

Data Description: Assets and Health Dynamics Among the Oldest Old (AHEAD)

I. Acknowledgments

The National Institute on Aging (NIA) provided funding for the first wave of data collection on the study of Asset and Health Dynamics among the Oldest Old (AHEAD) in the form of a supplement to a Cooperative Agreement with The University of Michigan for the Health and Retirement Study (HRS).

The Principal Investigator for the HRS and AHEAD Cooperative Agreement was Thomas Juster. Co-principal investigators on the first wave of the AHEAD project were (in alphabetical order):

John Henretta (Department of Sociology, University of Florida)
Regula Herzog (University of Michigan)
Martha Hill (University of Michigan)
Michael Hurd (Department of Economics, State University of New York-Stony Brook)
Willard Rodgers (University of Michigan)
Beth Soldo (Department of Demography, Georgetown University)
Robert Wallace (Dept. of Preventive Medicine & Environmental Health, University of Iowa)
Douglas Wolf (Syracuse University)

The design and content of AHEAD were largely based on the substantial work that went into the planning of the Health and Retirement Study (HRS) during 1990 and 1991. That work was done by six Expert Working Groups (Labor Force Participation and Pensions; Health Conditions and Health Status; Family Structure, Family Support and Mobility; Economic Status; External Record Linkages; and Survey Operations), with a total of 42 individuals from many different universities and academic disciplines.

The first wave of the AHEAD project had a Steering Committee consisting of the following members:

Richard Campbell (University of Illinois)
Larry Corder (Center for Demographic Studies, Duke University)
Walter Ettinger (Bowman Gray School of Medicine)
Al Hermalin (Population Studies Center, University of Michigan)
Powell Lawton (Philadelphia Geriatric Center)
Korbin Liu (Urban Institute)
Jennifer Madans (National Center for Health Statistics)
Tim Smeeding (Maxwell School of Citizenship and Public Affairs, Syracuse University)
Richard Suzman (NIA, ex officio)
Robert Willis (NORC, University of Chicago)

In addition, NIA has its own Data Monitoring and Design Committee for the HRS and AHEAD projects, which forms a second oversight group along with the AHEAD Steering Committee. The NIA committee had the following members representing both academic disciplines and relevant government agencies:

Jack Feldman (National Center for Health Statistics)
Linda Fried (Johns Hopkins Hospital)
Richard Hinz (Pension and Welfare Benefits Administration)
Alex Inkeles (Stanford University)
Lee Lillard (RAND)
Marshal McBean (Health Care Financing Administration)
Nancy McConnell (Population Resource Center)
George Myers (Duke University)
Janice Olson (Social Security Administration)
Dorothy Rice (University of California, San Francisco)
Dallas Salisbury (Employee Benefits Research Institute)
Steven Sandell (Assistant Secretary for Planning and Evaluation, DHHS)
James P. Smith (RAND)
Richard Suzman (NIA)
David Wise (National Bureau of Economic Research)
Barbara Atrostic (Congressional Budget Office)
Robert Clark (Assistant Secretary for Planning and Evaluation, DHHS)
Gillian Hunter (Treasury Department)
Richard Seefer (AARP)
Patricia Willis (Pension and Welfare Benefits Association)

In preparation for the fielding of this first national Computer Assisted Personal Interviewing (CAPI) study conducted by the Survey Research Center, the AHEAD research staff worked closely with SRC Survey Operations Administration, Computing Section, Sampling Section, and National Field Management and Interviewing staff. We wish to acknowledge the contributions of Rhonda Ash, Marcy Breslow, Judy Connors, Marshall Cummings, Steve Heeringa, Barbara Homburg, Kathy LaDronka, Jody Lamkin, Gary Munce, and Glenna Redmond.

Special thanks are extended to study staff members Lynn Dielman and Kathy Terrazas who worked from questionnaire design through data management to help bring this second public release AHEAD data set to completion.

II. Survey Overview

The focus of the AHEAD survey is to understand the impacts and interrelationships of changes and transitions for older Americans in three major domains: health, financial, and family. The questions included in the interview were designed to reflect as much as possible the analytic and policy interests of those from a variety of disciplines who are working in the area of aging. The constraint was the need to keep the interview burden reasonable for people in the oldest-old age range; concretely, our goal was to limit the average interview length to about 60 minutes (actual average length was 61.2 minutes). We believe that we have been successful in representing the main strands of thinking about how to model the aging process and in representing the most important policy issues.

