HEALTH AND RETIREMENT STUDY
1995 AHEAD Core
Final, Version 2.0
May 2002
(Revised January 2003)
Data Description and Usage

TABLE OF CONTENTS

_____

Data Description and Usage

1. Overview

The 1995 AHEAD Core (Final, Version 2.0) data release includes data obtained as part of the Asset and Health Dynamics among the Oldest Old Study (AHEAD). Funding was provided by the National Institute on Aging at NIH (U01 AG09740), with supplemental support from the Social Security Administration. AHEAD is a national longitudinal study that describes for older Americans the dynamics of economic, health, marital, and family statuses, as well as public and private support systems. It was companion study to the Health and Retirement Study (HRS), and was administered at the Institute for Social Research (ISR), the University of Michigan. These studies provide a rich source of longitudinal, cross-sectional data for researchers and policymakers who study aging.

By receiving the data, which have been freely provided, you must agree to use it for research and statistical purposes only and make no effort to identify the respondents therein. In addition, you must agree to send us a copy of any publications you produce based on the data. See Obtaining the Data for additional details.

1A.   The Sample Interviewed in 1995

The AHEAD sample, which consists of people who were born in 1923 or earlier and of their spouses or partners at the time of the initial interview, was interviewed previously in 1993. The data collection period for the 1995 interview was November 1995 through June 1996. In 1998, and every two years thereafter, the HRS, AHEAD and the two new samples – War Babies (WB), and Children of the Depression (CODA) - were interviewed. For more details about the samples, see The Sample Design below.

1B.   What's Included in Version 2

Version 1 of the 1995 AHEAD Core (Final) included all sections, but did not include variables identified as "family" variables most of which are located in sections D and J or "helper" variables in section E. Version 2 includes all sections and all variables. See Detailed Comparison of Data in Version 1 and Version 2 for details.

1C.   1995 Questionnaire Sections

The 1995 AHEAD data collection instrument or questionnaire contained the following sections.

```
     Section PR   Preload
     Section CS   Cover Screen
     Section A    Demographics
     Section B    Health Status
     Section PC   Proxy Cognition
     Section C    Cognition
     Section D    Family Structure
     Section E    Health Care Utilization
     Section F    Housing
     Section G    Employment (includes GA, GB and GC)
     Section H    Expectations
     Section J    Assets and Income
     Section N    Widowhood and Asset Changes
```

```
      Section R    Health Insurance and Event History
      Section M    Modules
      Section T    Other (Time Stamps, Thumbnails)
```

See Modules for the 1995 Data Collection below for details about the 1995
modules.

1D. Levels of Files

In the 1995 data collection instrument, most questions were asked of all
respondents. For two respondent households, some questions were asked just of a
designated financial respondent on behalf of the entire household. Other
questions were just asked of a designated family respondent on behalf of the
entire household. And some questions were just asked of the first respondent
interviewed, the coversheet respondent, on behalf of the entire household.

In addition to the familiar household-level and respondent-level files, the 1995
AHEAD Core (Final, Version 2.0) data release contains files at four other levels
-- helper, household-member-and-child, transfer-to-child and transfer-from-
child.

1D1. Household-Level Files

Household-level files contain questions that were asked about the household of a
designated coversheet, financial or family respondent. A coversheet respondent
answered family questions (section CS) on behalf of the entire household. A
financial respondent answered household-level financial questions (sections F,
J, and N) on behalf of the entire household. And, a family respondent answered
family questions (section D) on behalf of the entire household. The household-
level files contain one record for each household in which at least one
interview was obtained in 1995.

1D2. Respondent-Level Files

Respondent-level files contain questions that were asked of all respondents
about themselves (or asked of a proxy about the respondent if the respondent was
not able to give an interview). The files one record for each respondent or
proxy who gave an interview in 1995 AHEAD.

1D3. Household-Member-and-Child -level Files

These files contain information provided by the family respondent (sections PR,
CS and D) or financial respondent (section J) about each household member or
child of the respondent or the child of the respondent's spouse. The files
contain one record for each household member or child.

1D4. Helper-Level File

This file contains information provided by each respondent about helpers other
than the respondent's spouse or partner. A helper may be a person who or
organization that was reported by the respondent as providing help with ADLs or
IADLs. If the helper is a child, the record contains information about the child
and also about the child's spouse or partner, if any. The file contains one
record for each helper or, if the helper is a married or partnered child, the
helping couple.

1D5. Transfer-to-Child/Grandchild-Level File

This file contains information provided by the family respondent about transfers
of money to children or grandchildren. The file contains one record for each
transfer to a child or grandchild.

1D6.   Transfer-from-Child/Grandchild -Level-File

This file contains information provided by the family respondent about transfers
of money from children or grandchildren. The file contains one record for each
transfer from a child or grandchild.

2. File Naming Conventions

In the version 2 release, files are named beginning with "A95" for 1995 AHEAD.
Next a letter or two designating the questionnaire section. Next a separator,
"_" and then one or two letters designating the level,

        H for household-level,
        R for respondent-level,
        MC for household-member-and-child/grandchild-level,
        HP for helper-level,
        TC for transfer-to-child/grandchild-level, and
        FC for transfer-from-child/grandchild-level.

Putting it all together, files named A95A_R include variables from section A at
the respondent-level.

The following extensions are used for the six different types of files that are
distributed.

        .DA for data files,
        .SAS for SAS program statements,
        .SPS for SPSS program statements,
        .DO for Stata DO statements,
        .DCT for Stata dictionary statements, and
        .TXT for codebook files.

One of each of these file types is provided for each of 27 parts of the 1995
AHEAD Core (Final, Version 2.0) data release.

3. Data Files

The 1995 AHEAD Core (Final, Version 2.0) data release is distributed in 27
parts. The parts are listed below along with the number of cases (N), number of
variables (NV), and the primary identifiers (IDS).

1995 AHEAD data are provided in ASCII format, with fixed-length records. The
records in the data files are sorted in order by the primary identifiers.
You'll want to use associated SAS, SPSS or Stata program statements to read the
data into the analysis package of your choice.

Household-level-files

  A95CS_H   N=5222 NV=58   IDS=HHID DSUBHH
  A95D_H   N=5222 NV=80   IDS=HHID DSUBHH
  A95F_H   N=5222 NV=190   IDS=HHID DSUBHH
  A95J_H   N=5222 NV=561   IDS=HHID DSUBHH
  A95N_H   N=5222 NV=120   IDS=HHID DSUBHH

Respondent-level-files

  A95PR_R   N=7027 NV=146   IDS=HHID PN
  A95CS_R   N=7027 NV=124   IDS=HHID PN

```
A95A_R       N=7027   NV=74      IDS=HHID PN
A95B_R       N=7027   NV=183     IDS=HHID PN
A95C_R       N=7027   NV=113     IDS=HHID PN
A95PC_R      N=7027   NV=72      IDS=HHID PN
A95D_R       N=7027   NV=31      IDS=HHID PN
A95E_R       N=7027   NV=205     IDS=HHID PN
A95G_R       N=7027   NV=914     IDS=HHID PN
A95H_R       N=7027   NV=32      IDS=HHID PN
A95J_R       N=7027   NV=26      IDS=HHID PN
A95M_R       N=7027   NV=340     IDS=HHID PN
A95N_R       N=7027   NV=77      IDS=HHID PN
A95R_R       N=7027   NV=118     IDS=HHID PN
A95T_R       N=7027   NV=38      IDS=HHID PN


Household-member-and-child-level-files

A95PR_MC     N=15617 NV=27       IDS=HHID DSUBHH OPN
A95CS_MC     N=15617 NV=17       IDS=HHID DSUBHH OPN
A95D_MC      N=15617 NV=33       IDS=HHID DSUBHH OPN
A95J_MC      N=15617 NV=27       IDS=HHID DSUBHH OPN


Transfer-to-child/grandchild-level-file

A95D_TC      N=2829   NV=28      IDS=HHID DSUBHH OPN DTRANNUM


Transfer-from-child/grandchild-level-file

A95D_FC      N=829    NV=27      IDS=HHID DSUBHH OPN DTRANNUM


Helper-level-file

A95E_HP      N=2136   NV=32      IDS=HHID PN OPN DHLPTYPE
```

4. Identification Variables

Identification variables for 1995 AHEAD are stored in character format. For further information about merging 1995 data, see Examples of Merging below.

4A.   Primary Identification Variables for Datasets at Each of the Six Levels

Six variables, HHID, DSUBHH, PN, OPN, DTRANNUM, and DHLPTYPE are used in various combinations to uniquely identify the six different level datasets that comprise this data release. The primary identifiers that uniquely identify a record in the datasets are listed in the table above.

4A1.   HHID – Household Identifier.

At the initial wave of data collection in 1993 each sample household was assigned a Household Identifier. HHID is stable across waves of data collection and uniquely identifies the original household and any households derived from that household in subsequent waves of data collection. HHID has six-digits.

4A2.   DSUBHH – AHEAD 1995 Sub-household Identifier.

DSUBHH identifies a household within a HHID at the time of 1995 interview. Sub-household identifiers can be different at each wave. DSUBHH has one-digit. For more information, see Examples of Sub-Household and Respondent Person Number and Other Person Number Assignments below.

4A3.   PN – Person Number.

