

**THE SURVEY OF INCOME AND  
PROGRAM PARTICIPATION**

**TYPE OF OASDI BENEFIT AND YEAR  
OF DEATH BASED ON AN EXACT  
MATCH TO SOCIAL SECURITY  
ADMINISTRATION BENEFIT  
RECORDS, 1990 AND 1991 PANELS OF  
THE SURVEY OF INCOME AND  
PROGRAM PARTICIPATION (SIPP):**

**DESCRIPTION OF THE DEVELOP-  
MENT OF THE DATA FOR PUBLIC  
RELEASE AND A PRELIMINARY  
EVALUATION OF DATA QUALITY**

No. 239

Denton R. Vaughn  
Office of Research,  
Evaluation and Statistics  
Social Security Administration

U.S. Department of Commerce U.S. CENSUS BUREAU

**Type of OASDI Benefit and Year of Death based on an Exact Match to  
Social Security Administration Benefit Records, 1990 and 1991 Panels  
of the Survey of Income and Program Participation (SIPP):**

**Description of the Development of the Data for Public  
Release and a Preliminary Evaluation of Data Quality**

By

Denton R. Vaughan

Division of Housing and Household Economic Statistics  
Bureau of the Census, and  
the Division of Economic Research,  
Office of Research, Evaluation and Statistics  
Social Security Administration  
December 18, 2000

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion

## **Acknowledgements**

The author would like to thank Barry Bye, Russ Hudson, and Joel Packman for the patience they showed in responding to the many questions I directed their way; to David Weaver for arranging access to a more complete set of MBR records for the 1990 and 1991 panels; to Faye Aziz, Howard Iams, Dan Kasprzyk, Karen King, and Fritz Scheuren for a number of helpful suggestions; to Weltha Logan for editorial assistance; to both the Bureau of the Census and Social Security Administration for the forbearance that enabled me to complete this work, and especially to Henry Ezell for his superb data processing support.

## CONTENTS

### TEXT

Introduction .....	1
SSA record data on benefits .....	1
OASDI benefit information .....	1
Nature of Master Beneficiary Record data on type of benefit and timing of benefit receipt .....	1
OASDI benefit categories covered by the public release.....	3
Comparison of benefit estimates based on the survey match and program estimates.....	4
Consistency between survey and survey match representation of benefit receipt.....	7
Year of death information.....	9
Overview .....	9
Relationship between the type of benefit and year of death data.....	10
Comparison between year of death estimates based on the survey match and corresponding estimates from the U.S. Vital Statistics System.....	11
Nature of the NCHS estimates .....	11
The approach to comparing the survey match and NCHS estimates .....	11
Overall assessment and considerations in reconciling the estimates.....	12
A hypothesized institutionalization effect.....	13
The initial interview effect.....	13
Nonmatches.....	14
Deaths of nonbeneficiaries.....	14
Comparing the estimates corresponding to the initial two calendars of each panel.....	15
Comparing the estimates for the third and subsequent panel calendar years.....	15
Review of 1991 panel estimates .....	18
General assessment.....	19
Conclusions .....	19
References.....	22

### TEXT TABLES

A. Number of social security beneficiaries in current payment status, December 1990, by gender and selected type of benefit, based on independent and matched survey estimates, 1990 SIPP panel.....	25
B. Number of social security beneficiaries in current payment status, December 1991, by gender and selected type of benefit, based on independent and matched survey estimates, 1991 SIPP panel.....	27

C. Percent of persons in current pay as of December 1990 and 1991 for whom a survey social security amount is present for the subsequent January, 1990 and 1991 SIPP panels.....	29
D. Reconciliation of current pay status from the record match with survey reports of benefit receipt based the sample for whom benefit records were located.....	30
E. Reconciliation of estimates of the number of deaths of persons aged 65 and over as identified from the SIPP-SSA benefit record match and the U.S. death registration system, initial two calendar years of the 1990 and 1991 SIPP panels.....	31
F. Number of deaths 1990-1996, by year, sex, and age as reported to the NCHS and as identified from the 1990 SIPP panel based on a match to Social Security Administration benefit records, persons age 65 and over in year of death.....	32
G. Number of deaths 1991-1996, by year, sex, and age as reported to the NCHS and as identified from the 1991 SIPP panel based on a match to Social Security Administration benefit records, persons age 65 and over in year of death.....	33