III. Sample Design

A sample of community based individuals aged 70 and older (i.e., born in 1923 or earlier) was identified through the HRS screening of an area probability sample of households. This procedure identified a total of 9,473 households and 11,965 individuals in the target age range. Because of budget constraints, the number of primary sampling units from which the AHEAD sample was drawn was cut from the 93 that were selected for the HRS screening to 66. African Americans, Mexican-Hispanics, and residents of the state of Florida were sampled at about 1.8 times the probability of the general population.

AHEAD used a dual sampling frame for those aged 80 and older (the birth cohorts through 1913). Those in this age range in half of the sampling segments from HRS were dropped and replaced by an approximately equal number of selections from the Master Enrollment File maintained by the Health Care Financing Administration (HCFA) for Medicare enrollees. HCFA provided a tape with all enrollees aged 69 and older in the selected counties. From that list of several million names, the SRC Sampling Section selected samples that parallel the samples dropped from the HRS frame. As with the HRS frame, the sample selected from the HCFA frame was limited to those living in households at the time of the initial interview, thereby excluding those who were living in long-term care facilities or other institutions at baseline.

If more than one age-eligible individual was listed as living in a household, one person was randomly selected. In addition, if that selected individual was married, an interview was sought with the spouse regardless of his or her age. If the sampled individual married or started living with a partner by the time of the request for an interview, an interview was sought with the new spouse or partner. Similarly, if an individual selected from the HCFA Medicare enrollment file was married, an interview was requested with the spouse or partner regardless of age. If the spouse was also cohort-eligible, the spouse was part of the sample in his or her own right; but if the spouse was born in 1924 or later, the interview was conducted to provide additional information about the household of the selected (cohort-eligible) individual.

The long range study design calls for re-interviews with the surviving members of the sample every two years; for those who are deceased, enter nursing homes, or are unable to provide useful information, interviews are to be done by proxy.

IV. Weights for Wave I Data Analysis

The complex sampling design of the AHEAD Study--which includes oversamples of Mexican-American Hispanics, African-Americans, and households in the State of Florida--requires compensatory weighting in descriptive analyses of the survey data. Beyond simple compensation for unequal selection probabilities, weighting factors are also used to adjust for geographic and race group differences in response rates, as well as for the subsampling of households in a small number of locked buildings or dangerous areas. Post-stratification adjustments were made at both the household and person level in order to match sample demographic distributions with known 1990 Census totals. Please see (ahdsample.doc) for a detailed description of the development of the weights.

A. Household Analysis Weight: WTHHPOP or WTHHNORM

The household analysis weight, which is recommended for descriptive analyses of the 6047 households, is a composite weight which has been formed as the product of five component factors: (1) the housing unit selection weight, (2) an adjustment factor for non-listed segments, (3) an adjustment factor for sub-sampled areas, (4) a household non-response adjustment factor, and (5) a household post-stratification factor. This household weight should be used for descriptive analysis of household-level data from the AHEAD Study households. WTHHPOP is the population weight for households; when the centered WTHHPOP (WTHHNORM) is used, the weighted n is very close to the actual n and is appropriate for analyses that use the probabilities ("p values") of statistics.

B. Respondent Analysis Weight: WTRPOP or WTRNORM

The person-level analysis weight is the product of the Household Analysis Weight and the person-level post-stratification weight. Only age-eligible respondents have valid person-level weights. Age-ineligible respondents have a value of zero for the person weight. Age-eligible respondents incorporate the household weight as one of the multiplicative factors of the final person-level analysis weight.

V. Field Notes and Procedures

The AHEAD interviews were conducted by field interviewers who used CAPI/CATI: Computer-Assisted Personal/Telephone Interviewing. Most of the interviews with individuals aged 80 and older were done face-to-face in the respondent's home. Most interviews with those under age 80 were done by telephone, although interviewers were able to arrange face-to-face interviews if the respondent had difficulty doing a phone interview or preferred a face-to-face interview.

Approximately 130 interviewers worked on the data collection. Each interviewer attended one of three training sessions held during October, 1993, at hotels in the Detroit/Ann Arbor area. Each training session lasted for seven days (five days for interviewers who had already had general interviewer training and experience on other SRC studies.) The training involved instruction and experience in the use of the computers as well as training on study objectives, question content, and how to deal with respondent questions and difficulties.