PNs are unique within an original household (HHID). The PN assigned to a particular respondent does not change across waves. PN has three-digits.

4A4.   OPN – Other Person Number.

OPNs are unique within a sub-household (DSUBHH for 1995). OPN has three-digits. See Datasets Including DSUBHH and OPN as Primary Identifiers for additional information.

4A5.   DTRANNUM – 1995 TRANSFER NUMBER

DTRANNUM designates whether transfer is first, second, etc. from/to a particular child or child's child. DTRANNUM has one digit.

4A6.   DHLPTYPE – 1995 HELPER TYPE

DHLPTYPE designates whether help was from 1) a household member, child or other unlisted helper or from 2) a spouse of household member or child. DHLPTYPE has one digit.

4B.   Secondary Identification Variables

In addition to the primarily identification variables that uniquely identify records in a dataset, secondary identification variables that allow links to other datasets are provided. Two sets of secondary identification variables are provided. They are listed below.

Respondent-level and Helper-level datasets:

```
DSUBHH    1995 SUB-HOUSEHOLD IDENTIFIER
BSUBHH    1993 SUB-HOUSEHOLD IDENTIFIER
DPN_SP    1995 SPOUSE/PARTNER PERSON NUMBER
DCSR      1995 WHETHER COVERSHEET RESPONDENT
DFAMR     1995 WHETHER FAMILY RESPONDENT
DFINR     1995 WHETHER FINANCIAL RESPONDENT
DQNR      1995 QUESTIONNAIRE
```

Household-level, Household-member-or-child-level, Transfer-from-child/grandchild-level and Transfer-to-child/grandchild-level datasets:

```
BSUBHH    1993 SUB-HOUSEHOLD IDENTIFIER
DPN_CS    1995 COVERSHEET RESP PERSON NUMBER
DPN_FAM   1995 FAMILY RESP PERSON NUMBER
DPN_FIN   1995 FINANCIAL RESP PERSON NUMBER
DPN_NCS   1995 NON-COVERSHEET RESP PERSON NUMBER
DPN_NFAM  1995 NON-FAMILY RESP PERSON NUMBER
DPN_NFIN  1995 NON-FINANCIAL RESP PERSON NUMBER
DQNR_CS   1995 QUESTIONNAIRE - COVERSHEET RESP
DQNR_FAM  1995 QUESTIONNAIRE - FAMILY RESP
DQNR_FIN  1995 QUESTIONNAIRE - FINANCIAL RESP
```

4C.   Datasets Including DSUBHH and OPN as Primary Identifiers

When working with datasets including DSUBHH and OPN as primary identifiers -- household-member-and-child-level, transfer-from-child/grandchild-level, or transfer-to-child/grandchild-level datasets – it is essential to use DSUBHH to distinguish other persons.

Sometimes records with the same HHID and OPN will be separate reports about the same person. For example, in the example where the couple with children divorces (Couple with Children and Siblings Divorces.), the record where HHID is 89012, DSUBHH is 1, and OPN is 101 is the report about the child from the mother, and the record where HHID is 89012, DSUBHH is 2, and OPN is 101 is the report about the same child from the father.

At other times records with the same HHID and OPN will be reports about different persons. For example, in example where the couple divorces and both split-off households have new members (Couple Divorces, One Respondent Remarries, Both Split-off Households Have New Members.), the record where HHID is 90123, DSUBHH is 1, and OPN is 151 is for the ex-wife's mother, while the record where HHID is 90123, DSUBHH is 2, and OPN is 151 is for the ex-husband's new stepchild.

See also the section Linking Other Persons across Time below.

5. Distribution Files and Directory Structure

5A. Distribution Files

The files are packaged for download from our Web site in two different ways – as one big .zip file that contains seven smaller .zip files or the seven smaller .zip files available individually for separate download. The combined file is a95core.zip.

The individual files for separate download are:

Data files
    a95da.zip containing data files.

Program statement files
    a95sas.zip containing SAS data descriptors.
    a95sps.zip containing SPSS data descriptors.
    a95sta.zip containing Stata data descriptors.

Documentation files
    a95cb.zip containing codebook.
    a95qn.zip containing questionnaire.
    a95spc.zip containing Surveycraft programming specifications.

You'll typically want to download at least the data files, one set of program statement files, and the codebook files. Decompress the selected .zip files into the appropriate subdirectories. You will need about 51 MB of free space on your storage device to store the 27 .da files.

5B. Directory Structure

While a particular setup is not required for using AHEAD files, we have traditionally suggested a directory structure. By using this directory structure, you will not have to change the path name in your data descriptor files. If you use a different structure, just change the directory references in the distribution files.

Directory   Contents

c:\ahd1995   Files downloaded from Web site

```
c:\ahd1995\codebook       Unzipped files from a95cb.zip
c:\ahd1995\data           Unzipped files from a95da.zip
c:\ahd1995\qnaire         Unzipped files from a95qn.zip
c:\ahd1995\sas            Unzipped files from a95sas.zip
c:\ahd1995\spss           Unzipped files from a95sps.zip
c:\ahd1995\stata          Unzipped files from a95sta.zip
c:\ahd1995\scraft         Unzipped files from a95prn.zip
```

6. Program Statements

Each data file comes with associated SPSS, SAS or Stata program statements to read the data. Files containing SPSS statements are named with a .SPS extension, those with SAS statements with a .SAS extension, and those with Stata statements with .DO and .DCT extensions.

The statement files are named beginning with the same prefix as the corresponding data file. For example, SAS statements in the file A95A_R.SAS go with the A95A_R.DA data file.

6A.   Using the Files with SAS

To create a SAS system file for a particular dataset, two file types must be present for that dataset -- .SAS program statement files and .DA data files.

To create a SAS system file, load the *.SAS file into the SAS Program Editor.

If the *.SAS file is located in "c:\ahd1995\sas" and the data file is located in "c:\ahd1995\DATA", you can run the file as is. A SAS system file (*.SD2 or *.SAS7BDAT) will be saved to directory "c:\ahd1995\sas".

If the files are not located in the specified directories, you will need to edit the *.SAS file to reflect the proper path names prior to running the file.

6B.   Using the Files with SPSS

To create an SPSS system file for a particular dataset, two file types must be present for that dataset -- .SPS program statement files and .DA data files.

To create an SPSS system file, open the *.SPS file in SPSS as an SPSS Syntax File.

If the *.SPS file is located in "c:\ahd1995\spss" and the data file is located in "c:\ahd1995\DATA", you can run the file as is. An SPSS system file (*.SAV) will be saved to directory "c:\ahd1995\spss".

If the files are not located in the specified directories, you will need to edit the *.SPS file to reflect the proper path names prior to running the file.

6C.   Using the Files with Stata

To use Stata with a particular dataset, the following three file types must be present for that dataset -- .DCT files, .DO files, and .DA data files.

The .DA files contain the raw data for Stata to read. The .DCT files are Stata dictionaries used by Stata to describe the data. The .DO files are short Stata programs ("do files"), which you may use to read in the data. Load the .DO file into Stata and then submit it.

If the *.DO and .DCT files are located in "c:\ahd1995\stata" and the data file
is located in "c:\ahd1995\DATA", you can run the .DO file as is.

If the files are not located in these directories, you must edit the *.DO and
*.DCT files to reflect the proper path names before you run the files.

7. Documentation

There are several types of documentation available for use with the 1995 AHEAD
Core (Final, Version 2.0) data release. In addition to this document, they
include a codebook, programming specifications, and a box-and-arrow
questionnaire.

7A. Codebook

The 1995 AHEAD Codebook is provided as a series of 27 ASCII text files and one
master code text file. There is a codebook file corresponding to each data
file. Each variable has its own codebook entry. Examples are included below.

```
=============================================================================
D1161      C1.RATE MEMORY
           Section: C              Level: Respondent       CAI Reference: Q1161
           Type: Numeric           Width: 1                Decimals: 0

           C1. Part of this study is concerned with people's memory, and ability
           To think about things. First, how would you rate your memory at the
           Present time? Would you say it is excellent, very good, good, fair or
           poor?

           .................................................................
            472          1. EXCELLENT
           1382          2. VERY GOOD
           2502          3. GOOD
           1403          4. FAIR
            361          5. POOR
                         7. Other
              3          8. DK (don't know); NA (not ascertained)
              1          9. RF (refused)
            903      Blank. INAP (Inapplicable); [Q218:CS1a] IS (2 OR 3)

=============================================================================
D2246      F6.HOME VALUE $
           Section: F              Level: Household        CAI Reference: Q2246
           Type: Numeric           Width: 8                Decimals: 0

           F6. What is its present value? I mean, about what would it bring if
           it were sold today?
           DO NOT PROBE DK
           0 NOTHING
           AMOUNT:
           .................................................................
           2759          0-99999996. Actual value (allowable range is shown)
                            99999997. Other
            745           99999998. DK (don't know); NA (not ascertained)
             35           99999999. RF (refused)
           1683              Blank. INAP (Inapplicable); [Q377:R FIN-NON V106]
                                    IS (NE 2); [Q240:CS11] IS (1) AND
                                    [Q420:CS26] IS (NE 1); [Q2226:F3] IS (2);
```

```
                                [Q2226:F3] IS (7 OR DK OR RF); [Q2229:F4]
                                IS (3); [Q2229:F4] IS (7 OR DK OR RF);
                                [Q2234:F5] IS (4); [Q2234:F5] IS (7 OR DK
                                OR RF); partial interview

===============================================================================
          ASSIGNMENT STATEMENTS
          *    If [Q210:R SEX] IS (1) then [D290.SEX OF NEW SPOUSE] = 2
          *    If [Q210:R SEX] IS (2) then [D290.SEX OF NEW SPOUSE] = 1
          *    If [Q1:HH ID] IS (GE 0 OR LE 0) then [D290.SEX OF NEW SPOUSE] = 0
          - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
D290      SEX OF NEW SPOUSE
          Section: CS           Level: Household         CAI Reference: Q290
          Type: Numeric         Width: 1                 Decimals: 0

          User note:  This value was used to update Q26.
          ......................................................................
                         0. Unknown
        20               1. Male
        19               2. Female
      5183      Blank. INAP (Inapplicable); [Q220:WHICH IW 1ST R-2ND R]
                        IS (NE 1); [Q263:NEW SP] IS (NE 1)


===============================================================================
```

7A1.   Variable and Question Information

A horizontal line of equal signs sets off codebook entries from each other.
Variable and question information is separated from codeframe information by a
line of periods.