**APPENDIXES**

A. Definitions and Explanations .....	A-1
B. Description of Methods Used to Characterize Type of OASDI Benefit and Identify Year of Death .....	B-1
General description of the record match .....	B-1
Creation of the type of OASDI Benefit code .....	B-2
Development of the type of benefit code .....	B-5
Identification of the type of benefit.....	B-5
Identification of current payment status.....	B-5
Identification of dual entitlement status .....	B-5
Presence in sample .....	B-5
Final assignment of type of benefit.....	B-6
Development of the year of death information .....	B-7
Source of year of death data .....	B-8
Consistency of information on fact and year of death .....	B-9
Use of the January cross-sectional weight .....	B-9
Explanation of the initial interview effect on deaths.....	B-10
C. Accuracy of Estimates.....	C-1
Use and calculation of standard errors .....	C-1

Standard error parameters and tables and their use.....	C-4
Standard errors of estimated numbers and percentages .....	C-4
Assessing comparisons between survey estimates and independent estimates .....	C-8

**APPENDIX TABLES**

B-1. Relationship between summary Type of Benefit Code (TOB) and the full Beneficiary Identification Code (BIC) from the Master Beneficiary Record..... B-13

B-2. Summary of type of benefit code assignment criteria..... B-14

B-3. Selected types of OASDI benefits identified by the SIPP - SSA benefit record match, 1990 and 1991 panels: Sample counts and weighted estimates ..... B-15

B-4. Principal instances of discrepant current pay status, by timing of MBR extract and representation of social security receipt in the survey, 1990 panel..... B-16

B-5. Variables used to establish the fact and year of beneficiary death..... B-17

B-6. Source of year of death by panel..... B-18

B-7. Year of death as identified by the SIPP - SSA benefit record match, 1990 and 1991 panels: Weighted and unweighted counts..... B-19

B-8. Number of deaths of persons age 65 and over by age, sex, and year of death, 1990-1996 as reported to the National Center for Health Statistics (NCHS) and as identified from the 1990 SIPP panel based on a match to Social Security Administration benefit records..... B-20

B-9. Number of deaths of persons age 65 and over by age, sex, and year of death, 1991-1996 as reported to the National Center for Health Statistics (NCHS) and as identified from the 1991 SIPP panel based on a match to Social Security Administration benefit records..... B-21

B-10. Presence of alternative survey weights with an identified year of death based on the survey match by panel ..... B-22

B-11. Rotation group and calendar months, January - April, initial calendar year of the 1990 and 1991 panels ..... B-23

C-1. Values for generalized variance parameters for SSA beneficiaries, 1990 and 1991 SIPP panels..... C-12

C-2. Standard errors for estimated population totals, SSA recipients, 1990 SIPP panel ..... C-13

C-3. Standard errors for estimated percentages of SSA recipients, 1990 SIPP panel ..... C-14

C-4.	Standard errors for estimated population totals, SSA recipients, 1991 SIPP panel .....	C-15
C-5.	Standard errors for estimated percentages of SSA recipients, 1991 SIPP panel .....	C-16





## **INTRODUCTION**

This report describes two items that are being made available for the 1990 and 1991 panels of the Survey of Income and Program Participation (SIPP) from the benefit records of the Social Security Administration (SSA): (1) a summary type of benefit code for the Old-Age-Survivors and Disability Insurance (OASDI) program, popularly known as Social Security for persons in current payment status as of December of the initial full calendar year of each panel, and (2) year of death for sample members identified as having died prior to 1997.

## **SSA RECORD DATA ON BENEFITS**

Two sources of administrative data on beneficiaries were accessed to obtain information on type of OASDI benefit and occurrence and year of death for the sample members with validated SSN's—the Master Beneficiary Record (MBR) and the Supplemental Security Income Record (SSR). The MBR contains benefit information about persons who have ever filed for OASDI benefits. The SSR contains similar information for persons who have ever applied for benefits under auspices of the Supplemental Security Income program. SSR information was obtained in order to insure that deaths of all SSI recipients, regardless of their representation in the MBR, would be identified.