The survey questions and the question-by-question instructions for interview flow were programmed, using the Surveycraft CAI system that SRC utilizes for CAPI and CATI interviewing. In addition to the English language version of the questions, a Spanish translation was incorporated into the computer program. A modification of the interview was prepared for use with proxy informants for those cases where the selected individual was unable to participate.

For married respondents, interviewers were instructed to divide the reporting task between the two spouses. The interviewers asked which spouse would be the most knowledgeable about the household financial situation (income sources, assets, medical expenditures, and insurance), and that person was designated as the "Financial Respondent;" the other spouse was then designated as the "Non-financial Respondent." In addition, the first respondent interviewed in a household was asked questions about other household members and about all children living elsewhere. To keep the interviews approximately equal in length, the interviewer suggested that the Non-financial Respondent be interviewed first. In practice, this preferred sequence was not always followed, especially when one spouse was in poor health and the other spouse wanted to ease his or her reporting burden.

Data collection began in October, 1993 and continued through July, 1994. The number of individuals in the HRS-based sample was 9854, of whom 1268 were identified as ineligible (i.e., institutionalized or deceased) with 6954 interviewed. A total of 2058 selections were made based on the HCFA frame (including spouses of the original selections), and these were released to the interviewers in February, 1994. Of these selections, 416 were identified as ineligible and 1268 were interviewed. The total number of interviews at the close of the data collection period was 8,222 for a response rate slightly over 80% of the eligible persons.

Table 1 shows the number of interviews with the various types of respondents.

Table 1: Type of Respondent	Number
ONLY Respondent (If sampled person was not married/partnered and living with spouse/partner)	3762
MARRIED OR PARTNER:	
Lead, Non-Financial Respondent	1115
Second, Non-Financial Respondent	1088
Lead, Financial Respondent	1155
Second, Financial Respondent	1102
TOTAL	8222

VI. Interview Content

The interview consists of 11 sections, as described below (sections are asked of all respondents unless noted.)

Section A. Demographics; Year of birth, education, education of parents, marital status and history, veteran status.

Section B. Health conditions: Whether R has ever seen doctor for each of 12 conditions; assessment of vision and hearing; pain; smoking; drinking; weight; height; depression.

Section C. Cognition: Self-assessment of memory; immediate and delayed recall of ten words, plus other questions from the TICS ("Telephone Interview for Cognitive Status"); for proxy respondents, assessments of level and change in cognitive abilities.

Section D. Family structure: (asked of Only and Lead Respondents): List of other household members, with details of their age, education, employment status, earnings, whether moved in with Respondent and if so why; list of children and children-in-law living elsewhere with details of their age, relationship to Respondent and spouse, marital status, number of children, education, employment status, home ownership, distance from Respondent, financial situation relative to Respondents; financial help given to children and others.

(Asked of all Respondents) Number and marital status of siblings; if parents not living then age when died, if parent(s) living then their current age and whether Respondent has provided help with ADLs.

Section E.

E1. Health care utilization and costs: (For all Respondents) Previous twelve months: hospitalizations, nursing home stays, doctor visits, outpatient surgery, dental care, prescription drugs; bed days; whether covered by Medicare A/B and Medicare number.

(For Only and Financial Respondents) Any out-of-pocket costs for each type of health care listed, amount of out-of-pocket expenditure for nursing home stays and other medical expenses for self and spouse; whether a child or anyone else has helped with health care costs.

E2. For all Respondents: For six ADLs, whether R gets help; uses equipment; and degree of difficulty. Degree of difficulty with several other activities. For each of five IADLs, whether R able to do without help, and difficulty.

E3. For each helper (accumulated across ADLs and IADLs): gender, frequency, hours, whether paid, out-of-pocket costs, whether anyone helps pay those costs and if so, who.

Section F. Housing (asked of Only and Financial Respondents): Type of housing, whether part of a condominium or housing project, whether income limit, whether age limit, whether entry fee or association payments; services offered to residents; number of stories; special features for physically impaired; ownership, mortgage, others on deed; home value (if owner) or rent; amount paid for property taxes, insurance, utilities.

Section G. Job status: Current employment status, whether worked in last two years, whether ever worked for at least 10 years; occupation, earnings and hours last calendar year; most ever earned per year, and at what age that occurred. If widowed or divorced: similar job history questions for former spouse.

Section H. Expectations: Chances (on 0 to 100 percent scale) of giving major financial assistance to family members in next ten years; of receiving such help; of leaving an inheritance and amount; of entering a nursing home in next five years; of medical expenses depleting savings in next five years; of income keeping up with inflation; of living to a specified age; of moving in next five years and if so, type of move and which child may move near.