7A1a.   Assignment Instructions

Assignment statements are included for variables where no question was asked and
the CAI program assigned a value based on other variables. If present, they are
followed by a line of dashes and precede any other variable information.

7A1b.   Variable Information

The variable information is contained on three lines. The first eight columns
of the first line contain the variable name. Variable names that are specific
to the 1995 AHEAD cross-section begin with an "D", i.e., "D1161". Other
variable naming conventions are described below. To the right of the variable
name is the variable label.

The next two lines contain six fields with additional variable information.

7A1b1.   Section

This is the section letter of the data files. This may differ slightly from the
letter the question text. For instance included in data file for section G are
questions labeled GA, GB, and GC. Questionnaire section EV is found in data
file section R. All modules are found in section M.

7A1b2.   Level

This indicates whether the data were collected at the household-level,
respondent-level or another level.

        "Household" for household-level files,

> "Respondent" for respondent-level files,
> "HHMemberChild" for household member/child-level files,
> "ToChild" for transfer-to-child/grandchild-level files,
> "FromChild" for transfer-from-child/grandchild-level files, and
> "Helper" for helper-level files.

7A1b3.   CAI Reference

The CAI (Computer Assisted Interviewing) Reference begins with the letter "Q".
For data collected as part of the interview and not masked, the variable name
and CAI name differ only in this first letter and in any extension to the
variable name distinguishing multiple mention and looped variables. These
numbers are in the range 1 to 5991. The 10000 and 30000 series are used for
masked variables, and derived variables with alterative formats. The 9000
series is used as the CAI Reference for identification variables. The 20000
series is used for data release version variables.

7A1b4.   Type

Variable type -- numeric or character. Most variables are stored as numeric
with the exception of the identification variables, which are stored as
characters.

7A1b5.   Width

Variable width in the ASCII data file.

7A1b6.   Decimals

The number of decimals. Usually zero.

7A1c.   Question Text, Interviewer Instructions and User Note

This is followed, for most variables, by the question text. Also included for a
few variables are interviewer instructions, a user note or specification
changes.

Question text in upper and lower case was that read to the respondent. Text
fills, e.g., answers to previous questions that were part of the question text
are designated with a Q-number and question reference or brief description,
e.g., "since Q218-PREV WAVE IW MONTH". See also Banked-screen Questions.

Interviewer instructions that appeared on the CATI screen but were not read to
the respondent and rules that determined what text was read to the respondent
are in upper case.

User notes contain supplemental information provided by HRS staff.

Specification changes note changes made in the data collection instrument during
the course of the data collection period. For a fuller context, see the
Programming Specifications.

7A2.   Codeframe Information

After a line of periods, codeframe information for each variable is provided.

7A2a.   Code Categories or Ranges

For categorical values single code values are provided. For continuous
variables, ranges are designated in the codebook, e.g., "1992-1995"; however,

actual, continuous values, e.g., 1992, 1993, 1994, etc., are provided in the data. Blanks are used throughout for INAP.

7A2b.   Frequencies

To the left of each code is its frequency. The frequencies are unweighted. They can be used to gain an initial overview of the data and to verify that data read into your statistical package accurately.

7A2c.   Code Text

In general, code text picked up from the data collection instrument specifications is in all capital letters. Code text not from the data collection instrument is in upper and lower case letters. These include codes added as the result of the comment review process. See Changes in Data because of Respondent Comments for further description of comment review process.

For values that were determined to be unlikely, the data collection instrument required verification of the value. These are designated in the codebook as "Actual value (verified)".

7A2c1.   INAP Code Text

Extensive documentation of the INAP category derived from the data collection instrument program specifications is provided. The rules specify the CAI reference and do not include loop or multiple mention information.

In a series of variables derived from multiple-mention question, the INAP category of mentions other than the first will include respondents who provided no second, third, etc. response even though the INAP text does not explicitly state this.

When used in the codebook, typically in the INAP text, but also in question text that includes rules for alternative wording, the (A) symbol means "answered". For example, "[Q1264:C6A] IS (A)" can be read as "if Q1264 was answered".

A vertical bar as part of the question reference in the INAP text, e.g., [Q513:CS # CHILDREN│512U(1-20)=], indicates the variable was assigned a value as the result of a CAI calculation.

The rules in the INAP text do not reference codes added to the codeframe as a result of the comment review process. Nor to they reflect occasional, minor, changes made to variable labels in processing.

7A3.   Special Considerations

Several special codebook conventions and types of variables are described below.

7A3a.   Multiple-response and Looped Variables

Variables from questions that allowed multiple-mentions have a "M" in their name and a suffix that designates the mention, e.g., D702M1, D702M2, D702M3.

Variables from questions that were in a loop have an "_" in their name and a suffix that designates the loop, e.g., D2969_1, D2969_2, D2969_3.

Null multiple mention variables and variables from null loops are not included in the data. User notes indicating the number of mentions or loops possible and the number with non-null data are provided, e.g., "User note: Up to twelve

responses were allowed for this question, the actual maximum number of responses was six."

Variables from questions in loops that allowed multiple mentions end with a three-digit number indicating the loop and the mention. For example, fourteen variables would result for a question that was in a loop of two and allowed up to seven mentions. The resulting variables would be, for example, D3168001 to D3168007 for the seven mentions from the first loop, D3168008 to D3168014 for the seven mentions from the second loop.

A user note is provided indicating the maximum and the actual number of loops and mentions, e.g., "User note: Up to two loops with seven mentions were allowed for this question. The maximum number of loops used was one and the maximum number of responses was two."

7A3b.   Masked Variables

Because we are dedicated to protecting the anonymity of our respondents and to protecting the confidentiality of the information they provide, a number of variables have been masked or are simply not included in this public dataset. Some of these variables may be made available to analysts as restricted data. See our Web site for details.

Names, addresses, days of birth, information on geographical relocation and similar variables are not included in publicly released files. See also Variables Not Included in This Release

Geographical locations are recoded to a level no more detailed than U.S. Census Region and Division. Data on the highest educational degree earned have been further grouped together to increase cell sizes. Industry and occupation codes have been recoded into a limited number of categories from the original three-digit U.S. Census code.

The names of variables that were masked for confidentiality end in the letter "M"; for example, variable "D640M" is the masked version of variable "D640". In all, the following variables have been masked for confidentiality.

```
    D640M     A2A. REGION - US BORN                       A95A_R
    D644M     A2B. REGION - NON US BORN                   A95A_R
    D652M     A3C.R HIGH DEGREE - MASKED                  A95A_R
    D665M     A7A.TYPE HISPANIC - MASKED                  A95A_R
    D665AM    A7A.TYPE HISPANIC - MASKED COMBINED         A95A_R
    D667M     A8. R RACE - MASKED                         A95A_R
    D667AM    A8. R RACE - MASKED COMBINED                A95A_R
    D715M     A27.WHERE LIVE WHEN IN SCH - REGION         A95A_R

    D2820M    GA39/42. OCCUPATION                         A95G_R
    D3254M    GB15/16. OCCUPATION                         A95G_R
```

7A3c.   Variables with Alternative Format

A few variables have been derived. They have an "A" extension and are listed below. They appear in the codebook following the associated variable.

Much of Section A was asked only of new respondents. For your convenience, past wave data were merged for a few variables for previously interviewed respondents.

```
D636A      A1. BIRTHDATE MONTH – COMBINED              A95A_R
D638A      A1B. BIRTHDATE YEAR – COMBINED              A95A_R
D649A      A3. R EDUCATION – COMBINED                  A95A_R
D664A      A7. HISPANIC – COMBINED                     A95A_R

D2135A     HELPER RELATIONSHIP COMBINED SOURCE         A95E_HP

D5991A     VQ FINAL VERSION                            A95T_R
```

In this release, the corresponding variable without the "A" extension contains data only for the new respondents who were asked the question.

In the prior release of this data, 1995 AHEAD Core (Final, Version 1.0), a "flag" variable, ending in an "F", e.g., D636F, indicated whether past-wave data had been merged for select background variables, e.g., D636.

