PUBLIC USE VERSION

Technical Description of the Asset and Health Dynamics (AHEAD) Survey Sample Design

Steven G. Heeringa

Survey Design and Analysis Unit
Institute for Social Research
University of Michigan
Ann Arbor, MI

October 1995

Table of Contents

Asset and Health Dynamics (AHEAD) Among the Oldest Old: Public Use Sample Documentation

1. Introduction

The following technical memorandum describes the sample design, sampling procedures, and sample outcomes for Wave 1 of the Study of Aging and Health Dynamics (AHEAD). This document is divided into eight sections. The introduction describes the purpose and organization of the AHEAD study. Sections 2, 3, and 4 provide an overview and a detailed description of the AHEAD dual-frame probability sample design. The fifth and sixth sections describe the AHEAD Wave 1 sample control procedures and sample outcomes. Sections 7 and 8 contain descriptions of the construction and use of the analysis weights and the codes and procedures for computation of sampling errors for the AHEAD Wave 1 data.

AHEAD is funded by the National Institute on Aging (NIA) through a cooperative agreement. Although the initial AHEAD funding was for two and one-half years beginning in January 1993, the study is expected to continue for at least 10-12 years and possibly longer. The initial funding included a planning year and one more of data collection, October 1993 - July 1994. Dr. F. Thomas Juster at the Institute for Social Research (University of Michigan) is the Principal Investigator for this national program of research. In addition, many researchers and professionals from the ISR and other universities and government agencies have collaborated on the AHEAD study design and content.

As the proportion of the population living to older age increases, it is important for policy makers to understand the changing needs of that population in order to guide planning and policy decisions. AHEAD is intended to provide policy-makers with up-to-date information on changes in health and financial status of older-age households and to provide scientists with data to generate more accurate and realistic models of the health and financial status of the older-age population of the United States. AHEAD is a longitudinal study of the U.S. population cohorts born prior to January 1, 1924. From October 1993 to April 1994, AHEAD Wave 1 interviews were conducted with national samples of these age 70-plus individuals and their spouses about major transitions in their health and how financial, family and social resources are used when important health transitions occur. The longitudinal study plan specifies a full-scale reinterview of the AHEAD panel every second year beginning in 1995.

1.A Survey Population

The target population for the AHEAD survey consists of United States household residents who were born in 1923 or earlier. AHEAD uses a national probability sample of U.S. households with supplemental oversamples of Blacks, Hispanics and residents of the state of Florida. The majority of the target population is retired from the work force, but the sample also includes some individuals who are still currently working as well as those who have never worked outside the home.

The AHEAD observational unit is an eligible **household financial unit**. The AHEAD household financial unit must include at least one age-eligible member from the pre-1924 birth

year cohorts. The eligible AHEAD household financial unit might be: 1) a single unmarried age-eligible person; 2) a married couple in which both persons are age-eligible; or 3) a married couple in which only one spouse is age-eligible. Throughout this document, the convenient term "household" will be used interchangeably with the more precise "household financial unit" definition. For most AHEAD-eligible households, the terms are interchangeable. However, the reader should note that some households may contain multiple household financial units. If a sample housing unit (HU) contains more than one unrelated age-eligible person (i.e., financial unit), one of these persons is randomly selected to determine the financial unit to be observed. If an age-eligible person has a spouse, the spouse is automatically included in the financial unit even if he or she is not age-eligible.

Persons in institutions at the time of the Wave 1 survey are ineligible for AHEAD. Therefore, the following types of individuals who are otherwise age-eligible at the time AHEAD eligibility is established are excluded from the study population: people in nursing homes, long-term medical care or dependent care facilities, prisons, and jails. AHEAD eligibility was established at the time of first contact for the 1993-1994 Wave 1 data collection. If a selected person was found to be institutionalized at Wave 1, he or she was declared out-of-scope for AHEAD. At subsequent waves, Wave 1 respondents will be followed to institutions and interviewed. Persons who were temporarily hospitalized at the time of the Wave 1 AHEAD contact were eligible for interview upon recovery. Proxy interviews were sought for all eligible AHEAD respondents who were unable to complete the interview themselves.

A discussion of AHEAD sample design and interview methodology requires the definition of two subgroups of eligible households in the survey population. The two groups of AHEAD-eligible households are defined solely for purposes of the sample design and determination of the primary mode of interview -- phone for younger households, face-to-face for older households. The assignment of households to the two groups is based on the age of the oldest person in the household financial unit. If the single adult or either spouse in a married couple was born prior to 1914, the household financial unit is assigned to Group 2. If the single adult or both persons in a married couple were born after 1913 the household financial unit is assigned to Group 1. The full national sample of AHEAD-eligible households is divided approximately 60% to Group 1 and 40% to Group 2. Under the AHEAD sample design, Group 1 households are selected exclusively from the area probability (AP) frame component. Group 2 households are selected using a dual-frame design, roughly 50% of the Group 2 sample originating with the AP frame and the remaining 50% from a stratified sampling from a list frame of Medicare enrollees. (See Section 4.)

Beginning with Wave 1, AHEAD Group 1 households -- those age 70-79 in 1993 -- were interviewed by telephone except in cases where there was no telephone in the household or the respondent was unable to complete the interview by telephone. Their spouses were also interviewed by telephone. Most AHEAD respondents in Group 2 were interviewed face-to-face in their homes, although telephone interviews were permitted in cases where the respondent preferred the telephone mode. Face-to-face interviews were also the primary mode of Wave 1 data collection for the spouses of these respondents, irrespective of the spouse's age. The percent of persons in each major age group who were interviewed in person or by telephone is summarized in Table 1-1 below.

Table 1-1: Distribution of Wave 1 Responses by Respondent Age and Interview Mode

Respondent Type	In-Person Interview	Telephone Interview
Age 70 - 79	28%	72%
Age 80 +	70%	30%
Age-ineligible Spouse	28%	72%

1.B. Oversamples of Special Populations

In addition to the nationally-representative, dual-frame probability sample (the **core** sample), the AHEAD design includes three **oversamples**. The oversamples are introduced as supplements to the core sample component and are designed to increase the numbers of Black and Hispanic respondents as well as the number of AHEAD respondents who are residents of the state of Florida. Sampling weights are provided on all AHEAD data sets to compensate for the unequal probabilities of selection between the core and oversample domains (see Section 7).

2. The AHEAD Sample Design

2.A Overview of the Design

A national area probability sample of AHEAD-eligible households was identified in 1992-93 in conjunction with the extensive household screening required to obtain the Health and Retirement Survey (HRS) sample of the population cohorts born 1931 to 1941. The original plan was to use the HRS-screening sample as the sole basis for the AHEAD samples of both Group 1 and Group 2 households. This sample plan was subsequently changed. Group 1 AHEAD households (with eligible persons age 70 to 79) were selected by subsampling from the pool of eligible households identified in the course of the HRS screening, but a dual-frame sample design was used for the sampling of Group 2 households (with eligible persons age 80+). The proposed dual-frame design combined a subsample of Group 2 households screened by the HRS interviewers with an independent and equal-size sample of Group 2 households selected from the Health Care Finance Administration's (HCFA) Enrollment Data Base (EDB) file.

2.B AHEAD Dual-Frame Sample Design

Both the area probability (AP) and the HCFA EDB file components of the Group 2 dual-frame approach employed multi-stage probability sampling. The AP design component employs conventional multi-stage area probability sampling down to the selection of addresses from second stage unit (SSU) listings generated by SRC enumerators. Anticipating that the sample would also be used as the basis for a national study of the oldest old, the HRS screening forms completed during contact with the approximately 69,000 sample housing units selected for HRS were designed to identify households with members who would be eligible to participate in AHEAD.

The HCFA EDB file list sample was selected from Medicare enrollees whose listed addresses were linked to a primary stage unit of the AP sample design. EDB file addresses were linked to AP primary stage units (PSUs) using county and ZIP Code identifiers that were present on each enrollee's record. Within PSUs, geographic clusters (based on ZIP Code areas) of persons born in 1913 or earlier (i.e., age 80 or older in 1993) were then linked to the AP SSUs (area segments) of the AP sample component. A sample of ZIP areas was selected and individual enrollees in sampled ZIPs were subsampled with probabilities that yielded an equal overall probability of selection for each eligible Group 2 enrollee. The union of these two independent samples, the AP sample and the HCFA EDB file list sample selections, produced a dual frame probability sample of the Group 2 AHEAD population.

Table 2-1 provides a summary of the AHEAD sample outcomes for Group 1 and Group 2 households and individual respondents. As the table shows, the total sample of age 80+ respondent households was divided between the AP and the HCFA EDB file list frames. The difference in the expected 80+ interview counts for the two frames was due to the fact that the AP frame sample includes supplements of Blacks, Hispanics and Florida residents that were not replicated in the HCFA EDB file selection. A summary of final dispositions for the actual dual frame sample selection is given in Section 4.

Table 2-1: Summary of AHEAD Wave 1 Household and Respondent Samples

Sample Frame/ Sample Group	Eligible Households	Eligible Persons	Respondent Interviews	Unweighted Response Rate
Area Probability				
Age 70-79 (Group 1)	4,603	6,605	5,323	80.6%
Age 80+ (Group 2)	1,570	1,982	1,631	82.3%
HCFA EDB File List				
Age 80+ (Group 2)	1,336	1,643	1,268	77.2%
Total Sample	7,509	10,229	8,222	80.4%

2.C Objectives for the Group 2 Dual-frame Design

Survey methodologists share a concern that conventional area probability household sample designs do not provide optimal coverage of the oldest old (Rodgers, 1995). The HCFA EDB file is often proposed as an alternative list frame for high-coverage sampling of populations above the age of Medicare eligibility; however, there has been no systematic study of its coverage properties in relation to area probability sampling of household populations. Waldo and Lazerby (1984) discuss the population coverage of the HCFA medicare enrollment files. The dual-frame design of the Group 2 sample was proposed with two objectives in mind. First, it is intuitive that the combined samples from the two independent and overlapping frames provided a coverage of the survey population that is equal to or better than that offered by any one frame. The prospect of de facto improvement in population coverage would be a sufficient justification for the dual frame design, but it is also important to attempt to measure the relative coverages and properties of the two sample frame components. The second objective is therefore a methodological and statistical one -- through proper survey design and control, to compare the sample coverage properties, response rates, and cost per unit of the two alternative frames.

The second objective imposes several requirements on the sample design and survey process. The sections which follow describe the design and development of the AP (Section 3) and HCFA EDB file (Section 4) samples of AHEAD-eligible persons. A discussion of the survey costs and errors for the two frames and a general design for the methodological investigations is presented in Appendix A.

3. AHEAD Area Probability Sample Component

3.A HRS Area Probability Sample Screening

To identify the probability sample of the 1931-1941 birth year cohorts for the HRS, SRC interviewers conducted a household screening of a national area probability sample of approximately 69,000 housing unit addresses. Approximately 18% of the HRS area probability sample households contained one or more persons in the eligible age range for the HRS. The large probability sample of households that are ineligible for the HRS provided a unique opportunity to efficiently pursue surveys of other populations, such as the older birth cohorts. A significant portion of the Wave 1 HRS survey activity was the collection of complete household member listings for the entire sample of addresses to determine the eligibility of the household for participation in the HRS. **This HRS sample screening process served equally well for the identification of a national area probability sample of households with individuals age 70+ for AHEAD.**

During the HRS screening, SRC interviewers identified and obtained complete tracking information for all members of sample households who were age 68 or older (born in 1923 or earlier). Names, addresses and telephone numbers of these older cohort members were recorded on the sample coversheet, along with two reference contacts who would know their whereabouts should they move. Sample individuals who were eligible for AHEAD were informed that SRC might contact them the following year to request their participation in a research project. Figure 1 contains relevant excerpts from the HRS coversheet which define the procedure used to identify household members age 69+ and obtain the desired recontact and tracking information. This special tracking information was obtained for all AHEAD-eligible individuals, whether or not the household also contained an HRS-eligible individual.

The area probability component of the AHEAD sample design was therefore a probability subsample of the full 1992 HRS area probability sample (Heeringa & Connor, 1994). The AHEAD sample of the pre-1924 birth year cohorts which was derived from the HRS sample household screening shared all of the area probability sample design features of the Health and Retirement Survey: 1) a core national area probability sample of households (which included an oversample of Black households); 2) a supplemental area probability sample of households in the state of Florida; and 3) a supplemental area probability sample of Hispanic households.

For the most part, the household populations for HRS and AHEAD did not overlap. Based on 1991 CPS household composition data, 91.5% of households with one or more persons in the AHEAD study population were <u>not</u> expected to include persons who were also eligible for HRS. In the remaining 8.5% of AHEAD households, an HRS interview could also be taken with one or more younger members of the household. Prior to the AHEAD Wave 1 data collection, the sample of HRS/AHEAD "overlap" households was randomized 60% to the HRS panel and 40% to AHEAD (see Section 7.A below).

Figure 1

Z2. RECRUITMENT FOR OLDEST OLD SAMPLE

1. PEOPLE AGE 69 OR OLDER IN HH LISTING AND NOT HRS R

2. ALL OTHERS ---> GO TO Z3

Right now, we are only interviewing people who were born between 1931 and 1941.) We (also) expect to be doing an important study of older adults next year, and we would like to talk with (NAME/PERSON(S) 69 OR OLDER [BORN BEFORE 1924]) then.