### **OASDI benefit information**

*Nature of MBR data on type of benefit and timing of benefit receipt.*—The MBR data used to define the type of OASDI benefit and to identify year of death is organized on an entitlement basis, that is, at any given time, the file is intended to represent the

agency's view of what an individual ought to have received at various points in the past. This aspect of the MBR is more significant with respect to representing type of benefit at a given point in the survey reference period than with respect to representing the fact and year of death. Since the survey attempts to measure income and benefit receipt as they actually occur, to the extent that the Agency's view of benefit entitlement for a particular month changes after the fact, the type and amount of benefit income actually received in that month may not be what is represented in the MBR by the time the record is typically accessed (generally a few years after the relevant survey period). The difficulties which likely arise from this inconsistency between the representation of program participation and benefit receipt in the record system and the survey are not fully understood, but is thought be particularly troublesome during the period that certain types of beneficiaries are first coming on the rolls. In creation of the MBR extract, ancillary information in the record was employed to reduce the effect of the contrast between the MBR and survey representation of the timing of program participation. In conjunction with steps taken in the development of the type of benefit code for public release, additional efforts were made to deal with this issue, especially with respect to disabled-worker beneficiaries who may receive retroactive benefits for as many as 12 months prior to the first actual month of payment.<sup>1</sup> As will be shown, direct comparison between benefit receipt as represented in the MBR and as reported in the survey, suggests that for the most part such problems have been overcome, at least with respect to the representation of benefit receipt. Still, it is well to keep in mind that the survey and record system do not share precisely the same approach to the representation of benefit receipt.

---

<sup>1</sup> In such instances an individual would not actually have received a social security disability benefit *in* any of the months prior to award but would receive a payment *for* those months at the time of award. However,

*OASDI benefit categories covered by the public release.*—The type of benefit code prepared for public release identifies six separate benefit types plus a seventh residual category<sup>2</sup>, as follows:

- (1) retired worker benefit,
- (2) disabled-worker benefit,
- (3) aged wife benefit without dual entitlement to a retired worker benefit (aged wife only benefit),
- (4) aged wife with dual entitlement to a retired worker benefit (dually entitled aged wife benefit),
- (5) aged widow without dual entitlement to a retired worker benefit (aged widow only benefit)
- (6) aged widow with dual entitlement to a retired worker benefit (dually entitled aged widow benefit), and
- (7) all other types of benefits, including principally dependent and survivor benefits received by minor children or adult children disabled in childhood, benefits received by spouses of retired, disabled or deceased workers with minor children or disabled children in their care, disabled widows, and all benefits received by men as a husband or widower of a retired, disabled or deceased worker.

For purposes of the type of benefit code developed for public release, type of benefit was represented as the benefit, if any, in current payments status for the month of

---

the MBR would indicate a benefit in current pay status for months of entitlement prior to date of award.  
<sup>2</sup> The number appearing in parenthesis before the benefit category description represents the code value assigned to that category in the public release file.

December 1990 (1990 panel) or December 1991 (1991 panel).<sup>3</sup> Since benefits in current payment status as of a given month are actually received (either by check through the mails or via electronic deposit) early the following month<sup>4</sup>, the survey calendar month corresponding to a benefit in current payment status as of December is the following January. Consequently, the code was defined only for sample members considered to be in sample for that month, as determined by the presence of a positive final person weight (FNLWGT) for January 1991 or January 1992 depending on the panel involved.

Persons considered as not in sample as of January 1992, as determined by the absence of a positive final persons weight (FNLWGT) for that month are coded as zero (0). The remaining sample cases, those with an MBR record but not in current pay as of 12/90, those with a valid SSN for whom no MBR record was located, and sample persons for whom no operationally valid SSN was located are coded as nine (9).

*Comparison of benefit estimates based on the survey match and program*

*estimates.*-- Estimates of the number of persons in current payment status as of December 1990 and 1991 based on the public release type of benefit code are provided in tables A and B, respectively. The estimates employ the public use file cross-sectional weight for the subsequent January, given that December current pay benefits were received in that month.<sup>5</sup> Independent estimates by type of benefit and gender, derived from program

---

<sup>3</sup> For earlier work related to type of social security benefit in the SIPP context based on matched and unmatched data see Bye and Gallicchio (1988, 1993) Grad (1989), Vaughan (1989), and Social Security Administration (1987-1992).

<sup>4</sup> During the time period that the type of benefit code is being defined for this public release (early 1991 and early 1992), social security benefits were being made on the third of each month.