7A3d.   Banked-screen Questions

For some questions one question resulted in two variables. For instance, there are many questions that asked about amounts where the answers were given in the form of "amount" "per" period. For example, the question F7A asking "About how much are the payments on the (first) mortgage or land contract?" results in two variables D2252 -- F7A.$ PAYMENT ON MORTGAGE, and D2253 -- F7B.MORTGAGE PER where D2252 is a dollar amount and D2253 is a time period. In the codebook you'll find the question text with the first of these two variables and a line "AMOUNT: PER:". The second of these two variables includes no question text in the codebook. See the previous variable for the question text.

Similarly, through out the codebook you'll find questions that asked about the date of an event that generated two variables, month and year. For example, the question CS6 asking "In what month and year did you stop living together?" followed in the codebook with "MONTH: YEAR:" generated two variables, D228 -- CS6.MO/YR SEPARATE/DIE and D229 -- CS6A.YEAR SEPARATE/DIE. In the codebook the first variable of this pair includes the question text; the second does not.

7A4.   Master Code

The master code contains the detailed occupation codeframe, which is used for several variables.

7B.   Other Types of Documentation

In addition to this document and the codebook, two other types of documentation are available.

7B1.   Programming Specifications

1995 AHEAD data were collected using the Surveycraft CAI (Computer Assisted Interviewing) program. Surveycraft uses program specification files to determine what the program is supposed to do. The file "A95SPC.TXT" is a basic translation of the 1995 AHEAD Surveycraft program specifications.

This file retains much of the Surveycraft program language syntax and is, in places, not easy to decipher. However, it is the ultimate source for how the data were collected and the flow of the data collection instrument.

7B2.   Box and Arrow Questionnaire

The research community has referred to the type of documentation that describes the questions asked in the interview as a "questionnaire". Since the 1995 AHEAD

data were collected using a CAI program, a traditional hard-copy questionnaire
was not produced as part of the data collection phase. However, for 1995 we
have provided a version of the traditional box-and-arrow questionnaire to help
document the asking sequence of the questions.

8. Additional Notes

Found here are miscellaneous additional notes regarding 1995 AHEAD Core (Final,
Version 2.0) data release. If we become aware of additional issues, they will
be posted on our Web site in the Data Alerts section.

8A.   Basic Demographic Variables

We have included many variables in the preload and coversheet sections because
they determine the flow of the subsequent interview. Some of these variables
were updated by information obtained later in the interview. We have noted this
by including "UPDATED" in the label. However, all this can become confusing.
In addition to the "A" variables listed above, Variables with Alternative
Format, we recommend you use these variables for your analysis. Note that D256
and D674 have identical values.

```
        D257      NEW SP COUPLENESS                          A95CS_R
        D374      R SEX                                      A95CS_R

        D256      MARRIAGE STATUS                            A95CS_R
        D674      A21. R MARITAL STATUS                      A95A_R
```

8B.   Selecting the Correct Variable for the Respondent from Preload Information

The preload section contains multiple instances of a number of variables. The
information below will allow you to select the variable with the information the
respondent.

For the variables in the list below, one of eight variables contains information
about the respondent. For instance, there are eight variables that may contain
information about the respondent's birth year, D25_1 to D25_4 for respondents in
household 1 and D66_1 to D66_4 for respondents in household 2.

```
        D21_1-D21_4          HH1 R INDEX
        D22_1-D22_4          HH1 W1 R INTERV
        D25_1-D25_4          HH1 R W1 BIRTH YR
        D26_1-D26_4          HH1 R W1 SEX - V108 - updated
        D27_1-D27_4          HH1 R MAR STAT - updated
        D33_1-D33_4          HH1 LIVING - updated

        D62_1-D62_4          HH2 R INDEX
        D63_1-D63_4          HH2 W1 R INTERV - updated
        D66_1-D66_4          HH2 R W1 BIRTH YR - updated
        D67_1-D67_4          HH2 R W1 SEX - V108 - updated
        D68_1-D68_4          HH2 R MAR STAT - updated
        D74_1-D74_4          HH2 LIVING – updated
```

In order to know which variable to use, you should use two additional variables
from the coversheet section the first, D203, to determine whether to use the
variables for household 1 or household 2, and the second, D197, to determine
which of the four variables to use.

```
D203 - WHICH HH
     1. Household 1
     2. Household 2

D197 - R INDEX
     1. 010 One of the 1st Rs
     2. 020 2nd of the 1st Rs
     3. 011 Spouse of 010
     4. 021 Spouse of 020
```

For the variables listed below, one of two variables contains information about
the respondent. For instance, there are two variables, D37 and D78, that may
contain information about the respondent's previous wave second residence. You
should use D203 to determine which of the two variables to use.

```
     D20                HOLD 2 HH1 RANDOM MODULE/WORDLIST – updated
     D34                HH1 1ST R IN NURS HOME - updated
     D37                HH1 HAS 2ND RESIDENCE
     D88                HOLD 19 HH1 1st R MOVED - updated

     D61                HOLD 11 HH2 RANDOM MODULE/WORDLIST – updated
     D75                HH2 1ST R IN NURS HOME - updated
     D78                HH2 HAS 2ND RESIDENCE
     D89                HOLD 20 HH2 1st R MOVED - updated
```

8C.   Households with No or Two Coversheet or Family or Financial Respondents

As noted earlier in this document, the data collection design was to have asked
most questions of all respondents, some questions of just a designated
coversheet or family or financial respondent on behalf of the household.
However, occasionally that is not what happened. For some households we did not
obtain an interview from a family or financial respondent. And for a few other
households we obtained interviews with two coversheet or family or financial
respondents.

8C1.   Households with No Coversheet or Family or Financial Respondent

Household records are provided for all households. Thirteen households had no
coversheet respondent; 17 households had no family respondent; 21 households had
no financial respondent. The household records for these households contain
null values for the missing information.

Households missing a coversheet or family or financial respondent can
identified, respectively, by values of

```
     DPN_CS – 1995 COVERSHEET RESP PERSON NUMBER
     DPN_FAM – 1995 FAMILY RESP PERSON NUMBER
     DPN_FIN – 1995 FINANCIAL RESP PERSON NUMBER
```

in the household record.

Some households had two respondents, neither of whom, in error, was designated
as a family or financial respondent. There was one household with two
respondents neither of whom was a family respondent; there were two households
with two respondents neither of whom was a financial respondent.

Of these respondents the one whose PN appears as the FPN_FAM or FPN_FIN and the one whose PN appears as the FPN_NFAM or FPN_NFIN in the household record is indicated by the values of

        DFINR - 1995 WHETHER FINANCIAL RESPONDENT
        DFAMR - 1995 WHETHER FAMILY RESPONDENT

in the respondent record.

8C2. *Households with Two Family or Financial Respondents*

One hundred and one households had two coversheet respondents; 30 households had two family respondents; eight households had two financial respondents. The responses of one of these two respondents were selected for inclusion in the household-level files.

The respondents whose responses were used in the household records as well as those whose responses were not used are indicated, respectively, by the values of

        DCSR - 1995 WHETHER COVERSHEET RESPONDENT
        DFINR - 1995 WHETHER FINANCIAL RESPONDENT
        DFAMR - 1995 WHETHER FAMILY RESPONDENT

in the respondent record.

8D.   Detailed Comparison of Data in Version 1 and Version 2

Version 1 of the 1995 AHEAD Core (Final) included all sections, but did not include "family" variables from sections PR, CS, D and J and "helper" variables from section E. Version 2 includes all sections and all variables. Version 2 obsoletes previously released data in:

  1995 AHEAD Core (Final) (v.1.0)
  1995 AHEAD Core Helper (Early) (v.2.0)
  1995 AHEAD Core Other Person (Early) (v.2.0)

Also in version 2, CS variables with information provided by the coversheet respondent previously released at a respondent level are released at a household level. Some section N variables previously released at a household level are released at a respondent level. All modules are released in one file. The EV variables may be found in along with the section R variables. Variables from module 1 and module 2 have been combined.

In addition, the variables included in version 1 and version 2 varied slightly. Bracket components are only included in version 2. Bracket summary variables and continue variables are only included in version 1. See Data Alert at our Web site for a detailed listing of variables included in each version of the 1995 AHEAD Core (Final) data.

The effect of these changes is presented in the table below.

Version 1			Version 2		
File name	N	NV	File name	N	NV
AHD2PR	7027	67	A95PR_R	7027	146
			A95PR_MC	15617	27
AHD2CS	7027	85	A95CS_R	7027	124
			A95CS_H	5222	58

			A95CS_MC	15617	17
AHD2A	7027	68	A95A_R	7027	74
AHD2B	7027	185	A95B_R	7027	183
AHD2PC	7027	72	A95PC_R	7027	72
AHD2C	7027	69	A95C_R	7027	113
AHD2D_H	5222	34	A95D_H	5222	80
AHD2D_R	7027	31	A95D_R	7027	31
			A95D_MC	15617	33
			A95D_TC	2829	28
			A95D_FC	829	27
AHD2E	7027	182	A95E_R	7027	205
			A95E_HP	2136	32
AHD2F	5222	177	A95F_H	5222	190
AHD2G	7027	961	A95G_R	7027	914
AHD2H	7027	32	A95H_R	7027	32
AHD2J	5222	440	A95J_H	5222	561
			A95J_MC	15617	27
			A95J_R	7027	26
AHD2MD1	7027	50	A95M_R	7027	340
AHD2MD10	7027	63			
AHD2MD2	7027	51			
AHD2MD3	7027	57			
AHD2MD4	7027	25			
AHD2MD5	7027	35			
AHD2MD6	7027	64			
AHD2MD7	7027	46			
AHD2N	5222	164	A95N_H	5222	120
			A95N_R	7027	77
AHD2R	7027	122	A95R_R	7027	118
AHD2EV	7027	15			
			A95T_R	7027	38

8E.   Unfolding Bracket Variables and Imputations

If a respondent did not give an exact report for a question, a series of
questions was often asked that allowed us to bracket a response to within a
certain range. These series of questions are included in the 1995 AHEAD Core
(Final, Version 2.0) data release for all of the unfolding bracket sequences.