The purpose of the study is to learn more about how and when people's health changes as they age. Most of what is currently known about this comes from studies done in hospitals, and not from talking with people in their own homes.

SECOND PERSON AGE 44-50 OR 68+

Z12. I need to verify your full legal name as it appears on official documents (VERIFY SPELLING OF FULL NAME AND WRITE CLEARLY)

NAME REFUSED | TITLE: | MR | MS | MISS | MS | DR | REV |

Z13. In what year were you born? _____ YEAR OF BIRTH

Z14. Do you have another place of residence or somewhere else you live during different times of the year?

1. YES → Z14a. We may wish to contact you at the other residence. May I have that address and phone number?

ADDRESS REFUSED

5. NO

_____
STREET ADDRESS

_____
CITY STATE ZIP

TELEPHONE NO: _____/_____ R HAS NO PHONE PHONE REFUSED
AREA CODE

Z15. If for any reason we should have difficulty contacting you, could you give me the name, address, and telephone number of two close friends or relatives who will know how to get in touch with you? [And what is this person's relationship to you?]

1. NAME:_____ RELATIONSHIP TO R:_____
ADDRESS:_____
TELEPHONE:_____

2. NAME:_____ RELATIONSHIP TO R:_____
ADDRESS:_____
TELEPHONE:_____

3.B The Primary Stage Sample

With the exception of Florida and a special Hispanic supplement sample, the AHEAD AP core sample used the 2/3 partition of the SRC National Sample design. The original HRS AP sample design used a larger sample of PSUs that included the full complement of National Sample PSUs in the Census South region and non-MSA strata of the Northeast, Midwest, and West regions. The 2/3 partition of the SRC National Sample included the 16 largest self-representing MSA primary areas and a stratified subsampling of 45 of the 68 nonself-representing PSUs for a total of 61 unique primary stage sample locations.

In addition to the national, multi-stage area probability sample (the **core** sample), the AHEAD AP sample design included all or part of the original HRS **oversamples**. The oversamples were designed to increase the numbers of Black and Hispanic respondents as well as the number of residents of the state of Florida. Within the 61 PSUs which comprised the first stage of the AHEAD AP design, a supplemental sample of second-stage units (SSUs) had been selected for HRS screening from second stage strata of Census tracts containing 10% or more 1990 Census households with a Black head of household. Thus, households living in residential areas eligible for the second stage sample supplement (more than 10% Black households per block) had a greater probability of selection than those in SSUs which had less than 10% Black households.

A design objective for the HRS that was carried forward to the AHEAD design was to obtain a two-fold oversampling of Mexican-American households. The HRS Hispanic Supplement, concentrated in 21 PSUs, had required additions to the PSU sample, especially in the West and Southwest. Five of the 16 core sample self-representing PSUs were included in the Hispanic PSU supplement: Los Angeles CA, Chicago IL, San Francisco CA, Dallas TX, and Houston TX. In addition to expanding the primary stage of the sample, supplemental sampling of SSUs with Hispanic population density of 10% or more was used to assure sufficient sample size to permit subgroup analysis.

In addition to the oversamples of Blacks and Hispanics, the AHEAD design inherited an oversample of Florida residents (across all race and ethnic groups) from the HRS AP design. For HRS, the number of primary stage strata for Florida PSUs had been expanded from five to 12. This expanded primary stage sample was retained in the AHEAD AP design.

For a detailed description of the original HRS sample design from which the AP sample for AHEAD was derived, see Heeringa and Connor (1995).

3.C Secondary-Stage Selection of SSUs

3.C.1. Core Sample

The SSUs of the HRS/AHEAD multi-stage area probability sample were selected directly from computerized files that were prepared from the 1990 Census PL 94-171 CD-ROM file. The designated second-stage sampling units (SSUs) or "area segments" are comprised of Census blocks or groups of blocks. Each SSU was assigned a measure of size equal to the total 1990 housing unit count for the area. A minimum of 72 housing units was required for SSUs. If a block had no housing units or fewer than 72 housing units, a computer program developed at SRC was used to group the ordered file of Census blocks into SSUs of minimum measure of size (72 housing units). The final sample of SSUs was a systematic selection with probabilities

proportional to the assigned measures of size.

The number of secondary selections varied across the PSUs but was designed to achieve an approximately proportionate allocation to the primary stage strata. The number of SSUs in the self-representing PSUs was proportional to the size of the PSU (stratum) and ranged from a high of 61 in New York to a low of 16 in the six smallest self-representing PSUs. Table 3-1 shows the number of core SSUs in each AHEAD PSU.

Table 3-1: AHEAD Wave 1 Area Probability Sample Primary Stage Strata and Second Stage Sample Allocation

AHEAD Stratum	Total SSUs	Core SSUs	Black SUs	Hisp SSUs
1	75	61	14	·
2	85	50	7	28
3	60	43	10	7
4	35	29	6	·
5	34	27	7	·
6	31	24	3	4
7	27	20	7	·
8	31	23	3	5
9	35	25	3	7
10	19	18	1	·
11	17	16	1	·
12	19	16	3	·
13	17	16	1	·
14	19	16	3	·
15	17	16	1	·
16	19	16	3	·
17	27	24	3	·
18	29	24	5	·
21	27	24	3	·
23	28	24	4	·
24	24	24	·	·
26	27	24	3	·
27	28	24	4	·
28	27	24	3	·
29	24	24	·	·
31	24	24	·	·
32	25	24	1	·
33	24	24	·	·
34	28	24	4	·
36	21	18	3	·
39	24	18	6	·
40	27	24	3	·

Table 3-1, continued

AHEAD Stratum	Total SSUs	Core SSUs	Black SUs	Hisp SSUs
41	25	24	1	·
42	27	24	3	·
43	27	24	3	·
44	24	18	·	6
45	21	18	3	·
47	18	18	·	·
49	19	18	1	·
50	22	18	4	·
52	5	·	·	5
53	24	24	·	·
55	27	24	·	3
56	30	24	·	6
57	30	24	·	6
58	30	24	·	6
59	24	24	·	·
60	30	24	·	6
63	12	12	·	·
64	12	12	·	·
65	12	12	·	·
66	12	12	·	·
68	12	12	·	·
70	12	12	·	·
73	16	12	4	·
74	15	12	3	·
75	12	12	·	·
76	12	12	·	·
77	18	12	6	·
78	12	12	·	·
80	12	12	·	·
81	16	12	4	·
82	12	12	·	·
84	12	12	·	·

Table 3-1, continued

AHEAD Stratum	Total SSUs	Core SSUs	Black SUs	Hisp SSUs

85	12	12	·	·
86	12	12	·	·
87	12	12	·	·
88	18	18	·	·
89	12	12	·	·
90	12	12	·	·
91	12	12	·	·
92	6	·	·	6
93	6	·	·	6
94	7	·	·	7
95	4	·	·	4
96	6	·	·	6
97	6	·	·	6
98	6	·	·	6
99	6	·	·	6
100	6	·	·	6
101	6	·	·	6
Total	**1695**	1400	147	148

3.B.2 Black Supplement

The Black Supplement to the HRS AP sample consisted of 166 additional SSU selections. Because the AHEAD study did not use all HRS AP PSUs, only 147 of the 166 HRS Black oversample SSUs were included in the AHEAD AP sample. At the primary stage of sampling, the Black supplement was fully integrated with the core National Sample design -- both the core and the Black Supplement shared the same set of primary stage sample locations. However within each PSU location, the selection of the Black Supplement SSUs was independent of the core SSU selection.

The first step in the original HRS sampling process was to allocate the 166 HRS Black Supplement SSUs to the PSUs. Since the primary purpose of the Black Supplement is to improve the precision of survey estimates for the Black population, the supplemental sample of SSUs was allocated to the primary stage sample locations in proportion to the total Black population of the stratum which each sample PSU represents. (In a standard national household sample -- such as the AHEAD core sample -- this allocation would be proportional to total population or housing counts.) Table 3-1 shows the SSU allocation by PSU for the Black Supplement.

A special second stage sampling frame was then constructed for each PSU which had been allocated one or more supplemental SSUs. This frame consisted of SSUs having at least ten percent Black population. Through the use of appropriate weights in the analysis of the survey data, Black households not covered by the supplemental frame (but covered by the core National

Sample frame) receive unbiased representation in survey estimates. Excluding low density Black areas from the supplemental frame greatly increased the efficiency of the Black Supplement.

Because the minimum measure of size for the SSUs was based on Black households, the total SSU size could vary depending on its Black household density. Based on the predetermined allocation to the PSUs, the Black Supplement SSUs were selected with probability proportionate to size measured in expected Black households. Although the Black Supplement was intended primarily to increase the number of eligible Black respondents, there was no race screening in the Black SSUs. All households with at least one person born prior to 1924 were eligible regardless of race. However, the average proportion of Black households in the Black supplement was expected to be about 75% (compared to 10% in the core SSUs). Thus, the Black Supplement introduces variation in selection probabilities for AHEAD-eligible households regardless of race or ethnicity. The sampling weight compensates for differential in household selection probabilities.

3.C.3 Hispanic Supplement

Within the HRS Hispanic Supplement PSUs, a special sample of 150 Hispanic Supplement SSUs was selected. The AHEAD AP sample excluded two Hispanic oversample segments which were from a non-MSA HRS PSU that was not retained in the AHEAD primary stage sample. The SSUs in the Hispanic Supplement were selected using the 1990 Census PL 94-171 file. For each PSU which was part of the Hispanic supplement, a file was constructed of all Census blocks which were part of the PSU definition. The file of census blocks was then ordered geographically as described above. A computer program was used to cluster the blocks into SSUs with a minimum measure of size of 96 Hispanic persons. A sampling frame was then formed from SSUs having at least 10% Hispanic population. From this frame the predetermined number of SSUs was selected from each Hispanic supplement PSU with probability proportional to Hispanic population. The AHEAD SSU allocation to Hispanic Supplement PSUs is shown in Table 3-1.

In the Hispanic Supplement SSUs, households were screened to include only those which had at least one age-eligible Hispanic person. Therefore, selection probabilities for AHEAD-eligible non-Hispanic households are not affected by the Hispanic Supplement sample. The average proportion of Hispanics in the Hispanic supplement SSUs was expected to be about 20% (versus about 5% in the core SSUs). Although the allocation of the Hispanic supplement PSUs and SSUs was based on Mexican-American Hispanic population, all Hispanic groups were eligible for the Hispanic supplement. However, because the supplement was concentrated in areas with high Mexican-American population density, the Hispanic respondents in the supplement were more likely to be Mexican-American than other groups such as Puerto Ricans or Cuban-Americans. The sampling weight compensates for the differential probabilities of selection for Hispanic households. (See Section 7.)

3.C.4 Florida Sample

The Florida oversample was completely integrated with the core sample at both the PSU and SSU levels. However, the AHEAD Florida sample had seven more Florida PSUs than the regular SRC National Sample. The expanded set of PSU selections was accomplished by subdividing the five original Florida strata into twelve new strata and making new PSU selections from each. Within each of the 12 Florida PSUs, a conventional sample of SSUs was

selected. Table 3-1 shows the allocation of SSUs to the Florida PSUs. A sampling weight which compensates for the oversampling in Florida is required for analyzing the core sample. The sampling weights are described in Section 7.

3.D Third-Stage Selection of Housing Units

For each original HRS AP SSU, a listing had been made of all housing units located within the physical boundaries of the SSU. For SSUs with a very large number of expected housing units or a very large geographic area, all housing units in a subselected part of the SSU were listed. Within each sample domain, the final equal probability sample of housing units for the HRS survey was systematically selected from the housing unit listings for the sampled SSUs. The equal probability sample of households within each sample domain was achieved by using the standard multi-stage sampling technique of setting the sampling rate for selecting housing units within SSUs to be inversely proportional to the PPS probabilities used to select the PSU and the SSU. The number of selected lines took into account the expected occupancy rate, the screening required to find age-eligible households, and the expected response rate.

3.E Fourth-Stage: AHEAD AP Frame Respondent Selection

Within each original HRS sample housing unit, the SRC interviewer prepared a complete listing of all household members. The full name, sex, age, and relationship to informant was recorded for each member of the household. The informant was then asked the year of birth of any person in the housing unit age 69 and older. The full name, telephone number and address information was recorded for each household member born prior to 1924. Names, addresses and phone numbers of two contact persons (i.e., relatives, friends) were also recorded on the form (see Figure 1). When the HRS cover sheets were returned to SRC at the conclusion of HRS field work, the name, address, and contact information for each AHEAD-eligible household member was entered into a computerized data base.

When the final specifications for the AP component of the AHEAD dual-frame sample design were complete, two major edit steps were performed:

1) Persons in the 12 HRS PSUs that were not included in the AHEAD AP primary stage sample were removed from the data base; and

2) Household financial units (married couples) that included both the HRS-eligible and AHEAD-eligible persons were identified. Through a random subsampling, 60% of such "overlap" units were assigned to the HRS panel and 40% to the AHEAD longitudinal data collection. Records for AHEAD-eligible persons residing in overlap units that were allocated to HRS were removed from the AHEAD AP sample data base.