<sup>5</sup> Note that the public use file weight has not been adjusted to compensate for sample persons for whom no social security number was located.

data, after nominal adjustment to account for differences between the program and survey universes,<sup>6</sup> are also provided.

In general, the matched estimates of benefits in current pay by type of benefit based on the 1990 panel appear to be reasonably complete with respect to program data. According to the survey-MBR match, nearly 35 million people were in current payment status in December 1990 and interviewed in the following January. This estimate of the total number of benefits in current pay from the survey amounts to 95 percent of the corresponding independent estimate of about 36 million. All told about 33 million individuals, representing 97 percent of the corresponding independent estimate, were identified as receiving one of the six types of benefits separately identified by the public release type of benefit code. According to the matched survey estimates, approximately 18 million persons were receiving benefits as retired workers (about 12 million men and 4 million women, excluding women dually entitled to a wife or widow benefit)<sup>7</sup> and about 1.8 million men and 1 million women were receiving benefits as disabled workers. Nearly 4.7 million women were receiving benefits as aged wives and 6 million as aged widows. Of these, about 2.0 million aged wives and 2.2 million aged widows were also dually entitled to retired worker benefits. Finally, the survey – MBR match identified an

---

<sup>6</sup> There are two principal differences between the program and survey universe definitions. The survey excludes persons living in institutions and outside the 50 States and the District of Columbia while program data include these population subgroups. Also included in the program data, but excluded from the survey, are individuals who were in current payment status for a given month but who died prior to interview. The most important factor affecting comparisons between survey and program estimates of the number of persons in current payment status by type of benefit, particularly for women over the age of eighty, is institutionalization. The adjustments employed to provide a nominal reconciliation between the survey and program beneficiary universes are documented in Vaughan (1992).

<sup>7</sup> All estimates involving gender in this report are principally based on gender as it is represented in the MBR. Estimates by gender based on the survey representation of gender will differ slightly from those given here.

additional 2.6 million persons receiving benefits of other kinds, about 2/3 of whom were minor children (under age 18).

Of the eight benefit type/gender combinations identified individually, differences between survey and independent estimate for aged wives not dually entitled, retired worker benefits being received by women who were not dually entitled and disabled-worker benefit being received by men are statistically significant at the .10 level.<sup>8</sup> However, even in these three instances, the survey point estimate accounts for between 86 and 94 percent of the corresponding independent estimate. Finally, the residual category, which includes all benefit types not identified separately, falls considerably short of the independent estimate. There is evidence of shortfall of similar magnitude in the matched estimate for beneficiaries under age 18.

Matched estimates of benefits in current pay based on the public release type of benefit code for December 1991 stemming from the 1991 panel present basically the same pattern with respect to both the number of beneficiaries in current pay by type of benefit and comparison to independent estimates. At the level of all beneficiaries in current pay, and adult beneficiaries in current pay, matched estimates from the 1991 panel are nominally slightly less complete with respect to the independent estimates than those stemming from the 1990 panel, and evidence for differences between the matched survey and independent estimates is somewhat stronger for the later panel. Finally, there is consistent evidence across panels that estimates of benefit receipt by male disabled workers, aged wives not dually entitled to retired worker benefits, and individuals

---

<sup>8</sup> All sampling errors for SIPP estimates presented and discussed in this note are derived from generalized variance parameters developed by Bye and Gallicchio (1993) for SSA beneficiaries (see also Bye and Gallicchio (1988) and Jabine (1990, pp. 207-208). More detail on the Bye and Gallicchio estimates is provided in Appendix C.

receiving one of the benefit types not separately identified by the public release code are understated with respect to their corresponding independent estimates. It is worth noting that with the exception of this residual category, even these more problematic of the matched estimates reach 85-90 percent of their respective independent estimates. Thus with allowance for these few caveats, the estimates of social security receipt by type of benefit based on the matched SIPP data appear to quite complete.