In the 1995 AHEAD Core (Final, Version 1.0) data release, rather than providing
all of the unfolding bracket questions, we provided just a summary variable,
which contained summary information. These summary variables had a "B" at the
end of the variable name. For example, variable "D2247B" was a summary of
bracket questions Q2247 through Q2250, which were removed from the publicly
distributed data.

More recently, and for this data release, instead of creating these summary
bracket variables, imputations for most unfolding bracket variables have been or
will be derived and provided separately. See the section Imputation File for
more details about the contents of this file.

8F.   Changes in Data because of Respondent Comments

The interviewer recorded comments the respondent made during the course of the
interview. HRS staff reviewed these comments for selected questions. The coded
answer was changed if it was determined that the comment changed the substance
of the recorded answer. Occasionally additional codes were added to an existing
codeframe. The text of added codeframes appears in the codebook in upper and

lower case. Double reporting of income or assets in Section J was corrected wherever possible.

When a change was made as a result of the comment review process, inconsistencies with subsequent variables may result. A limited number of changes for consistency were made to immediately subsequent variables within the section. The INAP text in the codebook does not include codeframes added in the comment review process. No consistency changes were made to variables in other sections. If any comment change affected a subsequent branch-point, the branch-point was not changed.

8G.   Modules for the 1995 Data Collection

The 1995 instrument contained seven modules. These are released in version 2 in one data file.

Module 1 and 2: AHEAD 1993 ADL Questions. This module contains questions about difficulty and the use of equipment and help in activities of daily living that were asked in AHEAD 1993 were modified in 1995 AHEAD. These questions were part of the core questions in AHEAD 1993. To facilitate analysts who wish to take account of these wording changes when examining changes in responses across waves, a random sub-sample of respondents were asked the AHEAD 1993 version of the ADL question (as well as the revised ADL questions that were asked in the main interview).

Most questions in module 1 and module 2 were identical. Variables from identical questions have been combined in the version 2 data release. Question text and INAP text from module 1 and module 2 is included in the codebook. The names of the variables are based on the module 1 questions. The entry points for the unfolding questions in module 1 and module 2 were different. These variables have not been combined.

                         Combined Variables
          Module 1   Module 2
     5327 M1-0.INTRO   5409 M2-0.INTRO
     5328 M1-32.ADL - PREAMBLE   5410 M2-32.ADL - PREAMBLE
     5331 M1-33.ADL WALK   5413 M2-33.ADL WALK
     5332 M1-33A.WALK OFTEN HELP   5414 M2-33A.WALK OFTEN HELP
     5333 M1-33B.WALK EQUIPMENT   5415 M2-33B.WALK EQUIPMENT
     5334 M1-33C.WHAT WALK EQUIP   5416 M2-33C.WHAT WALK EQUIP
     5335 M1-33D.WALK OFTEN EQUIP   5417 M2-33D.WALK OFTEN EQUIP
     5336 M1-33E.WALK DIFF   5418 M2-33E.WALK DIFF
     5337 M1-33F.WALK MUCH DIFF   5419 M2-33F.WALK MUCH DIFF
     5342 M1-35.ADL DRESS   5424 M2-35.ADL DRESS
     5343 M1-35A.DRESS OFTEN HELP   5425 M2-35A.DRESS OFTEN HELP
     5344 M1-35B.DRESS DIFF   5426 M2-35B.DRESS DIFF
     5345 M1-35C.DRESS MUCH DIFF   5427 M2-35C.DRESS MUCH DIFF
     5350 M1-37.ADL BATHE   5432 M2-37.ADL BATHE
     5351 M1-37A.BATHE OFTEN HELP   5433 M2-37A.BATHE OFTEN HELP
     5352 M1-37B.BATHE DIFF   5434 M2-37B.BATHE DIFF
     5353 M1-37C.BATHE MUCH DIFF   5435 M2-37C.BATHE MUCH DIFF
     5358 M1-39.ADL EAT   5440 M2-39.ADL EAT
     5359 M1-39A.EAT OFTEN HELP   5441 M2-39A.EAT OFTEN HELP
     5360 M1-39B.EAT DIFF   5442 M2-39B.EAT DIFF
     5361 M1-39C.EAT MUCH DIFF   5443 M2-39C.EAT MUCH DIFF

```
5366 M1-41.ADL BED            5448 M2-41.ADL BED
5367 M1-41A.BED OFTEN HELP    5449 M2-41A.BED OFTEN HELP
5371 M1-43.BED EQUIPMENT      5453 M2-43.BED EQUIPMENT
5372 M1-43A.BED, WHAT EQUIP   5454 M2-43A.BED, WHAT EQUIP
5373 M1-43B.BED OFTEN EQUIP   5455 M2-43B.BED OFTEN EQUIP
5374 M1-43C.BED DIFF          5456 M2-43C.BED DIFF
5375 M1-43D.BED MUCH DIFF     5457 M2-43D.BED MUCH DIFF
5377 M1-44.ADL TOILET         5459 M2-44.ADL TOILET
5378 M1-44A.TOILET OFTEN HELP 5460 M2-44A.TOILET OFTEN HELP
5379 M1-44B.TOILET DIFF       5461 M2-44B.TOILET DIFF
5380 M1-44C.TOILET MUCH DIFF  5462 M2-44C.TOILET MUCH DIFF
5385 M1-45.FIN GET ALONG      5467 M2-45.FIN GET ALONG
5393 M1-46.SPENT LAST MONTH   5475 M2-46.SPENT LAST MONTH
```

Module 3: ADL Measures Used for the Longitudinal Studies of Aging (LSOA). This module contains the ADL questions used in AHEAD 1993, which were in turn based on those proposed for (and subsequently used in) the second LSOA. This module was based on AHEAD 1993, wave 1, module 3 and was asked of the same respondents in both waves.

Module 4: Security And Safety. This module contains questions about the subjective probability of being the victim of a crime and steps taken out of concern about crime. This module was new for 1995.

Module 5: Sleep. This module contains questions about trouble falling asleep and staying asleep through the night, problems of falling asleep during the day, and the use of medications to aid sleep. It also includes two questions about the sense of personal control. This module was new for 1995.

Module 6: Living Wills. This module contains questions about end-of-life directives, whether or not the person has named anyone to make health care decisions for them if they are unable to do so themselves. The respondents were also presented with two scenarios for someone with a life-threatening illness who is presented with the choice of taking an experimental treatment with randomly varied cost and probability of success. This module was new for 1995.

Module 7: Detailed ADL Measures. This module contains questions about various adaptive strategies that may be involved in carrying out tasks such as bathing and with managing money. This module was based on AHEAD 1993, wave 1, module 7 and was asked of the same respondents in both waves.

Each of the modules includes questions designed to assess the importance of anchoring effects in unfolding questions about dollar amounts. Respondents were asked about the amount of money they had in savings accounts and about total household consumption in the past month. The entry point varied across the modules.

Also included in the module section, although not actually a module, is a section labeled module 10. It is a section of the interview that is administered to sample persons for whom proxy interviews were done, but are willing and able to do this section themselves.

Module 10: Cognitive Section for Proxy Interviewed Sample Members. The cognitive section which the interviewer asked the sample member to do when a proxy interview had been done on their behalf.

