TECHNICAL REPORT AND USER'S GUIDE

 FOR THE

 NEW BENEFICIARY SURVEY (NBS)

 Submitted in response to:

 Phase II, Task 7 of SSA Contract #600‑81‑0080

 Submitted by:

 Institute for Survey Research

 Temple University

 Philadelphia, PA 19122

 Submitted to:

 Social Security Administration

 Baltimore, MD 21241

 January, 1986

 FOREWORD

 Tables referenced in text appear at the end of the report.  Due to the nature of the media involved in producing this document, formulae appearing in Sections 3, 6, 7, and 10 have been modified as follows:

 Meaning:  Appears as:

‑ square root SQRT:, as in SQRT:[4] = 2

‑ summation (usually indicated SUM:, as in SUM:[xi] = 5

 by the Greek letter sigma) where x1 = 2 and x2 = 3

‑ probability P:, as in P:head = .5

‑ exponentiation EXP:, as in EXP:[4+6] = e10

 TABLE OF CONTENTS

 Page No.

 TABLE OF TABLES . . . . . . . . . . . . . . . . . . . 6

1.  INTRODUCTION . . . . . . . . . . . . . . . . . . . . .  9

2.  THE NBS UNIVERSE . . . . . . . . . . . . . . . . . . .  11

 2.1.  The MBR Listing . . . . . . . . . . . . . . . .  11

 2.2.  Beneficiary Categories . . . . . . . . . . . . 11

 2.2.1.  Retired Workers . . . . . . . . . . . .  12

 2.2.2.  Disabled Workers . . . . . . . . . . . .  12

 2.2.3.  Spouses . . . . . . . . . . . . . . . .  12

 2.3.  The Nonretired Category . . . . . . . . . . . .  13

 2.4.  Individuals versus Couples . . . . . . . . . . 13

 2.5.  Cleaning the MBR Listing . . . . . . . . . . . 14

3.  SAMPLING . . . . . . . . . . . . . . . . . . . . . . .  16

 3.1.  Sample Size . . . . . . . . . . . . . . . . . .  16

 3.2.  Identification of the Sampling Frame and

 Selection Strategy . . . . . . . . . . . . .  17

 3.3.  Defining and Stratifying the Original Primary

 Sampling Units . . . . . . . . . . . . . . .  18

 3.4.  Selecting Primary Original Sampling Units . . .  19

 3.5.  Modification of the ISR/MPR Primary Sampling

 Units for the NBS . . . . . . . . . . . . . .  19

 3.5.1.  Reducing the Sizes of Some Primary

 Sampling Units . . . . . . . . . . . .  19

 3.5.2.  The Redefinition of Primary Sampling

 Unit Boundaries . . . . . . . . . . .  20

 3.5.3.  The Creation and Selection of Primary

 Sampling Units in Self‑Representing

 Areas . . . . . . . . . . . . . . . .  21

 3.6.  Selection of Respondents Within Primary Sampling

 Units for the NBS . . . . . . . . . . . . . .  22

 3.7.  Sample Results . . . . . . . . . . . . . . . .  23

4.  RESPONSE RATES . . . . . . . . . . . . . . . . . . . .  25

5.  DATA PROCESSING . . . . . . . . . . . . . . . . . . .  26

 5.1.  ISR/MPR Roles . . . . . . . . . . . . . . . . .  26

 5.2.  Check‑In Procedures . . . . . . . . . . . . . .  26

 5.3.  Edit Procedures . . . . . . . . . . . . . . . .  27

 Page No.

 5.4.  Coding/Check‑Coding Procedures . . . . . . . .  27

 5.5.  Key Entry . . . . . . . . . . . . . . . . . . .  29

 5.6.  Data Cleaning . . . . . . . . . . . . . . . . .  29

 5.7.  Tape Construction . . . . . . . . . . . . . . .  30

6.  WEIGHTING THE NBS DATA . . . . . . . . . . . . . . . .  31

 6.1.  Introduction . . . . . . . . . . . . . . . . .  31

 6.2.  Sampling and Post‑Stratification Weights . . .  31

 6.3.  Nonresponse Adjustment Weights . . . . . . . .  32

 6.4.  Generalizability to the NBS Sample Universe . .  33

7.  SAMPLING ERRORS FOR THE NBS . . . . . . . . . . . . .  34

 7.1.  Introduction . . . . . . . . . . . . . . . . .  34

 7.2.  Algorithms for Sampling Error Calculations . .  34

 7.2.1.  Proportions, Contrasts and Totals . . .  34

 7.2.2.  Means, Medians and Quartiles . . . . . .  36

 7.2.3.  Design Effects . . . . . . . . . . . . .  37

 7.3.  Generalized Sampling Errors . . . . . . . . . .  37

 7.3.1.  Sampling Errors for Proportions . . . .  37

 7.3.2.  Sampling Errors for Contrasting

 Proportions  . . . . . . . . . . . . .   42

 7.3.3.  Sampling Errors for Totals . . . . . . .  45

 7.4.  BRR Half‑Sample Indicators . . . . . . . . . .  45

8.  THE QUESTIONNAIRE  . . . . . . . . . . . . . . . . . .  48

 8.1.  Development . . . . . . . . . . . . . . . . . .  48

 8.2.  Organization and Format . . . . . . . . . . . .  48

 8.2.1.  Household Composition . . . . . . . . .  48

 8.2.2.  Employment History . . . . . . . . . . .  48

 8.2.3.  Employment and Pension Detail . . . . . 49

 8.2.4.  Noncovered Employment . . . . . . . . .  50

 8.2.5.  Health . . . . . . . . . . . . . . . . .  51

 8.2.6.  Income and Assets . . . . . . . . . . .  51

 8.2.7.  Marital History . . . . . . . . . . . .  52

 8.2.8.  Child Care . . . . . . . . . . . . . . .  52

 8.2.9.  Program Knowledge . . . . . . . . . . .  52

 8.2.10. Spouse . . . . . . . . . . . . . . . . .  52

9.  FIELDING PROCEDURES . . . . . . . . . . . . . . . . .  54

 9.1.  Interviewer Recruitment and Training . . . . .  54

 9.2.  Callback Procedures and Refusal Conversion . .  55

 Page No.

 9.3.  Quality Control . . . . . . . . . . . . . . . .  55

 9.3.1.  Editing and Critical Items . . . . . . .  55

 9.3.2.  Validation . . . . . . . . . . . . . . .  56

10.  IMPUTATION IN THE NEW BENEFICIARY SURVEY . . . . . . 57

 10.1.  Introduction . . . . . . . . . . . . . . . . .  57

 10.2.  Imputation Methodology in the NBS . . . . . . 58

 10.2.1.  Imputation of Earnings . . . . . . . .  58

 10.2.2.  Imputation of Income Types . . . . . .  61

 10.2.2.1.  Imputation of Income Type

 Receipt . . . . . . . . . .  61

 10.2.2.2.  Imputation of Income Type

 Amounts . . . . . . . . . .  62

 10.2.3.  Imputation of Holdings, Amounts and

 Incomes from Assets . . . . . . . . . .  64

 10.2.3.1. Imputation of Asset Holdings . . 64

 10.2.3.2. Imputation of Asset Amounts . . 66

 10.2.3.3. Imputation of Income Flows . . . 69

 10.2.4. Imputation of Expected Future Pension and

 Miscellaneous Income Items . . . . . . 70

 10.2.4.1. Imputation of Expected Future

 Pension . . . . . . . . . . . 70

 10.2.4.2. Imputation of Miscellaneous

 Income Receipt Flags . . . . . 70

 10.2.4.3. Imputation of Miscellaneous

 Income Amounts . . . . . . . . 71

 10.3. Estimating Net Rent . . . . . . . . . . . . . . 72

TABLES . . . . . . . . . . . . . . . . . . . . . . . . . . . 73

 TABLE OF TABLES

 Page No.

Table 2.1. New Beneficiary Survey Universe and

 Completed Interviews . . . . . . . . . . . . . 73

Table 3.1. Sampling Rates for Categories of Respondents . . 74

Table 3.2. Final Disposition of the New Beneficiary Survey

 by Sample Subdomain . . . . . . . . . . . . . 75

Table 4.1. Response Rates for the New Beneficiary Survey

 by Sample Subdomain . . . . . . . . . . . . . 76

Table 6.1. Basic and Post-Stratification Weights in the

 New Beneficiary Survey . . . . . . . . . . . . 77

Table 6.2. Nonresponse Weighting Cells Employed in

 Calculating Final and Interim Weights

 in the New Beneficiary Survey . . . . . . . . 78

Table 7.1. Variates Employed to Calculate Sampling Errors of

 Percentages for the New Beneficiary Survey by

 Group . . . . . . . . . . . . . . . . . . . . 84

Table 7.2. Subgroups for Which Sampling Errors of Percentages

 Were Generated for the New Beneficiary Survey. 87

Table 7.3. Average DEFTs by Subgroup and Variable Group for

 the New Beneficiary Survey . . . . . . . . . . 88

Table 7.4. Unweighted Sample Sizes by Subgroup and Variable

 Group for the New Beneficiary Survey . . . . . 89

Table 7.5. Weighted Totals Within Subgroups by Variable Group

 for the New Beneficiary Survey . . . . . . . . 90

Table 7.6. Conservative Generalized Sampling Errors of

 Estimated Proportions in the New Beneficiary

 Survey Excluding Retired Workers Aged 65 and

 Over and Disabled Workers . . . . . . . . . . 91

Table 7.7. Conservative Generalized Sampling Errors of

 Estimated Proportions in the New Beneficiary

 Survey for Retired Workers Aged 65 and Over

 and Disabled Workers . . . . . . . . . . . . . 92

Table 7.8. Weighted Distribution of Current Working Status

 for Three Subdomains in the New Beneficiary

 Survey . . . . . . . . . . . . . . . . . . . . 93

 Page No.

Table 7.9. Weighted Totals, Unweighted Sample Sizes and DEFTs

 by Contrast Group for Six Asset Variables . . 94

Table 7.10. Conservative Generalized Sampling Errors of

 Estimated Percentages for Differences

 Between Two Subgroups . . . . . . . . . . . . 95

Table 7.11. Conservative Generalized Sampling Errors of

 Estimated Percentages for Differences

 Between Two Subgroups . . . . . . . . . . . . 96

Table 7.12. Conservative Generalized Sampling Errors of

 Estimated Percentages for Differences

 Between Two Subgroups . . . . . . . . . . . . 97

Table 7.13. Conservative Generalized Sampling Errors of

 Estimated Percentages for Differences

 Between Two Subgroups . . . . . . . . . . . . 98

Table 7.14. Conservative Generalized Sampling Errors of

 Estimated Percentages for Differences

 Between Two Subgroups . . . . . . . . . . . . 99

Table 7.15. Conservative Generalized Sampling Errors of

 Estimated Percentages for Differences

 Between Two Subgroups . . . . . . . . . . . . 100

Table 7.16. Weighted Distribution of Current Working Status

 for Female Retired Workers and Wives in the

 New Beneficiary Survey . . . . . . . . . . . . 101

Table 7.17. New Beneficiary Survey Sampling Errors of

 Estimated Totals . . . . . . . . . . . . . . . 102

Table 7.18. New Beneficiary Survey Estimates and Sampling

 Errors of the Deceased and Institutionalized

 Universe Population . . . . . . . . . . . . . 103

Table 7.19. Illustration of the Use of Half-Sample Replicates

 to Produce Estimates of Sampling Error in the

 New Beneficiary Survey . . . . . . . . . . . . 104

Table 7.20. Illustration of Half-Sample Means and Sampling

 Error of the Average Quarterly Social Security

 Benefit Among Retired Male Workers in the New

 Beneficiary Survey . . . . . . . . . . . . . . 105

Table 9.1. List of Critical Items for the New Beneficiary

 Survey . . . . . . . . . . . . . . . . . . . . 108

 Page No.

Table 10.1. Income and Income-Related Items Imputed in the

 New Beneficiary Survey . . . . . . . . . . . . 111

Table 10.2. Summary of Responses and Imputations to Income

 Recipiency Flags . . . . . . . . . . . . . . . 113

Table 10.3. Methodology Employed to Impute Missing Income

 Amounts in the New Beneficiary Survey by

 Income Type . . . . . . . . . . . . . . . . . 114

Table 10.4. Summary of Responses and Imputations for Income

 Amounts: Cases with Reported or Imputed

 Flags . . . . . . . . . . . . . . . . . . . . 115

Table 10.5. Response Disposition to Asset Holdings Flags in

 the New Beneficiary Survey . . . . . . . . . . 117

Table 10.6. Disposition of Imputed Asset Holdings in the

 New Beneficiary Survey . . . . . . . . . . . . 118

Table 10.7. Response Disposition to Asset Amounts in the New

 Beneficiary Survey . . . . . . . . . . . . . . 119

Table 10.8. Number of Cases and Percentage of Responses

 Imputed for Amounts of Fourteen Asset Types in

 the New Beneficiary Survey . . . . . . . . . . 120

Table 10.9. Response Disposition to Income Flows from

 Selected Assets in the New Beneficiary Survey. 121

Table 10.10. Number and Percentage of Imputed Income Flows

 Response Category for Eleven Asset Types in

 the New Beneficiary Survey . . . . . . . . . . 122

Table 10.11. Response Disposition to Expected Future Pensions

 from Current, Last and Longest Jobs . . . . . 123

Table 10.12. Response Disposition and Distribution of

 Imputations for Three Income Recipiency

 Flags in the New Beneficiary Survey . . . . . 124

Table 10.13. Response Disposition and Distribution of

 Imputations for Three Income Amounts in

 the New Beneficiary Survey . . . . . . . . . . 125

 1.  INTRODUCTION

 The New Beneficiary Survey (NBS) is a nationally representa­tive, cross‑sectional, household survey using samples randomly selected from the Social Security Administration's (SSA's) Master Beneficiary Record.[[1]](#footnote-1)  The NBS interviewed persons in October‑December, 1982, and linked their responses to administra­tive data on benefits status.  The NBS contains representative samples of new Social Security beneficiaries as retired workers, as disabled workers, as wives, widows, divorced wives, and surviving divorced wives.  The NBS also contains a representative sampling of persons aged 65 and over who were entitled to Medicare benefits but who had not yet received Social Security cash benefits as of July 1982.

 Proper analysis requires weighted data because separate strata differ substantially in sampling rate by age, gender, and type of benefits (see Section 6).  Persons who received their first retired‑work­er, disabled‑worker, wife's, or widow's cash benefit from mid‑1980 to mid‑1981 were eligible for sample selection.  The NBS separately sampled from strata of retired‑worker men aged 62, aged 63‑64, aged 65, and aged 66 or older, and from strata of retir­ed‑worker women aged 62, aged 63‑64, aged 65, and aged 66 or older.  The NBS separately sampled strata of dis­abled‑worker men and of disabled‑worker women.  The NBS also separatedly sampled from strata of women first receiving cash benefits solely as wives, as widows, as divorced wives, and as surviving divorced wives.  In addition, insured workers entitled to Medicare but not receiving benefits by July, 1982, were eligible for sample selection.  Variable number 1426 (FINLWGT) on the data tape contains the weights for sample cases needed in order to weight data for analysis.  Variable number 1425 (NEWSAMPT) contains the sample stratum identifica­tion.

 The survey questionnaire designed by SSA includes the following topics:

 1. Household composition

 2. Employment history

 3. Job characteristics, including pension status of the current job (if employed), the last job (if not currently employed or the current job if relatively recent), and the longest job (if different from the current or last job)

 4. Other employment not covered by Social Security

 5. Health

 6. Topics 2. through 5., above, for the respondent's spouse if married

 7. Sources of income and amounts of income received in the last three months (asked separately for unmarried respondents and for married respondents and their spouses)

 8. Asset holdings and income from assets (asked jointly for married respondents and their spouses)

 9. Marital history

 10. Child care

 Because different groups of respondents answered varying sections of the questionnaire, the questionnaire and codebook should be studied carefully before processing the data.  The New Beneficiary Survey Question‑by‑Question Specifications instructed the interviewers on the meaning of questions and also should be studied before processing the data.  The Questionnaire and the Question-by-Question Specifications are contained in this document­ation.

 The NBS data are available as an IBM standard label tape with the following characteristics:

 Data Set Name:  SSA.RBSPAS03.NBS.PUB.EBCDIC.

 Parity: odd

 Label:  IBM Standard Label (9 track tape)

 Tape Density: 6250 BPI

 Format:  EBCDIC

 Record Format:  Fixed length and blocked (FB)

 Record Count: 18,599 records

 Record Length: 6202 characters per record

 Block Size:  31010 (6202 times 5)

 The codebook is a second file on the IBM standard label tape with the following characteristics:

 Data Set Name: SSA.RBSPAS03.NBS.PUB.CODEBOOK

 Parity:  odd

 Label:  IBM Standard Label (9 track tape)

 Tape Density:  6250 BPI

 Format: EBCDIC

 Record Format:  Fixed length and blocked (FB)

 Record Count:  26,569

 Record Length:  90 characters per record

 Block Size:  32760 (90 times 364)

 2.  THE NBS UNIVERSE

2.1. The MBR Listing

 The goal of the new beneficiary portion of the NBS was to represent the situations of living noninstitutionalized persons in late 1982 (October‑December) who had begun receiving retirement or disability benefits under the Social Security program for the time period between mid‑1980 and mid‑1981.  The goal of the nonbene­ficiary portion of the survey was to provide comparable information about persons in late 1982 who were at least aged 65, but who had not retired.  Because of cost considerations, the NBS was restricted to the contiguous 48 United States and the District of Columbia.

 SSA initially created two listings from SSA's Master Beneficiary Record (MBR):  a listing of new beneficiaries in a recent 12‑month period; and a listing of nonretired, nonbeneficiaries aged 65 and older shortly before the survey.  The first listing, drawn in March, 1982, identified persons who first received cash benefits based on an individual earnings record as retired workers, wives, widows, divorced wives or surviving divorced wives for the time period between June, 1980, and May, 1981, and persons who first received cash benefits for a disability between July, 1980, and June, 1981.  The second listing, drawn in July, 1982, identified all insured workers aged 65‑71 who were entitled to Medicare and to retired‑worker benefits but who had not received cash benefits.

2.2. Beneficiary Categories

 The size of the initial universe of new beneficiaries selected from the MBR differs for several reasons from counts of new benefit awards that are regularly published in the Social Security Bulletin and its Supplement.  First, persons may receive more than one benefit award if they are entitled to more than one type of benefit or if they shift from one benefit category to another.  Second, timing of benefit awards does not always coincide with the timing of first payable benefits, especially in the case of retired workers.[[2]](#footnote-2) For the NBS, the general criteria for selecting new beneficiaries from the MBR was to allow each person only one chance of selection, when his/her first benefit became payable either because of retirement (in the case of retired workers), disability (in the case of disabled workers), or because of old age (in the case of aged wives, widows, divorced wives, and surviving divorced wives).  For the NBS, persons who shift from one of the above types of benefits to another are not usually counted as new beneficiaries when they shift to a second benefit category.  Using these criteria, the beneficiary categories include the groups described below.

2.2.1.  Retired Workers

 This group consists of individuals aged 62 and over receiving retired‑worker benefits based on their own work record (including persons who are dually entitled to retired‑worker benefits and to a partial supplemental benefit as a spouse or survivor).[[3]](#footnote-3)  It includes retired‑worker beneficiaries who may have received benefits as dependents or survivors, or disabled workers who subsequently recovered and later entered the benefit rolls as new retired workers during the NBS sample period.  This category excludes disabled workers converting to retired workers at age 65.

2.2.2.  Disabled Workers

 This group consists of individuals aged 18 to 64 receiving Social Security disabled‑worker benefits.  It includes workers with a first payment for a new period of disability entitlement, even though they may have received prior benefits as disabled workers, spouses, or survivors.  Persons who shift from retired‑worker to disabled‑worker benefits at ages 62‑64 are not included, as they are included in the retired‑worker category.

2.2.3.  Spouses

 Wives. This category consists of women aged 62 and over who receive benefits solely based on a retired or disabled husband's work record.  Women dually entitled to retired‑worker benefits and to a partial supplement as a wife beneficiary are not included, as they are included in the retired‑worker category.  The wives category includes wives with prior benefit receipt in a nonaged dependent or survivor category and excludes wives who have previously received benefits as retired workers.

 Widows. This category consists of widow beneficiaries aged 60 or over whose first benefits on account of old age are based solely on a deceased husband's work record.  It includes widows with prior benefit receipt as young mothers with dependents or as disabled workers and excludes widows with either dual entitlement as retired workers but does include widows who previously received benefits.

 Divorced Wives. This category is identical to the wives category above except that a divorced wife's benefits are based on the work record of a former husband.

 Surviving Divorced Wives. This category is identical to the widows category except that the surviving divorced wife's benefits are based on the work record of a former husband who is deceased.

2.3.  The Nonretired Category

 The nonretired, or Medicare‑only, group consists of persons aged 65 to 71 in July, 1982, who were entitled to Medicare and to retired‑worker benefits, but they had not yet received benefits because they were not yet retired.  Their benefits payments generally were in suspended status because they had earnings that could cause their benefits to be withheld under the Social Security earnings test.  Under the Earnings Test applicable to persons aged 65‑71 in 1982, $1 in annual benefits was withheld for each $2 in annual earnings above $6,000.

 This nonretired, or Medicare‑only group, was included in the survey for comparison with retired‑worker beneficiaries.  The age range of the Medicare‑only group is similar to that of the retired‑worker beneficiary category.  When retired‑worker beneficiaries began to receive benefits at age 62‑72 in June, 1980 through May, 1981, the Medicare‑only group was aged 63‑70.[[4]](#footnote-4)

 Because of the way in which the groups were specified, the Medicare‑only group cannot be combined with the new beneficiaries to represent a meaningful universe.

2.4.  Individuals versus Couples

 The NBS represents new beneficiaries or nonbeneficiaries as individual respondents.  If the respondent was married and living with a spouse at the time of the interview, the survey obtained information about the spouse and about the joint income and asset holdings of the respondent and spouse.  The husband, the wife, or both could have been selected into the sample, however.  For example, the husband could have been selected as a retired worker, while his wife was selected as a wife beneficiary.  Such a couple would be represented twice on the data file as individuals.  It is difficult to identify the weighted number of these couples where both partners enter retirement by receiving benefits. Thus, if all married respondents were grouped to form a category of "couples," the total count would be too high.  As a result, married men and women in the NBS should not be combined in an attempt to represent couples.

 "Couples entering retirement" can be redefined (or defined more precisely) in a way that is easily implemented using NBS data, by defining them either by the husband's entry onto the benefit rolls, or by the wife's entry onto the benefit rolls.  In the first case, married, male, retired workers and their spouses represent couples in which the husband first received benefits during the 12‑month period (NEWSAMPT = 1‑4).  The wife may or may not have been a beneficiary.  Alternatively, "couples entering retirement" could be defined as married women and their spouses, and should include both married, female, retired workers and wife beneficiaries (NEWSAMPT = 5‑8, 11).  In this case, the husband may or may not have been a beneficiary.

 To add married, female, retired workers or wife beneficiaries to the first alternative above would allow double counting of those couples where both spouses entered the benefit roles and were interviewed.  Similarly, adding married, male, retired workers to the second alternative would also result in double counting.

2.5. Cleaning the MBR Listing

 For the purposes of the NBS, the listings selected by SSA were unclean.  The MBR identifies names, addresses, Social Security numbers, and benefit information.  The MBR includes not only living persons but institutionalized and deceased persons who were not wanted for the NBS universe.  The MBR does not identify whether persons are institutionalized and, thus, they could not be eliminated from the listings.  Although the MBR identifies deceased persons who could be eliminated from the listings, additional persons died between extraction and interview.  Thus, the initial MBR listing contains persons ineligible for the NBS because of institutionalization or death, and the NBS had to oversample the listings in order to obtain the sample size desired.