*Consistency between survey and survey match representation of benefit receipt.--*

The additional question of consistency between the matched estimates of benefit receipt and corresponding measures stemming solely from the survey was addressed by comparing the indication of a benefit in current pay status for December as determined by the match and the presence of a social security benefit amount for the following January in the survey. The comparisons indicate a very high level of consistency for all benefit types separately identified by the public release type of benefit code. As shown in table C, considering results for both panels, 96-97 percent of beneficiaries age 18 or over<sup>9</sup> and identified as in current pay status for December had social security benefit income present for the following January in the survey. For five of the six principal types of benefits identified separately by the public release type of benefit code current pay status as indicated by the record match is accompanied by presence of a social security amount in the survey at least 97 percent of the time. The somewhat lower percentage of current pay disabled workers with a corresponding social security amount present in the survey (88-90 percent) may reflect a tendency for persons recently awarded social security

---

<sup>9</sup> Benefit amounts for persons under age 16 are included on a parent's survey record. This accounts for the finding that only 4-5 percent of sample members under age 18 and identified as in current pay status, have a social security benefit amount present (on their own person record) in the survey. While social security



benefits in the survey to misreport their social security income as supplemental security income. This situation may arise because during the required 5 month waiting period prior to award of disabled-worker benefits, some individuals receive SSI benefits that then terminate with the award of the disabled-worker benefit. Apparently, not all survey respondents understand and properly represent the programmatic sequencing that takes place in such circumstances.<sup>10</sup>

While the information presented in table C demonstrates the high probability that persons identified as in current payment status on the basis of the survey SSA record match also report a benefit for the corresponding month in the survey, it is also the case that 11-14 percent of persons reporting a benefit in the survey for that month were not identified as in current pay status on the basis of the record match. Clearly, such individuals represent some mix of “true” recipients and “true nonrecipients” whose reciprocity status had not been confirmed by the record match. Such individuals can plausibly be divided into two groups: 1) Those for whom no SSN was located, or for whom an SSN was available, but a benefit record, though existing, was not located,<sup>11</sup> and 2) those for whom both an SSN and benefit record were located, but the record provided no evidence of a current pay benefit in force for the month in question. As shown in table D (bottom panel), for both panels about five sixths of those reporting a survey benefit that is not confirmed by the record system match fall into the first group and the

---

benefit receipt for children under age 16 is identified on the child’s survey record, social security reciprocity indicators for children were not accessed in the context of this study.

<sup>10</sup>Other work carried out at the Social Security Administration using SIPP panel data matched to the MBR has shown that failure to report social security in the survey by current pay disabled workers is sometimes associated with the a report of SSI in the survey.

<sup>11</sup> For example, in development of the current type of benefit code the original procedure employed to obtain benefit records based on the roster of SSN’s developed for SIPP sample members missed, respectively, 14 and 23 percent of aged wife only and aged widow only benefit records for the 1990 panel

balance in the second. While nothing may be inferred about the “true” social security beneficiary status of those in the first group, those in the second group might be taken to represent persons who were actual nonrecipients were it not for the known problems of representing the timing of benefit receipt by means of the record system. However, if this caveat is ignored for the sake of argument, and only persons with matched benefit records are considered, then about 2 percent of survey reporters may be considered to be actual nonrecipients and about 3-4 percent of those with reciprocity confirmed by the match may be considered to nonreporters.<sup>12</sup> Thus, among persons for whom a benefit record was located, were age 18 and over, and reported a survey benefit, there is a net shortfall in survey reciprocity reporting of just 1-2 percent.

The importance of understanding the quality of social security reciprocity reporting in the SIPP context is underscored by evidence on the role that the identification of additional social security recipients has played in the 20-30 percent lower poverty rate for the elderly as measured in the SIPP as compared to the Current Population Survey during much of the 1980’s and 1990’s (Martini and Dowhan 1997).

### **Year of death information**

*Overview.*--The second characteristic to be included in the public release is year of death as represented in the Master Beneficiary Record and the Supplemental Security Record. Given SSA’s need to insure that benefit payments do not continue after death of a beneficiary, the agency goes to considerable lengths to identify and verify the fact and

---

and similar proportions for the 1991 panel. An alternative procedure was subsequently employed to identify the more complete set of benefit records underlying the current public release.

<sup>12</sup> It is certainly plausible, if not likely, that some of these cases are associated with imputations employed to deal with item nonresponse. Social security reciprocity status was imputed for up to 1.6 percent of

timing of beneficiary deaths and of covered workers whose dependents subsequently file for survivor benefits (Aziz and Buckler, 1992). Given that the social security and SSI programs were paying benefits to approximately 94 percent of the population age 65 and over in the early to mid 1990s, information with respect to the fact and year of death contained in SSA beneficiary record systems may be considered to be relatively comprehensive for the general population aged 65 and older as well as being essentially complete for beneficiary population regardless of age.