8H.   Special Considerations for Transfer To and Transfer From Files

In the 1995 data collection, four separate sets of questions were asked about transfers to and from children and grandchildren. The questions about children and grandchildren have been combined to make the 1995 transfer files more like the transfer files from later waves of data collection. Question text and INAP text from transfers involving both children and grandchildren is included in the codebook. The variable names are based on the children questions. The variable DTYPTRAN, 1995 TYPE OF TRANSFER, indicates whether the transfer involved a child or grandchild.

```
                        Combined Variables
        Transfer to Child                   Transfer to Grandchild
1474                              1496 D56 LOOP
1475 D52.TRANSFER TO CHILDREN $ WHICH-1 1497 D56.TRANSFER TO GRANDKID WHICH-1
1476 D52.TRANSFER TO CHILDREN $ WHICH-2 1498 D56.TRANSFER TO GRANDKID WHICH-2
1479 D53.TRANSFER TO CHILD $ AMOUNT    1501 D57.TRANSFER TO OFFSPRING $ AMOUNT
1480 D53a-c. DK-10               1502 D57a-c. DK-10
1481 D53b.DK D13-2               1503 D57b.DK D13-2
1483 D53c. DK D13-3              1505 D57c. DK D13-3
1484 |1474AM1479(L100+G0+NK1+NK2)=    1506 |1496AM1501(L500+G0+NK1+NK2)=
1485 |1474AM1483(5.K1.K2)=       1507 |1496AM1505(5.K1.K2)=
1486 D54.NEXT CHILD TRANSFER TO       1508 D58.NEXT CHILD OFFSPRING


                        Combined Variables
        Transfer from Child                 Transfer from Grandchild
1521                              1546 D66 LOOP
1522 D62.TRANSFER FROM CHILDREN LARGEST-1 1547 D66.TRANSFER FROM OFFSPRING
1523 D62.TRANSFER FROM CHILDREN LARGEST-2 1548 D66.TRANSFER FROM OFFSPRING
1527 D63.TRANSFER FROM CHILD $A       1551 D67.TRANSFER FROM OFFSPRING $
1529 D63a-c. DK-10               1552 D67a-c. DK-10
1530 D63b.DK D13-2               1553 D67b.DK D13-2
1532 D63c. DK D13-3              1555 D67c. DK D13-3
1533 |1521AM1527(L500+G0+NK1+NK2)=    1556 |1546AM1551(L500+G0+NK1+NK2)=
1534 |1521AM1532(5.K1.K2)=       1557 |1546AM1555(5.K1.K2)=
1535 D63d.  RECEIVE FUTURE       1558 D67d.  RECEIVE FUTURE
1536 D64.NEXT CHILD              1559 D65.NEXT CHILD OFFSPRING
```

Uniquely for the 1995 transfer to grandchildren questions, more than one grandchild could be mentioned as a recipient of a given amount. We have created duplicate records when more than one grandchild was mentioned. The variable DTOTMEN, 1995 TOTAL NUMBER OF MENTIONS, indicates the number of grandchildren mentioned as receiving the reported amount. The analyst, after reviewing the questionnaire, will need to decide whether to split the reported amount among the grandchildren mentioned or to infer that each grandchild got the reported amount.

Some respondents mentioned more than one transfer involving a particular child or children of a particular child. The data collection instrument did not prevent this. The variable DTRANNUM, 1995 TRANSFER NUMBER, uniquely

distinguishes the transfer records involving a particular child or children of a particular child.

8I.   Variables Not Included in This Release

A few variables have not been included in 1995 AHEAD Core (Final, Version 2.0) data release.

- o   Some variables have not been included for reasons of confidentiality. They may be made available to analysts as restricted data. See our Web site for details.

- o   Some open-ended questions were not coded during the initial round of coding. These variables are not included in the 1995 AHEAD Core (Final, Version 2.0) data release.

```
419   CS25b.FACILITY WHERE LIVE
430   CS33.CURRENT RESIDENCE
434   CS35.OTH RESIDENCE CITY/STATE

733   A36a.DENOMINATION

778   B1d.WHY BETTER
810   B5g.  CANCER START
918   B19a. MENTION HEALTH

1343  C.COGNITIVE ACTION

2884  GA52q. HOW PAID
3078  GA112a. HOW JOB DIFFERENT
3441  GB72a. HOW JOB DIFFERENT
3699  GC30A. HOW JOB DIFFERENT

4575  J62b.WHAT OTHER SOURCE
4655  J74b.  USE ASSETS OPEN

5883  M10-13a.REPAIR/CONCEAL
5886  M10-13c.FABRIC/ENORMOUS
5889  M10-13e.DOMESTIC/PERIMETER
5892  M10-13g.REMORSE/COMPASSION
5895  M10-13j.PLAGIARIZE/AUDACIOUS
```

9. Longitudinal Issues

9A.   The Sample Design

The AHEAD sample, first interviewed in 1993, consisted of people who were born in 1923 or earlier, were household residents of the conterminous U.S. in the spring 1992, and were still household residents at the time of their first interview; and of their spouses or partners at the time of the initial interview. The AHEAD sample was interviewed in 1993, 1995, 1998 and every two years thereafter.

In 1998, three additional samples were interviewed.

The HRS sample consists of people who were born 1931 through 1941 and were household residents of the conterminous U.S. in the spring 1992; and of their spouses or partners at the time of the initial interview in 1992. The HRS sample was interviewed in 1992, 1994, 1996, 1998 and every two years thereafter.

The War Babies (WB) sample consists of people who were born 1942 through 1947, were household residents of the conterminous U.S. in the spring 1992, and were still household residents at the time of the first interview in 1998; and their of spouses or partners at the time of the initial interview. The War Baby sample was interviewed in 1998 and every two years thereafter.

The Children of the Depression (CODA) sample consists of people who were born 1924 through 1930, were household residents of the conterminous U.S. when first interviewed in 1998, and who, at that time, did not have a spouse or partner who was born before 1924 or between 1931 and 1947. The Children of the Depression sample was interviewed in 1998 and every two years thereafter.

For more details about the sample, see our Web site.

9B. Two Versions of Masked Occupation Code

For AHEAD data collections in 1993 and 1995, occupation was coded using the SRC 2-Digit Occupation Code (adapted from the 1980 U. S. Census Occupation Code) and masked for public release. For HRS data collections in 1992, 1994 and 1996, occupation was coded using the 3-digit 1980 U. S. Census Occupation Code and masked for public release. The AHEAD and HRS masked occupation codeframes are not the same.

For 1998, the first year that the AHEAD and HRS samples were merged for data collection, we released occupation masked in two ways for the AHEAD sample – using both the AHEAD and HRS masking schemes. Variables masked in the AHEAD tradition were named with a "AM" at the end of the variable name and variables masked in the HRS tradition were named with a "HM" at the end of the variable name, e.g., "F3185AM" and "F3185HM". For data collections after 1998 we will use just the HRS masking scheme for all samples. See our Web site for more detailed information about how occupation and industry were coded in each wave of the study.

9C. Linking Respondents across Time

Respondent records from prior and subsequent waves may be linked to respondent records from 1995 by HHID and PN. The sub-household identifiers, BSUBHH for 1993, DSUBHH for 1995, and FSUBHH for 1998 can be used to link cross-sectional household data with the cross-sectional respondent data.

9D. Linking Other Persons across Time

Linking other person records across time is more complex. You can use the previous wave's SUBHH to link one wave's other person record with the corresponding other person record from the previous wave. For example, to link 1995 other person records with 1993 other person records use HHID, BSUBHH, and OPN.

For some 1995 other person records this will result in two 1995 records matching one 1993 record. For example, in the example where the couple with children divorces (Couple with Children and Siblings Divorces.), the 1995 record where HHID is 89012, DSUBHH is 1, and OPN is 101 is the report about the child from the mother and the record where HHID is 89012, DSUBHH is 2, and OPN is 101 is

the report about the same child from the father. The value of BSUBHH for each of these 1995 records would 0. Each of these 1995 records would match the 1993 record about the child before the couple divorced as the 1993 record would have a HHID of 89012, a BSUBHH of 0 and a OPN of 101.

For other 1995 other person records this will result in no 1995 record matching a 1993 record. For example, in the example where the couple divorces and both split-off households have new members (Couple Divorces, One Respondent Remarries, Both Split-off Households Have New Members.), the record where HHID is 90123, DSUBHH is 1, and OPN is 151 is for the ex-wife's cohabiting mother, while the record where HHID is 90123, DSUBHH is 2, and OPN is 151 is for the ex-husband's new stepchild. The value of BSUBHH for each of these 1995 records would 0. However neither of these 1995 records would match a 1993 record before the couple divorced because no 1993 record would have a HHID of 89012, a BSUBHH of 0 and a OPN of 151.

9E.   Differences from Other Waves

Over the course of the study there have been a number of changes in key variables. There are no new changes of this type for 1995 AHEAD Core (Final, Version 2.0). However, since the analyst may be working with some or all of the other waves of data, changes of this type are documented below.

9E1.   Character Type Identification Variables

In some prior releases, all variables were stored in NUMERIC format. Releasing identification variables in numeric format caused some problems.

- o All digits of an identifier matter. The character string "000031010" is not the same thing as the numeric string "31010".

- o When we begin to deal with complex respondent identifiers such as those in HRS and attempted to create one identification variable by combing several, we run the risk of inadvertently corrupting the identifier due to loss of precision as a side effect of treating it as a numeric value.

Thus starting with 1995 AHEAD, we are releasing all primary and secondary identification variables in character format and are no longer creating combined identification variables.

Also released as character variables are variables that include an other person number that may be used for merging with a primary or secondary identification variable. Typically these are variables from questions asking who or which person did something.

Most software packages will require that the identification variables in each dataset be of the same type. We suggest that analysts of older HRS or AHEAD datasets convert the old, numeric identifiers to character type.

For the most part, we leave the method of this transformation to you. However, we do have a couple of suggestions.

- o Each software package has different ways of converting variable types in an existing dataset.