Section J. Income (asked of Only and Financial Respondents): Income from each of several sources (Social Security, SSI, food stamps, pensions, veterans benefits, annuities, interest income) for self (and spouse), follow-up questions specific to the various types of sources. Financial assistance from children or from others in last year. Total income of Respondent (and spouse) last calendar year. Whether have a will, and provisions made for children.

Section K. Net worth (asked of Only and Financial Respondents): Current value of various assets (if any): Real estate (other than home); automobiles or other means of transportation; family business; IRA or Keogh accounts; shares of stocks or mutual funds; checking, savings, or money market accounts; CDs, government savings bonds; bonds or bond funds. Whether assets were used to pay expenses or additions made to savings or investments last year. Whether any assets are in trusts, and if so, beneficiaries, value of those assets, and whether those assets have already been listed. Other assets and liabilities and lump sum payments in past year (insurance, pension, or inheritance).

Section R. Insurance: Current coverage by Medicaid, other government insurance programs, or other health insurance. Any coverage for long term care, and if so, whether have received payments, covers home care, payments increase with inflation. (Asked of Only or Financial Respondents) Life insurance, whole and term: amount, beneficiaries, for Respondent (and spouse).

Modules. In addition to these "core" questions asked of the entire sample, there were additional topics that were important for cross-walking, or "experimental" in the sense of not having well-developed measures and clear relevance to aging processes or policy issues. Eight such "modules" of questions were developed, and asked of randomly assigned sub-samples.

1. Resiliency: This module was administered as a paper and pencil addition to CAPI Module 4 beginning January 20, 1994. This series of questions is about recent major life events and how much impact those events had on the Respondent. The questions were developed by Robert Kahn and colleagues of the MacArthur Program on Successful Aging.

2. Time use: A set of questions on unpaid but economically productive activities (home maintenance, volunteer work, and informal help to others). These questions permit a more balanced assessment of the utilization and provision of human resources by the Respondents than would otherwise be possible.

3. ADLs: A set of ADL questions that was planned for the second Longitudinal Study of Aging, by NCHS, so that comparisons can be made with the answers given to the ADL questions asked in the core of AHEAD.

4. ADLs: A set of ADL questions from the screener for the National Long Term Care Study.

5. Similarities: A measure of abstract reasoning taken from the WAIS. This module also has two ADL questions that were on the 1990 Census long form.

6. Quality of life: Quality of life has recently obtained much attention in the medical community as a means of assessing the broad impact of medical treatments and procedures beyond their effect on specific physiologic functions. A focus on the essential quality of life issue -- whether life is still worth living -- underlies the questions in this module, which were adapted from unpublished work by Powell Lawton and from the purpose-in-life subscale of Ryff's Subjective Well-Being Scale. In addition, there are a few items of mastery and personal control taken from work by Pearlin and Schooler.

7. In-depth ADLs: Research on cognitive, psychomotor, and psychological functioning has documented an enormous potential for adaptation to and compensation for declining functioning by the elderly. A number of specific mechanisms seem to be involved in such compensation, including a change of specific procedures when performing the activity, increased time allotted for completing the activity, lowered standards, and changes in the immediate environment to ease performance. In order to explore whether such adaptive mechanisms may account for a lack of reported difficulty with bathing and with managing money, questions in this module probe various adaptive strategies that may be involved in carrying out these activities.

9. Financial Pressure: Questions in this module ask whether any of several things have happened in the last 12 months because the Respondent was short of money. These

include, for example, not paying bills or rent on time, eating less expensive foods, not purchasing prescribed medications, postponing seeing a doctor, skipping a vacation, or skipping needed home repairs. Other questions ask about the perceived fairness of several alternatives proposed with respect to provision of long-term care.

VII. Orientation to the Data

A. Terminology

Before beginning analysis, it is helpful to understand how the data were collected and to grasp some of the keywords or terms. In brief, before conducting an interview in a household the interviewer first determined if the household contained only one eligible Respondent or an eligible Respondent and a spouse/partner. Next, regardless of whether a Self or Proxy interview was required, the interviewer selected Type of Respondent:

ONLY R	Only one eligible Respondent residing in household
LEAD R	The first person of a couple to complete an interview. The Lead (or Family) Respondent answers questions about household members and non-resident children.
2ND R	The 2nd person interviewed in the household. The 2 nd Respondent skips all of the detailed questions about the household members and non-resident children.