  Table 2.1 presents information on the MBR listings, the NBS sample, and the NBS survey universe.  As can be seen in Table 2.l, the total beneficiaries receiving first benefits in the 12‑month selection period and who were alive on the date of the extraction numbered 1,860,347.  The nonbeneficiary population numbered about 257,286 at the date of extraction.  The NBS completed a total of 18,599 interviews.  These interviews represent about 1.8 million new beneficiaries and about a quarter of a million nonbeneficiaries who were alive and noninstitutionalized in October to December, 1982.

3.  SAMPLING

3.l.  Sample Size

 Within the class of new recipients of Social Security benefits, 14 subdomains were targeted for special analyses.  These included eight categories of retired workers subdivided by age and sex, two categories of disabled workers subdivided by sex, wives of retired workers, widows of insured workers, divorced wives of insured workers, surviving divorced wives of disabled or retired workers.  Adding the Medicare beneficiaries who are eligible nonrecipients of Social Security benefits brings to 15 the number of analytic subdomains in the NBS.

 The survey objectives called for a national probability sample of 16,350 new beneficiaries of Social Security benefits plus 1,500 Medicare beneficiaries who are eligible for Social Security but do not receive it.  Thus, 17,850 interviews in total were desired.

 In order to meet the objectives of the survey, separate sample sizes were specified for each subdomain.  These are presented in Column 1 of Table 3.l.  Population totals are exhibited in Column 2; this permits the calculation of the ratio of population to sample size as seen in Column 3.

 The sampling intervals in Column 3 of Table 3.l. would be appropriate for the NBS if everyone appearing on the universe tape was eligible (e.g., noninstitutionalized and living) and nonresponse did not occur.  Since this was likely not to be the case, higher sampling fractions were calculated as follows.  The desired numbers of interviews were first multiplied by 1.75, and the products were divided into the corresponding population sizes (obtained from SSA) to derive original sampling intervals. These were intended to provide sample sizes 75 percent greater than the specified number of interviews.  A preliminary sample was drawn using these rates of selection.  (See Column 4, Table 3.l.)

 Thirty percent of the preliminary sample was then set aside in 10 equal‑sized replicate "reserves" to be used only in case unexpectedly large proportions of sample selections were found to be ineligible for the study, impossible to contact, or unwilling to be interviewed.  The original sampling intervals (Column 4) were therefore 70 percent of the originally computed rates.  This meant that the final sample sizes were (1.75 x .70) = 1.225 times as large as the numbers of interviews needed.  If 81.7 percent of all sample selections provided interviews, the requisite number of interviews would be obtained without use of the reserve sample.  If completion rates were lower than this, random selections could be made from the reserve sample as necessary.

 The NBS went into the field with an original sample of 70 percent of that selected using the rates in Column 4 of Table 3.1.  During the course of the survey, one replicate reserve was drawn to supplement the original sample for all but the Medicare subdomain.  The final sampling intervals in the last column of Table 3.1. reflect the use of 10 percent of the reserves for all subdomains except the Medicare-only cases.

3.2.  Identification of the Sampling Frame and Selection Strategy

 A major design issue concerned the best choice of sampling frames for the NBS.  Recall that a universe tape of beneficiaries was available for sampling.  When such lists are available, simple random sample designs are attractive.  However, personal interviews were required.  Once the time and expense of travel between interviews was considered, it was clear that a simple random sample was not cost effective.  In consequence, a clustered sample was designed.

 Given that a cluster sample was desired, the NBS used the Institute for Survey Research/Mathematica Policy Research (ISR/MPR) National Sample of Primary Sampling Units, based on the 1980 Census.  This sampling frame is representative of the populations of the 48 continental states and the District of Columbia and is appropriate both for general population studies and for surveys of special groups such as new Social Security beneficiaries.

 The sampling strategy was to adapt the NBS sample to the ISR/MPR frame by selecting beneficiaries only within the 100 sampling points of the ISR/MPR national sample.  However, before this took place, the 100 sampling points were modified in an objective, well‑defined manner to correspond to clusters of zip code areas encompassing the primary areas.  This allowed the subselection of beneficiaries on the basis of the zip code number associated with their address on the universe tape.  Beneficiaries were then selected with probabilities inversely proportional to the primary sampling unit selection within a study subdomain.  To state it precisely, the sampling proceeded in two stages:  selection of the ISR/MPR primary sampling units (psu's) and selection of respondents within the ISR/MPR sample of psu's using the MBR.  For all members of a given group h:

 1/fh = P:j x P:ihj

where fh is the sampling interval for group h shown in the last column of Table 3.1., P:j is the selection probability of psu j, and P:ihj is the selection probability of individual i in group h and psu j, given that psu j was selected.

 Since P:ihj = 1/[fh x P:j], second stage sampling rates varied across groups and psu's.  The fact that overall sampling rates within groups (fh) are equal minimized the likelihood that weights will be needed for within‑group data analysis.  However, the fact that the sample is clustered means that variance calculations must take the clustered nature of the sample into account, and the variances of sample estimates are likely to be somewhat larger than they would be under an assumption of simple random sampling.

 The description of psu selection is divided into two phases:

(i) original selection of psu's into the ISR/MPR national sample (Sections 3.3. and 3.4.), and

(ii) modification of psu's to be more suitable for the NBS (Sections 3.5. and 3.6.).

 This is followed by a discussion of the selection of respondents within sample psu's.  These procedures are discussed in the order in which they occurred during sampling.

3.3.  Defining and Stratifying the Original Primary Sampling Units

 Using results from the 1980 and 1970 Censuses of the United States, population projections for 1985 were derived and utilized for the creation and selection of the 100 psu national sample.  The projection was a simple extrapolation of growth from that observed in 1970‑1980.  The actual calculation was:

 MOS = 1980 population + 1/2 [1980 population ‑ 1970 population].

With a small amount of rounding, these measures of size (MOS), when summed over all counties, equal 235 million.

 Psu's were defined as follows.  Counties were first subdivided into two groups:  1) those which would be included in "self‑representing areas," and 2) the rest of the country.  Self-representing areas were defined as Standard Metropolitan Statistical Areas (SMSAs) or Standard Consolidated Areas (SCAs) with projected populations of two million or more.  There are 18 self‑representing areas, and these include a total projected population of 84.6 million, 36 percent of the national total.

 For the rest of the country, primary sampling units were constructed from SMSAs and counties outside SMSAs in one of two ways:  where an SMSA or county had a population of 150,000 or more, it was defined to be a psu; where an SMSA or county had a smaller population, it was combined with adjacent counties or a nearby SMSA to form psu's which were clusters of counties having populations of 150,000 or more.[[5]](#footnote-5)

 A total of 32 strata, each with total projected populations of 4.2 to 5.2 million were created by combining the primary sampling units.  Collectively, these strata included a projected population of 150.4 million.  Criteria for creating strata were established with the aim of increasing their homogeneity.  The first two criteria used were region and metropolitan/nonmetropolitan status, and within these categories, strata were defined using one or more of the following variables:  degree of urbanization, economic growth rates, racial composition, and the proportion of the population which was Hispanic.

 A few of these strata included metropolitan psu's having populations of 500,000 or more.  At the other extreme, there were several rural strata where most psu's had populations under 200,000 and where individual psu's included many counties covering a vast land area.  In some cases, individual psu's covered over half the nonmetropolitan area of western or plains states.  The next step was to select two psu's from each stratum, which provided a total of 64 nonself‑representing psu's.

3.4.  Selecting Primary Original Sampling Units

 The primary sampling units were selected into the sample with probability proportional to size. This can be written precisely as:

 P:j = MOSj/stratum size,

where MOSj is the measure of size assigned to a given psu j, and P:j is the probability of psu selection.

The two selections were made independently, or with replacement.  As a result, the same psu was selected twice in 2 of 32 strata, which means that there are 64 selections, but 62 actual areas included in the sample.  The psu's selected twice are Hillsborough County, New Hampshire, and the San Antonio, Texas SMSA, and within these psu's, two independent sets of respondent selections were made.  (See Section 3.6.)

3.5.  Modification of the ISR/MPR Primary Sampling Units for the NBS

 There were three modifications of the ISR/MPR primary sampling units:  (1) some of the psu's covering large land areas were reduced in size through another stage of selection; (2) the psu boundaries were redefined according to postal service zip code areas; (3) primary selections were drawn within the self‑representing areas.  These modifications are described in turn.

3.5.1.  Reducing the Sizes of Some Primary Sampling Units

 Two conflicting objectives were faced in specifying primary sampling units for the NBS.  First of all, they should be selected with probabilities greater than the rates given in Table 3.1.  Should a psu be selected at a lower rate, the implied within‑psu sampling rate for the second stage would be greater than 1.0, which is highly undesirable.  In this case, weights would have to be applied for data analysis which in turn would increase the variances of sample estimates.  The probabilities of selection of psu's (P:j) were all above 1 in 16.5, so the psu's as selected for the ISR/MPR sampling frame satisfied this requirement.  However, subselection within psu's (as described below) caused the probabilities of selection to fall below this amount for a few psu's.

 The second objective was that travel distances among selected respondents not be too great.  For some of the larger psu's, the average distance between respondents would have been very great, which would have increased travel time and, consequently, survey costs.  To reduce travel time, primary sampling units covering vast land areas were subdivided into two or three smaller secondary sampling units (ssu's), each of which had a minimum population of 70,000.  One of these smaller secondary sampling units was selected. This naturally changed the probabilities of selection to P:j\* :

 P:j\* = P:j x MOSj\*/MOSj,

where MOSj is the total measure of size in psu j and MOSj\* is the sum of the measures of size assigned to the counties included in ssu j.

Another way of saying this would be to write:

 P:j\* = MOSj\*/stratum size.

For ease of presentation, ssu's will be called psu's for the remainder of this report.

 The new nonself‑representing psu's had probabilities of selection greater than all the fh's given in Table 3.1. in all but a few cases.  There are four groups where implied within psu selection probabilities (P:ihj) are sometimes greater than 1.0:  retired female workers aged 66 and over, disabled workers (both male and female), and divorced wives.  The need to weight the data was minimal, however, and the land area covered by sparsely populated psu's was significantly reduced.  Only a small number of psu's covered large land areas and the populations of these tended to be concentrated in one or two large towns.

3.5.2.  The Redefinition of Primary Sampling Unit Boundaries

 The SSA beneficiary record includes complete address information, supplemented by a county code.  The address information, including a zip code, was likely to be accurate since it was used to send checks, but the county codes might be inaccurate in some cases.  This was especially likely to occur when a beneficiary changed address, and the county code was not updated.

 As a result, the operational definition of a primary sampling unit was altered to be the collection of zip code areas assigned to post offices within psu counties.  The definitions of self‑representing areas were similarly changed.  Because an individual metropolitan psu included both the central and suburban counties of SMSAs, psu boundary changes typically occurred in the sparsely populated, rural boundaries of psu's, and did not affect many people.  It is reasonable to assume that these changed boundaries, on average, moved equivalent numbers of people out of and into the sample psu's.  Changing the operational definition of a psu did not affect its probability of selection, and had only a minimal effect on the size of its population.

 By redefining psu's in this fashion, the implementation of the sample design was facilitated considerably.  Sample selection was operationalized by drawing from people who resided in "eligible" zip code areas.

3.5.3.  The Creation and Selection of Primary Sampling Units in

 Self‑Representing Areas

 The self‑representing areas in the ISR/MPR national sampling frame included 149 counties, some of which, even though they were in large SMSAs, were sparsely populated and located at some distance from population centers.  Because it was specified that random sampling be used within primary selections, it would have increased travel times and distances and, hence, survey costs to have included the entire populations of the 149 counties in the NBS sample.  As a consequence, the zip codes in these counties were subdivided into 77 primary sampling units, stratified into 18 strata, and two selections per stratum were made to provide a total of 36 psu's from the original ISR/MPR self‑representing areas.  Combined with the 64 nonself‑representing psu's, a total of 100 psu's for the entire sample was achieved.

 There was no need for the zip code clusters to respect county boundaries, so they were defined in such a way as to make them heterogeneous with respect to race and income.  In particular, Black areas of central cities were subdivided and the various parts were combined with suburban counties.  To give an example, the city of Philadelphia was subdivided into three parts, one of which was combined with Delaware and Chester Counties, one with Montgomery, and one with Bucks to form three psu's.  Each of these primary sampling units included significant numbers of central city and suburban, Black and White, affluent and poor populations.

 For these 77 psu's, the measures of size, MOSj, assigned were the total numbers of recipients eligible for the 14 subgroups.  These were simply summed across all 14 categories of beneficiaries (not including Medicare beneficiaries).  The 77 psu's were grouped into strata according to region and the racial composition of beneficiaries.  Two selections per stratum were made, so the probability of psu selection (P:j = MOSj/stratum size) is written in the same way as that for nonself‑representing psu's except that both the MOSj and the stratum size refer to numbers of recipients rather than projected populations.

3.6.  Selection of Respondents Within Primary Sampling Units for the NBS

 In order to reduce the effects of clustering on the variances of sample estimates, the beneficiaries were stratified within the psu's.  Stratifiers likely to be related to important NBS variables were employed.  This increased the heterogeneity within psu's of the NBS variables and consequently reduced the increases in variance due to clustering.  Since 14 of 15 subdomains were homogeneous by sex, remaining variables thought to be useful were race, age, and Primary Insurance Amount (PIA), an imperfect indicator of income.  (Sex was the primary stratifier for the Medicare group.)  No information was available to determine Hispanic origin of beneficiaries.  Because the PIA is only an indirect measure of income, it may not be as effective a stratifier as race or age.  The strategy, therefore, was to stratify first by race, then by age, and finally by PIA.  The eight groups of retired workers were stratified only by race and PIA, since these groups were already defined by narrow age ranges.  For the Medicare group, cases were stratified by sex, age, race, and PIA.

 The sample was drawn in three steps.  First, "eligible" beneficiaries were extracted from the MBR tape.  A beneficiary was deemed eligible for sample selection if he/she resided in an eligible zip code and was assigned to one of

the 15 eligible subdomains in this survey.  Next, the resultant data set was sorted by design variables and stratifiers.  Design variables included psu (i.e., prespecified clusters of zip codes) and sample subdomain (since each had a distinct sampling rate); stratifiers have already been described above.  The sample was then drawn systematically within each psu.  The selection scheme utilized measures of size which were proportional to the overall desired sampling rates for each sample subdomain and inversely proportional to the first stage selection probability.

 As indicated in Section 3.1., the selected sample was partitioned into a "main" and "reserve" sample.  The preliminary or main sample was 70 percent of the original selections.  This was immediately assigned to the field.  The 30 percent reserve sample was randomly split into l0 equal‑sized replicates and held at ISR.  It was to be used if unexpectedly large proportions of main sample selections were found ineligible, impossible to contact, or unwilling to be interviewed.  Early survey results suggested that reserve samples be allocated to attain desired sample sizes.  Consequently, one replicate reserve was drawn to supplement the main sample for all subdomains except the Medicare‑only group.

 Selection of a beneficiary into the sample did not guarantee eligibility in the NBS universe.  The reasons were threefold.  First, a beneficiary selected into the sample could be deceased by the time a first contact was attempted.  Secondly, a beneficiary could have been institutionalized by the time of the first contact.  Finally, the MBR record could be incorrect.  In this case, no eligible respondent was found at a given address.

3.7.  Sample Results

 The sample design yielded 22,434 selections from the universe tape.  Interviews were conducted with 18,599 respondents.  After deleting ineligibles from the base, this represents an overall unweighted response rate of 85.9.  Table 3.2. presents the final disposition of the sample by subdomain (i.e., beneficiary status).  Column 5 provides the number of persons selected into the sample; entries in this column are the sum of those in Columns 1‑4.

 Columns 1‑4 show the final disposition of the sample.  Column 1 displays the number of interviews conducted for each subdomain.  The desired number of interviews was exceeded in the survey for all but Medicare recipients and female retired workers aged 66 and over.  The number of interviews conducted in these two subdomains was less than 4 percent below the targeted goal.

 Columns 2 and 3 show nonresponse.  Column 2 exhibits the number of known eligibles who were not interviewed.  A selected beneficiary was defined as a "known eligible" if contact was made with the selected person and a set of screening questions was completed.  The screening questions ascertained whether or not the person contacted was, in fact, the selected beneficiary.  Also, some information from the MBR for that person was checked for accuracy.  When the screening was not completed (e.g., refusals, not at home), the eligibility of the selected individual was unknown.  These cases are displayed in Column 3.  Note that the entries in Column 3 are two to three times those of Column 2.  A more detailed account of nonresponse in the NBS is furnished in Section 4.

 Finally, Column 4 presents the number of cases found to be ineligible in the New Beneficiary Survey.  (See Section 2 for definition of eligible universe.)  The majority of ineligible cases were identified as deceased or institutionalized beneficiaries.  Overall, 3.5 percent of the sample was found to be ineligible.  The incidence of ineligibility varied considerably across subdomains.  The highest rates occurred for Disabled Workers; both male and female disabled workers experienced a 6.6 percent ineligibility rate.  This seems plausible, since the disabled are more likely to be institutionalized or to die than beneficiaries in other subdomains.  The lowest incidence of ineligibility occurred for the Medicare group; their rate was only 0.9 percent.  The ineligibility rates of all other subdomains ranged from 1.2 to 4.5 percent.  In sum, the incidence of ineligible beneficiaries in the NBS was negligible and should not affect inferences from the survey.

4.  RESPONSE RATES

 A conservative estimate of the overall response rate is 85.9 percent.  This conservatively assumes that all unscreened beneficiaries (i.e., eligibility is unknown) are eligible.  (See Section 3.7. for a discussion of eligibility.)  The overall response rate is the product of screening and interview response rates.  An overall screening rate of 90.1 percent was attained and the interview rate was 95.3 percent.

 Table 4.1. (Column 5) exhibits the overall response rates by sample subdomain.  Response rates ranged from 80 to 89 percent.  The lowest response (79.5%) occurred with retired females aged 66+, while the highest rate (89.4%) was realized with the divorced wives group.  As the age of beneficiaries increased, the tendency to respond decreased.  Male beneficiaries responded at about the same rate as females.  Generally speaking, the overall response rates were fairly consistent across subdomains.

 The NBS achieved a 90.1 percent screening response rate.  Column 2 delineates screening rates by subdomain.  Screening response ranged from about 86 to 94 percent.  Medicare recipients and retired females aged 66 and over had the lowest screening response.  The divorced wives and retired males seemed to respond best.  The response rates for screening also seem inversely related to age for retired workers.

 The NBS interview response rate was 95.3 percent.  Column 4 presents interview rates by subdomain.  With few exceptions, interview rates were 4 to 7 percentage points higher than screening rates.  The most notable exception regards the Medicare group, for whom the interview rate was about 9 percentage points higher than the screening rate.  Interview rates ranged from 93 to 96 percent; these rates were quite consistent across sample subdomains.

 In concluding, it should be noted that the weighted screening and interview response rates are essentially the same as the unweighted estimates within each of the 15 subdomains.  This is due to the equal probabilities of selection within each subdomain.  Overall, the weighted response rates were almost identical to the unweighted values:  86.0 percent overall response, 95.4 percent interview response, and 90.1 percent screening response.

 5.  DATA PROCESSING

5.1.  ISR/MPR Roles

 ISR and MPR played separate roles in data processing.  ISR was responsible for keeping records on in‑coming forms and questionnaires (check‑in), complete editing of the first five questionnaires of each interviewer, editing for missing critical items (defined in Section 9.3.1.) on 10 percent of each interviewer's work, coding open‑ended questions, and processing validation letters for all completed questionnaires.

 MPR received precoded questionnaires from ISR and key‑entered the data.  Simple and complex range and logical checks were performed at MPR and, when necessary, missing critical item memos were forwarded to ISR for disbursal to interviewers in the field.  MPR was responsible for producing a clean edited data tape with imputations for table production, and for recoding of several variables to SSA specifications.  Due to delays in data collection and editing, and the resultant back‑log of questionnaires, ISR assisted in the final data cleaning.

5.2.  Check‑In Procedures

 In Check‑In, the receipt of all completed interviews was recorded in the log books printed from the computerized sample file.  At this stage, the information reported on the screening form was checked against the data from the sample file.  The major concern was that the correct respondent had been interviewed.  The Social Security account number, birth date and sex designation printed in the log book were expected to agree with the information provided by the respondent and recorded on the Screening form.

 At the time the completed interviews were checked in, a routing slip was completed and attached to the front of the questionnaire.  This slip stayed with the interview throughout processing at ISR and was removed only at the time the questionnaire was being boxed to send to MPR.

 The last step in the initial check‑in procedure was to mail a validation letter to each respondent.  The respondent was asked to answer the questions in the letter and then to return it in a postage‑paid envelope to ISR.  At the time the validation letter was mailed, a card was completed giving the responses to comparable questions in the interview.  Later, returned validation letters were compared with the validation card and any discrepancy between the two sets of answers was checked out by the Field staff.

 After all of the Check‑In steps were completed, the questionnaires in the detailed edit sample were transmitted to the Edit section.  Remaining questionnaires were packaged for Coding.  Before those interviews in the detailed edit sample were sent to the Coding section, they were brought back to Check‑In and also put in packs for coding.

 When all coding steps were completed, questionnaires ready for data entry were again returned to the Check‑In section where the questionnaires were boxed for transmittal to MPR.

5.3.  Edit Procedures

 Interviews scheduled for detailed edit (the first 5 interviews and 10 percent of all remaining interviews for each interviewer) were transmitted to the Edit section and then separated so that all first interviews would be given top priority.  Next in priority were second through fifth interviews and after these, those falling in the 10 percent sample.

 For completed questionnaires in the priority sample, procedures called for the editor to check the interview thoroughly and note on a tri‑part Edit Report Form (ERF) any problems found.  Space was provided on this form for the editor to indicate any action the interviewer should take for resolution of the problem or for the editor to provide some general instruction to the interviewer.  For any cases where problems required communication with the interviewer, one copy of the ERF was sent to the interviewer, a second copy to the regional Coordinator, and the third copy retained with the questionnaire.  When needed to indicate clearly the nature of the problem, copies of appropriate pages from the questionnaire were sent along with the interviewer's copy of the ERF.

 For each questionnaire edited, an Edit Evaluation Form was completed with an overall rating assigned.  Records of these ratings were kept in the Interviewers' Evaluation Books where they were easily accessible to the Field staff when they might have questions about the quality of individual interviewers' work.

 An interview with missing information was stored in the Edit section until a response to the missing information memo was received from the interviewer.  Initially, if no response had been received within a month, a second memo was sent.  Toward the end of the study, the waiting period was shortened to two weeks.

 All interviews that passed the initial edit or ones for which the missing information had been retrieved and entered into the questionnaire were returned to Check‑In and packaged there for coding.

5.4.  Coding/Check‑Coding Procedures

 Initially, the steps in the coding process consisted of coding and check‑coding.  In the first stage of coding, codes were entered for all the occupational questions and for the relationship of all household occupants to the respondent, leading zeros were entered wherever the interviewer had omitted them, dollar amounts were recoded when necessary for data entry, and codes that MPR had requested to expedite processing were also entered.  It was soon determined, however, that not all of the coders who had been trained were able to handle occupational coding and, as a consequence, it was decided to separate the coding process into four different steps:

(i)  First round coding, where all of the coding operation was to be performed except for entering occupation and industry codes.

ii)  First round occupation coding, where codes from the 1980 Alphabetical Index of Industries and Occupations were to be entered for all questions where civilian occupation and industries were reported.  Special codes indicate the armed forces.

(iii) Check coding of occupation and industry codes, where the coded occupational data from 20 percent of the interviews (one out of each pack of five) were checked by a second coder.

(iv) Final round of coding where the total edited and coded interview was checked for completeness.