     In SAS, for example, to convert numeric variable HHID to character variable HHID, you can use this code.

```
            *      numeric to character;
         CHHID=put(HHID,z6.0);
         attrib
              CHHID label=' HOUSEHOLD IDENTIFIER'
              length=$6 format=$char6.;
         drop HHID;
         rename CHHID=HHID;
```

In STATA to convert numeric variable HHID to character variable HHID, you can use this code.

```
         *        numeric to string
         replace hhid=hhid+1000000
         gen str7 chhid=string(hhid)
         drop hhid
         gen str6 hhid=substr(chhid,2,6)
         drop chhid
```

o Before reading in datasets, you can change the variable type in the data descriptor statements themselves.

```
   Package    Type         Statement for variable V1
   -------    ---------    -------------------------------
   SAS        numeric      V1 1-4
              character    V1 $ 1-4
   -------    ---------    -------------------------------
   SPSS       numeric      V1 1-4
              character    V1 1-4 (A)
   -------    ---------    -------------------------------
   Stata      numeric      _column(1) byte V1 %4.0f "LABEL"
              character    _column(1) str4 V1 %4s   "LABEL"
```

9E2. INAP Codes Stored as Blanks

Beginning with 1995 AHEAD data, INAP codes were stored in the ASCII data files as blanks. In 1993, INAPs were assigned to a system missing value. If you are merging the 1995 data with data from 1993, you will want to be aware of these differences.

10.   Other Data Files

10A. Tracker File

We attempt to aid you in tracking the sample longitudinally through use of our "tracker file". The tracker file is available at our Web site.

The current version of the tracker file, Cross-Wave: Tracker File (Version 2.0), is based on information through the 1998 data collection. Version 3.0 of the tracker file, based on information through the 2000 data, will be available soon.

The tracker file contains a record for every individual who was interviewed one or more times during the period 1992 through 1998 (or 2000) as well as a record for the non-responding spouse or partner of such a person. The primary identification variables for the tracker file are HHID and PN.

The tracker file contains basic demographic information as well as some basic information obtained from the National Death Index (NDI). In addition, it contains basic information about the interview year such as whether a person gave an interview in a particular wave. The tracker file also contains cross-sectional weights for each interview year 1992 through 1998 (or 2000) for both respondent and household level analysis.

10B. Imputation File

Imputed variables based on the 1995 AHEAD Core (Final, Version 1.0) (December 1999) data may be found in the 1995 AHEAD Core Imputations (Final, Version 1.0) (July 2000). These files contain imputations for all bracket variables from all sections satisfying our ad hoc selection rules listed below.

1) The Bracket Rule: In general, all the variables with brackets will imputed, and included in the imputation file.
2) The Income/Asset Component Rule: A variable without brackets will be imputed if it is determined to be an integral component of household income, housing equity, or net wealth.
3) The Imputability Rule: A variable eligible for imputation based on Rule 1 or 2 will not be imputed if there are no enough valid observations.

Also included are four aggregate variables -- main home equity, and second home equity, total assets, and household income. The files are available at our Web site.

11.   Obtaining the Data

11A. Registering and Downloading the Data

AHEAD data are available for free to researchers and analysts at the HRS Web site. In order to obtain public release data, you must first register at our Web site and get a username and password. Your username and password are required to download any data files.

11B. Conditions of Use

By registering, you agree to the Conditions of Use governing access to Health and Retirement public release data. By registering all users, we are able to document for our sponsors the size and diversity of our user community allowing us to continue to collect these important data.

Registered users receive user support, information related to errors in the data, future releases, workshops, and publication lists. The information you provide will not be used for any commercial use, and will not be redistributed to third parties. For more information concerning privacy issues and conditions of use, please read "Conditions of Use for Public Data Files" and "Privacy and Security Notice" at the Public File Download Area of the HRS Web site.

11C. Publications Based on Data

As part of the data registration process, you agree to inform HRS of any papers, publications, or presentations based on AHEAD data. Please send a copy of any publications you produce based on HRS data, with a bibliographical reference, if appropriate, to the address below.

        Health and Retirement Study
        Attn: Papers and Publications

The Institute for Social Research, Room 3050
P.O. Box 1248
Ann Arbor, MI (USA) 48106-1248

Alternately, you may contact us by e-mail at hrsquest@isr.umich.edu with "Attn: Papers and Publications" in the subject line.

12.   If You Need to Know More

This document is intended to serve as a brief outline and to provide an approach to using the 1995 AHEAD Core (Final, Version 2.0) data. If you have questions or concerns that are not adequately covered here or on our Web site, or if you have any comments, please contact us. We will do our best to provide answers.

12A. HRS Internet Site

Health and Retirement Study public release data and additional information about the AHEAD and HRS studies are available through the Internet. To access the data and other relevant information, point your Web browser to the HRS Web site:

http://www.umich.edu/~hrswww/

12B. Contact Information

If you need to contact us, you may do so by one of the methods listed below.

E-mail: hrsquest@isr.umich.edu

Postal service:
        Health and Retirement Study
        The Institute for Social Research, Room 3050
        The University of Michigan
        P.O. Box 1248
        Ann Arbor, MI 48106-1248

FAX: (734) 647-1186

Appendices

1. Examples of Merging

Many analyses require variables that appear in separate files. Sometimes you will need to obtain variables from files at different levels that contain different numbers of records. Before you can do your analysis work, the files will need to be merged in an appropriate manner.

Prior to doing any data management you should ask yourself several questions. What should be the level of the analysis file – will it have one record per respondent, one record per household or what? What are the variables of interest? What identification variables will be required to merge the various files that contain variables needed for your analysis? Will the merge be a one-to-one matching of records or a one-to-many matching?

We've provided sample setups for several types of merging. We hope that these will be of help to you as you think of your particular analytic needs.

1A.   Respondent-level Files

To create a respondent-level file with data from two or more respondent-level files, merge the respondent-level files using HHID and PN. This is a one-to-one match. Each input file contains 7,027 records. A respondent-level output file with 7,027 respondent records results.

1A1.   SAS Example of Respondent to Respondent Merge

```
proc sort
     data=in.a95a_r
     out=a95a_r(keep=hhid pn d636 d638);
     by hhid pn;
run;
proc sort
     data=in.a95b_r
     out=a95b_r(keep=hhid pn d772 d776);
     by hhid pn;
run;
proc sort
     data=in.a95c_r
     out=a95c_r(keep=hhid pn d1162 d1168);
     by hhid pn;
run;
data resp;
     merge a95a_r a95b_r a95c_r;
     by hhid pn;
run;
```

1A2.   Stata Example of Respondent to Respondent Merge

```
set prefix "ahd1995"
* This is an optional statement
* Make sure all the data files are in directory "ahd1995"

use HHID PN D772 D776 using a95b_r
sort HHID PN
```

```
save tmp1, replace

use HHID PN D1162 D1168 using a95c_r
sort HHID PN
save tmp2, replace

use HHID PN D636 D638 using a95a_r
sort HHID PN
merge HHID PN using tmp1
drop _m
sort HHID PN
merge HHID PN using tmp2
drop _m
sort HHID PN
save resp, replace
```

1B.   Household-level Files

To create a household-level file with data from two or more household-level
files, merge the household-level files using HHID and DSUBHH. This is a one-to-
one match. Each input file contains 5,222 records. A household-level output
file with 5,222 household records results

1B1.   SAS Example of Household to Household Merge

```
proc sort
     data=in.a95f_h
     out=a95f_h(keep=hhid dsubhh d2223 d2224);
     by hhid dsubhh;
run;
proc sort
     data=in.a95j_h
     out=a95j_h(keep=hhid dsubhh d4753 d4752);
     by hhid dsubhh;
run;
data hhold;
     merge a95f_h a95j_h;
     by hhid dsubhh;
run;
```

1B2.   Stata Example of Household to Household Merge

```
use HHID DSUBHH D2223 D2224 using a95f_h
sort HHID DSUBHH
save tmp, replace

use HHID DSUBHH D4753 D4752 using a95j_h
sort HHID DSUBHH
merge HHID DSUBHH using tmp
drop _m
sort HHID DSUBHH
save hhold, replace
```

1C.   Respondent-level Files with Household-level Data

To create a respondent-level file including data from a household-level file,
merge the respondent-level file(s) and the household-level file(s) using HHID

and DSUBHH. This is a one-to-many match (one household-to-many respondents). Household-level input files contain 5,222 records; respondent-level input files contain 7,027 records. A respondent-level output file with 7,027 respondent records results.

Since you are creating a respondent-level output file, be sure to keep PN, as well as HHID and DSUBHH, in the output file even though PN is not used for the merging.

1C1. SAS Example of Household to Respondent Merge

```
proc sort
      data=in.a95a_r
      out=a95a_r(keep=hhid dsubhh pn d636 d638);
      by hhid dsubhh;
run;
proc sort
      data=in.a95j_h
      out=a95j_h(keep=hhid dsubhh d4753 d4752);
      by hhid dsubhh;
run;
data resp;
      merge a95a_r a95j_h;
      by hhid dsubhh;
run;
```

1C2. Stata Example of Household to Respondent Merge

```
use HHID DSUBHH D4753 D4752 using a95j_h
sort HHID DSUBHH
save tmp1, replace

use HHID PN DSUBHH D636 D638 using a95a_r
sort HHID DSUBHH
merge HHID DSUBHH using tmp1
tab _m
drop _m
save resp, replace
```

1D. Household-level Files with Respondent-level Data

To create a household-level file including variables from both respondents requires several steps.

First, separate the respondent-level records into two groups, one for the family respondent, the first person, and one for the non-family respondent, the second person.