The third and fourth edit steps prepared the data base for final sample selection:

3) In six Florida PSUs, a 7 in 10 subsample of the original HRS SSUs was chosen for the AHEAD sample. Cost considerations were the basis for this reduction in the available sample size;

4) Within the AHEAD AP data base, households were classified as Group 1 or Group 2 according to the following rules: 1) If only one age-eligible person was listed, the household was

Group 1 if the person was born between 1914 and 1923, and Group 2 if the person was born prior to 1914; 2) If a married couple had at least one spouse born prior to 1914, the household was classified as Group 2; 3) In a household with more than one single person or a single person and a married couple, the classification of the household depended on the year of birth of the person(s) in the randomly selected household financial unit.

A fifth step introduced the subsampling of Group 2 households that is required by the dual-frame design:

5) Subject to the edits described in Steps 1 to 3 above, all eligible persons in Group 1 AP households were eligible to be selected as an AHEAD primary respondent. However, to accommodate the dual-frame design for the older age group, a 1 in 2 sample of AHEAD AP Group 2 households was selected. The 1/2 subsample was achieved through a stratified sampling of 50% of the core sample SSUs. All Group 2 households originally screened in the AHEAD Black and Hispanic Supplement SSUs were included in the AHEAD AP sample. The subsampling of SSUs was conducted in a way that preserved the original stratification for the full AP sample of SSUs (see Section 3.B above). If the sample address for the Group 2 household belonged to a subsampled SSU the AHEAD-eligible persons in the household were eligible for final selection as a primary respondent. Group 2 households in the complementary 50% subsample of SSUs were excluded from the final sample of AHEAD AP primary respondents. Field travel and associated cost savings motivated the decision to subsample Group 2 households at the SSU as opposed to individual household level.

3.F Selection of the Primary Respondent, Household Contact

Through these five steps, the original HRS AP data base of AHEAD-eligible households and persons was reduced to a final sample frame of unique households containing eligible persons from the pre-1924 birth year cohorts. A sample of primary respondents for the AHEAD AP sample was then selected from the frame (one per eligible household).

Each selected primary respondent was contacted by telephone or in-person for the AHEAD Wave 1 interview. If the selected primary respondent was married, his or her spouse was also automatically designated for the AHEAD interview regardless of the spouse's age. Although the data base of eligible households and persons provided marital status and spouse's name for many married primary respondents, SRC interviewers verified the primary R's marital status at the time of contact for interview. Care was taken to reflect all new marriages, divorces, deaths that had taken place since the HRS AP screening and to pick up age-ineligible spouses who because of their age (birth year >1923) were not included in the AHEAD AP sample frame data base that was compiled from the HRS screening questionnaire.

If the AHEAD Wave 1 primary respondent was deceased or institutionalized, the case was coded as nonsample. No attempt was made to conduct a proxy interview or to interview a surviving age-ineligible or new spouse who resided in the household. In subsequent waves of AHEAD, a close-out interview will be conducted with the spouse or other proxy whenever a panel member has died. If an AHEAD panel member has become institutionalized after Wave 1, interviews will be continued in future waves either with the respondent or with a proxy.

4. HCFA Enrollment Data Base (EDB) Sample

The EDB file is a data base which contains names, addresses, demographic and benefit-related information for all persons who are currently enrolled in Medicare. The file is a special subset of the Social Security Administration's (SSA) Master Beneficiary record. SSA maintains the EDB file for the Health Care Finance Administration (HCFA), which in turn makes the file available to governmental agencies for use as a sampling frame for surveys of older age, beneficiary or disabled populations.

The HCFA EDB file list is updated regularly and contains the following information:

- name of enrollee
- mailing address (street address, city, state, ZIP)
- county of record
- Medicare claim number (extended SSN)
- date of birth
- sex of beneficiary
- race (white, Black, unknown, other)

As the federal-government sponsor of AHEAD, the National Institute on Aging (NIA) requested HCFA to provide confidential access to these basic elements of the EDB file for use in AHEAD sample selection. Standard procedures for such sample selections limited the sampling rate/ sample size to no more than 5% of the EDB file population.[1] To a large extent the 5% "limitation" may have been a practical decision that was linked to the existence of the Health Insurance Skeleton Eligibility Write-off (HSKEW) data base. Selected quarterly, the HSKEW represented a 5% sampling of the complete EDB file. From a computational standpoint the sample-based HSKEW data base was much easier to process and was routinely used by HCFA for the analysis and tabulation of Medicare enrollment data.[2] Were it not for the methodological study component, a 5% sample of Medicare beneficiaries drawn from a quarterly HSKEW data base would have met the sample development requirements of the AHEAD dual-frame design. As discussed in more detail in Appendix A, a major objective behind using a dual-frame design for the sampling of Group 2 AHEAD households was to compare the relative population coverage of the EDB file and standard area probability sampling methods currently used in the Health Interview Survey (HIS) and other major national studies which include older age populations. To conduct the coverage comparison in a statistically efficient way, it must be possible to conduct an exact match of the AP sample frame and the HCFA EDB file list of Medicare enrollees. The exact match could only be performed if universal access was gained to the EDB file.

Since the EDB file data base contained millions of records and since it required special programming efforts, HCFA was initially asked to allow NIA and SRC confidential access to a copy of the full data base file. This request would have resulted in a database that was far too

[1]For example, in selecting the sample for the 1991 round of the Medicare Current Beneficiary Survey (MCBS), HCFA provided Westat, Inc., with a 5% sample of beneficiaries who are included in the Health Insurance Skeleton Eligibility Write-off (HSKEW) file.

[2]The Continuous Medicare History Sample (CMHS) file is a HSKEW companion data base that contains beneficiary characteristic data and information on hospital and nursing home stays, surgeries and other Medicare-covered treatment.

large to process efficiently. Alternative approaches permitted some form of match to assess the relative coverages of the AP and HCFA EDB file frames. The selected alternative was to limit the sample frame to a geographic sample such as all EDB file entries for specific ZIP Codes or for the counties that belonged to a primary stage unit of the HRS AP sample design.

4.A Acquiring and Developing the HCFA Frame

4.A.1 HCFA file request

Given the size and cost of the entire data base of enrollment records, HCFA agreed to build a subset of the full file for SRC. For them to do this, SRC provided HCFA with 1) a demographic definition -- persons 77+ years of age, and 2) a geographic definition -- a complete list of the counties used in the Health and Retirement Study.[3] The list of counties included all HRS oversample PSUs and the 1/3 sample in the South and non-MSAs. Counties included in a stratified sample of six of the 12 HRS Florida PSUs were included. The list sent to HCFA included county name, FIPS state and county codes for the 274 counties in HRS/AHEAD PSUs. The HCFA EDB file arrived on 10 nine-track tapes with a total of 4,373,198 records that matched the FIPS county code for one of the designated sets of HRS/AHEAD PSUs.

4.B HCFA EDB Frame Sample Selection

A three-stage probability sample of Medicare enrollees was drawn for this study. At the first stage of selection, all enrollees residing within AHEAD PSUs were extracted from the EDB file. Prior to the second stage of selection, ZIP Code clusters were formed by pooling together all EDB enrollees within the same ZIP Code area. The counts of enrollees formed a measure of size for each second stage ZIP Code cluster. At the third and final stage of sampling, a subsample of enrollees was drawn within each selected second stage unit. Subsampling utilized probabilities inversely proportional to size (number of enrollees) in a fashion which yields an overall equal probability sample of enrollees. (See Section 4.B.3.)

4.B.1 Primary Stage

With two minor exceptions, the primary stage of the HCFA EDB frame sample selection is identical to that for the AHEAD AP sample component (see Section 3.A). In the AHEAD AP component, the state of Florida is represented by PSU selections from 12 primary stage strata. The HCFA EDB sample is restricted to a subset of six of the 12 AHEAD AP strata, and the primary stage selection probabilities for the retained PSUs are adjusted accordingly. The second exception to the exact comparability of the primary stage designs for the two AHEAD dual-frame sample components is that the 10 added PSUs of the AHEAD AP Hispanic Supplement are not included in the HCFA EDB sample.

4.B.2 Second Stage Selection

The second stage of the HCFA EDB sample has two special features. First, within the designated sample of primary stage units, all enrollee records were grouped into ZIP Code geographic clusters. This ensured a degree of cost-saving clustering of selected respondent addresses within the PSUs. Second, a special probability sampling technique was used to link

[3]Although the HCFA sample was drawn from only the 2/3 sample PSUs and from people aged 80 years or older, the additional oversample and 1/3 sample PSUs were included for use in a methodological study on coverage.

the sampling of ZIP Code areas to the prior sampling of SSUs for the AHEAD AP design component. The linkage of second stage units for the two designs has several advantages: 1) it reduces travel costs for interviewers working in the larger geographic areas of sample PSUs; 2) it facilitates the coverage comparison for the AHEAD AP and HCFA EDB sample design components.

The special probability sampling technique used to link the HCFA EDB SSUs to the previously selected AHEAD AP SSUs is best described by first considering a conventional equal probability three-stage PPS sampling design: 1) selection of PSUs with probability proportionate to total population; 2) selection of ZIP areas within PSUs with probability proportionate to population; 3) and selection of beneficiaries with selected ZIPs with probability inversely proportionate to the probabilities of stages (1) and (2). In theory, the ZIP area measure of size used in stage 2 of this three-stage design can be viewed as an aggregate of measures of size for a set of smaller SSUs which together comprise the geographic ZIP Code area. Assuming there is a 1:1 correspondence between the smaller SSUs and a specific ZIP area (that is, we assume SSUs do not include parts of more than one ZIP area), a PPS sample of the smaller units also identifies a PPS sample of the larger ZIP areas. Under such an approach, the total second stage probability of selection for the larger ZIP area is the sum of the second stage probabilities for its smaller SSU divisions.

Therefore, the prior PPS selection of SSUs for the AHEAD AP design component can be (and was) used as a device for also designating the sample of ZIP Code areas for the HCFA EDB design component. To compute the second stage probability of selection for the selected ZIP areas, it was assumed that the total count of HCFA enrollees in a ZIP area was equal to the sum of enrollment counts for all AHEAD AP SSUs that were implicitly linked to that ZIP area.

4.B.3 Third-Stage Probability of Selection

The overall AHEAD Wave 1 equal probability of selection for HCFA EDB sample cases was .0000235 or 2.35 in 100,000, the rate required to obtain an equal probability sample of n=2000 eligible enrollees. The third-stage probability sample of HCFA beneficiaries was achieved by using the standard multi-stage sampling technique of setting the sampling rate for selecting HCFA beneficiaries within ZIP Codes to be inversely proportional to the PPS probabilities (above) used to select the PSU and ZIP Code. The final HCFA EDB selection equation is: where MOS_α is the measure of size of the AHEAD AP PSU, MOS_h is the measure of size of the

$$\begin{cases} f_{HCFA} = f_1 - f_2 - f_3 - f_4 \\ .000235 = \dfrac{MOS_\alpha}{MOS_h} \bullet \dfrac{MOS_B - b_a}{MOS_\alpha} \bullet \dfrac{MOS_{ZIP}}{MOS_B} - \dfrac{K}{MOS_{ZIP}} \end{cases}$$

primary stage stratum, MOS_B is the measure of size of the AHEAD AP SSU, MOS_{ZIP} is the measure of size of the ZIP Code linked to that AHEAD AP SSU, and b_a is the number of AHEAD AP SSUs selected from the PSU. K is the expected number of cases selected from the ZIP Code.

Prior to selecting the third stage sample, the HCFA file was sorted by sex within ZIP Code and a sample of approximately 2000 cases was selected to yield 1700 sample cases + 300 reserve cases. The sample was checked to ensure that approximately equal numbers of sample cases were selected from each NSR PSU. Each of the 1700 cases in the main sample was then assigned a Sample ID, and the Medicare No. (claim number) was removed to safeguard confidentiality while in the field. This final sample file was given to the Field Control Office. To ensure that the sample file had the most current known address for the beneficiary, the sample address file was compared to the NCOA (National Change of Address) File and updated as appropriate.

5. AHEAD Sample Release and Survey Monitoring

Within each AHEAD AP PSU, the full sample of AP SSUs was randomly divided into two parts. The first subsample of SSUs was introduced in October 1993 and the second set in January 1994. The HCFA EDB frame arrived at SRC in December of 1993. Therefore the HCFA EDB sample entered the field as a third release in late February of 1994. This staged sample release was designed to control sample size and cost. By releasing the full AHEAD sample as a sequence of probability subsamples, adjustments could be made as needed without affecting the overall representativeness of the sample. For large studies, the sequential sample release provides an opportunity to examine eligibility rates, response rates, and survey costs and to make appropriate adjustments in sample sizes for subsequent sample release.

The timing of the release of the AHEAD Wave 1 sample differs for Group 1 and Group 2 households (see Section 1.A). In October 1993, the Group 1 household sample in a 70% subsample of AHEAD AP SSUs was released to the field for data collection. At the same time, the Group 2 household sample in a 50% subsample of AHEAD AP SSUs was released for data collection. (Note: The status of the HCFA EDB sample of Group 2 households was not finalized by October, 1993.)