For couples, the interviewer next decided upon the Type of Interview:

FINANCIAL	Answers questions about the finances of the couple, including income, assets, housing, medical expenses and insurance.
NON-FINANCIAL	Skips detailed questions about family finances.

The variable 'TYPE' in the Respondent file can be used to identify which one of the four possible IW/R types applies to each respondent in the data:

1. Lead R/Non-Financial R
2. 2nd R/Non-Financial R
3. Lead R/Financial R, Only R
4. 2nd R/Financial R

B. Data Structure

The data are contained in four separate files which can be merged by using various identification variables. The four data files contain the following types of records:

1. *Household records* (BHH21--6047 cases, 678 variables): This file consists of all information pertaining to the household (obtained from only one respondent in two-respondent households.) More specifically, it includes questions asked of the Lead respondent about other members of the household and Non-resident children (most of Section D). Questions asked only of the Financial respondent were about costs of various types of health care incurred in the last 12 months (a subset of questions from Section E); about housing (section F); about income from various sources (most of section J); about assets (section K); and about life insurance (the latter part of Section R).
2. *Individual respondent records* (file BR21--8222 cases, 638 variables): This file consists of most of the remaining information from the interviews: demographic questions (Section A), individual health conditions (Section B), cognitive status (Section C); health care; ADLs and IADLs (most of Section E); current and past employment (Section G); expectations about the future (Section H); a few questions about wills (Section J); and the set of experimental modules asked of random subsets of the respondents.
3. *Other person records* -- household member/non-resident child records (file BOP21--17424 cases, 137 variables): This file consists of all information pertaining to non-resident children and each household member or couple other than the respondent and spouse. More specifically, it includes information from questions D20b - D35 for each non-resident child and his or her family, and information from questions D4 - D18d for each household member or couple other than the respondent and spouse. In addition, 89 variables have been recoded or merged from Household, Respondent and Helper files.
4. *Helper records* (file BHP21--3160 cases, 20 variables): This file consists of records for each person or organization listed as providing help to a Respondent with an ADL or IADL (information from questions E59 - E69).

Table 2 illustrates who gets asked which questions and which dataset contains the information.

C. Special Situations

1. HHIDs for households with no Financial Respondent. There are 34 households in which there is no Financial Respondent. The HHID numbers for these households are:

200174, 200237, 200527, 200551, 200817, 201171, 201363, 201625, 201644,
201691, 201737, 201863, 202116, 202183, 202934, 203061, 203623, 204282,
204569, 204645, 204731, 205755, 206030, 206723, 207019, 207174, 207254,
207958, 208202, 208250, 208279, 208287, 208395, 208580

2. HHIDs for households with no Family Respondent. There are 15 households in which there is no Family Respondent. The HHID numbers for these households are:

200072, 200159, 202283, 202709, 203096, 203658, 203752, 204067, 204318,
204883, 206758, 207339, 207958, 208560, 208869

Table 2: Content Overview			
SEC	CONTENT:	COMPLETED BY:	AHEAD Class FILES:
A	Demographics	All Rs	BR2 (Resp)
B	Health Status	All Rs	BR2 (Resp)
C	Cognition	All Rs	BR2 (Resp)
D	1. Family Structure and Transfers	FAMILY R/ONLY R	1. BHH2 (HH)
	2. Parent/Sibling info	All Rs	2. BR2 (Resp)
	3. HH member/Child Demographics	FAMILY R/ONLY R	3. BOP2 (Person)
E	1. Health care costs	1. FINANCIAL R/ONLY R	1. BHH2 (HH)
	2. Health care Services/Medicare information	2. All Rs	2. BR2 (Resp)
	3. ADLs and IADLs	3. All Rs	3. BR2 (Resp)
	4. ADL/IADL Helper info	4. All Rs	4. BHP2 (Helper)
F	Housing	FINANCIAL R/ONLY R	BHH2 (HH)
G	Job Status/Work History	All Rs	BR2 (Resp)
H	Expectations	All Rs	BR2 (Resp)
J	1. Income	1. FINANCIAL R/ONLY R	1. BHH2 (HH)
	2. Wills	2. All Rs	2. BR2 (Resp)
K	Net Worth	FINANCIAL R/ONLY R	BHH2 (HH)
R	1. Health Insurance	1. All Rs	1. BR2 (Resp)
	2. Life Insurance	2. FINANCIAL/ONLY R	2. BHH2 (HH)
MOD	Experimental Modules	All Rs	BR2 (Resp)

VIII. Derived Variables

Each derived variable is referenced at the origination point in the codebook. Two types of derived variables have been produced at this point (June 1994). All components of these derived variables have been left in the dataset and codebook so that users may create their own recodes or imputations if they wish. (Components of the "C variables" have been removed.)