 Check coding continued throughout the study.  The first packs were checked by supervisors.  After that, the responsibility for check coding was assigned to those coders whose performance in the practice exercises and in production coding of the first packs indicated that they understood the mechanics of the interview schedule and had the substantive knowledge required for making difficult coding decisions.  A notebook was used by check coders to record evaluations of coder performance.  Specific problems occurring on individual questionnaires and errors being made consistently by one or more coders were noted in this book.  The supervisors used the information contained in this notebook along with their own evaluations from spot checking to determine when individual coders needed to be alerted to errors and when general reminders were required for the entire coding staff.  General reminders took the form of memos circulated to all the coders or staff meetings where instructions were updated and errors occurring most frequently were discussed.

 Midway through data processing, procedural changes were implemented in order to increase efficiency.  In addition to their check‑coding functions, check coders were trained to perform all the duties of the editors except for writing memos to the interviewers.  The new procedures called for all completed questionnaires, except for first and second interviews and interviews for couples where both spouses had been selected as respondents, to go directly to the Coding section.  Interviews would be processed through first round of coding, occupation coding and occupation check coding, and a final check‑code edit.  If the check coder found any items missing on questionnaires in the 1‑5 sample or critical items missing on any other questionnaires, the problems would be noted on the Routing Slip and the questionnaire sent to the Edit section where a memo was written to the interviewer.  The check coder was also given responsibility for completing evaluation forms for all interviews in the priority sample that were not routed to the Editing section.  This change in procedure meant that a quality control check was being made on every interview but that interviews (except for first and second interviews) without any problems bypassed the Edit section.

5.5. Key Entry

 Data were key entered at MPR using an "intelligent" data entry system.  As the data were entered, prespecified checks were made for illegal code errors.  Key operators were notified of errors as they were entered.  The error could be corrected or flagged for later cleaning as required.  Key verification by a second operator was also performed.

5.6.  Data Cleaning

 Checks for missing information, range, and consistency were made by the data entry control program.  All interviewer instructions and questionnaire skip logic were also part of the program.

 There were three basic levels of checks:

(i)  If applicable answer fields were missing, operators entered codes which called them to the cleaner's attention.  The cleaner examined the hard copy for a potential answer (entry operators also entered flags if there was ambiguity or confusion present).  If no answer was apparent, the cleaner noted the problem on an ERF.

(ii)  If keyed answers fell outside the programmed allowable range, operators keyed in flag codes.  Range checks were made to ensure that only allowable codes existed on the data file and that outlying numeric values were checked.  In the latter case, the flags alerted the cleaners to possible interviewer recording errors.  The cleaner examined the hard copy for evidence of recording errors and looked for explanatory interviewer notes.  Questions for the interviewer were noted on the ERF.

(iii)  Some consistency checks were made part of the control program where information from two or more places in the interview were compared by algorithm for a pass or fail result.  Flags for this reason alerted the cleaner to look for explanations and, if none were found, note a question for the interviewer.

 Cleaners did not begin recording problems on an ERF until a critical item was missing or found inconsistent with another critical item.  Otherwise, the document was considered to have passed cleaning and was forwarded to the verification stage.

 Inconsistent noncritical fields were changed unless they interfered with the skip logic.  For example, when skip logic was incorrectly followed, the cleaner removed any data which should have been skipped.  If the data removed by the cleaner should have been entered in a different series of questions, the cleaner transfered that data into the appropriate answer categories.

5.7.  Tape Construction

 Development of the final NBS data tape began with the cleaned data tape resulting from the above procedures.  Imputations for several income and asset items were calculated and placed on the file to form the NBS public use data tape. See Section 10 for a discussion of imputation methods.

 6.  WEIGHTING THE NBS DATA

6.1.  Introduction

 Because sample estimates which are obtained from weighted survey data are less biased than those produced from unweighted data, weights have been furnished with this survey data file.  The NBS public use data tape comes with weights which are the product of sampling, post‑stratification and nonresponse adjustment factors.  The "sampling" weight is necessary in order to yield unbiased estimates of the survey population.  "Post‑stratification" uses ancillary information from the Master Beneficiary Record file to produce weights which improve the precision of sample estimates.  The "nonresponse" adjustment factor is designed to reduce biases due to differential response rates among population groups.  Thus, the overall weights can be used to produce the best possible estimates of the NBS survey population.

 This section documents the methodology employed in the creation of weights for the NBS data tape.  The data set contains final weights which represent the product of the three weight adjustments:

 (i) a basic sampling weight,

 (ii) a post‑stratification weight, and

 (iii) a nonresponse adjustment weight.

6.2.  Sampling and Post‑Stratification Weights

 The basic sampling weight is simply the inverse of the overall selection probability.  The sampling weight varies for each subdomain and ranges in value from 22.6 to 236.2.  The first column of Table 6.1. presents the basic sampling weights employed in the NBS.  The basic sampling weight also incorporated weight adjustments required when the second stage sampling rates within psu's were larger than one.  However, this was necessary for only five subdomains (i.e., those with the highest sampling rates).

 The post‑stratification weight adjusts the survey data so that the sample and universe distributions match precisely across the subdomains of study.  Without this adjustment, one can expect these distributions to match only on the average.  The reason is that the sample distribution is subject to a small amount of sampling variability.  The variability is a result of first‑stage selection probabilities (for nonself‑representing psu's) being based on 1985 projected population counts rather than counts of eligible beneficiaries.

   Post‑stratification adjustments were calculated by dividing the universe counts by the weighted sample counts separately within each subdomain.  The resultant weights are exhibited in the second column of Table 6.1.  Most adjustments are negligible.  Larger adjustments range from 1.04 to 1.08; the largest occurs for divorced wives (1.12).  The post‑stratification weight was applied to all persons selected into the NBS regardless of whether or not they were interviewed.

6.3.  Nonresponse Adjustment Weights

 In order to best understand the procedures employed to adjust for nonresponse in the NBS, it is useful to envisage the total sample as being composed of four groups:

 Group Composition

 A Eligible respondent interviewed

 B Eligible respondent not interviewed

 C Respondent not interviewed, eligibility unknown

 D Respondent not eligible

Screening information obtained at the time of first contact was used to establish the eligibility of a selected individual.  For instance, an individual who was deceased or instituionalized at the time of first contact was not considered to be an eligible member of the survey population.  Group C was the only subset for which screening information, and thus eligibility, could not be ascertained.

 The weighting strategy for nonresponse compensation takes the expected proportion of eligibles in Group C plus the population in Group B (known eligible) and allocates their weight to Group A (interviewed eligible respondents) within weighting classes.  The weighted counts are based on the product of the sampling and post‑stratification weights.  The weighting classes are constructed from auxiliary data related to income items, eligibility status and the propensity to respond (i.e., the response rate).

 The following variates were considered for the creation of weighting cells:

 (i) Primary Insurance Amount (PIA)

 (ii) Age

 (iii) Sex

 (iv) Race

 (v) Subdomain (which incorporates age and sex)

 (vi) Primary Sampling Unit (psu)

Various tabulations and regressions were employed to investigate the potential usefulness of these items.  The results of this effort are the weighting cells presented in Table 6.2.  The cells are various combinations of PIA quartiles, psu categories and subdomain.  Four psu categories were used:  nonself‑representing metropolitan psu's (NSR SMSA); nonself‑representing nonmetropolitan psu's (NSR RURAL); self‑representing psu's demonstrating high response rates (SR GOOD); and self‑representing psu's demonstrating poor response rates (SR POOR).  In total, 98 weighting cells were utilized; these are serially numbered 1 to 98 in Column 4 of Table 6.2.

 The weight adjustments were calculated as follows.  Let A, B, C and D represent the weighted totals for Groups A to D, given above.  Furthermore, let T denote the weighted total of all groups (T = A + B + C + D).  Then

CE = C{[A+B]/[T‑C]} is the expected number of eligibles in Group C;

TOTE = A+B+CE is the total expected number of eligibles in the universe;

and TOTE/A = T[A+B]/A[T‑C] is the weight adjustment used for Group A members.

These adjustments were calculated and applied separately within each weighting class.  The final weight is the product of the sampling weight, the post‑stratification weight, and the nonresponse adjustment.

6.4.  Generalizability to the NBS Sample Universe

 Using the weighted data set, the NBS provides valid inferences to the nondeceased, noninstitutionalized household population of new Social Security beneficiaries (plus the Medicare‑only group) as defined in Section 2 of this report.

 Note that in its entirety, the NBS data set does not form a conventional population.  The reason is due to the Medicare‑only group.  It was included in the survey as a comparison group to the other sample types, and therefore is not a member of the SSA beneficiary universe.  Analysts are therefore urged to exclude the Medicare‑only group when analyzing the data, except when specific intergroup comparisons are desired.

 7.  SAMPLING ERRORS FOR THE NBS

7.1.  Introduction

 Since the data presented on the NBS file are based on a sample, they may differ somewhat from the figures that would have been obtained from a complete census based on the same schedules, instructions, and interviews.  Particular care should be exercised in the interpretation of figures based on relatively small numbers of cases as well as small differences between figures.  The standard error is primarily a measure of sampling variability, that is, of the variations that occur by chance because a sample rather than the entire population is surveyed.  Assuming a normal sampling distribution, the chances are about 68 out of 100 that an estimate from the sample would differ from a complete census figure by less than the standard error.  Chances are about 95 out of 100 that the differences would be less than twice the standard error. The more commonly used cut-off points 90, 95, and 99 chances out of 100 represent deviations of less than 1.65, 1.96, and 2.56 standard errors, respectively.

 The NBS employed a complex (clustered) multistage area probability sample design.  Thus, the usual simple random sample estimates of sampling error are inappropriate for these data.  Sampling error of sample‑based statistics must reflect the complex nature of the sample design.  This section delineates the technical approach used to calculate sampling errors of estimated totals, percentages, means, medians and quartiles.

 All sampling errors are based on variables which have been imputed for item nonresponse.  If the assumptions inherent in the imputation scheme are correct, there should be little impact on the values of the sampling errors.  A full discussion of these issues is found in Santos.[[6]](#footnote-6)

7.2.  Algorithms for Sampling Error Calculations

7.2.1.  Proportions, Contrasts and Totals

 Because the NBS utilized a design whereby two psu's were independently selected (with replacement) per stratum, a paired selections model is used to calculate sampling errors.  Psu's are paired in accordance with their original selection stratum.  In all, 50 paired strata are employed.

 In this section, the formulae for calculating sampling errors of proportions, contrasts (differences of proportions), and totals are considered.  To begin, note that proportions are estimated using ratio estimators of the form

 y SUM:[yh1 + yh2],

 r = - = ---------------

 x SUM:[xh1 + xh2]

 where yhj denotes the weighted total of individuals with characteristic y in psu j (j = 1,2) of paired stratum h (h = 1,..., 50); xhj is simply the sum of weights in psu j, stratum h, over all sample elements.  In the case of a proportion, the variate y takes a value of 1.0 if characteristic 'y' (e.g., receives food stamps) is present, and a value of zero otherwise.

 A Taylor Series Expansion is used as a method of approximating the variance of a ratio mean:

 varr = {vary + r2 varx ‑ 2r covxy}/x2,

 where vary and varx denote variance estimators of y and x, respectively, and covxy is the covariance between x and y.

 Under a paired selection model, the formula for varr simplifies considerably:

 varr = {SUM:[Dyh2] + r2 SUM:[Dxh2]

 ‑ 2r SUM:[Dxh Dyh]} / x2 (1)

 where Dyh = [yh1 ‑ yh2] and Dxh = [xh1 ‑ xh2].

Estimates of precision using (1) tend to be conservative because the finite population correction is ignored.

 Equation (1) also yields sampling errors of subclass ratio means.  This is accomplished by simply ignoring those sample elements which are not members of the subclass.

 A word of caution is in order when the variance of r is approximated via the Taylor expansion method.  The approximation is appropriate when the coefficient of variation of x (cvx) is less than 20 percent and preferably less than 10 percent.  Large coefficients of variation of x are usually associated with small subclasses.  Since most if not all sampling errors here will be calculated on the basis of a relatively large subdomain[[7]](#footnote-7) this problem should not arise.

 The formula in equation (1) may be extended to estimate the variance of the difference of ratio means (r‑r'):

 varr-r' = varr + varr' ‑ 2covrr'

 where r = y/x, r' = y'/x' and

 covrr' = {covyy' + rr' covxx' ‑ r' covyx' ‑ r covy'x}/xx'.

 Finally, variances of population totals (y=yi) are easily calculated using the formula:

 vary = SUM:[Dyh2].

7.2.2.  Means, Medians and Quartiles

 A pseudo‑replication technique called Balanced Repeated Replications (BRR) is used to calculate sampling errors of means, medians and quartiles.  BRR calculations are especially amenable to paired selection sampling designs.  The algorithm for calculating variances calls for the formation of K replicate half‑samples (where K is a number divisible by four and greater than the number of paired strata).  Each half‑sample is formed by selecting one psu from each paired stratum and assigning it to that replicate.  Half‑sample replicates are formed to meet certain orthogonality requirements (see Kish and Frankel or Plackett and Burman).[[8]](#footnote-8)

 The formula used to calculate the variance of any statistic t is simple.  Let t denote the statistic (e.g., mean annualized total income) calculated using the entire sample.  Furthermore, for half‑sample i, let ti denote the statistic t based on the ith half‑sample.  Then the BRR variance of t is simply

 vart = {SUM:[ti ‑ t]2} / K

 where K denotes the number of generated half‑samples.

For the New Beneficiary Survey, 52 replicate half‑samples were employed.  The replicate design matrix used to generate the half‑samples was taken from Plackett and Burman.[[9]](#footnote-9)

7.2.3.  Design Effects

 Design Effects (DEFF) are used to produce generalized sampling error tables.  They also measure the efficiency of a specific sample design relative to what would have been obtained under a simple random sample (SRS) of the same size.  For a given statistic, t, and design, d, DEFF is defined as the ratio of the variance of t divided by the variance of t assumed under SRS:

 DEFF t = vart(d) / vart(SRS)

When DEFF exceeds 1.0, the design used was not as efficient as SRS; a value of 1.25, for example, indicates a 25 percent increase in variance over SRS.

 Two related quantities are useful in the calculation of generalized sampling errors.  First, DEFT is defined as the square root of DEFF.  It is useful because it is the factor by which the sampling error is affected by the complex sample design.  The second quantity is the effective sample size, n\*, defined as follows:

 n\* = n/DEFF,

 where n denotes the unweighted sample size.

The following sections show how the square root of n\* was employed to produce generalized sampling errors.

7.3.  Generalized Sampling Errors

 The most appropriate way to assess sampling variability is to calculate the sampling error of each statistic employed in analysis.  This is an impossible task when preparing sampling errors for public use survey data because the potential analyses of such data are innumerable.  Instead, the common practice is to generate sampling errors for a variety of statistics and subclasses, and then assemble the results into generalized sampling errors.  The generalized sampling errors are considered to be adequate approximations to the true values.

 This section delineates the methodological approach to forming generalized sampling errors in the NBS.  The main objective is succinctness‑‑to produce as few sampling error tables as possible with the widest applicability.  It was felt that fewer and simpler sampling error tables would increase the likelihood that analysts would utilize them when conducting analyses with NBS data.  The only drawback is that the resultant sampling errors tend to be conservative.

7.3.1.  Sampling Errors for Proportions

 A four‑step process was used to obtain generalized sampling errors for proportions.  First, sampling errors were calculated for 50 variables within each of 12 subgroups, producing 50 x 12 = 600 values.  Next, the average square root of design effects (DEFT) was obtained for six groupings of the 50 variables; this was done separately for each of 12 subgroups, yielding 6 x 12 = 72 average DEFTs.  Using weighted and unweighted sample sizes, a linear model was fitted between the square root of the weighted sample total and the square root of the effective sample size (SQRT:[n\*]).  The fitted line was used to generate conservative sampling errors applicable to all variables and most subgroups considered in this analysis.  The details are now presented.

 Table 7.1. displays the variables for which sampling errors of percentages were derived.  The variables roughly comprise six variable groupings:  Assets, Health, Income, Job Characteristics, Work History and Pension.  Table 7.2. presents the subgroups for which sampling errors were generated.  Twelve subgroups were used, representing various combinations and cross‑classes of the original 15 sampling subdomains.  Table 7.3. shows the average DEFT for each variable‑subgroup combination.  The DEFTs range from about 0.92 to 1.47, averaging 1.15.  The Assets variable tends to have higher DEFTs, while those for Job Characteristics seem to be lower than the DEFTs for the remaining variables.  Retired workers exhibit larger DEFTs on average, and wives and widows seem to be lower on average.

 Tables 7.4. and 7.5. present the unweighted sample sizes and the weighted population totals for each variable‑subgroup cell, respectively.  Table 7.4. demonstrates that a wide range of raw sample sizes was employed in the generation of sampling errors (89‑5,307).  Within a subgroup, changes of sample size across variable groups reflect both differential rates of missing data and cases not applicable (mostly the latter).  This is most evident in the Pension variable where sample sizes are reduced 40 to 90 percent over those of Assets, Health, and Income.  The corresponding weighted totals of Table 7.5. represent an even wider range of values than the Table 7.4. entries.  Here, values range from 3,800 to about one million.  The strategy used in producing generalized sampling errors assumes that the variation in DEFTs, sample sizes and weighted totals captures much of the sampling error variation across variables and subgroups.

 The next step in the production of generalized sampling errors was the calculation of square roots of the effective sample sizes and square roots of weighted totals.  This was done for each variable‑subgroup cell in Table 7.1.  The motivation behind these calculations lies in the method of calculating generalized sampling errors.  The following approximation is used (the finite population correction factor is ignored):

 sep = DEFT x SQRT:[p(l‑p)/n] = SQRT:[p(l‑p)]/L,

 where L = square root of the effective sample size.

If a strong linear relationship can be found between L and the square root of the weighted total, t\*, then generalized sampling errors can be easily generated using the formula

 sep = SQRT:[p(1‑p)]/L

 with L = a + bt\*.

Here, a and b are Ordinary Least Squares (OLS) solutions to the simple regression of L on t\*.  The slope and intercept can be chosen to yield more conservative generalized sampling errors, as well.

 A scattergram of t\* by L was derived from the average DEFTs, weighted totals and unweighted sample sizes reported in Tables 7.3.‑7.5.  Apart from a small collection of points on the middle right side of the scattergram, a linear relationship seemed to exist.[[10]](#footnote-10)  Upon further examination of the outlying points, it was found that they represented two distinct subpopulations:  Disabled Workers and Retired Workers Ages 65+.  Deleting them from the scattergram yielded a strong linear relationship.  A simple regression of L on t\* using OLS produced the prediction equation:

 L = 7.108 + (.0585)t\*.

The r2 statistic was 0.943; a student's t test of the slope and intercept indicated that both were significantly different from zero.  A slightly more conservative equation, L = 7 + (.058)t\*, was employed to generate sampling errors.

 Table 7.6. presents the generalized sampling errors of estimated proportions for all but the Disabled Workers and the Retired Workers Aged 65+.  Because the sampling errors in this table are much too conservative for these two subgroups, a separate table (7.7.) was generated.  The prediction equation used to generate this table was L = 9.35 + (.060)t\*.  The equation was obtained by fitting a conservative line by hand to the scattergram of actual values of L and t\* (not the averages for each of six variable groups), and estimating the slope and intercept.

 To illustrate the use of the sampling error tables, consider the data in Table 7.8.  Suppose that one wished to construct a 95 percent confidence interval for the percentage of male retired‑worker beneficiaries age 65 who are currently employed.  From Column l, observe that the point estimate is 3l.4 percent and that the estimated total retired‑worker mean age 65 is about l09 thousand.  Conservatively using the fourth row (size of base = 100,000) and sixth column (estimated percentage = 30 or 70) of Table 7.7. (for retired workers age 65), one obtains a sampling error of l.6 percentage points.

 Recall that the lower and upper limits of a (1‑a)100 percent confidence interval for a percentage is given by:

 lower limit = C ‑ (S x Z1-a/2) and

 upper limit = C + (S x Z1-a/2),

 where

 C denotes the point estimate;

 S denotes the sampling error of C;

 Z1-a/2 represents the (1‑a/2)100 percentage point of a standard normal density function.

In the example being discussed:

 (1‑.05)100 = 95 percent level of confidence

 C = 31.4 percent

 S = 1.6 percent

 Z1-.05/2 = Z.975 = 1.96.

Thus,

 lower limit = 31.4 ‑ (l.6)(l.96) = 28.3,

 upper limit = 31.4 + (1.6)(1.96) = 34.5.

A 95 percent confidence interval for the percentage of currently working male retired‑workers age 65 lies between 28.3 and 34.5 percent.

 Suppose, on the other hand, a 90 percent confidence interval was to be constructed for the percentage of retired‑worker females age 65 who are not currently working.  From Column 2 of Table 7.8., 68.1 percent are not currently working and the estimated total female retired‑workers age 65 is about 63 thousand.  Turning to Table 7.7., a conservative estimate of sampling error is found in the third row (50 thousand), sixth column (30 or 70)‑‑2.0 percent.  In this case,

 (1‑.10)100 = 90 percent level of confidence

 C = 68.1

 S = 2.0

 Z1-.10/2 = Z.95 = 1.65.

Thus,

 lower limit = 68.1 ‑ (2)(1.65) = 64.8,

 upper limit = 68.1 + (2)(1.65) = 71.4.

A 90 percent confidence interval for the percentage of retired‑worker females age 65 not currently working lies between 64.8 and 71.4 percent.

 Finally, suppose a 68 percent confidence interval is desired for the percentage of currently working divorced wives.  Column 3 of Table 7.8. shows that C = 9.8 percent and that the estimated total population of new divorced wife beneficiaries is about 5,300.  The sampling error of this estimate is shown in the third column (10 or 90), first row (5 thousand) of Table 7.6.  Conservatively, it is 2.7 percent.  In this case,

 (1‑.32)100 = 68 percent level of confidence

 C = 9.8 percent

 S = 2.7 percent

 Z1-.32/2 = Z.84 = 1.0.

Thus,

 lower limit = 9.8 ‑ (2.7)(1) = 7.1,

 upper limit = 9.8 + (2.7)(1) = 12.5.

A 68 percent confidence interval for the percentage of divorced wives who are currently working lies between 7.1 and 12.5 percent.

 The above examples illustrate the use of generalized sampling error tables.  When the size of the base falls between rows in these tables, interpolation yields adequate approximations of sampling error.

 In closing, note that the confidence intervals described above are only suitable for inference about a single percentage.  When inferences are desired for two or more (i.e., multiple) percentages simultaneously, different methodological approaches are warranted.  Interested readers are referred to Miller.[[11]](#footnote-11) (The simplest approach to simultaneous confidence interval construction is via the Bonferonni method.)

7.3.2.  Sampling Errors for Contrasting Proportions

 Generalized sampling errors were also calculated for contrasts of proportions between two subgroups.  A methodology analogous to that described above was employed.  The same set of variables was used to contrast the following subgroups:

 ‑ Retired Workers Aged 62‑64 versus Disabled Workers Aged 62‑64;

 ‑ Wives versus Married Female Retired Workers;

 ‑ Retired Worker Males versus Medicare‑only Males.

With three subgroups and 50 variables, 150 sampling errors were calculated.  These data were then utilized in a regression‑type approach to yield a prediction equation; in turn, the equation produced generalized sampling errors for contrasts.  Before presenting the details, it should be noted that, technically, the generalized sampling errors for contrasts are applicable only to the specific contrasts and variables considered in their generation.  Their applicability to other contrasts or variables is either a matter of assumption or future investigation.

 Table 7.9. illustrates the data obtained from the calculation of sampling errors for contrasts.  DEFTs ranged from 0.767 to 1.393, averaging 1.045.  Although the range of DEFTs was large, the variation was modest.  About 83 percent of all DEFTs calculated were within 15 percent of the overall average (0.91 ‑ 1.20).