We suggest you use DFAMR – 1995 WHETHER FAMILY RESPONDENT to create the two groups of respondents. The first group will contain 5,205 records (17 households did not have a family respondent); the second group will contain 1,822 records.

If you choose to use another variable to create two respondent groups, be sure each respondent group does not have more than one person from any one household (otherwise data will be lost).

Second, rename all variables for the second respondent group (except HHID and DSUBHH) to avoid overlap in the combined file.

Third, merge the two respondent files and the household-level file using HHID
and DSUBHH.

Although the respondent groups contain fewer than 5,222 records, this is a one-
to-one match. A household-level file with 5,222 household records results. The
household-level output file includes household variables, a set of variables for
the first respondent and a set of variables for the second respondent.

We recommend that you keep PN for the first and second respondent even though PN
is not used for the merging and the output file is a household-level file.

1D1.   SAS Example of Respondent to Household Merge

```
*      file of family respondents;
proc sort
      data=in.a95a_r(keep=hhid dsubhh dfamr pn d636 d638
      rename=(dfamr=dfamra pn=pna d636=d636a d638=d638a)
      where=(dfamra in(1 2 3 4)))
      out=a95a1;
      by hhid dsubhh;
run;
*      file of non-family respondents;
proc sort
      data=in.a95a_r(keep=hhid dsubhh dfamr pn d636 d638
      rename=(dfamr=dfamrb pn=pnb d636=d636b d638=d638b)
      where=(dfamrb not in(1 2 3 4)))
      out=a95a2;
      by hhid dsubhh;
run;
*      variables from a household-level file;
proc sort
      data=in.a95j_h(keep=hhid dsubhh d4753 d4752)
      out=a95j;
      by hhid dsubhh;
run;
*      household-level file with information from two respondents;
data hhold;
      merge
      a95a1 a95a2 a95j;
      by hhid dsubhh;
run;
```

1D2.   Stata Example of Respondent to Household Merge

```
use HHID DSUBHH D4753 D4752 using a95j_h
sort HHID DSUBHH
save tmp1, replace


use HHID DSUBHH DFAMR PN D636 D638 using a95a_r
keep if DFAMR==1|DFAMR==2|DFAMR==3|DFAMR==4
rename DFAMR DFAMRa
rename PN PNa
rename D636 D636a
rename D638 D638a
sort HHID DSUBHH
save tmp2, replace
```

```
use HHID DSUBHH DFAMR PN D636 D638 using a95a_r
keep if DFAMR~=1 & DFAMR~=2 & DFAMR~=3 & DFAMR~=4
rename DFAMR DFAMRb
rename PN PNb
rename D636 D636b
rename D638 D638b
sort HHID DSUBHH
merge HHID DSUBHH using tmp2
tab _m
drop _m
sort HHID DSUBHH
merge HHID DSUBHH using tmp1
tab _m
drop _m
sort HHID DSUBHH
save hhold, replace
```

2. Examples of Sub-Household and Respondent Person Number and Other Person Number Assignments

In the first year of data collection, all households, consisting of either a single respondent or of two married or partnered respondents, were assigned a SUBHH of 0.

In subsequent waves, a SUBHH of 0 indicates that the original household has not split due to divorce or separation of spouses or partners, although one member of a couple may have died or a single respondent may have become married or partnered.

A value of 1 or 2 indicates a household in which the original couple split, divorced or separated. One of the original couple is assigned a SUBHH of 1; the other is assigned a SUBHH of 2.

A value of 5 or 6 indicates a previously split household split a second time. One of the couple from a SUBHH 1 or 2 retains a SUBHH of 1 or 2; the other is assigned a SUBHH 5 or 6.

A value of 7 indicates respondents from split household reunited[1].

It is important to understand these assignments when you merge records from different waves of the study. In the examples listed below the made-up HHIDs have only five digits; in the real data HHIDs have six digits.

2A.   Married Couple Stays Married.

Two respondents in a sample household are married at the time of the first cross-section. Each respondent is assigned a HHID of 12345 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

At the time of the second cross-section the two respondents are still married, and each retains their HHID of 12345 and their SUBHH of 0 and his and her PN of 010 and 020, respectively.

2B.   Couple Divorces.

Two respondents in a sample household are married at the time of the first cross-section. Each respondent is assigned a HHID of 23456 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

By the time of the second cross-section, the couple has divorced. Both respondents retain the HHID of 23456, but one is assigned a SUBHH of 1 and the other is assigned a SUBHH of 2. Each original respondent retains his and her PN of 010 and 020, respectively.

2C.   One or Both Respondents Die.

Two respondents in a sample household are married at the time of the first cross-section. Each respondent is assigned a HHID of 34567 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

---

[1] In addition, a SUBHH of 3 or 4 indicates the "household" a deceased respondent who is considered to be in a household of his or her own. These values do not occur in these files because all records in these files are from living respondents.

One respondent dies before the next wave. At the next wave, both respondents retain their HHID of 34567. The living respondent retains her SUBHH of 0; the deceased respondent is assigned a SUBHH of 3. If both respondents die, one is assigned a SUBHH of 3 and the other is assigned a SUBHH of 4. Each original respondent retains his and her PN of 010 and 020, respectively.

2D.   Single Respondent Marries.

A respondent who has never been married is in the first cross-section. The respondent is assigned a HHID of 45678 and a SUBHH of 0 and a PN of 010.

At the time of the second cross-section, the respondent has married. Both the respondent and her new spouse are assigned a HHID of 45678 and a SUBHH of 0 because the household was not divided. The original respondent retains her PN of 010. Her new spouse is assigned PN of 011.

2E.   Couple Divorces, One Respondent Remarries and Divorces.

Two respondents in a sample household are married at the time of the first cross-section. Each respondent is assigned a HHID of 46789 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

By the time of the second cross-section, the couple has divorced and he has remarried. Both original respondents retain the HHID of 46789, but she is assigned a SUBHH of 1 and he is assigned a SUBHH of 2. His new spouse is also assigned the HHID of 56789 and the SUBHH of 2. Each original respondent retains his and her PN of 010 and 020, respectively. His new spouse is assigned PN of 011.

By the time of the third cross-section, that new couple has gotten divorced. All respondents retain the HHID of 56789. The original sample member ex-wife has the SUBHH of 1. The original sample member ex-husband has a SUBHH of 2, and the non-original sample member, his second ex-wife, is assigned the SUBHH of 5. Each original respondent retains his and her PN of 010 and 020, respectively. His second ex-wife retains her PN of 011.

2F.   Couple Divorces and Marries Again.

Two respondents in a sample household are married at the time of the first cross-section. Each respondent is assigned a HHID of 67890 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

By the time of the second cross-section, the couple has divorced. Both respondents retain the HHID of 67890, but one is assigned a SUBHH of 1 and the other is assigned a SUBHH of 2. Each original respondent retains his and her PN of 010 and 020, respectively.

By the time of the third cross-section, the respondents have remarried each other. Both are assigned the HHID of 67890 and the SUBHH of 7. Each original respondent retains his and her PN of 010 and 020, respectively.

2G.   Married Couple with Children and Siblings.

At the time of the first cross-section, sample household with a HHID of 89012 contains two respondents assigned PNs of 010 and 020, respectively. Associated with the household are three children with OPNs of 101, 102, and 201, and two siblings with OPNs of 051 and 052. All seven persons will keep those same PNs and OPNs across time. A friend who lives with the respondents is assigned an

OPN of 080. The friend will keep her OPN of 080 across time only if she is a household member at each wave.

2H.   Couple with Children and Siblings Divorces.

At the time of the first cross-section, sample household with a HHID of 89012 contains two respondents assigned PNs of 010 and 020, respectively. Associated with the household are two children with OPNs of 101, and 102, and three siblings, her two brothers with OPNs of 051 and 052 and his sister with an OPN of 061. A friend who lives with the respondents is assigned an OPN of 080.

By the time of the second cross-section, the couple has divorced. Both respondents retain the HHID of 89012, but he is assigned a SUBHH of 1 and the she is assigned a SUBHH of 2. Each original respondent retains his and her PN of 010 and 020, respectively. The two children appear in both their father's SUBHH 2 and their mother's SUBHH 1 with their respective OPNs, 101, and 102. The three siblings appear in the SUBHH of their respective siblings and maintain their respective OPNs. The ex-wife's brothers appear as part of her SUBHH 1 with their OPNs of 051 and 052, respectively. The ex-husband's sister appears as part of his SUBHH 2 with her OPN of 061. The friend has moved out and does not appear in either household.

2I.   Couple Divorces, One Respondent Remarries, Both Split-off Households Have New Members.

Two respondents in a sample household are married at the time of the first cross-section. Each respondent is assigned a HHID of 90123 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

By the time of the second cross-section, the couple has divorced. She has moved in with her mother. He has married a woman with two children. At the second cross-section, both original respondents retain the HHID of 90123, but she is assigned a SUBHH of 1 and he is assigned a SUBHH of 2. Each original respondent retains his and her PN of 010 and 020, respectively. Her mother is assigned the HHID of 90123 and the SUBHH of 1 and an OPN of 151. His new spouse and new stepchildren are assigned the HHID of 90123 and the SUBHH of 2. His new spouse is assigned an OPN of 011. His new stepchildren are assigned OPNs of 151 and 152.