1. *Summary variables* based on recodes of existing variables. (Examples: Body Mass Index based on an algorithm of height and weight; degree of difficulty for ADLs; Cognition; amount of income from a variety of sources in the last calendar year.)

2. *Imputed dollar amounts* mostly based on unfolding DK/RF categories. (Examples: Family income (V1648X), household income (HHINC), ASSET (and components), DEBT, NETWORTH (and components); and flags for each imputed amount.

Note that if a specific dollar amount has been imputed, the variable names follow this convention:

V1648 Original dataset variable (family income)
V1648C Ranges for DK/RF category follow-up
V1648X All values of the amount including imputation (Use the "X" variable version for analysis if you want to avail yourself of our preliminary imputations.)
V1648F Flag indicating the degree of information available for imputation of that variable in a case, or indicating no imputation necessary (0).

Also note that components of income received in the last calendar year have usually NOT been imputed as yet (although the single family income question in J52. has been imputed). Further, derived variables ending in "Y" do contain recodes of the amount of each source received in the last calendar year which still contain unimputed DK or Refused codes.

IX. Derived Financial Variables: Money Amounts (Sections F, G, J, K)

Imputations have been completed for many of the money amount variables in the survey as indicated below and referred to by banners in the core codebook.

A. Assets

All asset questions have been fully imputed, including the lead-in "holding variable" of whether the R has the asset. Sum variables of ASSET, DEBT, and NETWORTH (ASSET-DEBT) are thus complete with imputation done on the component parts.

B. Income

Most individual income questions have been recoded to amount in the last calendar year, and have been imputed if the answer was DK, RF. Income components have NOT been added together for a total household income. However, the direct question about income of R (and spouse/partner) in the last calendar year [J52, V1648], has been imputed and thus V1648X may be used for family income, and HHINC [V1648X + V1681X] used for household income.

C. Money Amount

All money amount answers which had follow-up unfolding questions for DK and refused were imputed by first using a "hot deck" to randomly assign final DK or refusals to one of the follow-up categories. The pool of possible donors for final DKs and NAs consisted of those who said DK to the open ended question, but answered the unfolding questions; while the pool of possible donors for final RFs consisted of those who refused to answer the open-ended question, but answered the unfolding questions. All respondents who did not answer the open-ended question were now in a range defined by the unfolding questions. These respondents were assigned a specific value by imputation, using as potential donors all who gave an answer to the open-ended question that was in the appropriate range.

Money amount variables which did not have follow-up unfolding questions have also been imputed. For these variables, the hot deck method was used for all imputations. This method uses a "donor" case with real (non-missing) data for a variable to impute data for a case with missing data for the variable. Prior to imputation, the data set was sorted by characteristics of the respondents which were expected to be related to the variable being imputed - such as marital status, age group, years of education, or household income category. A case with missing data will then receive imputed data from a donor who is similar on characteristics thought to be related to the variable being imputed.

Flags have been created for each imputed money amount variable. For variables with unfolding categories, these flags have the following codes:

0	no imputation
1-9	degree of information available for imputation from unfolding categories
."	INAP, question not asked of R.

For variable without unfolding categories the flag codes are:

0	no imputation
1	imputation
."	INAP if question not asked of R.

The convention for naming derived and imputed variable amounts is:

V... Actual answers. (Amount given .D, .R, INAP=.)
V...C Categories of unfolding if DK/RF (0=R gave amount)
V...F Flag, 0=not imputed, 1-9=imputed, or degree of information available
V...X Total answers: original answer or imputed answer
V...H Holding variable [whether have asset] (including imputed)
V...R Monthly amount derived for some variables, missing data not imputed

X. Merging Data Files

The AHEAD data are organized into four relational data sets which have been sorted by HHID (Household Identifier). The following "identification" variables permit the information from these different files to be match-merged:

Household identification number (HHID): This is a sequential number that identifies households at the baseline data collection. It appears on the records in each of the four files. Note that all of the files are sorted first by HHID.

Respondent identification number (HHIDPN) : This combines the person number (PN) with the household identifier (HHID) to uniquely identify the respondent across all future waves of the study as well as in the baseline data.