 Weighted totals and unweighted sample sizes were about as dispersed as those in the previous subsection.  Weighted totals ranged from about 8,000 to about 700,000; unweighted sample sizes ranged from 120 to 5,307.

 Generalized sampling errors for the differences of proportions (p1‑p2 = p) are based on a formula which ignores the finite population correction.  Letting p1‑p2 = p, and

 p\* = [n1 p1 + n2 p2] / [n1 + n2],

 sep = {SQRT:[p\*(1‑p\*)]} DEFT {SQRT:[1/n1 + 1/n2]},

 where n1 and p1 are the sample size and proportion of group 1, respectively; n2 and p2 are defined analogously.

 To operationalize the regression type approach in producing generalized sampling errors, analogues are needed for (1) the square root of the effective sample size, L, and (2) the square root of the weighted total, t\*.  The obvious choice for L is

 L = 1/{DEFT x SQRT:[1/n1 + 1/n2]}.

Two candidates for t\* were considered.  They were

 t\*1 = SQRT:[t1] + SQRT:[t2]

 t\*2 = SQRT:[t1 + t2]

 where t1 represents the weighted total of the first contrasted subgroup and t2 denotes that of the other.

Regressions of L on t\*i showed that t\*1 was the best predictor of L.  (The symbol t\* will refer to t\*1 in the following discussion.)  The correlation between t\* and L was 0.782.  An OLS regression yielded the following predictive equation:

 L = 4.161 + (0.569)t\*.

The r2 statistic was 0.612; the t‑tests for the slope and intercept indicated significant differences from a value of zero.  In producing generalized sampling errors, the simpler equation L = 4 + (.57)t\* was employed.

 Generalized sampling errors for contrasts are presented in Tables 7.l0.‑7.15.  They were generated using the formula

 sep = {SQRT:[p\*(1‑p\*)]}/L

 where L = 4 + (.57)t\*,

 and t\* = SQRT:[t1] + SQRT:[t2],

 where t1 and t2 are the weighted totals of the two groups being contrasted.

To use Tables 7.10.‑7.15. note that one must first calculate p\* = [n1p1 + n2p2]/[n1 + n2], where p1 and p2 are the proportions in groups 1 and 2, respectively; also, n1 and n2 are the weighted totals of the groups being contrasted.  Once p\* is calculated (in percent), the proper table is found by matching (as closely as possible) the value of p\* with those presented in the upper right corner of Tables 7.10.‑7.15.

 To illustrate the use of the generalized sampling error tables, consider the following example.  Suppose one wished to contrast the percentages of currently working wife and female‑retired worker beneficiaries, and produce a 95 percent confidence limit.  Table 7.16. provides weighted percentages of current workers in Row 1.  About 3.8 percent of wives currently work, while 21.1 percent of retired females are current workers.  Thus, p2 = 3.8, p1 = 21.1, and p = 21.1 ‑ 3.8 = 17.3 percent.  Next, we calculate p\* by noting from Row 3 that n2 = 209,213 and n1 = 551,632.  So,

 p\* = [(21.1)(551,632) + (3.8)(209,213)]/[551,632 + 209,213]

 = 16.3 percent.

To get a conservative estimate of the sampling error of p, use Table 7.13. (25 or 75 percent).  The sizes of the two populations being contrasted are 551,632 and 209,213.  From Row 5 (600 thousand), Column 5 (175 thosuand), the sampling error is estimated to be 1.7 percent.  (More accurate sampling errors could be obtained through interpolation between Tables 7.12. (10 or 90 percent) and 7.13. (25 or 75 percent), since p\* = 16.3 percent.)  Using the same notation as in the previous subsection,

 (1‑.05)100 = 95 percent confidence interval

 C = p = 17.3 percent

 S = 1.7 percent

 Z1-.05/2 = Z.975 = 1.96.

Therefore,

 lower limit = 17.3 ‑ (1.7)(1.96) = 14.0,

 upper limit = 17.3 + (1.7)(1.96) = 20.6.

A 95 percent confidence interval for the difference between the percentage of currently working female‑retired and wife beneficiaries contains values ranging from 14.0 to 20.6 percent.

7.3.3.  Sampling Errors for Totals

 Sampling errors of estimated totals were requested for a specific set of subgroups.  As such, generalized sampling errors were not calculated.  Tables 7.17. and 7.18. provide estimated totals of sampling errors of totals and coefficients of variation (CV), in percent, for each subgroup.  Table 7.17. presents data for the eligible (i.e., living, household) beneficiary population.  Table 7.18., on the other hand, provides data for the portion of the new beneficiary population which is deceased or institutionalized.  Note that all CVs are under one percent in Table 7.17.  In Table 7.18., CVs vary considerably, ranging from 2 to 24 percent.

7.4.  BRR Half‑Sample Indicators

 This section presents the appropriate alternative to generalized sampling errors for producing sampling errors on means, medians, and other statistics.  Using 52 half-samples, the analyst can estimate an approximate sampling distribution and standard error for any statistic. This standard error should be used for significance testing and confidence limit estimates. The NBS public use data tape comes with a set of 52 half‑sample indicators.  These variables have two possible values‑‑zero and one.  A value of one indicates that a case is included in a particular half‑sample.  These indicators were formed by selecting one PSU from each paired stratum in the NBS sample design.  In total, the 52 half‑samples meet the orthogonality requirements as specified in Plackett and Burman.[[12]](#footnote-12)  Moreover, the half‑sample indicators may be employed to yield sampling errors via the BRR pseudo‑replication technique.

 The formula for the variance of any statistic, t, is simple.  Let ts denote the value of the statistic based on the entire sample.  Furthermore, let ti denote the statistic based on the ith half‑sample indicator.  The variance of ts is estimated as follows:

 vart(s) = {SUM:[ti ‑ ts]2}/52.

The sum of the squared differences is divided by the total number of half‑sample replicates (in this case, 52).

 To illustrate the use of half‑sample replicates for sampling error computation, the data in Table 7.19. have been fabricated.  For simplicity, 10 cases and eight replicate half‑samples are employed.  Suppose an estimate of the mean monthly pension income and its standard error were desired.  Then ts = mean = 89.7, and the half‑sample means are presented on the eleventh row of Table 7.19.  The squared differences between the half‑sample mean and the overall mean are displayed on the next row.  Finally, the variance and sampling error are given at the bottom of the table.

 As the illustration shows, the algorithm for calculating the variance is easy.  It can be displayed in a five‑step procedure:

 (i) calculate the estimate from the entire sample (ts);

 (ii) calculate the estimate based on each half‑sample (ti);

 (iii) sum the squared differences of the half‑sample

 and total sample estimates (SUM:[ti ‑ ts]2);

(iv) divide this sum by the total number of half‑samples to yield the variance of the estimate;

 (v) take the square root of the variance to get the sampling error.

Other estimates of variance may be computed from the half‑sample indicators.  The simplest one has been presented in this report.  Alternatives can be found in Kalton.[[13]](#footnote-13)

 Calculating sampling errors for the NBS is slightly more complicated.  The reasons are two‑fold.  First, allowances must be made for missing data.  Although many of the missing entries were imputed, a small number remained when there were insufficient auxiliary data.  Secondly, the data must employ weights to yield estimators which reflect unequal probability sampling, post‑stratification and nonresponse adjustment.

 Table 7.20. employs actual NBS data to illustrate the use of replicate sample indicators in producing the sampling error of mean quarterly Social Security benefits (Variable QSS) among retired male workers who reported nonzero amounts.  The first two columns of the table display the weighted and unweighted number of valid cases for this variate.  The last column presents the average quarterly Social Security benefit for each replicate half‑sample.  At the bottom of the table, the average benefit among retired male workers who receive nonzero amounts is about $1,881, with a sampling error of roughly $109.

 To construct a 95 percent confidence interval of the mean quarterly benefit, the unweighted number of cases is needed, n = 5,182.  This is necessary to determine the degrees of freedom (df) associated with the t statistic.  In the example,

 (1‑.05)100 = 95 percent level of confidence

 C = 1,881

 S = 109

 t{(1-.05/2), df} = t{.975, 5182} = 1.98 (conservatively),

where t{.975, 5182} represents the 97.5 percentile of a student's t distribution with 5l82 degrees of freedom.  (Here, a conservative value (1.98) corresponding to 120 degrees of freedom was used.)  Thus,

 lower limit = 1881 ‑ (109)(1.98) = 1,665

 upper limit = 1881 + (109)(1.98) = 2,097.

A 95 percent confidence interval for mean nonzero quarterly Social Security benefit for retired male workers includes values $1665 to $2097.

 In closing, the replicate half‑samples provide the analyst with the necessary tools for calculating the approximate standard error of proportions, means, quartiles, regression coefficients, and other statistics.  In this respect, the NBS public use data tape is rather unique.

 8.  THE QUESTIONNAIRE

8.1.  Development

 Items from the questionnaire were initially developed from the experiences with the data gathered from three prior studies‑‑the Survey of Newly Entitled Beneficiaries administered in 1969‑1970, the 1978 Survey of Disability and Work, and the Income Survey Development Program.  Once an initial draft had been developed, it was subjected to an extensive series of pretests, each pretest with nine or fewer respondents.  In all, six pretests were conducted, resulting in extensive modifications to the instrument.  Following the final pretest, the questionnaire was then completely precoded and precolumned in order to facilitate data entry.

8.2.  Organization and Format

 The survey instrument was divided into the following sections:

 1. Household composition 6. Income and assets

 2. Employment history 7. Marital history

 3. Employment and pension detail 8. Child care

 4. Noncovered employment 9. Program knowledge

 5. Health 10. Spouse

On average, the questionnaire was administered in 65 minutes, with a range of 35 to 90 minutes depending on the individual's work history and marital status.  A brief screening form first confirmed that the interviewer had contacted the correct individual.  The screening form was also used to eliminate institutionalized individuals and to record any deaths of beneficiaries that had occurred since the NBS universe selection was made in March 1982.  The questionnaire, as described below, was then administered.

8.2.1.  Household Composition

 The purpose of this section was to obtain basic demographic information about the sample person and members of his or her household.  The requested information included the ages and relationships to the respondent of all persons living in the household.  The beneficiary's current marital status, race, and number of years of school completed were also recorded.  These data were to be used to produce basic descriptions of the beneficiary population and their living situations, and to provide explanatory controls for analyses.  The sequence of questions was asked of all respondents.

8.2.2.  Employment History

 This section collected an employment chronology of each respondent from either 1951 or the year the respondent became age 21, whichever was later.  The year 1951 was chosen since SSA uses annual earnings amounts after 1951 (or the year the beneficiary became age 21, if later) to compute cash benefits.  Job beginning and ending dates were recorded for each job lasting at least 12 months.  Any secondary jobs held concurrently with the primary job were noted.  This section provided a descriptive history of the work experience of the beneficiary for use in analyzing job patterns and employment status.  Additionally, periods of no employment were of particular analytical interest in research on women's benefits and incomes.  Respondents who had 1) never worked, 2) last worked before 1951, or 3) never had a job which lasted at least one year were not asked questions from any of the employment sections.

 The employment history section also served as a screening device for the employment and pension detail section that followed.  After completing the chronological history, the interviewer selected the jobs that fit into three categories‑‑CURRENT, LAST, or LONGEST.  The CURRENT job, if any, was the job at which the beneficiary was employed at the time of the interview; the LAST job was that held at or immediately before the window (May 1980); and the LONGEST job was the job of longest duration since 1951.

 Often, a job fell into more than one category; in fact, a single job may have been in all three categories‑‑CURRENT, LAST, and LONGEST.  A separate sequence of questions was available for each of the CURRENT, LAST, and LONGEST jobs.  If a single job was identified in more than one category, only one of the sets of questions was asked about that job.  As a result, from one to three jobs may have been highlighted.  If the respondent was currently working, the CURRENT job sequence was always asked.  If not currently working, the CURRENT job sequence was never asked.  The LAST job sequence was always asked except when the job identified as LAST and the job identified as CURRENT were the same; then only the CURRENT sequence was asked.  The LONGEST job sequence was always asked unless the job identified as LONGEST was also identified as LAST (in which case, the LAST sequence was asked), or as CURRENT (in which case, the CURRENT sequence was asked), or as LAST and CURRENT (in which case, the CURRENT sequence was the only one asked).  As a result, the data for a respondent might not include answers to questions in each of the three sets.

8.2.3.  Employment and Pension Detail

 For each job highlighted in the employment chronology section (CURRENT, LAST, and LONGEST), a detailed set of questions was asked about that job's characteristics and pension plan provisions.  CURRENT job data defined the employment circumstances of the respondents who were working at the time of the interview.  A series of questions was also asked about changes in hours or weeks worked during the previous three years in order to ascertain the transition to retirement.  The LAST job described the employment characteristics of the beneficiary immediately before Social Security benefits were first received.  Reasons for leaving the LAST job were recorded to ascertain the perceived role of pension eligibility, family commitments, health, and mandatory retirement in the decision to leave the job.  The LONGEST job detailed the work environment at which the respondent spent the greatest period of time.

 For each job, information was obtained on industry and occupation, Social Security coverage, and self‑employment or employee status.  The hours and weeks of work and job‑end earnings level were used to determine full‑ or part‑time status, earned income, and the worker's hourly wage rate.  Self‑employed beneficiaries were asked about Keogh plans, and employees were asked questions concerning pension coverage and receipt.  The responses from these pension questions provided basic information on the types of job covered by pensions, eligibility ages, benefit amounts, changes in benefits, lump‑sum payments, and survivor benefits.  These data can be used to assess the role of pensions in the timing of retirement.  Within the sequence of questions for each job, the subseries of questions for self‑employed beneficiaries was separate from the nonself‑employed questions.  As a result, only responses for one or the other but not both of these subseries could be collected.

8.2.4.  Noncovered Employment

 In addition to the CURRENT, LAST, and LONGEST jobs, information on all additional jobs not covered by Social Security was also requested in order to evaluate proposals affecting Social Security benefits for workers with periods of noncovered employment.  This section of the questionnaire focused on any periods of noncovered work listed in the employment history section that were not CURRENT, LAST, or LONGEST jobs.  For each such period with a Federal, State, or local government, or a nonprofit or charitable organization, questions were asked concerning pension and Social Security coverage.

 The noncovered employment section concluded with questions to determine periods of active duty in the Armed Forces.  This section included questions about service‑connected disability and the disability percent ratings that determine Veterans' Administration benefits.  These questions were asked separately to ensure that military service was not overlooked if it occurred before 1951.  All respondents, whether asked the employment series or not, were included in this series.

8.2.5.  Health

 Respondents were asked to identify current health conditions and recent utilization of medical and inpatient hospital services to provide a partial measure of current health status.  A series of questions about self‑assessed work limitations and their influence on work activity elicited details about the nature, extent, and work‑related causes of reported health impairments.  This section also contained questions about physical mobility, functional limitations, and health care plan coverage.  All respondents were asked this question series. Please note that respondents skip some or all of the functional limitation questions if they reported being bedridden or wheelchair-bound.

8.2.6.  Income and Assets

 Social Security benefits alone are not designed to provide full retirement or disability income.  It is expected that beneficiaries will have available other sources of income such as pensions, savings, or retirement earnings.  Assets are an integral source of retirement income, both for the income they produce in the form of interest, dividends, and rent, and for their cash value if liquidated.  Both income and assets influence the relative importance of Social Security benefits in the total income of new beneficiaries.

 Information was obtained on the respondent's (and spouse's) income from earnings, pensions, means‑tested payments, and income received from other individuals.  Questions were asked about the amounts of income received from each source during each of three months preceding the month of interview.  Where relevant, a respondent was also asked if the benefit was derived from one's own work record or from that of a spouse, and whether it was a retirement or disability benefit.  Information on ownership and value of liquid assets, stocks, bonds, IRA or Keogh accounts, own home, real property, and businesses or professional practices was collected.  Questions on income of the respondent's own children under age 19 who were living in the household and any other person aged 19 or older were also asked.

 Two similar sets of income items were recorded:  one if the respondent was not currently married, i.e., widowed, separated, divorced, or never married, and one if the respondent was married.  Only one set was used for each respondent.  If the married set was used, the questions asked about income sources and amounts for the respondent and for the spouse or, when they received a single combined income from a particular source, both of them together.

 The asset questions were asked of all respondents.  Except for Keogh accounts or IRAs and life insurance policies, married respondents were to include their spouse's assets with their own.  The Keogh‑IRA and insurance policies were asked about for the respondent alone (whether married or not) and if married, separate questions were asked regarding the spouse.

8.2.7.  Marital History

 Marital status, past and present, can affect Social Security benefits.  This section gathered information on the duration of the current marital status and the number of times the respondent had been married.  For widowed, divorced, and separated beneficiaries, the beginning and ending dates for their terminated marriage were collected.  Parts of this section were asked of all respondents except those who had never married.

8.2.8.  Child Care

 Caring for children can affect a person's labor‑force participation, and, thereby, the earnings used to calculate Social Security benefits.  This section collected information on the dates of birth and the relationships to the respondent of children whom the respondent raised.  This section was asked of all respondents who raised any children, biological or otherwise, who were alive or deceased.  Information on up to five children was recorded.

8.2.9.  Program Knowledge

 This section investigated the beneficiary's knowledge of certain aspects of Social Security program provisions.  Questions designed for the retirement sample assessed the individual's knowledge of the earnings test.  Individuals in the disability sample were asked a separate set of questions to assess their knowledge of the provisions of the Social Security Disability Amendments of 1980.  The Medicare sample was asked part of the retirement set of questions.  If a proxy for a retirement sample member was acting as the respondent, no program knowledge questions were asked.

8.2.10.  Spouse

 Spouses of married beneficiaries were also asked to respond to several sections of the questionnaire.  If the spouse was unavailable, the respondent was asked the questions as a proxy for the spouse.  Those who agreed to participate were individually asked all questions in the employment history section, the employment and pension detail section, the noncovered employment section, and a much‑shortened health section.  Income and asset holding information of spouses was included in the income and asset section of the respondent.

 The employment history, employment and pension detail, and noncovered employment sections were identical to those described in Sections 8.2.2., 8.2.3., and 8.2.4., except that no separate history of military service was requested.  Those who had never worked, last worked prior to 1951, or had never had a job last at least one year were only asked the health section.  All others followed the same response sequence for CURRENT, LAST, and LONGEST as described for the respondent.

 9.  FIELDING PROCEDURES

 The 18,599 interviews were conducted in‑person and were administered between October and December 1982 in the 48 contiguous states and the District of Columbia.  The sample design used a clustered probability sample of 100 primary sampling units based on the 1980 Census.  A small number of interviews with beneficiaries who were previously unavailable were conducted in January 1983.

 Persons selected for the survey were sent an introductory letter asking for their participation and notifying them that an interviewer would be visiting them shortly.  Interviewers were instructed to attempt a first contact in person.  If the first attempt was unsuccessful, the interviewers were trained to make as many additional contacts as necessary in order to obtain a completed questionnaire.  Questions from persons concerning the authenticity of the survey were expected.  All NBS interviewers were trained to display their identification badges, to confirm the voluntary nature of each respondent's participation, and to encourage a call to the local Social Security office if the person had any remaining doubts.

9.1.  Interviewer Recruitment and Training

 Interviewer recruitment took place throughout the 100 primary sampling units in the ISR/MPR national sample.  Approximately one‑half of the interviewer staff was composed of previous ISR/MPR interviewers and persons referred by these interviewers.  The remainder were recruited through newspaper advertisements, State Job Services, referrals by the U.S. Bureau of the Census, university placement offices, Chambers of Commerce and other survey research organizations.  In all, 746 interviewers were initially recruited and, of these, 637 were trained.  Eighty‑two of these dropped out or were unacceptable.  An additional 84 interviewers were trained, resulting in a field force of 639.

 Interviewers participated in a three‑day training conference prior to receiving their assignments.  A total of 16 initial and three restaffing training conferences were held.  Training occurred in Philadelphia, Tampa, Houston, St. Louis, San Francisco, Chicago, Atlanta, Los Angeles, Boston, and Rochester (NY).  Prior to attending the conference, the interviewers received a home‑training packet containing a sample questionnaire, training manual, and appropriate forms.  Interviewers were to review these materials prior to coming to the training conference.

 The training conference covered general interviewing techniques, conventions concerning the questionnaire and general forms, techniques for reducing nonresponse, specifics of the questionnaire and practice interviews.

 Interviewers were to receive their assignments at the conclusion of the training conference or at their homes a few days later.  Delay in the preparation of materials, however, resulted in up to a four‑week delay in the field effort.

 The Field Department was composed of four in‑house Field Administrators at ISR (two from ISR and two from MPR).  Each Field Administrator was responsible for 13‑14 Field Coordinators, in the field, who in turn were responsible for 12‑18 interviewers.  Each set of individuals reported progress and problems to the next higher administrative level.  Problems which could be handled at the coordinator level were taken care of in the field.

9.2.  Callback Procedures and Refusal Conversion

 Interviewers were instructed to make unlimited recontacts or callbacks in order to obtain a completed questionnaire.  If no one was at home in the initial contact attempt (conducted in‑person), neighbors were to be contacted to try to determine when the respondent was likely to be found at home.  Additional contacts by telephone and/or return to the household were made on the basis of any such information obtained.  If no other information was available, the interviewer was instructed to vary calling times, emphasizing afternoons, evenings, and weekends.  Interviewers were not allowed to return a Screening/Call Report form until a final disposition had been obtained.

 In the event a refusal was encountered, the interviewer discussed the circumstances with the Field Coordinator.  Depending on the case, the interviewer may have made an additional contact, the Field Coordinator may have reassigned the case to another interviewer, the Field Coordinator may have contacted the respondent directly, or the case may have been forwarded to ISR.  At ISR, the case was again considered and whenever a conversion looked remotely possible, a letter was sent to the respondent requesting their participation.[[14]](#footnote-14)  Such contact was again followed by recontact on the part of the interviewer or a reassigned interviewer.  In all, only 1,677 or 7.7 percent of the eligible sample (21,657) had a final disposition of refusal.

9.3.  Quality Control

9.3.1.  Editing and Critical Items

 The first five of each interviewer's completed questionnaires were thoroughly edited upon receipt at ISR.  Additional interviewing work was not authorized until the questionnaires had been checked and memos for missing information and recording errors had been issued.  For the remainder of the assignment for each interviewer, 10 percent of the questionnaires completed were edited for missing responses to critical items (as defined by the Project Officer at SSA).  If responses to critical items were missing, those items and any other missing items were explained in a memo to the interviewer.  It was the interviewer's responsibility to recontact the respondent to obtain the missing information.  When unable to retrieve the information, telephone calls from ISR were made directly to the respondent.  Regardless of the outcome of recontact for missing information, a questionnaire was retained if, at a minimum, the employment history and the CURRENT, LAST, and LONGEST job sections had been completed.

 Unedited interviews with missing critical items were identified during coding at ISR and during machine editing at MPR.  Each resulted in an additional memo to the interviewer and/or a direct call to the respondent.  A list of the questions considered as critical items appears in Table 9.1.

9.3.2.  Validation

 A validation letter was mailed to all cases for which a completed questionnaire was received at ISR.  The letter contained a few demographic questions and a few questions concerning the administration of the interview.  The respondents were asked to complete the questions and return the letter in a postage‑paid return envelope.  When returned, the information was checked against the information on the questionnaire.  Any inconsistencies were forwarded to the Field Administrator for rectification.  If the discrepancies were due to deception on the part of an interviewer, that interviewer's entire work file was telephone validated and, when necessary, the respondents were reinterviewed by another interviewer.

 10.  IMPUTATION IN THE NEW BENEFICIARY SURVEY

10.1.  Introduction

 Imputation is a technique used in data processing to compensate for item nonresponse.  Item nonresponse occurs when a respondent answers some, but not all questions in the survey instrument.  Missing data are "imputed" when values estimating the true (but unknown) response are inserted in place of missing value codes.  Missing data were imputed in the NBS for a set of critical income and income‑related items.  This section is devoted to a brief description of the NBS imputation methodology.