All AHEAD-eligible households in the AHEAD Hispanic Supplement and Florida oversample SSUs were released in January 1994. The HCFA EDB extract file was released to SRC in mid-December of 1993, too late for the Group 2 sample from this frame to be included in the January release of the AHEAD sample. Essential data management steps and the final selection of the HCFA EDB sample were completed by mid-January of 1994. The HCFA EDB sample of Group 2 households was released to interviewers for data collection in late February of 1994.

Figure 2 shows the timing of the sample release for the various components of the total sample, i.e., the Group 1 AHEAD AP sample, the Group 2 AHEAD AP sample and the HCFA EDB Group 2 sample. Figure 3 shows that the sample rotation release schedule for the total sample produced an interview completion rate which facilitated monitoring of survey quality factors and cost factors. From October to January, interview completions accumulated at a relatively slow rate. Following the January sample release, interview completions began to rise more sharply. Immediately prior to the February release date, about 60% of all expected AHEAD interviews were completed. At this point, a decision was made to release a HCFA EDB sample of n=1700.

FIGURE 2
AHEAD WAVE 1 SAMPLE RELEASE FOR CONTROL OF SAMPLE SIZE AND COST

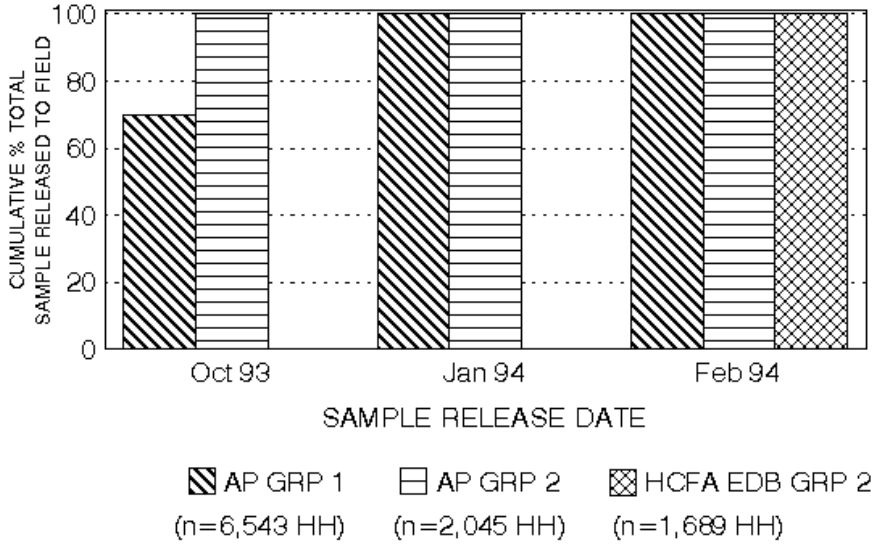
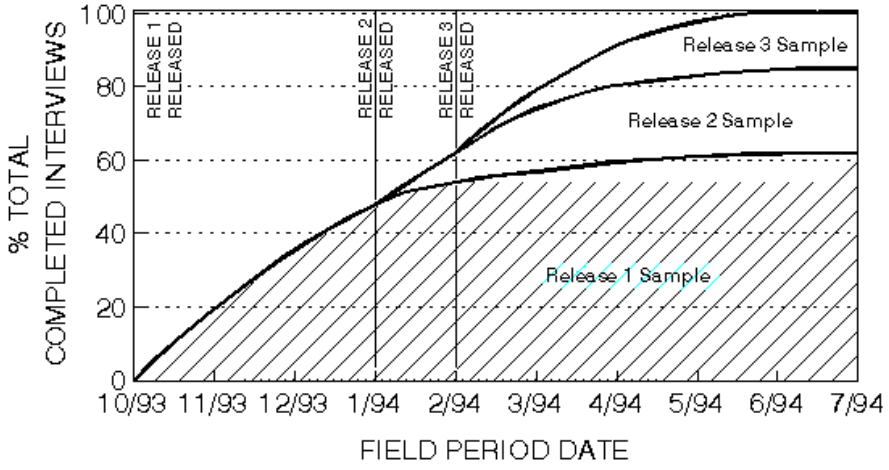


FIGURE 3
DISTRIBUTION OF AHEAD WAVE 1 INTERVIEWS BY SAMPLE RELEASE AND FIELD PERIOD



6. Survey Dispositions

This section summarizes the actual field experience with the AHEAD dual-frame sample design, reporting empirical outcomes for eligibility and response rates and comparing these outcomes to the original sample design specifications and assumptions. Section 6.A provides separate descriptions of the eligibility rate outcomes for the HRS AP and HCFA EDB components of the dual-frame sample design. Section 6.B presents weighted and unweighted response rate outcomes.

6.A Original HRS AP Occupancy, Screening and AHEAD Eligibility Outcomes

The original 1992 HRS area probability sample screening involved a total sample of n=69,377 sample housing units. Based on in-person contacts by SRC interviewers and supervisors, 59,918 (87.1%) were found to be occupied by a household unit. Through concerted efforts by the SRC field staff, screening information needed to establish HRS and AHEAD eligibility was obtained from 99.6% (all but 214) of these households. (See Heeringa and Connor (1995) for a more detailed discussion of HRS AP sample occupancy and screening response rate outcomes.)

Of the 59,880 households that completed HRS AP screening interview, 9,474 (15.9%) identified one or more persons who were eligible for AHEAD. Focusing on the nationally representative HRS core sample of 43,229 screened households, AHEAD eligible persons were identified in 7,826 (18.1%) of households. This compares favorably to an April 1990 Census Public Use Micro Sample (PUMS) estimate which suggests that 17.8% of U.S. households contain one or more eligible persons in this age range.

6.B Household-level and Person-level Response Rates

Table 6-1 summarizes the household-level response rate experience of the overall AHEAD survey and its sample components. Table 6-2 shows the corresponding person-level response rates. The sample design specifications called for an 80 percent response rate. The tables below show that this rate was met or exceeded by all sample components except the HCFA Group which had a household response rate of 74.3% and a person-level response rate of 76.8%.

Table 6-1: AHEAD Wave 1 Household-level Response Rates

Sample Component	Eligible HHs	Completed At Least One Interview	Response Rate	
			Unweighted	Weighted
Complete Sample	7,509	6,047	0.805	0.810
AP Group 1	4,603	3,753	0.815	0.822
AP Group 2	1,570	1,305	0.831	0.831
HCFA Group 2	1,336	989	0.740	0.743

Table 6-2: AHEAD Wave 1 Person-level Response Rates

Sample Component	Eligible Persons	Interviewed Persons	Response Rate	
			Unweighted	Weighted
Complete Sample	10,229	8,222	0.804	0.810
AP Group 1	6,605	5,323	0.806	0.816
AP Group 2	1,982	1,631	0.823	0.821
HCFA Group 2	1,643	1,268	0.772	0.768

7. Wave I (AHEAD) Study Weights for Data Analysis

This section describes two weight variables that have been developed for use by analysts of the AHEAD Wave 1 data. As its label implies, the **household analysis weight** is to be used for analysis of household characteristics such as housing characteristics, household income and assets. [Throughout this section, the term "household" is used to refer to the household financial units that are the units of observation for AHEAD -- see Section 1.A]

The **person-level analysis weight** is the correct weight for analysis of AHEAD data collected about the individuals in the sampled households. This includes both age-eligible respondents and age-eligible spouses in married couple households. For example, the person level analysis weight would be used for an analysis of the earnings history of women in the pre-1924 birth cohorts. Likewise, an analysis of the medical expenditures of men with chronic heart disease would use the person-level analysis weight. Since AHEAD data on age-ineligible spouses (birth cohorts 1924 and later) serves primarily to provide context for analysis of their age-eligible spouses, all AHEAD cases with birthdates after 1923 have a zero value person-level weight.

The development of the household and person-level analysis weights involves three general steps. The first step is the computation of the selection weight for each household and age-eligible person in the sample. The selection weight factor is simply the reciprocal of the probability that the household or person is included in the sample. As described in detail below, the computation of the selection weight factor must take into account not only the dual frame design of the AHEAD Wave 1 sample, but also each of the detailed sample allocation decisions that were made in forming the final sample design. Step two of the weight development process is the derivation of a nonresponse adjustment factor which is designed to adjust for geographic and race group differences in response rates. The final step in the development of the household and person-level analysis weight is post-stratification to adjust weighted AHEAD sample demographic distributions to known 1990 Census totals.

Combining the three factors, the final form of each analysis weight is the product:

$$W_i = W_{i,SEL} \times W_{i,NR} \times W_{i,PS}$$

where:

W_i = the composite analysis weight for unit i;
$W_{i,SEL}$ = the sample selection weight for unit i;
$W_{i,NR}$ = the nonresponse adjustment factor for unit i;
$W_{i,PS}$ = the poststratification factor for unit i.

The following sections describe each of the three factors in the composite analysis weight variable.

7.A. Household Selection Weight, $W_{i,SEL}$

As described in Section 2 and 3, the AHEAD Study sample is selected under a dual-frame sample design that includes both an area probability sample component (the HRS screening recruitment) and an independent sample from the HCFA EDB file. In general terms, the following equation defines the dual-frame sample selection probability for AHEAD sample households:

$$f_{i,AHEAD} = f_{i,AREA} + f_{i,HCFA} - f_{i,AREA} * f_{i,HCFA}$$

where:

$f_{i, AHEAD}$ = the joint probability of selecting the ith household from either the AP or the HCFA sample frame;

$f_{i, AREA}$ = the probability of selecting the ith AHEAD household from the AP sample frame; and

$f_{i,HCFA}$ = the probability of selecting the ith AHEAD household from the HCFA sample frame.

The corresponding household selection weight factor is the reciprocal of the joint probability that the household is sampled for AHEAD:

$$W_{i,SEL} = (f_{i, AHEAD})^{-1}.$$

In order to apply this general model for the dual-frame sample selection probability, it is necessary to make several ***important assumptions*** concerning the representation of households on the two frames:

1) Representation of Area Probability Sample Frame: All AHEAD Wave 1 households are assumed to be eligible for selection under the area probability sample design. (See Appendix A for a review of area probability frame coverage.)

2) Representation of the HCFA EDB Sample Frame: Each household member born prior to 1914 is Medicare eligible and is enrolled on the EDB. (See Appendix A for a review of HCFA frame coverage.) If a household has two persons in this age range, it is assumed that both are enrolled and each will have an independent chance of being selected from the frame (see below).

The following subsections outline the derivation of the frame-specific selection probabilities that are required for the computation of the household selection weight factor.

7.A.1 Selection Weight: Area Probability Frame Component, $f_{i,AREA}$

The area probability frame component of the AHEAD dual-frame sample design is a subsample of the area probability sample design for the Health and Retirement Survey. Therefore, the computation sequence for the AHEAD area frame selection probability starts with the HRS sample selection probability for the household and adjusts this probability to account

for subsampling steps that apply specifically to the AHEAD sample design.

$$f_{i,AREA} = f_{i,HRS} * f_{i,AHEAD\ SUB}$$

7.A.1.a The HRS area frame selection probability, fi,HRS

At the time of the original HRS household screening, each U.S. household had a known sample selection probability. This probability was the product of four distinct factors:

$$f_{i,HRS} = f_{i,BASE} * k_{i,DOMAIN} * k_{i,UNLIST} * f_{i,SUB}$$

where:

$f_{i,BASE}$ = .000432 = the HRS constant or base rate of sample selection;

$k_{i,DOMAIN}$ = The "oversampling factor" for geographic domains:

The original HRS area probability sample design divided the U.S. into four geographic domains: 1) General (not in oversample areas); 2) High Black density Census Tracts (Census Tract is ≥ 10% Black); 3) High Hispanic density (Census Tract is ≥ 10% Hispanic and the stratum was eligible for Hispanic oversample selections) and 4) State of Florida. Sample households in the first domain are not oversampled relative to the base rate of sample selection and these households are assigned an oversample factor of 1.0. Respondents in the geographic domains 2-4 were oversampled at two times the base rate -- their oversampling factor is 2.0. There was no race screening in the high Black density domain. Regardless of race or ethnicity, all households in Census tracts with at least ten percent Black population were oversampled at twice the base rate. In geographic domain 3, only Hispanic households were oversampled at twice the base rate. Therefore, in this domain only Hispanic households have an oversample factor of 2.0. The non-Hispanic households in that domain have an oversample factor of 1.0. A household was classified as Hispanic if at least one age-eligible person in the household was Hispanic. It is possible for HRS sampled households to belong to two oversample domains (e.g., the high density Black domain in Florida) and therefore have four times the base chance of selection. Sampled housing units in these overlapping domains have an oversample factor of 4.0.

$k_{i,UNLIST}$ = unlisted SSU adjustment factor:

There were six SSUs in Los Angeles (including both Black and Hispanic Supplement SSUs) which could not be listed because of the danger from the April 1992 riots which followed the Rodney King verdict. In addition, one SSU in New Haven, CT was not listed because it was in a very dangerous area and one SSU in Anaheim, CA was not listed because it was a locked and gated area. The strategy used to compensate for SSUs which were selected from the PSU but were not listed was to create a special adjustment factor equal to the ratio of the number of SSUs in a domain in a PSU which should have been listed to the number which actually were listed and to apply the adjustment to the base selection weight of all sample lines in the listed SSUs. For example, in Los Angeles, seven Black oversample SSUs were selected but only five were listed. Therefore an adjustment weight of 7/5 or 1.40 was applied to sample lines in the five listed SSUs. The SSU location was also taken into account in constructing this adjustment factor. In New Haven, the weight factor was applied only to the SSUs in the central city which were similar to the dangerous SSU which was not listed. In this case a weight factor of 8/7 or 1.14 was applied to the seven listed SSUs in the central city.