On the Respondent-level file this number identifies the respondent.

On the Helper-level file this number identifies the respondent who reported getting the help; if both spouses reported getting help from the same individual, there are two records for that helper, but with different respondent identification numbers.

Child/person identification number (PN): This is a three-digit number that identifies each child and each household member and their spouse other than the respondent and the respondent's spouse/partner within the household. This number appears in the Other Person file as variable PN (Person Number). Information on each non-resident child and their family is collected on one PN and information on each household member, and their spouse or partner, is collected on one PN. Values of PN from 120 to 390 refer to children who do not reside in the same household as the respondent(s), while values of PN from 410 to 600 refer to other members, including children or grandchildren of the respondent, of the same household as the respondent(s). (See APNDXA-B.TXT for details.)

Helper identification number (HN): This is a three-digit number which combines the person number of the helper (PN) with their relationship to the respondent (BRELATEP) to identify the specific individual who provides the help. If the helper is a child of the

respondent or a member of the respondent's household, the first two digits of HN are the same as the first two digits of PN in the record for that person in the Other Person file. If the helper is the spouse of that child (i.e., a child-in-law of the respondent), the first two digits of HN are also the same as in the PN number for the child, to permit assistance to be "credited" to the child's family, whether it was provided by the child him- or herself or by his/her spouse or even the child's child (R's grandchild). To identify the individual within that family, the last digit of HN has the value of 1 if the helper is the child him/herself (or a group), the value of 2 if the helper is a child-in-law; the value 3 or 4 if the helper is a grandchild. (A maximum of two grandchildren in one family are present as helpers in this dataset.) (See APNDXA-B.TXT for details.)

XI. Merge Examples

Some users may need information from only one of the four AHEAD datasets; most, however, will probably want to use data contained in two or more of the files. Merging variables from separate datasets should be planned carefully. Keep in mind the following:

1. Make sure each dataset is sorted by the match variable (our original datasets are already sorted by HHID)
2. When merging a data set that has more than one record with the same HHID (i.e., Respondent level data) with a data set that has only one record per HHID (i.e., Household level data), you must include a data statement in your Data Step which renames the variables of the duplicate (i.e., spouse) variables.

Here is an example of merging selected variables using a LAG function which employs the SAS variables FIRST.<any id variable> and LAST.<any id variable>:

```
libname ahead 'c:\xxx';
data tempr;
  set ahead.ahcr(keep=hhid id sex age v435 adlany iadlany);
by hhid;
/* Capture both records of a couple and rename spouse variables */
if not first.hhid and last.hhid then do;
  newid=lag(id);
  newsex=lag(sex);
  newage=lag(age);
  new435=lag(v435);
  nadlany=lag(adlany);
  niadlay=lag(iadlany);
  output;
end;
/* Capture single records and leave lag variables as INAP (.) */
if first.hhid and last.hhid then output;
run;

/* Sort the temporary dataset by HHID */
proc sort data=tempr out=temprs;
  by hhid;
```

```

run;

/* Do the merge */
data ahead.flat;
merge ahead.ahchh(keep=hhid v1648x networth)
      temprs;
  by hhid;
run;

```

This procedure creates a small temporary dataset from the Respondent file which keeps only selected variables, and then renames the spouse variables which are to be added with the R variables to the selected variables from the household level file. Finally it merges the temporary data set to the selected household data, creating a new dataset which has been sorted by HHID.

The next example adds household-level variables to the respondent-level record.

```

libname ahead 'c:\ahead\xxxx';
%let hhvars=V407 V435 V466;
%let rvars=V559 V562 V565 V576;
data RSPFILE1;
  merge AHEAD.AHDHH(keep=HHID &hhvars)
        AHEAD.AHDR(keep=HHID HHIDPN &rvars);
  by HHID;
run;

```

For an extended merge example see APPENDX.SAS in file APNDXA-B.TXT. This chunk of code adds variables to the child/other household member records that include whether the child provided help with each of the six ADLs, with IADLs, or money management, whether the child provides help with paying for medical expenses, whether the child is included in the will(s) of parents, and whether the child has received financial assistance of various types.

XII. Data Cleaning--Editing, Recoding, and Flagged Variables

Many records in this second public release have been cleaned to eliminate inconsistencies both within cases and across cases in households. Outliers and inconsistencies have been resolved and edited on the income and asset variables and extensive cleaning has been done on household and child variables. There are still some couples or individuals who were not asked the family information or not asked the financial information because, for example, the spouse refused to complete an interview, or the interviewer made a mistake in selecting type of interview and we were not able to send the case back for data retrieval.