 The items imputed in the NBS may be categorized into four groups:  earnings items, income types, assets, and miscellaneous income items.  Table 10.1. shows the components of each variable group.  In all, 358 variables were imputed, representing 3 earnings items, 19 income types, 14 asset types, and 6 miscellaneous income variables.  Implicitly, all quarterly or annualized variables were also imputed, since they were constructed from imputed data.

 The rates of imputed data varied widely among the four variable groups.  About 15 percent of earnings data were imputed.  Less than one percent of income type receipts were imputed; however, 1 to 14 percent of income type amounts were substituted.  Values were inserted for 0.3 to 4.5 percent of asset holdings and 11 to 38 percent of asset amounts.  Finally, the rates of imputed data for miscellaneous income items ranged from 9 to 25 percent for amounts and 0.2 to 1 percent for receipts.

 The effects of imputed data on statistical analyses are contingent upon several factors:

 .  the amount of data imputed;

 .  the appropriateness of the model employed in the imputation;

 .  the specific analysis being conducted.

If the amount of imputed data is small, say less than five percent, then the effects of imputation are small.  When rates of imputed data are nonnegligible, the possibility of bias arises.  To the extent that the imputation model correctly predicts the missing value, the potential for bias will be lessened.  Finally, imputation schemes can yield data sets which allow unbiased estimates of some population parameters (e.g., the mean) but biased estimates of other parameters (e.g., regression coefficient).  Interested readers should consult Kalton or Santos[[15]](#footnote-15) for a discussion of this topic.  Given that substantial amounts of data are imputed for several income‑related items in the NBS, analysts should exercise caution when making statistical inferences from these data.

10.2.  Imputation Methodology in the NBS

 A multitude of imputation techniques were employed to compensate for item nonresponse in the NBS.  This section reviews the general approaches taken to impute missing data.  Interested readers are referred to the methodological report on imputation by Czajka[[16]](#footnote-16) for full details.  A subsection for each variable group requiring imputation is presented below.

10.2.1.  Imputation of Earnings

 Missing data were imputed for 12 earnings fields which represented the annual incomes from the three earnings items (CURRENT, LAST, and LONGEST jobs) for the respondent and the spouse, for self‑employed and employee earnings.  Annual amounts were constructed from annual work effort and rate of pay.  (Annual work effort combined hours worked per week and weeks worked per year into a single item.)  Annualized earnings were defined as the product of the rate of pay and the annual work effort.  As necessary, one or both of these components (i.e., effort and rate) were imputed in the construction of annual earnings.  Whenever rate of pay was missing, an hourly rate was imputed, so that maximum utility could be gained from complete or partial annual effort information.  In total, earnings were imputed to roughly 3,300 fields, representing 2,735 cases in the NBS data set.  Imputations of earnings account for about 15 percent of the entire sample.

 Not all cases with missing data were imputed.  Whenever the response to the employee/self‑employment question was unanswered, imputation was not performed.  This affected 815 fields, and roughly 400 cases.

 Most imputed hourly rates were obtained using a technique called stochastic regression.  The term "stochastic" is used to describe a procedure where residual error terms are randomly generated and added to the best estimate of the missing value.  In this case, the "best estimate" is obtained via regression, and residual errors are generated from the (respondent) empirical distribution of residuals.  Stochastic imputations are often preferred to other techniques because they display more realistic distributional properties.

 The stochastic regression modelled the natural log of hourly earnings as follows:

 Yi = SUM:{Bj Xij} + Ei, (1)

 where the Xij's denote independent variables which are available for all (or most) cases, and Ei represents an error term.  Log hourly earnings, Yi, were indexed to constant 1967 dollars to adjust for time differentials in reported earnings.  The coefficients Bj were estimated using all cases with reported Y and X.

 For those cases with missing hourly rates, the regression yielded predicted hourly earnings:

 yi = SUM:{bj Xij} (2)

 where the bj's denote estimated regression coefficients.

For those cases with reported hourly rates, the regression produced a set of empirical residuals:

 ei = Yi ‑ yi.  (3)

The empirical residuals were sorted by the log of predicted annual earnings and split into deciles.  The log of predicted annual earnings is defined by

 pi = yi + Zi, (4)

 where yi is the predicted log of hourly pay and Zi is the log of the annual work effort (in hours).

 Imputed annualized earnings were derived in three steps.  First, an empirical residual, ei, was selected at random from the same decile as the nonresponder.  Next, imputed annualized earning, AEi, was calculated as the exponentiated sum of the predicted annualized earning and the error term:

 AEi = EXP:[pi + ei] = EXP:[yi + Zi + ei]. (5)

Finally, the imputed values were reindexed to original (time dependent) dollars.

 Separate regression equations were estimated for 16 of 18 subgroups[[17]](#footnote-17) in a cross‑classification of job type (CURRENT, LAST, LONGEST), employment status (self, employee), and sex/disability (three groups).  Separate equations were used to increase predictive power.  Regressions were not performed in four subgroups because of unduly small sample sizes.

 The independent variables tested and included in some or all of the regression equations were:  occupation, industry, primary sampling unit, education, race/ethnicity, marital status, employer type, primary insurance amount, beneficiary status, hours worked per week, weeks worked per year, job duration, age (at end of job), and respondent/spouse indicator.  Terms reflecting curvilinear and interaction relationships with the dependent variable were also investigated.

 The cases with missing hourly rates which did not receive a stochastic regression imputation may be partitioned into four distinct classes:

(i) cases in the 16 subgroups which employed regression, but which displayed missing data on one or more important prespecified predictors;

 (ii) cases in 2 of 18 subgroups discussed above;

 (iii) cases showing losses in earnings;

 (iv) cases which were not to be imputed.

Cell mean imputation was performed in classes (i) and (ii).  This technique involves the computation of the respondent mean log hourly rate within a specified subgroup (e.g., self‑employed females), and assigning this value to all cases with missing data in that group.  This was done with separate mean calculations for CURRENT, LAST, and LONGEST jobs.

 For cases showing losses in earnings, a respondent‑donor approach was implemented.  The cases in class (iii) were matched manually to responders who showed losses.  Matching was based on job type (2 classes), sex, occupation, industry, sample subdomain, hours worked per week, weeks worked per year, respondent/spouse indicator, and year of loss.  When a match was made, the donor responder's reported amount was imputed to the missing field.  Losses were imputed to 88 fields in this fashion.

 Finally, cases which exhibited missing data on one or both of hours worked per week and weeks worked per year were imputed via the cell mean approach.  The cells in this scheme were the same as those described above.  Due to relatively high response rates, the amount of imputation for work effort was small.

10.2.2.  Imputation of Income Types

 Missing data were imputed for 19 income types identified in Questions 165‑168, 171‑174, 182‑187, and 190‑194 of the NBS questionnaire.  The income types include specific earnings, pensions, and transfer payments presented in the second variable group of Table 10.1.  Receipt indicators and income amounts received in each of the past three months (from the time of interview) were elicited for both the respondent and spouse.  These items, as well as their quarterly counterparts, were processed for imputation.

 The imputation methodology is presented in two parts.  First, the methods for imputing missing receipt indicators are described.  Next, the imputation of missing amounts is detailed.

10.2.2.1.  Imputation of Income Type Receipt

 The rates of missing data of income type recipiency were small in the NBS.  This point is illustrated in the first 3 columns of Table 10.2.  Missing data ranged from 0.2 percent (42 cases) for Social Security receipt to 0.7 percent for earnings (121 cases).  Moreover, apart from five income types, the rates of recipiency among responders were all under five percent.  Consequently, subjective "best guess" imputation methodologies were employed.

 Two imputation schemes were devised:  an edit routine which was implemented for Social Security, earnings and Federal pension recipiencies, and a case‑by‑case transcript examination scheme used for all other income types.  The edit routines were simple.  "Yes" flags were imputed to missing Social Security receipt fields when the respondent sample subdomain (NEWSAMPT) was anything except a Medicare‑only beneficiary; a "no" flag was imputed otherwise.  Similarly, missing earnings flags were imputed "yes" if the respondent or spouse work history indicated employment within the past three months, with "no" imputed otherwise.  Finally, missing Federal pension recipiency fields were imputed "yes" if responses to the LAST or LONGEST jobs indicated Federal pension receipt, and "no" otherwise.

 For the remaining income types, imputation of recipiency was performed using a case‑by‑case inspection of the interview transcript.  Subjective imputations were made, incorporating responses to related items, patterns of nonresponse, and the overall recipiency rate among responders.  The last two columns of Table 10.2. depict the patterns of recipiency imputations for all income types.

10.2.2.2.  Imputation of Income Type Amounts

 Four methods were utilized to impute missing amounts:  benefit formula construction, cell mean, stochastic mean and stochastic regression.  Table 10.3. lists the methodology used for each income type amount.  Imputed values were calculated from approximations to the benefit formula for food stamp and Social Security benefits.  For food stamp benefits, the imputed amount was equal to the "allotment" (a function solely of household size) minus 30 percent of the midpoint of the income range in Q. 246 less $85.  A minimum value of $10 was imputed to households of size one or two, in accordance with food stamp regulations.  For Social Security benefits, estimation equations conformed to specifications issued by SSA.  These imputations utilized the primary insurance amount plus other relevant data.  In all, imputed food stamp benefits were required for 11 cases, while 507 cases received Social Security benefit imputations.

 Cell mean imputation was performed for nine of 19 income‑type amounts.  Ten cells were used in this process; they result from a cross tabulation of sex by recipiency type (single; married respondent only; married spouse only; respondent and spouse with separate amounts; respondent and spouse with combined payments).  Only small numbers of cases were imputed via the cell mean methodology, ranging from two cases for alimony to 17 cases for railroad retirement income.

 A smoothed‑mean/global‑mean approach was used for estate/trust/royalty payments.  Reported estate/trust/royalty payments were tabulated by reported total assets and total income and smoothed to produce monotonic trends.  The smoothed means were imputed to nonresponders who reported assets and income.  For those with missing data on these "predictors," the global respondent mean was assigned.  A total of 3l cases were imputed for missing estate/trust/royalty payments.

 A stochastic mean approach was used for annuities income.  Here, imputations were equal to the global respondent mean plus a randomly generated error term from a normal distribution with mean zero and variance equal to that obtained from the reported annuities data.

 The remaining six income types were imputed via stochastic regression.  Log earnings were predicted from a regression model describing the three month earnings sum as a function of:  earnings from CURRENT job (see Section 10.2.1.), hours and weeks worked, indicator of second job during reference period, proportion of reference period covered by CURRENT job and respondent/spouse indicator.  A different equation was estimated for each reference period (i.e., last month, two months ago, three months ago).  Imputations were defined as the exponentiation of the summed regression prediction and a residual error selected at random from the empirical residuals.  For persons who reported no income from their CURRENT job, a cell mean approach was used to impute earnings.  Respondent means were calculated and imputed separately for single respondents, married respondents and spouses.  Likewise, for cases reporting losses in the CURRENT job, a mean value imputation of earnings was employed.  Finally, when two of the three reference period earnings were reported, their average was imputed to the missing value (except when both earnings were reported to be zero, in which case the cell mean approach was used).  Earnings were imputed to a total of 925 households.

 Imputations of missing veterans benefits involved the estimation of three regression equations:  one for male respondents, one for male spouses, and one for single females.  The predictors for the male respondent equation included the VA disability rating, VA disability status, indicators of handicaps limiting and preventing employment, Social Security recipiency and income, age, education, and race/ethnicity.  The equation for spouses used all but the first two predictors mentioned above, since these data were not collected for spouses.  Finally, the single female equation incorporated Social Security recipiency and income, age and race/ethnicity as predictors of veterans benefits.  The stochastic error term added to the regression predictions was generated from a normal distribution with mean zero and standard error taken from that of the corresponding regression equation.  Fifty‑eight cases were imputed in all.

 The four remaining income types involve private, Federal, military, and State/local pensions.  The stochastic regression schemes for these items utilized the same general approach.  A "full" equation and a "fallback" equation were used to impute missing data, depending on the amount of predictor information available.  For each pension type, the full equation contained earnings, duration and age data regarding a job providing that pension type, plus Social Security income, education, race/ethnicity and area of residence.  The fallback equation excluded the job‑specific predictors but added age.  Each equation was estimated separately for males and females.  Again, the residual error term was generated from a normal distribution with mean zero and appropriate variance.

 The extent to which data were imputed for each income type is summarized in Table 10.4.  The second column represents the total number of households for which at least one item was imputed.  The third column presents the number of households for which all amount items were imputed.  Most imputation involved the complete replacement of missing data for all three reference periods.  The last column of Table 10.4. exhibits the percentage of imputed cases.  Imputed data account for roughly one to 14 percent of income types.  Five income types required rates of imputation exceeding 10 percent:  railroad retirement, earnings, Federal pension, annuities, and estate/trust/royalty payments.  Four income types were imputed at a rate under three percent:  Social Security, Supplemental Security Income, alimony, and food stamps.

10.2.3.  Imputation of Holdings, Amounts and Incomes from Assets

 Holdings indicators, values of assets and income from assets were processed for imputation in the NBS.  Questions 201 to 245 of the NBS questionnaire cover the 14 asset types treated in this section.  They include financial assets, retirement accounts, real property and business equity.  Table 10.1. lists the asset items subject to imputation.

 The imputation methodology for assets is divided into three subsections.  In the first, techniques for imputing asset holdings flags are described.  Once holdings were determined, asset amounts were imputed.  Thus, the second subsection reports the imputation methodology of amounts such as balances in checking accounts and equity in own home.  The third subsection discusses the various approaches used to impute missing income from assets.  Income returns from assets are called income flows for the remainder of this report.

10.2.3.1.  Imputation of Asset Holdings

 Asset holdings flags were imputed for the 14 asset types shown in Table 10.1.  The "other property assets" item was composed of four specific property asset components:  rental housing, vacation property, commercial property, and land.  Consequently, a total of 17 asset holdings flags were processed for imputation.

 Table 10.5. presents the response disposition to the holdings flags.  Nonresponse to these items was low, ranging from 0.3 percent for home ownership to 4.5 percent for certificates of deposit (CD).  Generally, the accounts flags incurred the highest nonresponse.

 The last eight asset items in Table 10.5. pertain to home ownership, other property and equity.  The response dispositions for these items are characterized by high response rates (above 99 percent), and either very high or very low rates of holdings among reporters.  Subjective "best‑guess" imputation schemes were thus used for these items.  Imputation was based on multi‑way crosstabulations.  Home ownership utilized various categorizations of sample subdomain, sex, and marital status; for equity assets, categorizations of employment status (self vs. employee) and occupation were used.  Cross-classifications of dwelling type, employment status, subdomain and home ownership were employed to impute other property asset holdings.

 The imputation routine was implemented in two steps.  The respondent data were sorted into the cells of multiway tables described above.  Next, the rate of holdings were calculated among reporters in each cell.  Nonresponders were then placed into their respective imputation cell.  In general, if the rate of holdings within a cell was high (e.g., 96 percent), then all nonresponders in that cell were imputed as holders of that asset type.  If the holding rate was low (e.g., 10 percent), then all nonresponders were imputed nonholder status.  Roughly, 700 flags were imputed home ownership, other property or equity holdings flags.  Their distributions are shown in the last eight rows of Table 10.6.

 Both respondent and spouse IRA holdings flags were imputed via a regression approach where the dependent variable was the dichotomous zero/one indicator of holding.  Separate models were estimated for five respondent groups:  retired married females, retired married males, retired single persons, disabled married persons, and disabled single persons.  The predictor variables for these models included home ownership, race/ethnicity, sample subdomain, employment status and occupation indicators, plus several continuous items such as age, education, earnings, and PIA.  Coefficients were estimated from the respondent data and used to calculate predictive probabilities of being a holder.  Holding status was determined by comparing the probability to a randomly generated number between 0 and 1.  If the random number was less than the expected probability of a nonresponder, then IRA ownership was imputed.  The eighth and ninth rows of Table 10.6. show the disposition of the 273 cases with imputed IRA ownership.

 Households which failed to provide information for one of the five accounts assets typically did not respond to all accounts questions.  As such, it was necessary to impute entire patterns of financial asset holdings.  Seven holdings patterns were constructed from all remaining assets types:

 i) neither checking nor savings,

 ii) savings only,

 iii) checking only,

 iv) checking and savings only,

v) savings and one of the other five assets (money market, CD, credit union, bonds, stocks),

 vi) checking and one of the other five assets,

 vii) savings, checking and one of the other five

 assets.

Multiple discriminant analysis was used to impute patterns of asset holdings.  Seven discriminant functions were estimated from the respondent data.  These functions were then applied to nonresponders to yield probabilities corresponding to each holding pattern.  Nonresponders were assigned the pattern which displayed the highest probability.  As with the imputation of IRA holdings, equations were constructed for each of five respondent groups.  Independent variables included home ownership, IRA holdings, race /ethnicity, employment status, education, age, earnings, and PIA.

 Once an asset pattern was assigned, it was converted into imputations of the specific components:  savings, checking, and one of the five remaining asset types.  When partial asset holdings were reported, the imputed pattern was sustained for the missing portion of the pattern.  Thus, reported data were preserved.  From Table 10.6., 767 cases were imputed checking holdings flags while 823 cases were imputed savings holdings flags.

 For the remaining five assets (namely money market, CD, and credit union accounts, stocks and bonds), holdings flags were imputed using a regression approach.  The only exception, however, occurred for cases imputed as no checking and no savings.  Here, nonholding status was imputed to all five remaining assets (unless one or more of the five reported yes flags).  The linear equations included such predictors as home ownership, IRA, checking and savings indicators.  Flags indicating ownership of stocks and bonds were also included.  Other independent variables were similar to those employed in the modelling of IRA flags.  Predicted probabilities based on these regression models were converted to imputed flags using the same approach as the IRA holdings imputation.  Table 10.6. presents the results of the imputation scheme.

10.2.3.2.  Imputation of Asset Amounts

 Response dispositions of 14 asset amounts are presented in Table 10.7.  From Column 5, missing data rates ranged 1 percent for professional practice equity to 16 percent for savings and checking balances.  However, these rates are based on the entire sample, including cases for which asset holdings flags were missing as well.  Among households reporting asset holdings, nonresponse was considerably higher.  These data are displayed in Column 6.  Nonresponse among holdings reporters ranged from 11 percent for value of own home to 36 percent for stocks and professional practice equity.  Such high rates of nonresponse should be borne in mind when analyzing asset amounts items.

 In general, the methodology employed in the imputation of missing assets amounts was stochastic regression.  A three‑step procedure was used.  First, the equity on own home was imputed (when necessary) by stochastic regressions on the market value of home and debt on home.  Secondly, the net worth excluding own home was imputed (whenever necessary) via stochastic regression.  Finally, the proportion of net worth excluding own home was determined for each asset type (apart from own home) using regression.  These models were used in addition to imputed asset holdings and reported asset amounts to create the final imputations.

 To impute home market value, separate regression models were estimated for each of five respondent groups:  retired married females, retired married males, retired single persons, disabled married persons, and disabled single persons.  The log of the market value was the dependent variable and the independent variables were those used in the IRA regression models.  A stochastic term was added to the regression prediction by drawing at random from the empirical distribution of the residuals resulting from the model estimation.  This procedure is identical to that used in the imputation of annual earnings in Section 10.2.1.

 The debt on the home was imputed by first imputing zero versus positive debt; then, for those with positive debt, an amount was imputed.  To impute zero versus positive debt, a dichotomous regression model was estimated using the same predictors employed in the market value models.  Separate equations were estimated for the five respondent groups specified above.  The regression model yielded a predicted probability of having a positive debt.  Comparison of this value to a randomly generated number determined which debt status was imputed.  If the predicted probability was larger than the random number, a positive debt was imputed; otherwise, a zero debt was imputed.

 To impute the amount of positive debt, regression models were estimated with the dependent variable being the log of the ratio of the reported debt to the market value of the home.  The independent variables included various categorizations of race/ethnicity, sample subdomain and age, plus value of home, total income and earnings variables.  Stochastic residual terms were added to predictions in the fashion described for market value of home.

 The results of this imputation scheme are shown in the tenth row of Table 10.8.  The number of cases requiring imputation of equity on home was 1,584.  This represents about 12 percent of all cases reporting positive equities on own home.

 The imputation of the remaining asset amounts was based on the imputation of net worth excluding own home.  The imputation method mimicked that of market home value.  The dependent variable was the log of net worth exclusive of own home.  Independent variables included most of those used in the market value equations plus home value, and pension income and indicators for the 13 remaining asset types.  Five equations were estimated for each respondent group used in the market value equations.  A stochastic term was added to predicted values in a slightly more detailed fashion than that used for market value of home.

 To construct imputed asset amounts, 13 models were employed to predict the proportion of net worth exclusive of home held in that asset.  Each model was estimated using a data set consisting of households with that particular asset and at least one other type of asset.  Here, the dependent variable was the log of the asset share.  Predictors included variables reflecting the numbers of assets held, types of assets held, combinations of assets, respondent type and net worth.  The models were not estimated separately by response group.

 Predicted asset shares were calculated for each asset held for which amounts data were missing. The patterns of missing and reported assets amounts influenced the use of predicted asset shares. If all amounts were missing for held assets, the predicted asset shares were prorated to unity, and imputed net worth exclusive of home was apportioned among the missing amounts proportional to these shares. A total of 2,546 cases was handled in this fashion.

 When one or more asset amounts were reported (with none of them negative), one of three imputation schemes was used. If the reported amounts were less than 80 percent of the total imputed net worth excluding home, then the difference was divided among the missing amount fields in proportion to the imputed asset shares. If the reported amounts were greater than 80 percent but under four times the imputed net worth excluding home, then the missing amounts were imputed solely on the basis of the predicted shares. When the sum of reported assets exceeded four times the imputed net worth, imputed amounts were proportionately allotted the predicted share based on a modified asset amount. (That amount equalled the average of the reported assets amounts and four times the imputed net worth.) In total, roughly 2,700 cases were imputed in this fashion.

 Finally, one of three methods was employed when any of the reported asset amounts were negative. If the reported asset amounts sum was negative, missing amounts were imputed according to the method used for households with no reported amounts. If the sum of reported assets amounts was positive but less than the imputed net worth excluding home, missing amounts were imputed as the product of predicted asset shares and imputed net worth. When the reported sum exceeded imputed net worth, the missing amounts were imputed as the product of predicted asset shares and the implied asset sum. The implied asset sum was computed as in the second of the three methods used for cases with one or more reported asset amounts.

 Table 10.8. shows the number of cases imputed for each asset type. The largest numbers of cases imputed correspond to checking and savings amounts. However, the largest percentage of imputed nonzero amounts occurred with professional practice equity. Here, almost half the cases were imputed. Between one-fifth and one-third of nonzero asset amounts were imputed for all other asset types except respondent IRA, and equity in home and other property. It is imperative to remember that such large amounts of imputed data could possibly affect the outcome of statistical analyses.

10.2.3.3. Imputation of Income Flows

 Income flows (i.e., income from assets) were imputed for the 11 asset types shown in Table 10.9. From this table, it can be observed that missing data ranged from 1.4 percent for other property to 25 percent for savings accounts. Among asset holders, however, nonresponse was much higher. (See last column.) One-third or more of holders failed to report income flows for financial assets, except for checking. Missing data among holders is under 10 percent for the remaining asset types.

 Income flows were imputed by applying expected rates of return to the asset amounts. Expected rates of return were averaged observed rates of return estimated within ranges of asset amounts. Stochastic components were added to them as well. For all assets, a zero versus positive income indicator was first imputed. Imputed income flows were then calculated among those with positive income flags.