$\mathbf{f_{i,SUB}}$ = special SSU subsampling probability:

There were 39 SSUs in which a subsampling procedure was used -- either for all or part of the sample lines in the SSU. Twenty-four of these SSUs were subsampled because of access problems such as locked buildings or gated subdivisions. Fifteen of the SSUs were subsampled because they were in dangerous areas. Interviewers could request subsampling of an SSU when normal procedures for interviewing in the SSU failed. Their requests were reviewed by their supervisor and if approved were sent to the Sampling Section for subselection. The Sampling Section then selected a systematic sample of one-third of the sample lines for attempted interviews. The remaining two-thirds received a special result code of "75," a non-sample code which did not count against the response rate.

The goal of the subselection process was to obtain at least some interviews from the difficult SSUs. Special efforts and resources were expended on the one-third of the sample lines retained, and the remaining two-thirds received a special non-sample result code. The sample probability adjustment for subsampled lines was spread across all sample lines in groups of similar SSUs in the same PSU. For example, there were two SSUs in Manhattan (New York City) which were subselected because of access problems. In order to create the probability adjustment to compensate for this subsampling, a list of all Manhattan SSUs was compiled together with a count of the original number of selected lines in each SSU. The number of sample lines which were "subselected out" was also determined. The probability adjustment weight factor which was applied to each sample line in the Manhattan SSUs was the total number of original sample lines divided by the total number of sample lines after subselection. In this case, 11 lines were removed from two SSUs by subselection and the total number of original sample lines in the 14 Manhattan SSUs was 388. Therefore the probability adjustment weight factor was 388/377 = 1.029.

This procedure of forming groups of similar SSUs within a PSU and calculating weight factors equal to the total original lines selected divided by the total lines after subselection was done for each of the 39 SSUs. In some cases, such as the Manhattan SSUs, more than one subselected SSU was in the same weighting group.

7.A.1.b AHEAD subsampling of the HRS area frame: fi,AHEAD SUB

The final AHEAD area probability sample is a subsample of the full design that was actually used for the original HRS AP sample screening. For the most part, the goal of subsampling was to reduce the geographic dispersion (i.e., the interviewing costs) of the AHEAD Wave 1 data collection effort. The following equation is the composite expression for the subsampling probabilities which determined the HRS to AHEAD transition of the area probability sample design:

$$\mathbf{f_{i,AHEAD\ SUB} = f_{i,PRIM} * f_{i,FLOR} * f_{i,BLACK} * f_{i,WIHH} * f_{i,OVER}}$$

where:

$\mathbf{f_{i,PRIM}}$ = AHEAD PSU subsampling probability

The primary stage sample for the HRS screening included the full complement of SRC National Sample PSUs in the Census South Region and the non-MSA domain of the Census Northeast, Midwest and West regions. For AHEAD, the primary stage sample in the South Region (excluding Florida) and the non-MSA domain was reduced to the 2/3rd partition of the full National Sample (see Section 3.A). Therefore, the following subsampling adjustment was applied to the basic HRS area sample probability for households in these domains:

Census Region	MSA/Non-MSA	$f_{i,PRIM}$
South	SR MSA	1.000
South (except Florida)	non SR MSAs non-MSAs	0.667
Florida	MSA, non-MSA	1.000
Northeast, Midwest, West	MSA	1.000
Northeast, Midwest, West	Non-MSA	0.667

$\mathbf{f_{i,FLOR}}$ = AHEAD Florida subsampling probability:

The original design plan for the AHEAD area probability sample specified a subsampling reduction in the number of Florida PSUs from the HRS total of 12 to a new total of six. Just prior to the start of the field period this design decision was reversed, and the full complement of 12 HRS AP PSUs was also used for the AHEAD. However, in six of the 12 PSUs a random sample of 70% of the pre-screened sample of AHEAD eligible households was released to the field to contact for an interview. Therefore a subsampling adjustment of .70 must be applied to AHEAD area probability sample households in these six PSUs (PSUs 86, 87, 88, 89, 90, 91).

$\mathbf{f_{i,BLACK}}$ = AHEAD Black Oversample Adjustment:

The HRS Black oversample involved an independent selection of high Black density SSUs across the full set of HRS primary stage units. The AHEAD reduction to the two-thirds sample of PSUs in the Census South Region and the non-MSAs (see above) disproportionately reduced the number of Black Supplement SSUs. Therefore, an additional subsampling correction of 0.907 is applied to all AHEAD households in the geographic domain which includes all Census tracts with at least 10% Black population.

$\mathbf{f_{i,WIHH}}$ = AHEAD Within Housing Unit Subsample Adjustment:

An HRS AP sample housing unit could contain more than one AHEAD-eligible individual or couple. In such cases a respondent unit (single adult or married couple) was randomly subsampled. [Since the HRS screening results were available on a computer database, this step was performed before the AHEAD sample was sent to the field.] The probability that an individual was subsampled as the respondent (or part of a married couple unit) depended on the marital status and numbers of age-eligible persons in the household. The subsampling probabilities for various combinations of AHEAD-eligible individuals in a housing unit are:

Household Composition (AHEAD eligible persons)	$f_{i,WIHH}$
1 single person living alone	1.0
2 single people living together (not a couple)	0.5
3 single people living together	0.33
2 people married, living together	1.0
3 people living together (married couple + single person)	
single person	0.5
married couple	0.5

$\mathbf{f_{i,OVER}}$ = the HRS/AHEAD "overlap" subsampling:

A small percentage of HRS AP married couple financial units included persons in both the HRS and AHEAD age ranges (e.g., a woman age 60 married to a man age 70). The decision was made that it would cause undue burden and other complications if such "overlap" financial units were asked to participate in both studies. The resulting action was the division of the overlap units between the two studies -- 60% being retained in the HRS panel and 40% transferred permanently to AHEAD. In computing the selection weight, overlap units retained for AHEAD are therefore given an additional .4 subsampling probability factor. Please note here that this subsampling probability produces a nontrivial inflation of the selection weight values for these household financial units.

7.A.2 Selection Weight: HCFA EDB Frame Component, $f_{i,HCFA}$:

Compared to the complex set of factors which determine the HRS AP frame selection probability for AHEAD households, the computation of the HCFA frame selection probability is the simple

product of two factors:

$$f_{i,HCFA} = f_{i,BASE} * k_{i,MARHH}$$

where:

$f_{i,BASE}$ = the basic equal probability HCFA sampling rate;
= .000200 if individual is born before 1914;
= 0 for persons born after 1913.

In theory, the AHEAD HCFA EDB sample selection gives each Medicare enrollee born prior to 1914 an equal chance of being selected. (See Section 4.B)

$k_{i,MARHH}$ = the HCFA EDB married household factor:

If both spouses in an AHEAD married couple household financial unit were born prior to 1914, the HCFA sample selection gives their respondent unit twice the basic probability of being selected for interview. The HCFA selection probability for such household financial units is therefore increased by a factor of 2.0.

7.B Household Nonresponse Adjustment Factor

Nonresponse is a potential source of nonsampling error in the AHEAD survey data. In an effort to counteract potential biases that may result from differential response across sample subclasses and domains, a nonresponse adjustment weight factor is incorporated as one of the multiplicative factors in the final household and person-level analysis weights. Several forms of nonresponse occurred in the AHEAD Wave 1 data collection. The first and most common form is nonresponse on the part of the complete respondent unit -- the single age-eligible adult or both spouses in a married couple. The second form of nonresponse could occur only in married-couple respondent units. Here one member of the couple could agree to provide the interview and the other refuse or be incapable of responding (non-interview). For the 2,768 married couples sampled for AHEAD Wave 1, both husband and wife cooperated 79% of the time. In 95% of couples with at least one age-eligible respondent, both parties gave interviews. Therefore, the nonresponse adjustment was made at the household rather than at the person level. (The household nonresponse adjustment is a factor in the final person-level analysis weight.)

To compute the nonresponse adjustment, households were assigned to nonresponse adjustment cells based on Census Division, MSA/non-MSA status, age group, and racial composition of the neighborhood. Florida households were assigned to a separate set of nonresponse adjustment cells in the Census South Atlantic division. Three race/ethnicity groups were defined: (1) non-Black/non-Hispanic, (2) Black, and (3) Hispanic. The first group consisted of households in Census tracts which were less than ten percent Black and less than ten percent Hispanic.[4] Households in the second or third group were in tracts which were at least ten percent Black or Hispanic respectively. If a household was in a tract which qualified for both the second and third group it was assigned to the group which had the highest proportion of population in the tract. The race of the respondent was not considered in the assignment of a household to race/ethnicity

[4]Only SSUs in PSUs which were eligible for the Hispanic oversample (those with significant Mexican-American population) were classified as Hispanic in forming the nonresponse adjustment cells.

group; only the proportion Black or Hispanic in the Census tract in which the SSU was located was considered.

The weighted response rate for each PSU/Race cell was determined by dividing the weighted total of households interviewed by the weighted total of known eligible households. The weight used in the household response rate calculation was the AHEAD selection weight described in Section 7.A. Households with unknown eligibility were excluded from the denominator of this calculation. The overall weighted household response rate for AHEAD Wave 1 was 81.0%. Table 7-1 shows the weighted response rate and household nonresponse adjustment factor for each adjustment cell.

Table 7-1: AHEAD Wave 1 Household Nonresponse Adjustment Factors

Division	MSA	Domain	Age Group	n	Resp. Rate	HHNRWGT
New England	MSA	Base	70-79	140	0.68931	1.45072
New England	MSA	Base	80+	95	0.64563	1.54887
New England	MSA	High Black	70-79	45	0.70841	1.41161
New England	MSA	High Black	80+	19	0.70383	1.42080
New England	non-MSA	Base	70-79	59	0.73554	1.35955
New England	non-MSA	Base	80+	32	0.67245	1.48710
Mid-Atlantic	MSA	Base	70-79	367	0.79019	1.26552
Mid-Atlantic	MSA	Base	80+	231	0.73637	1.35801
Mid-Atlantic	MSA	High Black	70-79	288	0.80628	1.24026
Mid-Atlantic	MSA	High Black	80+	176	0.78986	1.26605
Mid-Atlantic	non-MSA	Base	70-79	57	0.84426	1.18447
Mid-Atlantic	non-MSA	Base	80+	34	0.88224	1.13347
E. North Cntl	MSA	Base	70-79	468	0.80140	1.24782
E. North Cntl	MSA	Base	80+	260	0.86516	1.15586
E. North Cntl	MSA	High Black	70-79	221	0.76456	1.30795
E. North Cntl	MSA	High Black	80+	138	0.71148	1.40552
E. North Cntl	MSA	High Hisp.	70-79	34	0.71212	1.40426
E. North Cntl	MSA	High Hisp.	80+	17	0.62149	1.60904
E. North Cntl	non-MSA	Base	70-79	91	0.88360	1.13174
E. North Cntl	non-MSA	Base	80+	68	0.85765	1.16598
W. North Cntl	MSA	Base	70-79	133	0.82892	1.20639
W. North Cntl	MSA	Base	80+	75	0.76244	1.31158
W. North Cntl	MSA	High Black	70-79	60	0.87733	1.13982
W. North Cntl	MSA	High Black	80+	35	0.80041	1.24936
W. North Cntl	non-MSA	Base	70-79	122	0.88189	1.13393
W. North Cntl	non-MSA	Base	80+	83	0.91009	1.09879

Table 7-1, cont. Page 2

Division	MSA	Domain	Age Group	n	Resp Rate	HHNRWGT
South Atlantic	MSA	Base	70-79	136	0.83982	1.19073
South Atlantic	MSA	Base	80+	84	0.78548	1.27311
South Atlantic	MSA	High Black	70-79	171	0.92814	1.07742
South Atlantic	MSA	High Black	80+	115	0.78987	1.26603
South Atlantic	MSA	Florida	70-79	532	0.77802	1.28531
South Atlantic	MSA	Florida	80+	332	0.74848	1.33604
South Atlantic	non-MSA	Base	70-7	52	0.88889	1.12500
South Atlantic	non-MSA	Base	80+	39	0.83095	1.20344
South Atlantic	non-MSA	High Black	70-79	149	0.87261	1.14599
South Atlantic	non-MSA	High Black	80+	107	0.88539	1.12945
South Atlantic	non-MSA	Florida	70-79	86	0.80755	1.23832
South Atlantic	non-MSA	Florida	80+	48	0.78899	1.26744
E. South Cntl	MSA	Base	70-79	62	0.92063	1.08621
E. South Cntl	MSA	Base	80+	38	0.80838	1.23704
E. South Cntl	MSA	High Black	70-79	61	0.95276	1.04959
E. South Cntl	MSA	High Black	80+	49	0.81008	1.23445
E. South Cntl	non-MSA	Base	70-79	23	0.76471	1.30769
E. South Cntl	non-MSA	Base	80+	17	0.88164	1.13425
E. South Cntl	non-MSA	High Black	70-79	41	0.63529	1.57407
E. South Cntl	non-MSA	High Black	80+	22	0.85663	1.16737
W. South Cntl	MSA	Base	70-79	56	0.78778	1.26939
W. South Cntl	MSA	Base	80+	34	0.74413	1.34385
W. South Cntl	MSA	High Black	70-79	66	0.77644	1.28794
W. South Cntl	MSA	High Black	80+	47	0.74242	1.34695
W. South Cntl	MSA	High Hisp.	70-79	103	0.94927	1.05344
W. South Cntl	MSA	High Hisp.	80+	73	0.90473	1.10530
W. South Cntl	non-MSA	Base	70-79	26	0.89091	1.12245
W. South Cntl	non-MSA	Base	80+	27	0.88533	1.12952
W. South Cntl	non-MSA	High Black	70-79	169	0.84438	1.18430

