year 7 we collapsed variables GAMNRS and
GAMNRS4 down

```

* into one variable - GAMNRS. Lynn Harvey 10/14/05
 *****;

```
IF GADWQMYN=9 AND GAMNRS>0 THEN GADWQMYN=1;
*IF GADWQMYN=9 AND (GAMNRS>0 OR GAMNRS4 IN(1,2)) THEN GADWQMYN=1;
IF GADWQMEZ=8 and GADWQMYN ne 8 THEN GADWQMEZ=2;
IF GADWQMDF=8 and GADWQMYN ne 8 THEN GADWQMDF=2;
IF GADWQMYN IN (0,7,8) AND GADWQMEZ<0 THEN EASEQM=4;
  ELSE IF GADWQMYN IN (0,7,8) THEN EASEQM=7-GADWQMEZ;
IF (GADWQMYN=0 AND GADWQMEZ<0) THEN EASEQM=7-GADW1MEZ;
IF (GADWQMYN=0 AND GADWQMEZ<0) AND GADW1MEZ<0 AND (GADW1MYN=1 OR GADW1MYN=8
OR GADW1MYN<=.z) THEN EASEQM=4;
IF (GADWQMYN=0 AND GADWQMEZ<0) AND (GADW1MYN=0 AND GADW1MEZ<=.z)
THEN EASEQM=5;
IF GADWQMYN=1 THEN EASEQM=4-GADWQMDF;
```

```
IF GADWQMYN=1 AND GADWQMDF<0 THEN EASEQM=7-GADWQMEZ;
IF GADWQMYN=1 AND GADWQMDF<0 AND GADWQMEZ<0 THEN EASEQM=2;
IF GADWQMYN<0 AND GADWQMDF>0 THEN EASEQM=4-GADWQMDF;
IF GADWQMYN<0 AND GADWQMEZ>0 THEN EASEQM=7-GADWQMEZ;
```

```
if GADWQMYN=8 and GADWQMEZ=8 then EASEQM=4;
IF GADWQMYN=9 AND GAMNRS<0 THEN EASEQM=.;
*IF GADWQMYN=9 AND GAMNRS<0 AND (GAMNRS4<0 OR GAMNRS4=8) 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 GADW1MYN=1 THEN EASE1M=0;
  ELSE IF GADW1MEZ=3 THEN EASE1M=1;
  ELSE IF GADW1MEZ=2 THEN EASE1M=2;
  ELSE IF GADW1MEZ=1 THEN EASE1M=3;
```

```
*MISSING VALUE RECODES;
IF GADW1MYN=8 AND (GADW1MEZ<0 OR GADW1MEZ=8) THEN EASE1M=1;
IF GADW1MYN<=.z AND GADW1MEZ<=.z AND EASEQM=4 THEN EASE1M=0;
IF GADW1MYN<=.z AND GADW1MEZ<=.z AND EASEQM=5 THEN EASE1M=1;
IF GADW1MYN<=.z AND GADW1MEZ<=.z AND EASEQM=6 THEN EASE1M=2;
IF GADW1MYN=0 AND GADW1MEZ<=.z AND EASEQM=4 THEN EASE1M=0;
IF GADW1MYN=0 AND GADW1MEZ<=.z AND EASEQM=5 THEN EASE1M=1;
IF GADW1MYN=0 AND GADW1MEZ<=.z AND EASEQM=6 THEN EASE1M=2;
IF GADWQMYN=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 DOWNGRADES 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 GADW1MYN=8 and GADW1MEZ in (1,2,3) then EASE1M=4-GADW1MEZ;
if GADW1MYN=0 and GADW1MEZ=8 then EASE1M=EASEQM-4;
```

*CODE TO CREATE A SUMMARY MEASURE OF SELF-REPORTED WALKING ABILITY. THIS SCALE WILL ULTIMATELY RANGE FROM 0 TO 9, BUT AT BASELINE IT RANGES 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;

***TIREDQM;

*IF GADWQMT2<0 THEN TIREDQM=.;
*IF GADWQMT2=8 OR GADWQMT2=9 THEN TIREDQM=1;
*IF GADWQMT2=0 THEN TIREDQM=0;
*IF GADWQMT2=1 THEN TIREDQM=1;

*CODE FOR EASE OF CLIMBING STAIRS;

*EASE1F;
IF GADW10YN=9 THEN DO; EASE1F=.; EASE2F=.; END;
IF GADW10EZ=8 THEN GADW10EZ=2;
IF GADIF=8 THEN GADIF=2;
EASE1F=7-GADW10EZ;
IF GADW10EZ<0 THEN EASE1F=7-GADW20EZ;
IF GADW10YN=1 THEN EASE1F=4-GADIF;
IF GADW10YN=1 AND (GADIF=8 OR GADIF<0) THEN EASE1F=2;
IF GADW10YN=0 AND GADW10EZ<0 AND GADW20EZ<0 THEN EASE1F=4;

IF EASE1F=. AND GADW10YN NE 9 AND GADW10YN>0 AND
(GADW20YN=1 OR GADW20YN=8 OR GADW20YN<=.z)
THEN EASE1F=4;
IF EASE1F=. AND GADW20YN=0 AND GADW20EZ<=.z THEN EASE1F=5;

*EASE2F;
IF GADW20YN=1 THEN EASE2F=0;
ELSE IF GADW20EZ=3 THEN EASE2F=1;
ELSE IF GADW20EZ=2 THEN EASE2F=2;
ELSE IF GADW20EZ=1 THEN EASE2F=3;
IF (GADW20YN=8 OR GADW20YN<0)AND (GADW20EZ<0 OR GADW20EZ=8) AND EASE1F>0 THEN
EASE2F=1;
IF GADW20YN<=.z AND GADW20EZ<=.z AND EASE1F=4 THEN EASE2F=0;
IF GADW20YN<=.z AND GADW20EZ<=.z AND EASE1F=5 THEN EASE2F=1;
IF GADW20YN<=.z AND GADW20EZ<=.z AND EASE1F=6 THEN EASE2F=2;
IF GADW20YN=0 AND (GADW20EZ<=.z OR GADW20EZ=8) AND EASE1F>0 THEN EASE2F=EASE1F-
4;
IF GADW10YN=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 RANGE FROM 0 TO 9, BUT AT BASELINE IT RANGES FROM 4-9;
CSAINDEX=EASE1F + EASE2F;

*IF GADW10WX=8 OR GADW10WX=9 THEN TIRED1F=1;
*IF GADW10WX=0 THEN TIRED1F=0;

```
*IF GADW10WX=1 THEN TIRED1F=1;

LABEL
EASEQM='EASE WALKING 1/4 MILE, 6=VERY EASY'
EASE1M='EASE WALKING 1 MILE, 3=VERY EASY'
WKAINDEX='WALKING ABILITY INDEX, 9=BEST'
/*TIREDQM='GETS TIRED WALKING 1/4 MILE, 1=YES'*/
EASE1F='EASE CLIMBING 1 FLIGHT, 6=VERY EASY'
EASE2F='EASE CLIMBING 2 FLIGHTS, 3=VERY EASY'
CSAINDEX='CLIMBING STAIRS ABILITY INDEX, 9=BEST'
/*TIRED1F='GETS TIRED CLIMBING 1 FLIGHT, 1=YES'*/;

FORMAT EASE1F EASE2F EASEQM EASE1M CSAINDEX WKAINDEX SPMISS.
      /*TIRED1F TIREDQM YNDK.*;/;

RUN;
```