 For the IRA (respondent and spouse) and other property rent income flows, a regression approach was used to predict positive income flow. Predictors included 1) age, disability status and sample subdomain indicators for IRAs, and 2) income, PIA, occupation, race, property type, employment status and sample subdomain for other property assets. Predicted probabilities were compared to randomly generated probabilities to determine positive income status.

 The procedure to determine positive versus zero income flows for the remaining asset types was simpler. Proportions of households with no income from assets were calculated for ranges of asset amounts in each asset type. Income flags were then determined by comparing a randomly generated probability to the observed proportion in the same asset range with no income from those assets. If the random probability was less than the observed proportion, an asset income of zero was imputed. Otherwise, a positive asset amount was imputed.

 To impute positive income flows for all asset types, the expected log rate of return was calculated as the sum of a mean log rate and the product of a random normal deviate and a standard deviation. The imputed income flow is equal to the product of the reported asset amount and the exponentiation of the expected log rate of return.

 Table 10.10. presents the number of imputed income flows by response category. The table also shows the percentage of responses imputed by response category. Income flows from financial assets contained the largest number of cases imputed. Roughly, one to four thousand cases were imputed for these assets. Apart from checking, the accounts assets have nonzero income flows imputed for over 40 percent of the cases. Almost two-thirds of nonzero income and over two-thirds of "zero" income from bonds have been imputed. These rates of imputed data are high and should alert the analyst to take caution when drawing inferences from this data.

10.2.4. Imputation of Expected Future Pension and Miscellaneous Income Items

 This section describes the imputation of expected future pension and three miscellaneous income items. Expected future monthly payments from pensions were imputed for each of the CURRENT, LAST, and LONGEST jobs of the respondent and spouse. Expected recipiency flags for future pensions were not imputed. Moreover, future monthly payments were imputed whenever the individual was not currently receiving a pension from that job and expected to receive or did not know about a pension from that job in the future. Pensions were not imputed if future pension recipiency was missing.

 The three income types requiring imputation were income from boarders, income from repayments of a personal loan and other income. Both recipiency flags and amounts were imputed for these items. A single payment covering the 12 months preceding the interview was imputed for loan income. Monthly payments for each of the three months immediately preceding the interview were imputed for the other two income amount items.

10.2.4.1. Imputation of Expected Future Pension

 Table 10.11. presents the response disposition for the expected future pension recipiency questions asked of the respondent and spouse. Separate distributions are provided for each job type. The first column represents all cases for which imputation was performed. Pension amounts were imputed for roughly 2,000 respondents and spouses across three job types. Slightly over half of the imputations occurred for the expected pension from CURRENT job; under one-tenth occurred for the LONGEST job.

 Expected future pension payments were imputed using a model of wage replacement. Wage replacement was estimated among current pension recipients who reported pensions for LAST or LONGEST jobs. Wage replacement rates were defined as the ratio of monthly pension to the average monthly salary for that job. This was done separately for private, State/local, Federal civilian and military employer types. A regression equation was then estimated with the log wage replacement rate as the dependent variable. Predictors included industry, occupation, job duration, employer type, and education. Separate equations were employed for males and females. Expected future monthly pension was imputed as the product of the predicted wage replacement rate and the average monthly salary. A stochastic term was not utilized.

10.2.4.2. Imputation of Miscellaneous Income Receipt Flags

 The first three columns of Table 10.12. display the response disposition to the recipiency flags for three miscellaneous income types. Missing data rates were very small--one percent or lower. The third column represents the number of cases requiring imputation.

 The recipiency flag for income from boarders was imputed on the basis of a crosstabulation of home ownership, dwelling type, and a categorization of respondent status. Recipiency rates among reporters were observed in each cell. Next, nonresponders were sorted into these imputation cells. Nonreceipt was imputed to all cases except those falling into cells with relatively "high" rates of recipiency of 3.2 and 0.8 percent. The last two columns of Table 10.12. show that three cases were imputed recipients and 43 were imputed nonrecipients.

 Flags for loan income and other income were imputed via stochastic regression. Predictors included home ownership, financial asset holdings, respondent status, home value, income, pensions, and PIA. Estimated equations yielded predicted probabilities of being a recipient. Recipiency was imputed when a randomly generated probability was less than the predicted probability. The last two columns of Table 10.12. give the distribution of imputed recipiency flags.

10.2.4.3. Imputation of Miscellaneous Income Amounts

 The first four columns of Table 10.13. present the disposition of responses to three income amounts. Rates of nonresponse were low, ranging 0.3 to 1.5 percent. After imputation of recipiency status to the three income types, slightly over 200 fields required imputation of income amounts. (See Column 6.)

 The amounts for all three income types were imputed via stochastic regression. The dependent variables for the regressions were the logs of income amounts. The independent variables for loan income and other income were identical to those used in the recipiency flag equations. The income from boarders model employed home ownership, home values and income as predictors. A stochastic term was added to each regression prediction. The term was calculated as the product of a standard normal deviate and the standard error of the regression equation.

 Up to three monthly amounts were imputed for income from boarders and other income. The same regression based amount was imputed to each missing month. When one or two months had nonzero reported amounts, the average reported amount was assigned to the missing month(s).

 The last two columns of Table 10.13. display the distribution of amounts for each income type. Inapplicables represent cases with missing recipiency and amounts which were imputed as nonrecipients. Income from loans and other income accounted for most imputed values. Among reported and imputed recipients, imputations comprised 7.3 percent (18/247), ll.2 percent (90/804), and 9.5 percent (103/1079) of amounts from boarder income, loan income and other income, respectively.

10.3. Estimating Net Rent

 Question 234 asked for the respondent's best estimate of total income from rental properties. A question on income from rental properties net of expenses was inadvertently omitted from the questionnaire. A correction factor to transform gross revenue to net revenue was estimated by regressing net rent on gross rent using data from the Income Survey Development Program. The equation used was y = 69.2 - .344x where x is percentile of gross rent and y is the correction factor.

Table 2.1.  New Beneficiary Survey Universe and Completed Interviews

 MBR Universe1

 New Still

 Beneficiaries Alive

 in 12‑Month at Completed Survey Universe

 Window Extraction Interviews2 (in thousands)3

Total

Beneficiaries 1,927,195 1,860,347 17,155 1,805.1

 Retired

 Workers 1,298,047 1,274,789 9,519 1,244.0

 Men 734,129 716,510 5,307 692.4

 Women 563,918 558,279 4,212 551.6

 Disabled

 Workers 281,314 242,257 5,198 224.8

 Men 199,599 171,655 3,593 159.0

 Women 81,715 70,602 1,605 65.8

 Spouses

 Wives 216,361 213,332 1,041 209.4

 Widows 116,639 115,356 975 113.4

 Divorced Wives 5,618 5,517 210 5.3

 Surviving

 Divorced Wives 9,216 9,096 212 8.9

Nonbeneficiaries

(Medicare only) na 257,286 1,444 254.5

Total Completed Interviews 18,599

1New beneficiary universe was extracted in March 1982.  Nonbeneficiary universe was extracted in July 1982.  For all but disabled workers, 12‑month window is June 1980 ‑ May 1981.  For disabled workers, the window starts one month later, July 1980 ‑ June 1981. The total beneficiaries and the nonbeneficiaries, when combined, do not represent a meaningful total universe.

2Completed during October through December 1982.

3The size of the survey universe (living, noninstitutionalized persons at the time of the survey who were in the MBR universe) is estimated based on reasons for noninterview (death or institutionalized) obtained in the survey.  These numbers are based on weighted sample results obtained between October and December 1982.

 Table 3.1.  Sampling Rates for Categories of Respondents

 Ratio,

 Desired Popula‑ Sampling Intervals

 Number tion to (1 in):

 of Popula‑ Sample

Sample Subdomain Interviews tion Size Size Original Final

Retired Workers

 Male, Aged

 62 years 1,350 340,998 253.0 155.2 212.6

 63‑64 years 1,350 195,564 145.0 83.7 114.7

 65 years 1,300 112,418 86.5 48.9 67.0

 66+ years 1,000 67,540 67.5 38.5 52.7

 Female, Aged

 62 years 1,200 348,099 290.0 172.4 236.2

 63‑64 years 1,000 120,986 121.0 69.5 95.2

 65 years 1,000 64,505 64.5 36.8 50.4

 66+ years 800 24,689 30.9 17.4 23.8

Disabled Workers

 Male 3,450 171,655 49.8 29.5 40.4

 Female 1,550 70,602 45.5 28.4 38.9

Spouses

 Wives 1,000 213,332 213.0 124.4 170.4

 Widows 950 115,356 121.0 72.0 98.6

 Divorced Wives 200 5,517 27.6 16.5 22.6

 Surviving

 Divorced Wives 200 9,096 45.5 25.1 34.4

Medicare 1,500 257,286 171.5 98.0 140.0

NOTE:  The original sampling intervals were set approximately equal to the population size divided by 1.75 times the desired number of interviews.

 The final sampling intervals were equal to the original sampling intervals divided by 0.73 for all groups except the Medicare group.  The Medicare group has a sampling interval equal to the original one divided by 0.70; this reflects the use of only the main sample.  For the other groups, a reserve sample replicate equal to 10 percent of the total reserve was used (0.7 + 0.1 x 0.3 = 0.73).

 Table 3.2.  Final Disposition of the New Beneficiary Survey

 by Sample Subdomain

 DISPOSITION RESPONSE NONRESPONSE OTHER

 Sample Eligibility Not Total

 Subdomain Interviewed1 Eligible Unknown2 Eligible3 Selected

 (1) (2) (3) (4) (5)

Retired Workers

 Male, Aged

 62 years 1,442 52 142 37 1,673

 63‑64 years 1,466 65 114 53 1,698

 65 years 1,388 66 150 47 1,651

 66+ years 1,011 57 144 53 1,265

 Female, Aged

 62 years 1,319 67 133 19 1,538

 63‑64 years l,074 46 139 16 1,275

 65 years 1,045 61 152 25 1,283

 66+ years 774 59 141 35 1,009

Disabled Workers

 Male 3,593 160 333 290 4,376

 Female 1,605 60 144 127 1,936

Spouses

 Wives 1,041 57 123 22 1,243

 Widows 975 58 136 19 1,l88

 Divorced Wives 210 11 14 11 246

 Surviving

 Divorced Wives 212 14 26 6 258

Medicare 1,444 81 253 17 1,795

Total 18,599 914 2,144 777 22,434

1The distribution of sample type shown here differs from that of the final NBS data file due to revisions made in assigning sample type to respondents.

2Eligibility of a selected beneficiary is considered unknown if a screening form was not completed.

3Ineligibles include those persons who were deceased, who were institutionalized, who received first payment prior to that specified in the study population, etc.

 Table 4.1.  Response Rates for the New Beneficiary Survey

 by Sample Subdomain

 Overall

 Screening Interview Response

 Initial Rate2 Screened Rate3 Rate4

 Sample1 (in Sample (in (in

Sample Subdomain Size percent) Size percent) percent)

Retired Workers 11,107 90.0 9,992 95.3 85.7

 Male, Age:  6,097 91.0 5,547 95.7 87.0

 62 1,636 91.3 1,494 96.5 88.1

 63‑64 l,645 93.1 1,531 95.8 89.1

 65 1,604 90.6 1,454 95.5 86.5

 66+ 1,212 88.1 1,068 94.7 83.4

 Female, Age:  5,010 88.7 4,445 94.8 84.1

 62 1,519 91.2 1,386 95.2 86.8

 63‑64 1,259 89.0 1,120 95.9 85.3

 65 1,258 87.9 1,106 94.5 83.1

 66+ 974 85.5 833 92.9 79.5

Disabled Workers 5,895 91.9 5,418 95.9 88.2

 Male 4,086 91.9 3,753 95.7 87.9

 Female 1,809 92.0 1,665 96.4 88.7

Spouses 2,877 89.6 2,578 94.6 84.7

 Wives 1,221 89.9 1,098 94.8 85.3

 Widows 1,169 88.4 1,033 94.4 83.4

 Divorced Wives 235 94.0 221 95.0 89.4

 Surviving

 Divorced Wives 252 89.7 226 93.8 84.1

Medicare 1,778 85.8 1,525 94.7 81.2

Total 21,657 90.1 19,513 95.3 85.9

1Cases found to be ineligible are excluded from the sample sizes.

2Calculations based on initial sample sizes.

3Calculations based on screened eligible sample sizes.

4Calculations based on initial sample sizes which imply that all unscreened cases were eligible; in other words, a conservative rate is produced.

 Table 6.1.  Basic and Post‑Stratification Weights

 in the New Beneficiary Survey

 Basic Sampling Post‑Stratification

 Subdomain of Study Weight Weight Adjustment

 Male Retired Workers:

1. Age 62 212.6 1.048

2. Age 63‑64 114.7 0.998

3. Age 65 67.0 0.984

4. Age 66 and older 52.7 0.987

 Female Retired Workers:

5. Age 62 236.2 1.038

6. Age 63‑64 95.2 1.001

7. Age 65 50.4 1.003

8. Age 66 and older 23.8 1.045

 Disabled Workers:

9. Male 40.4 1.039

10. Female 38.9 1.075

 Spouses:

11. Wives 170.4 0.992

12. Widows 98.6 1.016

13. Divorced Wives 22.6 1.120

14. Surviving Divorced Wives 34.4 0.991

15. Medicare 140.0 1.024

 Table 6.2.  Nonresponse Weighting Cells Employed

 in Calculating Final and Interim Weights

 in the New Beneficiary Survey

 Final

 PSU PIA Weight

 Sample Subdomain Recode Cell (Quartiles) Cell

 MEDICARE NSR RURAL FIRST THREE 1

 MEDICARE NSR RURAL FOURTH 2

 MEDICARE SR GOOD ALL 3

 MEDICARE NSR SMSA FIRST THREE 4

 MEDICARE NSR SMSA ALL 5

 MEDICARE SR POOR FIRST THREE 6

 MEDICARE SR POOR ALL 7

 MALE RETIRED WORKERS NSR RURAL FIRST 8 (AGE‑62)

 MALE RETIRED WORKERS NSR RURAL SECOND 9 (AGE‑62)

 MALE RETIRED WORKERS NSR RURAL THIRD 10 (AGE‑62)

 MALE RETIRED WORKERS NSR RURAL FOURTH 11 (AGE‑62)

 MALE RETIRED WORKERS SR GOOD FIRST, SECOND 12 (AGE‑62)

 MALE RETIRED WORKERS SR GOOD THIRD, FOURTH 13 (AGE‑62)

 MALE RETIRED WORKERS NSR SMSA ALL 14 (AGE‑62)

 MALE RETIRED WORKERS SR POOR FIRST 15 (AGE‑62)

 MALE RETIRED WORKERS SR POOR LAST THREE 16 (AGE‑62)

 Table 6.2. (Continued)

 Final

 PSU PIA Weight

 Sample Subdomain Recode Cell (Quartiles) Cell

 MALE RETIRED WORKERS NSR RURAL ALL 17

 (AGE 63‑64)

 MALE RETIRED WORKERS SR GOOD FIRST 18 (AGE 63‑64)

 MALE RETIRED WORKERS SR GOOD SECOND, THIRD 19 (AGE 63‑64)

 MALE RETIRED WORKERS SR GOOD FOURTH 20 (AGE 63‑64)

 MALE RETIRED WORKERS NSR SMSA ALL 21 (AGE 63‑64)

 MALE RETIRED WORKERS SR POOR FIRST 22 (AGE 63‑64)

 MALE RETIRED WORKERS SR POOR SECOND 23 (AGE 63‑64)

 MALE RETIRED WORKERS SR POOR THIRD, FOURTH 24 (AGE 63‑64)

 MALE RETIRED WORKERS NSR RURAL FIRST THREE 25 (AGE 65)

 MALE RETIRED WORKERS NSR RURAL FOURTH 26 (AGE 65)

 MALE RETIRED WORKERS SR GOOD FIRST THREE 27 (AGE 65)

 MALE RETIRED WORKERS SR GOOD FOURTH 28 (AGE 65)

 MALE RETIRED WORKERS NSR SMSA FIRST THREE 29 (AGE 65)

 MALE RETIRED WORKERS NSR SMSA FOURTH 30 (AGE 65)

 MALE RETIRED WORKERS SR POOR FIRST THREE 31 (AGE 65)

 Table 6.2. (Continued)

 Final

 PSU PIA Weight

 Sample Subdomain Recode Cell (Quartiles) Cell

 MALE RETIRED WORKERS SR POOR FOURTH 32 (AGE 65)

 MALE RETIRED WORKERS NSR RURAL FIRST, SECOND 33 (AGE 66 AND OLDER)

 MALE RETIRED WORKERS NSR RURAL THIRD, FOURTH 34 (AGE 66 AND OLDER)

 MALE RETIRED WORKERS SR GOOD FIRST THREE 35 (AGE 66 AND OLDER)

 MALE RETIRED WORKERS SR GOOD FOURTH 36 (AGE 66 AND OLDER)

 MALE RETIRED WORKERS NSR SMSA FIRST, SECOND 37 (AGE 66 AND OLDER)

 MALE RETIRED WORKERS NSR SMSA THIRD, FOURTH 38 (AGE 66 AND OLDER)

 MALE RETIRED WORKERS SR POOR ALL 39 (AGE 66 AND OLDER)

 FEMALE RETIRED WORKERS NSR RURAL ALL 40 (AGE 62)

 FEMALE RETIRED WORKERS SR GOOD FIRST 41 (AGE 62)

 FEMALE RETIRED WORKERS SR GOOD LAST THREE 42 (AGE 62)

 FEMALE RETIRED WORKERS NSR SMSA ALL 43 (AGE 62)

 FEMALE RETIRED WORKERS SR POOR ALL 44 (AGE 62)

 FEMALE RETIRED WORKERS NSR RURAL ALL 45 (AGE 63‑64)

 FEMALE RETIRED WORKERS SR GOOD ALL 46 (AGE 63‑64)

 Table 6.2. (Continued)

 Final

 PSU PIA Weight

 Sample Subdomain Recode Cell (Quartiles) Cell

 FEMALE RETIRED WORKERS NSR SMSA FIRST, SECOND 47 (AGE 63-64)

 FEMALE RETIRED WORKERS NSR SMSA THIRD, FOURTH 48 (AGE 63-64)

 FEMALE RETIRED WORKERS SR POOR ALL 49 (AGE 63‑64)

 FEMALE RETIRED WORKERS NSR RURAL FIRST 50 (AGE 65)

 FEMALE RETIRED WORKERS NSR RURAL LAST THREE 51 (AGE 65)

 FEMALE RETIRED WORKERS SR GOOD FIRST THREE 52 (AGE 65)

 FEMALE RETIRED WORKERS SR GOOD FOURTH 53 (AGE 65)

 FEMALE RETIRED WORKERS NSR SMSA FIRST 54 (AGE 65)

 FEMALE RETIRED WORKERS NSR SMSA LAST THREE 55 (AGE 65)

 FEMALE RETIRED WORKERS SR POOR FIRST 56 (AGE 65)

 FEMALE RETIRED WORKERS SR POOR SECOND, THIRD 57 (AGE 65)

 FEMALE RETIRED WORKERS SR POOR FOURTH 58 (AGE 65)

 FEMALE RETIRED WORKERS NSR RURAL ALL 59 (AGE 66 AND OLDER)

 FEMALE RETIRED WORKERS SR GOOD FIRST 60 (AGE 66 AND OLDER)

 FEMALE RETIRED WORKERS SR GOOD SECOND, THIRD 61 (AGE 66 AND OLDER)

 Table 6.2. (Continued)

 Final

 PSU PIA Weight

 Sample Subdomain Recode Cell (Quartiles) Cell

 FEMALE RETIRED WORKERS SR GOOD FOURTH 62 (AGE 66 AND OLDER)

 FEMALE RETIRED WORKERS NSR SMSA FIRST, SECOND 63 (AGE 66 AND OLDER)

 FEMALE RETIRED WORKERS NSR SMSA THIRD, FOURTH 64 (AGE 66 AND OLDER)

 FEMALE RETIRED WORKERS SR POOR ALL 65 (AGE 66 AND OLDER)

 MALE DISABLED WORKERS NSR RURAL ALL 66

 MALE DISABLED WORKERS SR GOOD FIRST 67

 MALE DISABLED WORKERS SR GOOD LAST THREE 68

 MALE DISABLED WORKERS NSR SMSA FIRST THREE 69

 MALE DISABLED WORKERS NSR SMSA FOURTH 70

 MALE DISABLED WORKERS SR POOR ALL 71

 FEMALE DISABLED WORKERS NSR RURAL ALL 72

 FEMALE DISABLED WORKERS SR GOOD ALL 73

 FEMALE DISABLED WORKERS NSR SMSA ALL 74

 FEMALE DISABLED WORKERS SR POOR ALL 75

 WIVES NSR RURAL FIRST THREE 76

 WIVES NSR RURAL FOURTH 77

 WIVES SR GOOD FIRST, SECOND 78

 WIVES SR GOOD THIRD, FOURTH 79

 WIVES NSR SMSA FIRST THREE 80

 WIVES NSR SMSA FOURTH 81

 WIVES SR POOR FIRST, SECOND 82

 Table 6.2. (Continued)

 Final

 PSU PIA Weight

 Sample Subdomain Recode Cell (Quartiles) Cell

 WIVES SR POOR THIRD 83

 WIVES SR POOR FOURTH 84

 WIDOWS NSR RURAL FIRST, SECOND 85

 WIDOWS NSR RURAL THIRD, FOURTH 86

 WIDOWS SR GOOD FIRST, SECOND 87

 WIDOWS SR GOOD THIRD, FOURTH 88

 WIDOWS NSR SMSA ALL 89

 WIDOWS SR POOR ALL 90

 DIVORCED WIVES NSR RURAL ALL 91

 DIVORCED WIVES SR GOOD ALL 92

 DIVORCED WIVES NSR SMSA ALL 93

 DIVORCED WIVES SR POOR ALL 94

 SURVIVING DIVORCED WIVES NSR RURAL ALL 95

 SURVIVING DIVORCED WIVES SR GOOD ALL 96

 SURVIVING DIVORCED WIVES NSR SMSA ALL 97

 SURVIVING DIVORCED WIVES SR POOR ALL 98

 Table 7.1.  Variates Employed to Calculate Sampling Errors of

 Percentages for the New Beneficiary

 Survey by Group

I. Assets Variables

 1. Account holder of checking, savings or credit union account

 2. Respondent account holder of IRA or Keogh

 3. Spouse account holder of IRA or Keogh

 4. Home ownership

 5. Business, Farm or Professional Practice ownership

 6. Ownership of financial and real property

II. Health Variables

 1. Ability to work part‑time

 2. Duration of work limitation:  0‑12 months

 3. Duration of work limitation:  13‑36 months

 4. Duration of work limitation:  37+ months

 5. Cause of work limitation:  Accident on Job

 6. Cause of work limitation:  Bad Working Conditions

 7. Cause of work limitation:  Both Accident and Bad Working Conditions

 8. Medicaid coverage

 9. No health conditions

III. Income

 1. Earnings receipt

 2. Social Security receipt

 3. Interest receipt

 4. Veteran's Pension receipt

 5. Food Stamps receipt

 Table 7.1. (Continued)

IV. Job Characteristics

 1. Reason for leaving job:  Health

 2. Reason for leaving job:  Second Pension

 3. Longest job occupation:  Managerial

 4. Longest job occupation:  Technical, sales

 5. Longest job occupation:  Service

 6. Longest job occupation:  Precision Production

 7. Longest job occupation:  Operators

 8. Self‑employed workers for longest job

 9. Longest Job Industry category:  Manufacturing

 10. Longest Job Industry category:  Transportation, Communications

 11. Longest Job Industry category:  Wholesale and Retail Trade

 12. Longest Job Industry category:  Professional and Related Services

 13. Longest Job Industry category:  Other Industries

V. Work History

 1. 0‑9 years with Work since 1950

 2. 10‑19 years with Work since 1950

 3. 20‑32 years with Work since 1950

 4. No noncovered jobs (not covered under Social Security)

 5. 1‑7 noncovered jobs

 6. Federal and other noncovered employment

 7. Occupation if not currently employed:  Managerial, Professional

 8. Occupation if not currently employed:  Technical sales and administration

 support

 Table 7.1. (Continued)