Table 7-1, cont. Page 3

Division	MSA	Domain	Age Group	n	Resp. Rate	HHNRWGT
W. South Cntl	non-MSA	High Black	80+	118	0.88835	1.12568
W. South Cntl	non-MSA	High Hisp.	70-79	54	0.87578	1.14184
W. South Cntl	non-MSA	High Hisp.	80+	35	0.91646	1.09116
Mountain	MSA	Base	70-79	63	0.82813	1.20755
Mountain	MSA	Base	80+	37	0.69344	1.44209
Mountain	MSA	High Black	70-79	3	0.66667	1.50000
Mountain	MSA	High Black	80+	6	0.66667	1.50000
Mountain	MSA	High Hisp.	70-79	49	0.91919	1.08791
Mountain	MSA	High Hisp.	80+	31	0.82646	1.20998
Mountain	non-MSA	Base	70-79	14	0.85714	1.16667
Mountain	non-MSA	Base	80+	7	0.84038	1.18993
Mountain	non-MSA	High Hisp.	70-79	20	0.77193	1.29545
Mountain	non-MSA	High Hisp.	80+	21	0.84980	1.17675
Pacific	MSA	Base	70-79	223	0.82319	1.21479
Pacific	MSA	Base	80+	124	0.77693	1.28712
Pacific	MSA	High Black	70-79	63	0.72024	1.38842
Pacific	MSA	High Black	80+	45	0.68682	1.45599
Pacific	MSA	High Hisp.	70-79	249	0.82175	1.21692
Pacific	MSA	High Hisp.	80+	155	0.63726	1.56922
Pacific	non-MSA	Base	70-79	40	0.90361	1.10667
Pacific	non-MSA	Base	80+	25	0.91812	1.08918
Pacific	non-MSA	High Hisp.	70-79	7	0.64706	1.54545
Pacific	non-MSA	High Hisp.	80+	7	0.54768	1.82588

7.C Household Post-Stratification Factor

In spite of weighting corrections that reflect sample selection probabilities and nonresponse adjustments, weighted sample distributions of major demographic characteristics may not correspond exactly to those for the known household population. The departures of sample distributions from the underlying population are in part due to the variation that is inherent in the sampling process itself. Sample undercoverage, originating in the sampling frame or in the field sampling and updating procedures, also can cause sample distributions to deviate from known Census proportions. "Coverage" and estimation errors can also be introduced via the multiple weighting adjustments that are applied to the survey interview data. (Weights designed to attenuate one source of survey error may accentuate others.)

Post-stratification factors are adjustments to analysis weights that are designed to bring weighted sample frequencies for important demographic subgroups in line with corresponding population totals that are available from a source that is external to the survey data collection process. Beyond the simple appeal of the population controls, the post-stratification procedure is expected to reduce the mean square error of sample estimates.

The household post-stratification for the AHEAD sample was a simple control to 1990 Census totals for married couples and single households. Table 7-2 shows the household post-stratification factors for married couples and single person households (for households with persons ages 70+). The Census totals are from the 1990 PUMS file (Public Use Microdata Sample).

Table 7-2: Household Post-stratification Sample Factors

Household Type	1993 AHEAD Estimate	1990 PUMS Households	Poststratification Factor HHPSFACT
Single	9,941,336	10,827,408	1.08913
Married Couple	5,858,577	5,768,590	0.98464

The household analysis weight is the product of all of the factors described above -- the household selection weight, the household nonresponse adjustment, and the household post-stratification factor. This household weight should be used for descriptive analysis of household-level data from the interviewed AHEAD households.

7.D Person Level Post-stratification Weight

In addition to the household weight post-stratified to known 1990 Census household totals, the AHEAD person-level weight is post-stratified at the person level to 1990 PUMS totals for Census Region (4), by Sex (2), and by Age Group (2). In all, 16 post-stratification cells were formed (4 x 2 x 2 = 16). Each age-eligible respondent was given a basic weight equal to the AHEAD Household Analysis Weight. Weighted estimates of total persons were obtained for

each of the 16 poststratification cells. The person-level poststratification factor was then formed by dividing the 1990 PUMS total population for each cell by the AHEAD weighted estimate of total persons. Table 7-3 shows the definition for each cell, the PUMS and AHEAD estimated totals, and the person-level poststratification factor.

The person-level analysis weight is the product of the AHEAD household analysis weight and the person-level poststratification factor. Only age-eligible respondents have valid person-level weights. Age-ineligible respondents have a value of zero for the person weight.

Table 7-3: Person-level Poststratification Weights

Region	Sex	Age Group	1990 Census PUMS Estimate	1993 AHEAD Estimate	Person-level Poststratification Factor
Northeast	Male	70-79	1,266,345	1,228,442	1.0308
		80+	453,910	435,818	1.0415
	Female	70-79	1,917,535	1,761,819	1.0883
		80+	1,091,905	866,772	1.2597
Midwest	Male	70-79	1,422,225	1,622,431	0.8766
		80+	558,360	509,944	1.0949
	Female	70-79	2,066,210	2,199,795	0.9392
		80+	1,242,890	1,058,219	1.1745
South	Male	70-79	1,985,725	1,943,231	1.0218
		80+	742,690	821,629	0.9039
	Female	70-79	2,886,905	2,892,777	0.9979
		80+	1,561,055	1,450,482	1.0762
West	Male	70-79	1,129,365	1,105,306	1.0217
		80+	407,320	406,984	1.0006
	Female	70-79	1,498,310	1,394,932	1.0741
		80+	792,665	674,552	1.1750

Table 7-4 provides a national level comparison of the 1990 Census PUMS and 1993 AHEAD weighted estimates (before poststratification) of population by sex and age group. The ratio of the two estimates presented in the final column of Table 7-4 indicates close agreement in the two series of estimates except for the subpopulation of females age 80+ where the 1993 AHEAD estimate is only 86.4% of the corresponding 1990 Census PUMS total. The discrepancy might be explained by a number of factors including: 1) inclusion of institutional and group quarters populations in the PUMS estimates, and 2) AHEAD undercoverage of 80+ females in the household population. The possibility of undercoverage of 80+ females and other subpopulations is the subject of ongoing research into the coverage properties of the AHEAD dual-frame sample design (Rodgers, 1995).

Table 7-4: Comparison of AHEAD and PUMS Sex, Age Estimates of U.S. Population

Region	Sex	Age Group	1990 Census PUMS Estimate	1993 AHEAD Estimate	Ratio: 1993 AHEAD 1990 PUMS
U.S.	Male	70-79	5,830,660	5,899,410	1.0165
		80+	2,162,280	2,174,375	1.0056
	Female	70-79	8,368,960	8,249,323	0.9857
		80+	4,688,515	4,050,025	0.8638
Total U.S.		70-79	14,172,620	14,148,733	0.9983
		80+	6,850,795	6,224,400	1.9127
		Total	21,023,415	20,373,133	0.9691

8. Assets and Health Dynamics (AHEAD): Procedures for Sampling Error Estimation

This section focuses on sampling error estimation and construction of confidence intervals for survey estimates of descriptive statistics such as means, proportions, ratios, and coefficients for linear and logistic linear regression models.

8.A Overview of Sampling Error Analysis of AHEAD Sample Data

The AHEAD Survey is based on a stratified multi-stage probability sample of United States households. The AHEAD sample design is very similar in its basic structure to the multi-stage designs used for major federal survey programs such as the Health Interview Survey (HIS) or the Current Population Survey (CPS). The survey literature refers to the AHEAD, HIS and CPS samples as complex designs, a loosely-used term meant to denote the fact that the sample incorporates special design features such as stratification, clustering and differential selection probabilities (i.e., weighting) that analysts must consider in computing sampling errors for sample estimates of descriptive statistics and model parameters.

Standard analysis software systems such SAS and SPSS assume simple random sampling (SRS) or equivalent independence of observations in computing standard errors for sample estimates. In general, the SRS assumption results in underestimation of variances of survey estimates of descriptive statistics and model parameters. Confidence intervals based on computed variances that assume independence of observations will be biased (generally too narrow) and design-based inferences will be affected accordingly.

8.B Sampling Error Computation Methods and Programs

Over the past 50 years, advances in survey sampling theory have guided the development of a number of methods for correctly estimating variances from complex sample data sets. A number of sampling error programs which implement these complex sample variance estimation methods are available to AHEAD data analysts. The two most common approaches to the estimation of sampling error for complex sample data are through the use of a Taylor Series Linearization of the estimator (and corresponding approximation to its variance) or through the use of resampling variance estimation procedures such as Balanced Repeated Replication (BRR) or Jackknife Repeated Replication (JRR). New Bootstrap methods for variance estimation can also be included among the resampling approaches. [See Rao and Wu (1988).]

8.B.1 Linearization approach

If data are collected using a complex sample design with unequal size clusters, most statistics of interest will not be simple linear functions of the observed data. The objective of the linearization approach is to apply Taylor's method to derive an approximate form of the estimator that is linear in statistics for which variances and covariances can be directly estimated (Kish 1965; Woodruff, 1971).

Most univariate, descriptive analysis of survey data including the estimation of means and proportions involves the use of the combined ratio estimator:

$$\hat{r} = \sum_{1}^{n} w_i y_i \; / \sum_{i=1}^{n} w_i \; x_i = y / x$$

where:

r = the sample estimate of the ratio of population totals R = y/x;
y_i, x_i = variables for observation *i* (x_i = 1 for mean);
w_i = weight for observation *i*;
y, x = weighted sample totals for the variables y, x.

The linearized approximation to the variance of the combined ratio estimator is (see Kish and Hess, 1959)

$$var(\hat{r}) \_ \frac{1}{x^2} \left[var(y) + r^2 var(x) - 2r \bullet cov(y,x) \right]$$

Similarly, linearized variance approximations are derived for estimators of finite population regression coefficients, correlation coefficients and the coefficients of logistic regression models (Kish and Frankel, 1974). In these programs, an iteratively reweighted least squares algorithm is used to compute maximum likelihood estimates of model parameters. At each step of the model fitting algorithm, a Taylor Series linearization approach is used to compute the variance/covariance matrix for the current iteration's parameter estimates (Binder, 1983).

Available sampling error computation software that utilizes the Taylor Series linearization method includes: SUDAAN and PC SUDAAN, SUPERCARP and PC CARP, CLUSTERS, OSIRIS PSALMS, OSIRIS PSRATIO, and OSIRIS PSTABLES. PC SUDAAN and PC CARP include procedures for estimation of sampling error both for descriptive statistics (means, proportion, totals) and for parameters of commonly used multivariate models (least squares regression, logistic regression).

8.B.2 Resampling Approaches

In the mid-1940s, P.C. Mahalanobis (1946) outlined a simple replicated procedure for selecting probability samples that permits simple, unbiased estimation of variances. The practical difficulty with the simple replicated approach to design and variance estimation is that many replicates are needed to achieve stability of the variance estimator. Unfortunately, a design with many independent replicates must utilize a coarser stratification than alternative designs -- to achieve stable variance estimates, sample precision must be sacrificed. Balanced Repeated Replication (BRR), Jackknife Repeated Replication (JRR) and the Bootstrap are alternative replication techniques that may be used for estimating sampling errors for statistics based on complex sample data.

The BRR method is applicable to stratified designs in which two half-sample units (i.e., PSUs) are selected from each design stratum. The conventional "two PSU-per-stratum" design in the best theoretical example of such a design although in practice, collapsing of strata (Kalton, 1977) and random combination of units within strata are employed to restructure a sample design for

BRR variance estimation. The half-sample codes prepared for the HRS Wave 1 data set require the collapsing of nonself-representing strata and the randomized combination of selection units within self-representing (SR) strata. When full balancing of the half-sample assignments is employed (Wolter, 1985), BRR is the most computationally efficient of the replicated variance estimation techniques. The number of general purpose BRR sampling error estimation programs in the public domain is limited. The OSIRIS REPERR program includes the option for BRR estimation of sampling errors for least squares regression coefficients and correlation statistics. Research organizations such as Westat, Inc., and the National Center for Health Statistics have developed general purpose programs for BRR estimation of standard errors. Another option is to use SAS or SPSS macro facilities to implement the relatively simple BRR algorithm. The necessary computation formulas and Hadamard matrices to define the half-sample replicates are available in Wolter (1985).