Two types of data management were performed on household member and non-resident child variables in this data set. The first type was simple editing from interviewer notes, comments or other sources. Duplicate household member and non-resident child records were eliminated and full or partial records were added when information was missing because of data transmission problems or if child records had to be entered by hand from the Coversheet Roster (only 10 children could be entered into the CAPI application). Edited records are not flagged in HHMEMADD/NRCHDADD in the PERSONS file as imputed.

Another type of management involved the addition of padded records. If the Respondent indicated the number of household members or non-resident children but refused to answer questions about them, we added a partial record. Padded records contain PN, NRCHDADD/HHMEMADD, V417/V461. Sometimes records were added where no explicit interviewer comment or refusal response explained the inconsistency, but other related responses pointed to the existence of other household members or non-resident children. When records were imputed in this manner, the flag variables HHMEMADD/NRCHDADD have a value of '1'.

Other cleaning included reconciling relationship to Respondent and/or Spouse when children were coded as 7, "Other Relationship". Relationship was recoded to "Child or Step-Child" only when interviewer comments indicated that the child was adopted. When the interviewer comments indicated that the child was a foster child or a non-biological child raised as a child, or if no information was available, values of 7 were left as is. Please note that a code of 7 in these variables means "other relationship", not NO relationship,

XIII. Masking

For this second Public Release we have stripped the datasets of all names, Medicare, Medicaid, and Social Security numbers; we have removed or collapsed code values wherever these responses may violate confidentiality or allow the possibility of individual respondent identification. Variables which have been omitted are: Sample ID (PSU and Segment), month and day of Respondent birth, and state or country of birth. We have, in addition, collapsed occupation codes to broad categories.

XIV. Codebook Conventions

Except where noted, all variables are presented in the order that they were asked in the questionnaire.

(B)	(C)	(D)		
V325	[RESP]	PC2a. HOW MEMORY CHG		
(E)	PC2a.	Is (his/her) memory better or worse than two years ago?		

		20	BETTER.....	1
(F)		437	WORSE.....	2 PC2c (H)
		2	DK.....	.D PC3
(G)		1	RF.....	.R PC3
		7763	INAP, NOT PROXY [PROXY=1]; NO CHANGE IN MEMORY (DK/RF) [V324 NE 1]	(H)

B. Most variable names are of the form Vnnn in sequential order of their appearance in the questionnaire. Exceptions are some demographic variables such as SEX or MODE, and the derived variables.

C. The dataset in which the variable appears is indicated in brackets,

i.e., [HH] [RESP] [PERSONS] [HELPER].

- D. The variable label appears above each question.
- E. Question number and full question wording and interviewer instructions appear for each variable.
- F. Unweighted marginals (frequencies) for all cases (including age-ineligible) appear to the left of the code descriptions for most variables. For continuous variables, the unweighted n, mean, s.d., minimum and maximum are provided as well as frequencies of missing data.
- G. Missing data conventions in the dataset and codebook are as follows:

- 1. Don't know (DK) .D
- 2. Refused by R (RF) .R
- 3. Question Inappropriate (INAP) .

All three of these missing data codes are less than 0 and will not be used by SAS analyses unless deliberately requested. They can be addressed globally with: IF Vxxx lt 0
or individually: IF Vxxx = .D

Please note that CAPITAL .D and .R should be used in your setups.

- H. Skip patterns may be followed either from the "go-to indicators" (GO TO PC3), or from the INAP context description.

Category follow-up variables have been created for each series of unfolding bracketed range questions following missing data on an amount question. The original unfolding bracketed range variables have been removed.

- (A) If J46. = DK,RF go to J46a. Would the total be \$5,000 or more?
 - (B) If YES go to J46b. \$10,000 or more?
 - (C) If YES go to J46c. \$50,000 or more?
 - ALL go to J47-1. [END SEQUENCE]
 - (C) If NO,DK,RF go to J47-1. [END SEQUENCE]
- (B) If NO go to J46d. \$1,000 or more?
 - (C) ALL go to J47-1. [END SEQUENCE]
- (B) If DK,RF go to J47-1. [END SEQUENCE]

- A. First unfolding bracketed range question.
- B. Response to the first unfolding bracketed range question followed by the next unfolding bracketed range question.
- C. Response to the second unfolding bracketed range question followed by the next unfolding bracketed range question, if any.