 9. Occupation if not currently employed: Service

 10. Occupation if not currently employed:  Precision Production, crafts and repair

 11. Occupation if not currently employed:  Operators, fabricators and laborers

VI. Pension

 1. Receipt of Pension with Social Security

 2. Receipt of Federal pension and age at first eligibility is under 62

 3. Receipt of Federal pension and age at first eligibility is 62 and over

 4. Receipt of pension through profit‑sharing plan

 5. Receipt of pension decrease with Social Security increase

 6. Receipt of both private and military pensions

 Table 7.2.  Subgroups for Which Sampling Errors

 of Percentages Were Generated for

 the New Beneficiary Survey

1. Male Retired Workers

2. Female Retired Workers

3. Retired Workers Aged 62‑64

4. Retired Workers Aged 65+

5. Disabled Workers

6. Medicare‑only Recipients

7. Wives

8. Widows

9. Divorced Wives and Surviving Divorced Wives

10. Retired Worker Married Females plus Wife Beneficiaries

11. Retired Worker Widowed Females plus Widow Beneficiaries

12. Retired Worker Divorced Females plus Divorced Wife and Surviving Divorced Wife Beneficiaries

 Table 7.3.  Average DEFTs by Subgroup and Variable Group

 for the New Beneficiary Survey

 Variable Groups

 Job Work

 Assets Health Income Char's History Pension

Subgroup:

1.  Male Retired

 Workers 1.425 1.161 1.197 1.175 1.294 1.259

2.  Female Retired

 Workers 1.228 1.302 1.232 1.298 1.259 1.384

3.  Retired Workers

 Aged 62‑64 1.260 1.168 1.115 1.118 1.114 1.135

4.  Retired Workers

 Aged 65+ 1.465 1.142 1.116 1.163 1.193 1.369

5.  Disabled Workers 1.174 1.250 1.113 1.054 1.038 1.098

6.  Medicare‑only

 Recipients 1.282 1.098 0.946 1.063 1.044 0.934

7.  Wives 1.124 1.052 1.063 1.074 1.023 0.916

8.  Widows 1.049 0.971 1.022 1.067 1.086 1.063

9.  Divorced Wives &

 Surviving

 Divorced Wives 1.152 1.061 1.124 1.011 1.025 1.161

10. Retired Married

 Females, Wife 1.229 1.091 1.111 1.018 1.124 1.079

 Beneficiaries

11. Retired Widowed

 Females, Widow 1.206 1.167 1.134 1.102 1.264 1.465

 Beneficiaries

12. Retired

 Divorced

 Females,

 Divorced Wife 1.151 1.305 1.033 1.028 1.280 1.294

 and Surviving

 Divorced Wife

 Beneficiaries

 Table 7.4.  Unweighted Sample Sizes by Subgroup and

 Variable Group for the

 New Beneficiary Survey

 Variable Groups

 Job Work

 Assets Health Income Char's History Pension

Subgroup:

1.  Male Retired

 Workers 5,307 5,307 5,307 5,283 5,288 3,110

2.  Female Retired

 Workers 4,212 4,212 4,212 4,062 4,066 1,922

3.  Retired Workers

 Aged 62‑64 5,301 5,301 5,301 5,179 5,183 2,749

4.  Retired Workers

 Aged 65+ 4,218 4,218 4,218 4,166 4,171 2,283

5.  Disabled Workers 5,198 5,198 5,198 5,039 5,043 2,601

6.  Medicare‑only

 Recipients 1,444 1,444 1,444 1,444 1,444 329

7.  Wives 1,041 1,041 1,041 360 360 89

8.  Widows 975 975 975 753 753 263

9.  Divorced Wives &

 Surviving

 Divorced Wives 422 422 422 300 300 103

10. Retired Married

 Females, Wife 1,471 1,471 1,471 782 618 344

 Beneficiaries

11. Retired Widowed

 Females, Widow 2,106 2,106 2,106 1,868 1,869 842

 Beneficiaries

12. Retired

 Divorced

 Females,

 Divorced Wife 410 410 410 402 403 207

 and Surviving

 Divorced Wife

 Beneficiaries

 Table 7.5.  Weighted Totals Within Subgroups by Variable Group

 for the New Beneficiary Survey

 Variable Groups

 (Entries in

 Thousands) Job Work

 Assets Health Income Char's History Pension

Subgroup:

1.  Male Retired

 Workers 692.4 692.4 692.4 689.5 690.2 414.6

2.  Female Retired

 Workers 551.6 551.6 551.6 525.6 525.9 229.5

3.  Retired Workers

 Aged 62‑64 984.0 984.0 984.0 958.0 958.7 499.9

4.  Retired Workers

 Aged 65+ 260.0 260.0 260.0 257.1 257.3 144.2

5.  Disabled Workers 224.8 224.8 224.8 217.9 218.1 112.9

6.  Medicare‑only

 Recipients 254.5 254.5 254.5 254.5 254.5 59.1

7.  Wives 209.4 209.4 209.4 71.8 71.8 17.8

8.  Widows 113.4 113.4 113.4 87.8 87.8 30.7

9.  Divorced Wives &

 Surviving

 Divorced Wives 14.2 14.2 14.2 10.7 10.7 3.8

10. Retired Married

 Females, Wife 254.5 254.5 254.5 116.4 116.4 45.3

 Beneficiaries

11. Retired Widow

 Females, Widow 205.2 205.2 205.2 177.1 177.2 75.1

 Beneficiaries

12. Retired

 Divorced

 Females,

 Divorced Wife 46.6 46.6 46.6 45.4 45.4 22.7

 and Surviving

 Divorced Wife

 Beneficiaries

 Table 7.6.  Conservative Generalized Sampling Errors1

 of Estimated Proportions in the

 New Beneficiary Survey

 Excluding Retired Workers Aged 65 and

 Over and Disabled Workers

 Estimated Percentage

 Size of

 Base (in 2 or 5 or 10 or 20 or 25 or 30 or 40 or

Thousands) 98 95 90 80 75 70 60 50

 5 1.3 2.0 2.7 3.6 3.9 4.1 4.4 4.5

 15 1.0 1.5 2.1 2.8 3.1 3.2 3.5 3.5

 50 0.7 1.1 1.5 2.0 2.2 2.3 2.5 2.5

 100 0.6 0.9 1.2 1.6 1.7 1.8 1.9 2.0

 175 0.4 0.7 1.0 1.3 1.4 1.5 1.6 1.6

 250 0.4 0.6 0.8 1.1 1.2 1.3 1.4 1.4

 400 0.3 0.5 0.7 0.9 1.0 1.0 1.1 1.1

 600 0.3 0.4 0.6 0.8 0.8 0.9 0.9 1.0

 900+ 0.2 0.4 0.5 0.6 0.7 0.7 0.8 0.8

1Sampling errors are based on a paired selections model for calculating sampling errors from complex samples.  The sampling errors shown here are given in percentage points and are equal to one standard deviation of an estimated percentage.

 Table 7.7.  Conservative Generalized Sampling Errors1 of

 Estimated Proportions in the New Beneficiary

 Survey for Retired Workers Aged 65 and Over

 and Disabled Workers

 Estimated Percentage

 Size of

 Base (in 2 or 5 or 10 or 20 or 25 or 30 or 40 or

Thousands) 98 95 90 80 75 70 60 50

 5 1.0 1.6 2.2 2.9 3.2 3.4 3.6 3.7

 15 0.8 1.3 1.8 2.4 2.6 2.7 2.9 3.0

 50 0.6 1.0 1.3 1.8 1.9 2.0 2.1 2.2

 100 0.5 0.8 1.1 1.4 1.5 1.6 1.7 1.8

 175 0.4 0.6 0.8 1.1 1.2 1.3 1.3 1.4

 250 0.4 0.6 0.8 1.0 1.1 1.2 1.2 1.3

 400 0.3 0.5 0.6 0.8 1.0 1.0 1.0 1.1

 600 0.2 0.4 0.5 0.7 0.8 0.8 0.9 0.9

 900+ 0.2 0.3 0.5 0.6 0.7 0.7 0.7 0.8

1Sampling errors are based on a paired selections model for calculating sampling errors from complex samples.  The sampling errors shown here are given in percentage points and are equal to one standard deviation of an estimated percentage.

 Table 7.8.  Weighted1 Distribution of Current Working Status2

 for Three Subdomains in the New Beneficiary Survey

 Subdomain3

 Male Retired Female Retired Divorced

 Workers Age 65 Workers Age 65 Wives

Currently

 Working 34,127 20,139 516

 (%) (31.4%) (31.9%) (9.8%)

Currently Not

 Working 74,710 42,945 4,763

 (%) (68.6%) (68.1%) (90.2%)

Total Valid4

 Cases 108,837 63,084 5,279

 (%) (100%) (100%) (100%)

1Data are weighted using V1426, FINLWGT (final weight).

2Current working status is obtained from question 9 of the questionnaire, V57 on the public use tape, Q9YN.

3Subdomain is determined using V1425, NEWSAMPT (new sample type).

4Codes 96 and 98 in V57 have been excluded from this tabulation.

 Table 7.9.  Weighted Totals, Unweighted Sample Sizes and DEFTs

 by Contrast Group for Six Asset Variables

 Unweighted

Contrast Total #1 Total #2 Sample Sizes

 Group #1 (in thousands) (in thousands) n1 n2 DEFT

 1 126.9 18.2 758 418 0.9182

 1 126.9 18.2 758 418 1.0259

 1 126.9 18.1 758 416 1.0456

 1 126.9 18.2 758 418 1.981

 1 126.9 18.2 758 418 1.0339

 2 209.4 368.3 1041 2241 0.9308

 2 209.4 368.3 1041 2241 0.997

 2 196.6 368.3 978 2241 1.0766

 2 205.4 365.6 1020 2225 0.8605

 2 209.4 368.3 1041 2241 1.327

 2 209.4 368.3 1041 2241 0.9183

 3 692.5 211.5 5307 1202 1.1136

 3 692.5 211.5 5307 1202 1.0775

 3 580.1 188.7 4483 1075 0.9612

 3 686.5 211.0 5264 1199 1.4109

 3 692.5 211.5 5307 1202 0.9126

 3 692.5 211.5 5307 1202 1.2973

1Contrast Group #1:  Retired Workers Aged 62‑64 versus Disabled Workers Aged 62‑64.

 Contrast Group #2:  Wives versus Married Retired Workers.

 Contrast Group #3:  Retired Worker Males versus Medicare‑only Males.

 The rows of this table correspond to specific variables used in calculating sampling errors; occasionally, a variable was deleted when either there were an insufficient number of cases for valid calcuation or the estimated contrast was too close to zero or one.

 Table 7.10.  Conservative Generalized Sampling Errors1

 of Estimated Percentages2 for

 Differences Between Two Subgroups

 n1 (in Thousands)

 n2 (in

 Thousands) 10 15 50 100 175 250 400 600 900+

 10 1.3

 15 1.2 1.2

 p\* = 1 or 99 percent

 50 1.0 1.0 0.8

 100 0.9 0.8 0.7 0.6

 175 0.7 0.7 0.6 0.6 0.5

 250 0.7 0.7 0.6 0.5 0.5 0.5

 400 0.6 0.6 0.5 0.5 0.4 0.4 0.4

 600 0.5 0.5 0.5 0.4 0.4 0.4 0.3 0.3

 900+ 0.4 0.4 0.4 0.4 0.3 0.3 0.3 0.3 0.3

1Sampling errors are based on a paired selections model for calculating sampling errors from complex samples.  The sampling errors shown here are given in percentage points and are equal to one standard deviation of an estimated percentage.

2Arbitrarily, we assume n2 > n1.  To use the proper table, you must calculate p\* = (n1p1 + n2p2)/(n1 + n2) where p1 and p2 are the proportions being contrasted; n1 and n2 are the weighted totals in groups 1 and 2, respectively.  Once p\* is calculated, turn to the table which has a value closest to p\* in the upper right corner (in percent).

 Table 7.11.  Conservative Generalized Sampling Errors1

 of Estimated Percentages2 for

 Differences Between Two Subgroups

 n1 (in Thousands)

 n2 (in

Thousands) 10 15 50 100 175 250 400 600 900+

 10 2.9

 15 2.7 2.6

 p\* = 5 or 95 percent

 50 2.2 2.1 1.8

 100 1.9 1.8 1.6 1.4

 175 1.6 1.6 1.4 1.3 1.1

 250 1.5 1.4 1.3 1.2 1.1 1.0

 400 1.3 1.2 1.1 1.0 1.0 0.9 0.8

 600 1.1 1.1 1.0 0.9 0.9 0.8 0.7 0.7

 900+ 1.0 0.9 0.9 0.8 0.8 0.7 0.7 0.6 0.6

1Sampling errors are based on a paired selections model for calculating sampling errors from complex samples.  The sampling errors shown here are given in percentage points and are equal to one standard deviation of an estimated percentage.

2Arbitrarily, we assume n2 > n1.  To use the proper table, you must calculate p\* = (n1p1 + n2p2)/(n1 + n2) where p1 and p2 are the proportions being contrasted; n1 and n2 are the weighted totals in groups 1 and 2, respectively.  Once p\* is calculated, turn to the table which has a value closest to p\* in the upper right corner (in percent).

 Table 7.12.  Conservative Generalized Sampling Errors1

 of Estimated Percentages2 for

 Differences Between Two Subgroups

 n1 (in Thousands)

 n2 (in

Thousands) 10 15 50 100 175 250 400 600 900+

 10 3.9

 15 3.8 3.6

 p\* = 10 or 90 percent

 50 3.1 2.9 2.5

 100 2.6 2.5 2.2 1.9

 175 2.3 2.2 1.9 1.7 1.6

 250 2.0 2.0 1.8 1.6 1.5 1.4

 400 1.7 1.7 1.5 1.4 1.3 1.2 1.1

 600 1.5 1.5 1.4 1.3 1.2 1.1 1.0 0.9

 900+ 1.3 1.3 1.2 1.1 1.0 1.0 0.9 0.9 0.8

1Sampling errors are based on a paired selections model for calculating sampling errors from complex samples.  The sampling errors shown here are given in percentage points and are equal to one standard deviation of an estimated percentage.

2Arbitrarily, we assume n2 > n1.  To use the proper table, you must calculate p\* = (n1p1 + n2p2)/(n1 + n2) where p1 and p2 are the proportions being contrasted; n1 and n2 are the weighted totals in groups 1 and 2, respectively.  Once p\* is calculated, turn to the table which has a value closest to p\* in the upper right corner (in percent).

 Table 7.13.  Conservative Generalized Sampling Errors1

 of Estimated Percentages2 for

 Differences Between Two Subgroups

 n1 (in Thousands)

 n2 (in

Thousands) 10 15 50 100 175 250 400 600 900+

 10 5.7

 15 5.4 5.2

 p\* = 25 or 75 percent

 50 4.4 4.2 3.6

 100 3.8 3.6 3.2 2.8

 175 3.3 3.2 2.8 2.5 2.3

 250 2.9 2.8 2.5 2.3 2.1 2.0

 400 2.5 2.5 2.2 2.1 1.9 1.8 1.6

 600 2.2 2.1 2.0 1.8 1.7 1.6 1.5 1.4

 900+ 1.9 1.9 1.7 1.6 1.5 1.4 1.3 1.2 1.1

1Sampling errors are based on a paired selections model for calculating sampling errors from complex samples.  The sampling errors shown here are given in percentage points and are equal to one standard deviation of an estimated percentage.

2Arbitrarily, we assume n2 > n1.  To use the proper table, you must calculate p\* = (n1p1 + n2p2)/(n1 + n2) where p1 and p2 are the proportions being contrasted; n1 and n2 are the weighted totals in groups 1 and 2, respectively.  Once p\* is calculated, turn to the table which has a value closest to p\* in the upper right corner (in percent).

 Table 7.14.  Conservative Generalized Sampling Errors1

 of Estimated Percentages2 for

 Differences Between Two Subgroups

 n1 (in Thousands)

 n2 (in

Thousands) 10 15 50 100 175 250 400 600 900+

 10 6.4

 15 6.1 5.8

 p\* = 40 or 60 percent

 50 5.0 4.8 4.0

 100 4.3 4.1 3.6 3.2

 175 3.7 3.6 3.1 2.8 2.6

 250 3.3 3.2 2.9 2.6 2.4 2.2

 400 2.8 2.8 2.5 2.3 2.1 2.0 1.8

 600 2.5 2.4 2.2 2.1 1.9 1.8 1.7 1.5

 900+ 2.1 2.1 2.0 1.8 1.7 1.6 1.5 1.4 1.3

1Sampling errors are based on a paired selections model for calculating sampling errors from complex samples.  The sampling errors shown here are given in percentage points and are equal to one standard deviation of an estimated percentage.

2Arbitrarily, we assume n2 > n1.  To use the proper table, you must calculate p\* = (n1p1 + n2p2)/(n1 + n2) where p1 and p2 are the proportions being contrasted; n1 and n2 are the weighted totals in groups 1 and 2, respectively.  Once p\* is calculated, turn to the table which has a value closest to p\* in the upper right corner (in percent).

 Table 7.15.  Conservative Generalized Sampling Errors1

 of Estimated Percentages2 for

 Differences Between Two Subgroups

 n1 (in Thousands)

 n2 (in

Thousands) 10 15 50 100 175 250 400 600 900+

 10 6.6

 15 6.3 6.0

 p\* = 50 percent

 50 5.1 4.9 4.1

 100 4.3 4.2 3.6 3.2

 175 3.8 3.6 3.2 2.9 2.6

 250 3.4 3.3 2.9 2.7 2.4 2.3

 400 2.9 2.8 2.6 2.4 2.2 2.0 1.9

 600 2.5 2.5 2.3 2.1 2.0 1.9 1.7 1.6

 900+ 2.2 2.1 2.0 1.9 1.8 1.7 1.5 1.4 1.3

1Sampling errors are based on a paired selections model for calculating sampling errors from complex samples.  The sampling errors shown here are given in percentage points and are equal to one standard deviation of an estimated percentage.

2Arbitrarily, we assume n2 > n1.  To use the proper table, you must calculate p\* = (n1p1 + n2p2)/(n1 + n2) where p1 and p2 are the proportions being contrasted; n1 and n2 are the weighted totals in groups 1 and 2, respectively.  Once p\* is calculated, turn to the table which has a value closest to p\* in the upper right corner (in percent).

 Table 7.16.  Weighted1 Distribution of Current Working Status2

 for Female Retired Workers and Wives in the

 New Beneficiary Survey

 Female Retired

 Workers3 Wives3

 Currently

 Working 116,215 7,901

 (%) (21.1%) (3.8%)

 Currently Not

 Working 435,417 201,312

 (%) (78.9%) (96.2%)

 Total Valid

 Cases4 551,632 209,213

 (%) (100%) (100%)

1Data are weighted using V1426, FINLWGT (final weight).

2Current working status is obtained from question 9 of the questionnaire, V57 on the public use tape, Q9YN.

3Subdomain is determined using V1425, NEWSAMPT (new sample type).

4Codes 96 and 98 in V57 have been excluded from this tabulation.

 Table 7.17.  New Beneficiary Survey Sampling Errors

 of Estimated Totals

 Estimated Total Sampling Error CV (%)

Retired Workers‑‑Men 694,900 1,660 0.2

 Age 62 332,800 1,440 0.4

 Age 63‑64 188,900 820 0.4

 Age 65 108,900 550 0.5

 Age 66 and older 64,300 410 0.6

Retired Workers‑‑Women 549,400 1,200 0.2

 Age 62 343,300 1,050 0.3

 Age 63‑64 119,400 410 0.3

 Age 65 63,000 360 0.6

 Age 66 and older 23,700 170 0.7

Disabled Workers 224,900 830 0.4

 Men 159,300 780 0.5

 Women 65,600 370 0.6

Wives 209,300 950 0.5

Widows 113,400 400 0.4

Divorced Wives and

Surviving Divorced Wives 14,100 120 0.8

Nonbeneficiaries

(Medicare only) 254,500 660 0.3

 Table 7.18.  New Beneficiary Survey Estimates and

 Sampling Errors of the Deceased and

 Institutionalized Universe Population

 Estimated Total Sampling Error CV (%)

Retired Workers‑‑Men 39,600 1,590 4.0

 Age 62 16,000 1,410 8.8

 Age 63‑64 11,800 840 7.1

 Age 65 6,100 450 7.4

 Age 66 and older 5,700 390 6.8

Retired Workers‑‑Women 14,800 1,050 7.1

 Age 62 8,200 990 12.0

 Age 63‑64 3,000 360 12.0

 Age 65 2,200 200 9.1

 Age 66 and older 1,400 140 10.1

Disabled Workers 56,000 830 1.5

 Men 39,900 780 2.0

 Women 16,100 370 2.3

Wives 7,100 890 12.5

Widows 3,300 420 12.5

Divorced Wives and

Surviving Divorced Wives 700 130 18.9

Nonbeneficiaries

(Medicare only) 2,800 660 23.6

 Table 7.19.  Illustration of the Use of Half‑Sample Replicates

 to Produce Estimates of Sampling Error

 in the New Beneficiary Survey

 Monthly Replicate Half‑Sample Indicators

 Case Pension

 # Income (in

 Dollars) 1 2 3 4 5 6 7 8

 1 90 1 0 0 1 0 1 1 0

 2` 75 0 0 1 1 1 0 0 1

 3 100 1 1 0 0 1 0 1 0

 4 110 0 1 1 0 0 1 0 1

 5 85 1 1 1 0 0 1 0 0

 6 93 0 0 0 1 1 0 1 1

 7 79 0 1 1 1 0 0 1 0

 8 87 1 0 0 0 1 1 0 1

 9 98 1 0 1 1 1 0 0 0

 10 80 0 1 0 0 0 1 1 1

ts = 89.7 ti 92.0 90.8 89.4 87.0 90.6 90.4 88.4 89.0

 (ti ‑ ts)2 5.29 1.21 0.09 7.29 0.81 0.49 1.69 0.49

Var(ts) = {SUM:(ti ‑ ts)2}/8 = 17.36/8 = 2.17

SE(ts) = (2.17)1/2 = 1.473

Note:  ts is the mean in this example.