With improvements in computational flexibility and speed, jackknife (JRR) and bootstrap methods for sampling error estimation and inference have become more common (J.N.K. Rao & Wu, 1988). Few general purpose programs for jackknife estimation of variances are available to analysts. OSIRIS REPERR has a JRR module for estimation of standard errors for regression and correlation statistics. Other stand-alone programs may also be available in the general survey research community. Like BRR, the algorithm for JRR is relatively easy to program using SAS, SPSS or S-Plus macro facilities.

BRR and JRR are variance estimation techniques, each designed to minimize the number of "resamplings" needed to compute the variance estimate. In theory, the bootstrap is not simply a tool for variance estimation but an approach to actual inference for statistics. In practice, the bootstrap is implemented by resampling (with replacement) from the observed sample units. To ensure that the full complexity of the design is reflected, the selection of each bootstrap reflects the full complexity of the stratification, clustering and weighting that is present in the original sample design. A large number of bootstrap samples are selected and the statistic of interest is computed for each. The empirical distribution of the estimate that results from the large set of bootstrap samples can then be used to obtain a variance estimate and a support interval for inference about the population statistic of interest.

In most practical survey analysis problems, the JRR and Bootstrap methods should yield similar results. Most survey analysts should choose JRR due to its computational efficiency. HRS and AHEAD data analysts interested in the bootstrap technique are referred to LePage and Billard (1992) for additional reading and a bibliography for the general literature on this topic.

One aspect of BRR, JRR and bootstrap variance estimation that is often pushed aside in practice is the treatment of analysis weights. In theory, when a resampling occurs (i.e., a BRR half sample is formed), the analysis weights should be recomputed based only on the selection probabilities, nonresponse characteristics and poststratification outcomes for the units included in the resample. This is the correct way of performing resampling variance estimation; however, in practice acceptable estimates can be obtained through use of the weights as they are provided on the public use data set.

8.C Sampling Error Computation Models

Regardless of whether linearization or a resampling approach is used, estimation of variances for

complex sample survey estimates requires the specification of a *sampling error computation model.* AHEAD data analysts who are interested in performing sampling error computations should be aware that the estimation programs identified in the preceding section assume a specific sampling error computation model and will require special sampling error codes. Individual records in the analysis data set might be assigned sampling error codes which identify to the programs the complex structure of the sample (stratification, clustering) and are compatible with the computation algorithms of the various programs. To facilitate the computation of sampling error for statistics based on HRS and AHEAD data, design-specific sampling error codes will be routinely included in all public-use versions of the data set. Although minor recoding may be required to conform to the input requirements of the individual programs,the sampling error codes that are provided should enable analysts to conduct either Taylor Series or Replicated estimation of sampling errors for survey statistics.

Table 8.1 defines the sampling error coding system for AHEAD sample cases. Two sampling error code variables are defined for each case based on the sample design PSU and SSU in which the sample household is located.

SESTRAT - The sampling error computation stratum code. SESTRAT is the variable which defines the sampling error computation strata for all sampling error analysis of the AHEAD data. With the exception of the New York, Los Angeles and Chicago MSAs, each self-representing (SR) design stratum is represented by one sampling error computation stratum. Due to their population size, two sampling error computation strata are defined for each of the three largest MSAs. Pairs of similar nonself-representing (NSR) primary stage design strata are "collapsed" (Kalton, 1977) to create NSR sampling error computation strata.

Controlled selection and a "one-per-stratum" design allocation are used to select the primary stage of the HRS/AHEAD national sample. The purpose in using controlled selection and the "one-per-stratum" sample allocation is to reduce the between-PSU component of sampling variation relative to a "two-per-stratum" primary stage design. Despite the expected improvement in sample precision, a drawback of the "one-per-stratum" design is that two or more sample selection strata must be collapsed or combined to form a sampling error computation stratum. Variances are then estimated under the assumption that a multiple PSU per stratum design was actually used for primary stage selection. The expected consequence of collapsing design strata into sampling error computation strata is the overestimation of the true sampling error; that is, the sampling error computation model defined by the codes contained in Table 8.1 will yield estimates of sampling errors which in expectation will be slightly greater than the true sampling error of the statistic of interest.

HALFSAM - Stratum-specific half sample code for analysis of sampling error using the BRR method or approximate "two-per-stratum" Taylor Series method (Kish and Hess, 1959). Within the self-representing sampling error strata, the half sample units are created by dividing sample cases into random halves, HALFSAM=1 and HALFSAM=2. The assignment of cases to half-samples is designed to preserve the stratification and second stage clustering properties of the sample within an SR stratum. Sample cases are assigned to half samples based on the SSU in which they were selected. For this assignment, sample cases were placed in original stratification order (SSU number order) and beginning with a random start entire SSU clusters were systematically assigned to either HALFSAM=1 or HALFSAM=2.

In the general case of nonself-representing (NSR) strata, the half sample units are defined according to the PSU to which the respondent was assigned at sample selection. That is, the half samples for each NSR sampling error computation stratum bear a one-to-one correspondence to the sample design NSR PSUs. The particular sample coding provided on the AHEAD public use data set is consistent with the "ultimate cluster" approach to complex sample variance estimation (Kish, 1965; Kalton, 1977). Individual stratum, PSU and SSU variables may be needed by AHEAD analysts interested in components of variance analysis or estimation of hierarchical models in which PSU-level and neighborhood-level effects are explicitly estimated.

Table 8-1: AHEAD Sampling Error Codes for HRS Wave 1

Sampling Error Codes		Number of SSUs with 1 or more AHEAD Rs
SE Stratum	**Half-Sample Code**	
1	1	15
	2	16
2	1	16
	2	16
3	1	16,2
	2	15
4	1	16
	2	16
5	1	13
	2	14
6	1	14
	2	13
7	1	17
	2	17
8	1	17
	2	16
9	1	12
	2	12
10	1	13
	2	12
11	1	11
	2	11

Table 8-1 (cont.): AHEAD Sampling Error Codes for HRS Wave 1

Sampling Error Codes		Number of SSUs with 1 or more AHEAD Rs
SE Stratum	**Half-Sample Code**	
12	1	15
	2	15
13	1	9
	2	9
14	1	7
	2	8
15	1	8
	2	8
16	1	8
	2	8
17	1	10
	2	9
18	1	8
	2	8
19	1	9
	2	9
20	1	13
	2	12
21	1	12
	2	11
22	1	8
	2	9
23	1	6
	2	5

Table 8-1 (cont.): AHEAD Sampling Error Codes for HRS Wave 1

Sampling Error Codes		Number of SSUs with 1 or more AHEAD Rs
SE Stratum	**Half-Sample Code**	
24	1	6
	2	6
25	1	8
	2	7
26	1	25
	2	26
27	1	26
	2	27
28	1	21
	2	27
29	1	27
	2	24
30	1	26
	2	21
31	1	24
	2	25
32	1	24
	2	18
33	1	21
	2	20
34	1	21
	2	19
35	1	7
	2	9

Table 8-1 (cont.): AHEAD Sampling Error Codes for HRS Wave 1

Sampling Error Codes		Number of SSUs with 1 or more AHEAD Rs
SE Stratum	Half-Sample Code	
36	1	26
	2	5
37	1	19
	2	20
38	1	11
	2	12
39	1	22
	2	22
40	1	22
	2	25
41	1	24
	2	22
42	1	12
	2	12
43	1	12
	2	12
44	1	12
	2	12
45	1	8
	2	8
46	1	15
	2	18
47	1	12
	2	9

.

Table 8-1 (cont.): AHEAD Sampling Error Codes for HRS Wave 1

Sampling Error Codes		Number of SSUs with 1 or more AHEAD Rs
SE Stratum	**Half-Sample Code**	
48	1	12
	2	15
49	1	12
	2	11
50	1	12
	2	12
51	1	6
	2	3
52	1	6
	2	5
53	1	2,3,2
	2	4,4,2

Appendix A: Costs and Errors of the Two Sample Frame Alternatives

The AHEAD project provides an opportunity to directly compare the properties of the HCFA EDB list and area probability sample frames for studies of the oldest old. Coverage of the survey population is one obvious point of comparison. This appendix provides an overview of the frame coverage topic and a proposed methodology for the AHEAD coverage study. More generally, the appendix also contrasts the two sampling frames using other important measures of survey errors (sampling variability, response rates) and survey costs (specifically, the time and travel required to locate and interview the designated respondent).

A.1 Sample Coverage

The target population for AHEAD includes all persons born prior to 1924 who reside in U.S. households in 1993. The dual-frame design is used in AHEAD only for sampling of households with one or more persons born before 1914 (80+ in 1993). The HCFA Enrollment Data Base (EDB) file is designed to include each individual currently enrolled for Medicare benefits. Apart from the special situations (e.g., disabled; renal failure; certain persons retired from federal, state, or local government or railroad retirees; certain family members such as a divorced spouse or widow), this frame contains enrollees who are ages 65 and over. This includes everyone currently receiving Social Security retirement benefits (i.e., has a Social Security number, and has worked at least 40 quarters of Social Security "covered" labor, or meets some special eligibility criteria).

Persons age 65 or over without SSNs do not appear in the EDB file data base unless they are enrolled through an eligible family member or through a special set of circumstances. While their number is probably small, excluded persons include subsets of the following groups: those of foreign nativity, former federal, state and local government employees, former military career service personnel, those (mostly female) who were never employed or worked a very short period (and do not have an eligible spouse). It is worth noting that some of the very old are eligible for Medicare through legislation "grandfathering" those without Social Security benefits. HCFA EDB file users report that the quality of the EDB file deteriorates with an increase in age of the population under study. Users of the EDB file speculate that undercoverage of the oldest old may not be negligible, reflecting the fact that (unlike younger cohorts) many of the oldest old **never** worked in jobs covered by Social Security, including those at the high and low extremes of socio-economic levels, as well as many immigrants and illegal aliens.

The strengths of coverage provided by the AP sample screening speak to the coverage gaps of the HCFA EDB list, and vice versa. The HCFA EDB frame excludes some individuals age 65 and older who do not have Social Security numbers or who worked a short period of time and are not covered by Medicare in other ways. The AP frame includes such individuals since all household members are listed in the screening process regardless of whether or not they have SSNs or were employed.

On the other hand, the AP screening process relies on a thorough, accurate reporting of household membership. Erroneous exclusions cannot be avoided, so noncoverage will occur. The noncoverage rate is expected to be small. Such response error at the time of screening can occur because the household informant forgets to include an elderly household members, intentionally excludes an eligible elderly member, or erroneously reports age in a way that

makes him or her appear ineligible. The HCFA EDB frame avoids this source of noncoverage. Note that the elderly have an incentive to be included on the EDB file (although not for survey participation); no such incentive exists for being included in a household listing for the HRS.

Rodgers (1995) provides a detailed comparison of the AHEAD survey outcomes and response distributions for the two frames.

A.2 Sampling Variation

Less critical to the comparison of total survey error for the two sample frames is the sampling variability (design effects) of the survey data that is collected in a cost effective way. The samples from the AP and HCFA EDB frames are both stratified, multi-stage designs and share design linkages at the primary and secondary stages. In such design, the final stage samples of households and respondents will be clustered to some degree -- the HCFA sample slightly less so than the AP sample.

A.3 Nonresponse Error

A source of nonsampling error in the AHEAD will be screening and interview nonresponse. In the AP sample screening, only 1.5% of the sample households were not successfully screened due to outright refusals or failure to contact a household member during the survey process. The screening nonresponse/no contact rate for the HCFA EDB frame sample of Group 2 households was 3%. The difference in screening/contact rates for the two frames can in large part be attributed to the quality of tracking information collected for AHEAD-eligible persons during the original HRS screening process. (See Section 3.A.)

A second form of nonresponse occurs at the stage where the AHEAD-eligible household is contacted for the actual interview. In AHEAD Wave 1, the unweighted response rates for Group 2 households were 83.0% for AP frame sample and 74.1% for HCFA EDB frame household financial units. (See Section 6.B.)

One advantage provided by the EDB file is the availability of auxiliary information for persons who do not respond to the AHEAD survey. Both the HCFA EDB file and HRS AP screening data provide information such as age, sex and geographic location of the nonrespondent. (This ignores the .3% screening nonresponse incurred at the time the HRS household rosters were collected.) Additionally, the HCFA EDB file provides race and beneficiary status. Although this is not available for the AP frame, the HRS screening data base includes information not available from the HCFA frame: complete household composition, including the age, sex, and relationship of all household members.

Both sets of information will prove useful in investigating nonresponse bias for the AHEAD survey.

A.4 Survey Cost Comparison

The choice of sample frame influenced AHEAD data collection costs in the following two ways:

i. Screening costs to identify eligible respondents covered by the sampling frame; and
ii. Contact costs to locate and interview eligible sample respondents at the baseline and subsequent waves.

A.4.1 Screening Costs

Setting aside the eventual need to update the panel, screening of sample elements for eligible AHEAD respondents is a one-time cost factor. In the absence of the special opportunity presented by the HRS AP sample screening, the HCFA EDB file would have a tremendous cost advantage over a one-time screening of a large area probability sample of households for AHEAD. Current estimates based on 1990 Census PUMS data suggest that approximately 17.8% of sample households in an equal probability national sample would contain one or more persons who would be eligible for AHEAD. The costs of obtaining a similar sample of persons age 70 and older from the HCFA EDB file frame are much smaller since the age of each individual listed on the frame element is known -- a sample of eligible persons can be selected directly from the EDB file without the need for large-scale face-to-face screening of sample households.

However, since a large scale screening of area probability sample households was required to identify the HRS sample of the 51 to 61 year-old study population, the concurrent identification of an area probability sample of persons 70 and older introduced only a modest marginal cost to collect the necessary recontact and tracking information. Therefore, the cost differential ordinarily associated with the two sample frame alternatives was leveled considerably.

It is obvious that the AHEAD dual-frame methodological study does not provide a true comparison of the HCFA list frame survey cost to the full true cost of a survey that is based on area probability sample household screening. It is more appropriate to view the comparison as one which looks at the survey costs of the HCFA list frame sample vs. a prescreened area probability sample. In its own right, this is a very valuable comparison to make. Federal survey programs such as the Health Interview Survey can use large area probability samples of households as a means to identify probability samples for special follow-up studies of rare subpopulations (e.g., pregnant women, arthritis patients, the oldest old). If the rare population of interest is the oldest old, the AHEAD methodological study should provide results which will inform the choice between the pre-screened area probability sample and the HCFA list sample alternative.

A.4.2 Contact Costs

To locate and interview eligible AHEAD respondents there were significant expenditures to: 1) locate correct respondent addresses and phone numbers; 2) track respondents who have moved from a known address; and 3) pay interviewer travel costs to conduct face-to-face interviews in respondent households.

A principal advantage to an AHEAD sample that was identified concurrently with the HRS household screening lies in the respondent information available to the field staff. The following information was collected for HRS sample households with AHEAD age-eligible members regardless of whether or not there was an HRS respondent in the household:

- full address of the household
- mailing address (if different)
- names, gender, ages and relationship (to informant) of all household members
- telephone number of the prospective AHEAD respondent(s)
- name, address, telephone number and relationship of two contact persons for the AHEAD respondent (for use in tracking).

This contrasts sharply with the simple name and address information provided by the HCFA EDB file. The AP frame provides more efficient respondent contact opportunities because of the

detailed tracking information which was collected as the HRS household sample was screened.

Moreover, there were several other cost efficiencies associated with the use of HRS screening. Many of the same interviewers employed for HRS were employed in the AHEAD Wave 1 data collection. Since interviewers returned to SSUs and housing units previously visited in the HRS, they were familiar with the area, had to deal less with gatekeepers (dealt with during HRS), had established ties with appropriate authorities (police, sheriff), community leaders and groups, etc. The SRC interviewer may have previously established rapport with the respondent. Prior establishment of rapport allowed an efficient "first call" by telephone. To maintain contact with eligible AHEAD respondents, a mailing was sent about nine months after HRS screening to update addresses, and a letter of introduction was sent prior to the call for the first AHEAD interview.

For EDB file sample individuals with residential addresses, directory assistance or various commercial services were used to obtain a telephone number. If a telephone number was obtained, SRC interviewers called the respondent to make an appointment for the AHEAD interview. If no phone number was found, the initial contact was in person.

Another advantage of the HRS-based AP frame is the availability of up to two contact persons when the respondent had moved. No such information was available from the EDB file.

Geographic clustering of sample respondents leads to reduced costs for interviewer travel. A sample of the oldest old identified through the HRS AP household screening retains the clustered sample properties of the multi-stage area probability sample design. However, since the prevalence of oldest old in the population is rather low, the degree of clustering was modest, averaging 4-6 eligible households per SSU selection. The sample of Medicare enrollees from the HCFA EDB frame was clustered by ZIP Code prior to subselection and contact for the AHEAD survey. The ZIP Code units are much larger geographic units than the SSUs of the HRS AP sample design. Consequently, the clustering of sample respondents identified through the screening of HRS AP sample households led to slightly lower per-unit interviewer travel costs.

A.5 Theoretical Model of Frame Coverage

Figure A-1 schematically illustrates the coverages of the AP and HCFA EDB list frames and how they relate to the population of inference. As the figure shows, the intersection of the two frames with the target population produces five zones of coverage interest:

A. Noncoverage by both frames;
B. Coverage by the AP frame only;
C. Coverage by both frames;
D. Coverage by the HCFA EDB file only;
E. Coverage of elements outside the target population.

The HRS-based AP and HCFA EDB file list frames are both imperfect, and any comparison of the two presents the problem of the individual who wears two watches. The two frames can only be compared for consistency (inconsistency). Neither source or even their relative difference can inform us precisely of the true status of their individual or joint coverage of the target

population.

The underlying theoretical model with its defined coverage zones and corresponding populations is not complex. The complexity and complications enter when real survey processes and actual populations form the basis for quantifying the important properties of the model. For the AHEAD dual-frame design, the important properties to be quantified are: 1) the distribution of the total population to the defined coverage zones; and 2) conditional on this distribution, the nature of potential bias implied by the observed noncoverage of the component frames. From the theoretical standpoint, this could be accomplished through a complete element by element comparison of the two frames to each other and to the known universe of elements in the target population. The three-way comparison would assign each element to a unique zone in the coverage model diagram. That done, it is straight-forward to establish the contribution of each zone to a full and unbiased representation of the population of inference.

A.6 Practical Problems

In practice, there are several barriers to applying this theoretically simple process to the AHEAD dual-frame design:

1) The elements of both frames will not be completely enumerated. By its nature the multi-stage area probability frame does not provide us a complete enumeration or "listing" of the population elements which fall within its coverage domain. The HCFA EDB file is by definition a complete enumerative list of its domain.

   The consequence of the non-enumerative nature of the AP frame is that the size and population composition of the area probability frame population not covered by HCFA cannot be directly measured in this methodological investigation -- Zone B in Figure A-1. "Indirect" methods based on comparisons of distributions of frame-specific sample estimates provide an alternative, less precise means of examining area probability frame noncoverage.

   In effect, we are in the same position as researchers who conduct a post-Censal survey to estimate decennial Census undercount. The identity of sample elements can be "mapped" onto the enumerative census to determine which elements were not covered by the complete enumeration attempt, but the reverse "mapping" of census to sample has no practical utility for estimating the coverage of the sample procedure.

2) Standard or preferred procedures for drawing samples from the HCFA EDB file limit the first phase sample to at most 5% of data base entries. Typically this first phase 5% sampling is performed by a systematic sampling pass through the EDB file data base. HCFA granted the AHEAD study a special exception to this standard procedure. The protocol for HCFA frame access included a mechanism that permits the exact match of eligible individuals identified through the AHEAD area probability sample to the full enumerative EDB file data base for sample PSUs.

   Within PSUs, exact matching procedures can be used to "map" the area sample individuals onto the EDB file data file, and the evaluation of HCFA EDB file list frame coverage is greatly enhanced. If AHEAD had been able to access only a 5% sample of

the HCFA EDB file data base, the coverage analysis would have been limited to the less informative indirect method of comparing survey estimates of the marginal distributions of population characteristics.

Figure A-1 here

A.7 Direct vs. Indirect Evaluation of Frame Coverage

"Direct" and "indirect" statistical procedures may be used to evaluate the coverage of the two frames. Both types of procedures will be used to analyze the coverage properties of the AHEAD dual-frame sample design.

The **direct method** involves the exact match of each sample element from a frame to the alternative frame's enumerative list of its covered survey population. Based on the exact match, each sample element would be assigned a dichotomous indicator variable:

$$y_i = \begin{cases} 1 \text{ if HRS sample element is present on the HCFA EDB list frame;} \\ \\ 0 \text{ if the HRS sample element does not have a match.} \end{cases}$$

In this methodological investigation, the direct method can only be used to investigate which elements of the AP sample are covered/not covered by the HCFA EDB file list (Zones B and C of Figure A-1). Since the direct method established coverage/noncoverage for each individual case, logistic or probit models can then be used to analyze noncoverage as a function of a wide array of individual covariates that are measured in the course of the AHEAD survey interview. Covariates of potential interest include not only demographic variables but also other survey measures such as housing type, neighborhood characteristics, financial and health status of the eligible person, etc. If the exact match can be performed in a valid way, this is much more powerful form of analysis than is available under the indirect method of analyzing coverage. The disadvantage of the direct method is that its validity is no better than the validity of the exact match. Failure to identify a true match will cause a case to appear to be not covered. A false-positive match will make coverage appear better than it really is.

The direct analysis method described in the preceding paragraph has a shortcoming in that it provides no information about the noncoverage of the HRS AP frame relative to HCFA EDB file list (Zone D of Figure A-1). There is a partial solution to this problem if we can successfully geocode housing units addresses on the EDB file. Using existing geocoding on the EDB file (state, county, ZIP Codes), Geographic Information System (GIS) software could be used to assign 1990 Census tract and block codes to a substantial proportion of the HCFA EDB file addresses. Knowing the actual Census tracts and blocks, it is then possible to identify EDB file addresses on street segments that are included in the national area probability sample of area segments.[5] Further address matching within area segments would allow us to determine which household addresses in the EDB file were screened for the HRS and found eligible for AHEAD (and which were screened and found not eligible).

The **indirect** method for coverage analysis is the comparison of estimates of variable distributions computed from the independent samples from the HRS-based AP and HCFA EDB file frames -- e.g., proportions of respondents by age, race, or sex, proportion with a spouse, living alone, living with others. The advantage to the indirect method is that only probability

[5]Area segment is the term used by SRC for the second stage sampling units (SSUs) of its multi-stage sample design. The 1990 National Sample area segments are defined using Census tract and block boundaries. Typically, an area segment includes from one to five Census blocks and an average of approximately 100 housing units.

samples from the two frames are needed for the comparison; no exact matching of frame elements or matching of samples to alternate frames is needed. For complete samples of observations the indirect method is unbiased. There are two major disadvantages to the indirect method. The first is the power to detect any true differences between frames. The small amount of existing evidence suggests that noncoverage of both the HCFA EDB file and HRS-based area probability sample frames is probably small. This being the case, it is not reasonable to expect large true differences in the coverage of two frames. In the absence of the matching or statistical control for covariate factors, indirect analysis based on comparisons of estimates of univariate distributional sample statistics from the two frames may only be capable of detecting large differences between the two frames. Secondly, this indirect method of frame comparison is unbiased only if the complete samples are observed. In practice, each independent sample will be subject to nonresponse. The comparison of sample estimates is therefore one which confounds both true frame noncoverage with any additional bias arising from nonresponse.

A.8 Summary

Statistical sampling texts identify frame noncoverage of the survey population as an important potential source of survey error; yet the literature contains few published descriptions of empirical studies of this topic. One reason that empirical studies of frame coverage are so rare is that they are both costly and complex to design, execute, and analyze. Our best examples of both the costs and complexity are the post-Censal surveys which attempt to quantify decennial Census undercount.

The AHEAD project provides a special opportunity to compare the coverage properties of HCFA EDB file list samples and area probability samples of the oldest old members of the U.S. household population. The proposed sample sizes and analysis methods may not provide sufficient power to detect small true differences in the coverages of the two frames. However, if the coverage differences are truly small, their precise quantification may be of minor importance, particularly in the presence of major sources of survey error such as unit nonresponse. Certainly, the study will provide the ability to quantify large differences in frame coverage and a qualitative comparison of the procedures and costs of conducting surveys with samples from the HRS-based area probability and HCFA EDB file list frames.

References

Binder, D.A. (1983). "On the variances of asymptotically normal estimators from complex surveys," *International Statistical Review*, Vol. 51, pp. 279-292.

Heeringa, S., & Connor, J. (1995). *Technical Description of Health and Retirement Study Sample Design*. Ann Arbor: Institute for Social Research.

Kalton, G. (1977). "Practical methods for estimating survey sampling errors," *Bulletin of the International Statistical Institute*, Vol 47, 3, pp. 495-514.

Kish, L. (1965). *Survey Sampling.* John Wiley & Sons, Inc., New York.

Kish, L., & Frankel, M. R. (1974). "Inference from complex samples," *Journal of the Royal Statistical Society*, B, Vol. 36, pp. 1-37.

Kish, L., & Hess, I. (1959). "On variances of ratios and their differences in multi-stage samples," *Journal of the American Statistical Association*, 54, pp. 416-46.

LePage, R., & Billard, L. (1992). *Exploring the Limits of Bootstrap*. John Wiley & Sons, Inc., New York.

Mahalanobis, P.C. (1946). "Recent experiments in statistical sampling at the Indian Statistical Institute," *Journal of the Royal Statistical Society*, Vol 109, pp. 325-378.

Rao, J.N.K., & Wu, C.F.J. (1988). "Resampling inference with complex sample data," *Journal of the American Statistical Association*, 83, pp. 231-239.

Rodgers, W.L. (1995). "Comparison of two sampling frames for surveys of the oldest old." Ann Arbor: Survey Research Center, The University of Michigan. Manuscript in preparation.

Waldo, D.R., & H.C. Lazerby (1984). "Demographic characteristics and health care use and expenditures by the aged in the United States: 1977-1984," *Health Care Financing Review*, Vol. 6, pp. 1-29.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

Woodruff, R.S. (1971). "A simple method for approximating the variance of a complicated estimate," *Journal of the American Statistical Association*, Vol. 66, pp. 411-414.