 Table 7.20.  Illustration of Half‑Sample Means and Sampling Error

 of the Average Quarterly Social Security Benefit Among

 Retired Male Workers in the New Beneficiary Survey

 Replicate Weighted No.  Unweighted No.  Average Quarterly1

Half‑Sample Valid Cases2 Valid Cases Social Security Benefit

 1 332,986 2,562 1,833.040

 2 330,110 2,525 1,901.755

 3 334,304 2,564 1,884.310

 4 343,032 2,613 1,899.025

 5 330,243 2,533 1,888.202

 6 349,646 2,618 1,854.245

 7 337,660 2,620 1,866.787

 8 341,960 2,613 1,915.330

 9 330,865 2,536 1,879.694

 10 338,345 2,605 1,873.365

 11 345,656 2,631 1,894.561

 12 337,213 2,536 1,893.531

 13 338,277 2,584 1,879.818

 14 336,059 2,540 1,890.200

 15 335,403 2,550 1,907.820

 16 336,717 2,600 1,893.217

 17 342,390 2,600 1,872.482

 18 344,733 2,636 1,886.245

 19 336,829 2,587 1,881.819

 20 335,638 2,546 1,875.155

 21 337,353 2,555 1,882.475

 22 336,419 2,552 1,875.847

 23 341,164 2,609 1,884.312

 Table 7.20. (Continued)

 Replicate Weighted No.  Unweighted No.  Average Quarterly1

Half‑Sample Valid Cases2 Valid Cases Social Security Benefit

 24 340,747 2,597 1,880.774

 25 339,444 2,588 1,902.115

 26 338,115 2,584 1,894.136

 27 339,704 2,576 1,869.776

 28 343,557 2,608 1,893.176

 29 343,059 2,596 1,872.393

 30 342,316 2,603 1,881.808

 31 341,517 2,573 1,856.230

 32 352,284 2,662 1,843.900

 33 345,746 2,655 1,889.270

 34 346,525 2,641 1,863.071

 35 346,139 2,651 1,872.003

 36 340,306 2,571 1,868.848

 37 332,425 2,558 1,858.274

 38 337,767 2,550 1,872.923

 39 333,611 2,564 1,890.516

 40 343,613 2,588 1,891.287

 41 341,526 2,607 1,870.780

 42 338,831 2,606 1,887.859

 43 338,720 2,571 1,855.130

 44 341,130 2,612 1,892.467

 45 338,850 2,597 1,895.446

 46 344,013 2,611 1,877.722

 47 344,982 2,621 1,855.412

 Table 7.20. (Continued)

 Replicate Weighted No.  Unweighted No.  Average Quarterly1

Half‑Sample Valid Cases2 Valid Cases Social Security Benefit

 48 348,910 2,634 1,854.479

 49 347,986 2,654 1,870.969

 50 340,646 2,592 1,897.335

 51 336,216 2,573 1,910.518

 52 337,603 2,574 1,883.519

Total Sample 679,973 5,182 1,880.979

Vary = (1/52) {SUM:(yi ‑ ys)2} = 11,830.911

SEy = 108.77

where ys is the total sample average.

1This variable is defined as the quarterly Social Security Benefit (V1486, QSS).  Averages shown here are for those who received nonzero amounts and did not receive Social Security jointly with railroad retirement income.

2Valid cases include only those which showed nonzero, nonmissing amounts.  For illustrative purposes, retired‑worker men are defined by sample types (the variable number V1850 SAMPTYPE) equalling 1,2,3,4.

 Table 9.1.  List of Critical Items1 for the

 New Beneficiary Survey

Question # Content

 Marital Status/Household Composition

 1 Marital Status

 Job History

 9 Currently Working

 10 Year Last Worked

 12 Date R Stopped Work (Last or Longest Employer)

 13 Date R Started Working (Current/Last/Longest)

 Current Job

 17 Employed/Self‑Employed

 22 Hours Worked per Week

 23 Weeks Worked per Year (Self‑Employed)

 24 Salary (Self‑Employed)

 26 Keogh Account (Self‑Employed)

 27 Employment Status (Employee)

 32 Hours Worked per Week (Employee)

 33 Hours Worked per Year (Employee)

 34 Salary (Employee)

 40 Retirement Plan (Employee)

 43 Receiving Retirement Benefits (Employee)

 Last Employment

 51 Employed/Self‑Employed

 56 Hours Worked per Week (Self‑Employed)

 57 Weeks Worked per Year (Self‑Employed)

 58 Salary (Self‑Employed)

 60 Keogh Account (Self‑Employed)

 61 Employment Status (Employee)

 67 Weeks Worked per Year (Employee)

 68 Salary (Employee)

 69 F.I.C.A. Deducted (Employee)

 70 Covered by Pension Plan (Employee)

 74 Receiving Payments from Plan

 Longest Employment

 92 Employed/Self‑Employed

 98 Weeks Worked per Year (Self‑Employed)

 99 Salary (Self‑Employed)

 101 Koegh Account (Self‑Employed)

 102 Employment Status (Employee)

 108 Weeks Worked per Year (Employee)

 Table 9.1. (Continued)

Question # Content

 Longest Employment (Continued)

 109 Salary (Employee)

 111 Covered by Pension Plan (Employee)

 115 Receiving Payment from Plan (Employee)

 Health

 144 Health condition limiting work for pay

 145 Health condition limiting work at home

 Income

165 (a‑f) Receiving Social Security, etc., benefits

 (Not Currently Married)

171 (a‑m) Receiving other income

 (Not Currently Married)

172 (b‑e) Amount received last month2

 (Not Currently Married)

173 (b‑e) Amount received two months ago2

 (Not Currently Married)

174 (b‑e) Amount received three months ago2

 (Not Currently Married)

182 (a‑f) Receiving Social Security, etc., benefits

 (Currently Married)

190 (a‑m) Receiving other income

 (Currently Married)

191 (b‑e) Respondent or spouse benefit

 (Currently Married)

192 (b‑e) Amount received last month2

 (Currently Married)

193 (b‑e) Amount received two months ago2

 (Currently Married)

194 (b‑e) Amount received three months ago2

 (Currently Married)

 Assets

 201 Money in savings or other assets

 204 Holding any bonds

 207 Other stocks or shares in mutual funds

 210 Keogh Account or IRA

 212 Regular payments or withdrawals from account

 213 How much received in last three months

 214 Keogh Account or IRA‑‑spouse

 216 Regular payments or withdrawals‑‑spouse

 217 Amount spouse received in last three months

 220 Own or buying current residence

 Table 9.1. (Continued)

Question # Content

 Assets (Continued)

 246 Total monies received in reference month

 Marital History

 250 Spouse's full name

 251 Spouse's date of birth

 252 Spouse's Social Security number

 Disability Program Information

 276 Disability benefits and Medicare

 285 Trial work periods

 Spouse Work History

 S9 Currently working

 S10 Year last worked

 Spouse Current Employment

 S17 Employed/Self‑Employed

 S22 Hours Worked per Week (Self‑Employed)

 S27 Employment Status (Employee)

 S32 Hours Worked per Week (Employee)

 S33 Weeks Worked per Year (Employee)

 S144 Health condition or handicap limits paid work

 S145 Health condition or handicap limits housework

1If a critical item was incorrectly skipped or the response recorded was inadequate, the interviewer was directed to recontact the respondent to obtain the necessary information.

2Response was acceptable if amount was recorded for any one of the three months.

 Table 10.1.  Income and Income‑Related Items1 Imputed

 in the New Beneficiary Survey

1. Earnings Items (Amounts)

 (V2248 ‑ V2259)

 a) Earnings from Current Job

 b) Earnings from Last Job

 c) Earnings from Longest Job

2. Income Types (Recipiency and Amounts)

 (V2260 ‑ V2544)

 a) Social Security

 b) Supplemental Security Income

 c) Railroad Retirement

 d) Black Lung

 e) Veteran's Benefits

 f) Welfare

 g) Earnings

 h) State/Local Pension

 i) Military Pension

 j) Federal Pension

 k) Private Pension

 l) Annuities

 m) Worker's Compensation

 n) Unemployment Compensation

 o) Alimony

 p) Estate/Trust/Royalties

 q) Household Transfer

 r) Interhousehold Transfer

 s) Food Stamps

3. Assets (Holdings, Amounts and Income)

 (V2545 ‑ V2576, V2581 ‑ V2593)

 a) Money Market

 b) Certificates of Deposit

 c) Savings

 d) Credit Union

 e) Checking

 f) Bonds

 g) Stocks

 h) Respondent IRA

 i) Spouse IRA

 j) Own Home

 k) Other Property

 l) Business Equity

 m) Professional Practice

 n) Farm Equity

 Table 10.1. (Continued)

4. Future Pensions and Miscellany (Recipiency and Amounts)

 (V2928 ‑ V2933, V2577 ‑ V2580, V2594 ‑ V2599)

 a) Expected Future Monthly Pension from Current Job

 b) Expected Future Monthly Pension from Last Job

 c) Expected Future Monthly Pension from Longest Job

 d) Income from Boarders

 e) Income from Repayment of Personal Loan

 f) Other Income

1Both respondents' and spouses' amounts and recipiencies were computed for all items.

 Table 10.2.  Summary of Responses and Imputations

 to Income Recipiency Flags

 Responses Imputations

 Income

 Type Yes No Missing Yes No

Social Security 16,983 1,574 42 38 4

SSI 675 17,860 64 2 62

Railroad Retirement 141 18,398 60 0 60

Black Lung 84 18,455 60 1 59

Veteran's Benefits 1,144 17,393 62 3 59

Welfare 147 18,387 65 0 65

Earnings 7,479 10,999 121 64 57

State/Local Pension 1,666 16,866 67 12 55

Military Pension 481 18,046 72 3 69

Federal Pension 781 17,748 70 5 65

Private Pension 5,256 13,265 78 19 59

Annuities 718 17,784 99 8 91

Worker's Compensation 168 18,350 81 0 81

Unemployment Compensation 178 18,303 78 0 78

Alimony 98 18,420 81 0 81

Estate/Trust/Royalties 279 18,235 85 4 81

Household Transfer 332 18,186 81 1 80

Interhousehold Transfer 196 18,319 84 0 84

Food Stamps 839 17,682 78 0 78

 Table 10.3.  Methodology Employed to Impute Missing

 Income Amounts in the New Beneficiary

 Survey by Income Type

 Income Type Imputation Method

 Social Security Benefit Formula

 SSI Cell Mean

 Railroad Retirement Cell Mean

 Black Lung Cell Mean

 Veteran's Benefits Stochastic Regression

 Welfare Cell Mean

 Earnings Stochastic Regression

 State/Local Pension Stochastic Regression

 Military Pension Stochastic Regression

 Federal Pension Stochastic Regression

 Private Pension Stochastic Regression

 Annuities Stochastic Mean

 Worker's Compensation Cell Mean

 Unemployment Compensation Cell Mean

 Alimony Cell Mean

 Estate/Trust/Royalties Smoothed Mean/Global Mean

 Household Transfer Cell Mean

 Interhousehold Transfer Cell Mean

 Food Stamps Benefit Formula

 Table 10.4.  Summary of Responses and Imputations

 for Income Amounts:  Cases with

 Reported or Imputed Flags

 Households

 Total with Imputed Households Households Income as a

 Households with One with Percentage of

 with Fully or More Entirely All Households

 Reported Imputed Imputed with that

 Income Type Amounts Amounts Amounts Income Type

Social Security 16,514 507 344 3.0

SSI 600 17 13 2.5

Railroad

Retirement 121 20 17 14.2

Black Lung 79 6 6 7.1

Veteran's

Benefits 1,089 58 53 5.1

Welfare 141 6 6 4.1

Earnings 6,618 925 ‑‑1 12.3

State/Local

Pension 1,573 105 102 6.3

Military

Pension 441 43 42 8.9

Federal Pension 706 80 78 10.2

Private Pension 4,944 331 ‑‑1 6.3

Annuities 650 76 75 10.5

Worker's

Compensation 162 6 6 3.6

Unemployment

Compensation 170 8 7 4.5

Alimony 96 2 2 2.0

Estate/Trust/

Royalties 252 31 30 11.0

 Table 10.4. (Continued)

 Households

 Total with Imputed

 Households Households Income as a

 Households with One with Percentages of

 with Fully or More Entirely All Households

 Reported Imputed Imputed with that

 Income Type Amounts Amounts Amounts Income Type

Household

Transfer 313 20 17 6.0

Interhousehold

Transfer 182 14 12 7.1

Food Stamps 828 11 8 1.3

1Information not available.  For income types with substantial imputations the record dumps from which these columns were constructed generally did not contain sufficient data to permit derivation of these statistics.

 Table 10.5. Response Disposition to Asset Holdings Flags

 in the New Beneficiary Survey

 Holdings Flag

 Asset Type

 Yes No Missing % Missing

Money market 3,965 13,845 789 4.2

CD 5,100 12,655 844 4.5

Savings 10,320 7,456 823 4.4

Credit union 2,398 15,415 786 4.2

Checking 13,584 4,248 767 4.1

Bonds 2,591 15,682 326 1.8

Stocks 2,804 15,438 357 1.9

IRA (Respondent) 1,982 16,480 137 0.7

IRA (Spouse)1 1,158 11,026 136 1.1

Own home2 13,589 4,840 62 0.3

Other property3

 Rental housing 982 17,530 87 0.5

 Vacation 521 17,979 99 0.5

 Commercial 295 18,204 100 0.5

 Land 844 17,661 94 0.5

Business equity 1,186 17,333 80 0.4

Professional

 practice 265 18,244 90 0.5

Farm equity 861 17,652 86 0.5

1Counts exclude the 6279 single respondents.

2Counts exclude 108 households with a combined home and business. The equity of the home is included in the equity of the business.

3Other property assets were decomposed into four property asset types for the purpose of imputing holdings flags.

 Table 10.6. Disposition of Imputed Asset Holdings

 in the New Beneficiary Survey

 Imputed

 Asset Type Holdings Flag

 Yes No

 Money market 248 541

 CD 317 527

 Savings 575 248

 Credit union 150 636

 Checking 611 156

 Bonds 76 250

 Stocks 101 256

 IRA (Respondent) 25 112

 IRA (Spouse) 19 117

 Own home 52 10

 Other property

 Rental housing 5 82

 Vacation 4 95

 Commercial 3 97

 Land 4 90

 Business equity 8 72

 Professional practice 2 88

 Farm equity 6 80

 Table 10.7. Response Disposition to Asset Amounts

 in the New Beneficiary Survey

 Asset Amounts

 Holder

 Asset Type % %

 Positive Zero Inap. Missing Missing Missing1

Money market 2,912 13,845 1,842 9.9 26.6

CD 3,798 12,655 2,146 11.5 25.6

Savings 8,137 1 7,456 3,005 16.2 21.1

Credit union 1,953 15,415 1,231 6.6 18.6

Checking 11,068 300 4,248 2,983 16.0 16.3

Bonds 1,757 15,682 1,160 6.2 32.2

Stocks 1,781 8 15,438 1,372 7.4 36.2

IRA (Respondent) 1,685 16,480 434 2.3 15.0

IRA (Spouse)2 929 1 11,026 364 3.0 19.7

Own home3 12,0584 9 4,840 1,584 8.6 11.2

Other property 1,7215 13 16,435 430 2.3 16.6

Business equity 7396 77 17,333 450 2.4 31.2

Professional

 practice 1167 53 18,244 186 1.0 36.2

Farm equity 6138 5 17,652 329 1.8 28.2

1The "Holder % Missing" column includes only those who report asset holdings of that type in the base of the percentage calculation.

2Counts exclude the 6279 single respondents.

3Counts exclude 108 households with a combined home and business. The equity of the home is included in the equity of the business.

4Market value of the home (Q. 222).

5Includes 32 cases with reported negative equity.

6Includes 31 cases with reported negative equity.

7Includes 4 cases with reported negative equity.

8Includes 16 cases with reported negative equity.

 Table 10.8. Number of Cases and Percentage of Responses Imputed for

 Amounts of Fourteen Asset Types in the

 New Beneficiary Survey

 No. Cases Imputed % Responses Imputed

 Asset Type

 Positive Inap. Negative1 Appl.2 Inap.

Money market 1,301 541 30.9 3.8

CD 1,619 527 29.9 4.0

Savings 2,757 248 25.3 3.2

Credit union 595 636 23.4 4.0

Checking 2,827 156 20.3 3.5

Bonds 910 250 34.1 1.6

Stocks 1,116 256 38.5 1.6

IRA (Respondent) 322 112 16.0 0.7

IRA (Spouse)3 247 117 21.0 1.0

Own home4 1,574 10 11.55 0.2

Other property 351 70 9 17.3 0.4

Business equity 349 72 29 33.8 0.4

Professional

 practice 96 88 2 45.8 0.5

Farm equity 239 80 10 28.9 0.5

1These are cases that reported a loss flag but not an amount.

2This includes positive and negative imputed values (Columns l and 3).

3Counts exclude the 6,279 single respondents.

4Columns 4 and 5 exclude 108 households with a combined home and business. The equity of the home is included in the equity of the business; some of the joint amounts are imputed.

5Market value of the home.

 Table 10.9. Response Disposition to Income Flows from Selected Assets1 in the New Beneficiary Survey

 Income Flow

 Holder

 Asset Type % %

 Positive Zero Inap. Missing Missing Missing2

Money market 2,427 44 13,845 2,283 12.3 37.7

CD 3,025 104 12,655 2,815 15.1 38.6

Savings 5,991 464 7,456 4,688 25.2 37.5

Credit union 1,318 175 15,415 1,691 9.1 37.7

Checking 2,568 8,090 4,248 3,693 19.9 21.5

Bonds 718 202 15,682 1,997 10.7 64.5

Stocks 1,616 302 15,438 1,243 6.7 31.6

IRA (Respondent) 179 1,6723 16,480 268 1.4 6.6

IRA (Spouse)4 53 1,0073 11,026 234 1.9 8.5

Own home5 4,7056 8,488 4,840 458 2.5 2.9

Other property 8157 1,0903 16,435 259 1.4 8.3

1Equity assets are not considered in this table.

2The "Holder % Missing" column includes only those who report holdings of that asset type in the base of the percentage calculation.

3Includes cases responding "no" to the payment flag.

4Counts exclude the 6,279 single respondents.

5Counts exclude 108 households with a combined home and business. The equity of the home is included in the equity of the business.

6Total debt on home (Q. 221).

7Includes 53 cases with reported losses.

 Table 10.10. Number and Percentage of Imputed Income Flows Response

 Category for Eleven Asset Types in the

 New Beneficiary Survey

 Number of Cases Imputed % Responses Imputed

 Asset Type

 Positive Zero Neg.1 Inap. Positive Zero Neg. Inap.

Money market 1,734 8 541 41.7 15.4 3.8

CD 2,269 19 527 42.9 15.4 4.0

Savings 4,228 212 248 41.4 31.4 3.2

Credit union 955 100 636 42.0 36.4 4.0

Checking 978 2,559 156 27.6 24.0 3.5

Bonds 1,273 474 250 63.9 70.1 1.6

Stocks 863 125 255 34.8 29.3 1.6

IRA (Respondent) 833 734 112 31.7 4.25 0.7

IRA (Spouse)2 253 924 117 32.1 8.45 1.0

Own home6 167 281 10 3.47 3.2 0.2

Other property 138 41 10 70 15.3 3.65 15.9 0.4

1These are cases that reported a loss flag but not an amount.

2Counts exclude the 6,279 single respondents.

3Includes cases responding "yes" to the income flag but not giving an amount.

4Cases imputed "no" on the income flag.

5Includes cases imputed or responding "no" to the payment flag.

6Proportions exclude 108 households with a combined home and business. The equity of the home is included in the equity of the business; some of the joint amounts were imputed.

7Total debt on home.

 Table 10.11. Response Disposition to Expected Future Pensions from

 Current, Last and Longest Jobs

 Expected Future Pension Recipiency

Job and Person Expected or Expectation

 Don't Know1 Inappropriate2 Missing3

 Respondent

 Current 231 18,356 12

 Last 563 17,996 40

 Longest 75 18,484 40

 Spouse

 Current 884 17,524 191

 Last 210 18,183 206

 Longest 77 18,317 205

1These are the numbers of cases to whom expected future monthly pension payments were imputed.

2Persons already receiving a pension from that job; persons not expecting a pension; or persons not having a job in that field on the questionnaire.

3Persons refusing or missing a response to the expected future pension equation. These are largely persons who refused the entire set of questions for that job.

 Table 10.12. Response Disposition and Distribution of Imputations

 for Three Income Recipiency Flags

 in the New Beneficiary Survey

 Response Disposition Imputations

 Income Source Percent

 Yes No Missing Missing Yes No

Income from boarder 244 18,309 46 0.2 3 43

Income from loan 792 17,694 113 0.6 12 101

Other income 1,071 17,348 180 1.0 8 172

 Table 10.13. Response Disposition and Distribution of Imputations

 for Three Income Amounts

 in the New Beneficiary Survey

 Responses to Income1 Amounts Imputations

 Income Source Posi- Miss- Percent Posi-

 tive Zero Inap. ing2 Missing tive Inap.

Income from boarder 212 17 18,309 61 0.3 18 43

Income from loan 714 0 17,694 191 1.0 90 101

Other income 771 205 17,348 275 1.5 103 172

1The figures for income from boarder and other income refer to the third month prior to the interview.

2The missing column includes cases for which income recipiency was also missing.

1. See Linda Maxfield, "The 1982 New Beneficiary Survey," Social Security Bulletin, November, 1983, Vol. 46, No. 11, pp.3‑11. [↑](#footnote-ref-1)
2. See the Appendix in Linda Maxfield, op. cit.  Also, for further information about eligibility criteria and conditions for benefit entitlement and receipt, see the Social Security Administration, Social Security Handbook/1982, Seventh Edition, Washington, DC:  Government Printing Office, 1982. [↑](#footnote-ref-2)
3. Dual entitlement occurs when an individual is entitled to and receives two types of Social Security benefits concurrently.  Retired‑worker and disabled‑worker benefits are primary benefits, while auxiliary benefits as a spouse or survivor are secondary.  If the secondary benefit amount is higher than the primary benefit amount, the beneficiary is dually entitled and receives the secondary amount.  In this case, the auxiliary benefit supplements the primary benefit to equal the larger secondary amount.  The most common case exists when a retired‑worker or a disabled‑worker beneficiary also qualifies for a larger benefit as a spouse or a survivor. [↑](#footnote-ref-3)
4. Nondisabled persons covered by Medicare must be aged 65 or older.  The Medicare‑only nonbeneficiaries in the NBS were aged 65 to 71 in July 1982, implying that they were aged 63 to 70 during the window of June 1980 through May 1981.  Individuals under age 65 during the window would not then have been entitled to Medicare, but would have achieved entitlement on their 65th birthday. [↑](#footnote-ref-4)
5. At the time of sample frame construction, only preliminary SMSA definitions were available from the Census Bureau. Some of these have changed subsequently with the consequence that some of the "SMSA" psu's included counties which are now in fact on metropolitan.  Analyses incorporating metropolitan/nonmetropolitan residential location should rely on the current OMB definitions of SMSAs, but this was not possible for the public use tape. [↑](#footnote-ref-5)
6. Santos, R.L.  Effects of Imputation on Complex Statistics, Survey Research Center, University of Michigan, 1981. [↑](#footnote-ref-6)
7. The coefficients of variation of the estimated totals (excluding deceased and institutionalized beneficiaries) for each subdomain were all less than one percent. [↑](#footnote-ref-7)
8. Kish, L. and Frankel, M.R.  Balanced repeated replications for standard errors, Journal of the American Statistical Association, 1970, 65, 1071‑94.

 Plackett, R.L. and Burman, P.J.  The design of optimum multifactorial experiments, Biometrika, 1946, 33, 305‑25. [↑](#footnote-ref-8)
9. Ibid. [↑](#footnote-ref-9)
10. The scattergrams appear in the "Methodological Approach to Producing Generalized Sampling Errors in the New Beneficiary Survey" available from R. Santos at ISR. [↑](#footnote-ref-10)
11. Miller, Rupert G., Jr.  Simultaneous Statistical Inferences, Springer‑Verlog, New York, 1981. [↑](#footnote-ref-11)
12. Op.cit. [↑](#footnote-ref-12)
13. Kalton, G.  Practical methods for estimating survey sampling errors, Bulletin of the International Statistical Institute, 1977, 47(3), 495‑514. [↑](#footnote-ref-13)
14. The letter is reproduced on page 19 of The New Beneficiary Survey Interviewer's Manual and Question‑by‑Question Specifications. [↑](#footnote-ref-14)
15. Kalton, G.  Compensating for Missing Survey Data.  Survey Research Center, University of Michigan, Ann Arbor, 1981.

 Santos, R.  Effects of Imputation on Complex Statistics.  Survey Research Center, University of Michigan, Ann Arbor, 1981. [↑](#footnote-ref-15)
16. Czajka.  Imputation Methodology for the New Beneficiaries Survey, Mathematica Policy Research, Inc., Washington, DC, 1984. [↑](#footnote-ref-16)
17. In two of 16 subgroups, a different method of generating the stochastic error term was used.  For details, see the methodological report. [↑](#footnote-ref-17)