The death information involved in this public release pertains to all deaths identified from SSA's two beneficiary record systems regardless of age. However, since cash payments are received by only a minority of those under age 65, deaths identified based on the match for persons under age 65 cannot be held to characterize the mortality experience of the nonelderly. The balance of the discussion of the nature of the survey-match data on the fact and year of death will focus on the elderly.

*Relationship between the type of benefit code and year of death data.*— Since the MBR and SSR files were accessed in 1997 and 1995, year of death is available through calendar year 1996, well after the two panels left the field in mid 1992 and 1993, respectively.<sup>13</sup> Information on year of death obtained from the survey match is included regardless of the presence of a current pay social security benefit as of December of the first full year of the panel. Thus typically type of benefit will not be defined for persons identified via the match as dying during the first full calendar year of each panel. Neither will type of benefit be characterized for persons identified via the survey match as

---

January 1991 recipients in the context of the 1990 panel and 2.3 percent of January 1992 recipients in the 1991 panel. Imputations for type Z person noninterviews might also be involved.

<sup>13</sup> While a few deaths were identified on the basis of the MBR for 1997, they were determined to be too incomplete to warrant release.

decedents who became entitled to benefits subsequent to the initial full calendar year of each panel. Weighted and unweighted counts of deaths by calendar year as identified by the match are provided in appendix tables B-1 and B-2.

### **Comparison between year of death estimates based on the survey match and corresponding estimates from the U.S. Vital Statistics System**

Although, as noted, information on the fact and year of death for SSA beneficiaries that is available from agency benefit records is held to be quite complete, given the novel way of accessing mortality information via a survey record system match, a review of the completeness of the matched estimates was conducted. The assessment is based on comparisons of the survey match estimates to independent estimates of the number of deaths among U.S. residents age 65 and over by year of death, age, and gender. These estimates come directly from the U.S. Vital Statistics System as collected and processed by the National Center for Health Statistics (NCHS).

*Nature of the NCHS estimates.*—Estimates of the number of deaths of U.S. residents by age, sex, and calendar year are routinely published by NCHS in the National Vital Statistics Reports series. These data are based on all death certificates filed in the 50 States and the District of Columbia. It is believed that more than 99 percent of deaths occurring in the United States are registered (Anderson, et al. 1997, p. 3).

*The approach to comparing survey match and NCHS estimates.*—The comparison between the survey match estimates and those provided by NCHS is useful to the extent that it contributes to a better understanding of the nature of the mortality estimates stemming from the survey match. In order to identify patterns in the survey match—NCHS comparisons that are informative and assess which ones are likely to be credible from a statistical standpoint, consideration is given initially to the pattern of nominal

differences between the mortality estimates stemming from the two sources. Next, 90-percent confidence intervals were constructed for the survey estimates, to assess whether differences of level between the two sources are likely to exceed sampling error. Finally, variations in the completeness of survey estimates with respect to the NCHS estimates for different year of death, age and gender subgroups are assessed by testing whether observed variations from one subgroup to the next in the matched survey deaths expressed as a percentage of NCHS deaths may be considered to be statistically significant.

*Overall assessment and considerations in reconciling the estimates.*— An overall assessment of the comparisons between the number of deaths based on the survey match and the NCHX data is based on review of the information presented in tables F and G, respectively for the 1990 and 1991 panels.<sup>14,15</sup> These estimates will be considered in some detail, but considered generally in the light of certain features of SIPP survey design, three principal findings emerge.

- 1) The matched estimates are likely reasonably comprehensive for original sample members (so-called 100-level persons) age 65 or older for the years covered by the administrative data (1990-1996 for the 1990 panel and 1991-1996 for the 1991 panel).

---

<sup>14</sup> More detailed comparisons based on the full year of death, age and gender cross-classification are provided in tables B-8 and B-9 of Appendix B.

<sup>15</sup> The public use file cross-sectional weight for January of the first full calendar year of each panel is employed rather than the longitudinal weight because the longitudinal weight is not present for those individuals who died subsequent to the initial interview but were not identified as exiting the survey population. While the weighted estimates are similar regardless of which of these two weights is employed, the cross-sectional weight was chosen because it yields a somewhat larger sample size (see Appendix B for additional details) Note also, that the public use weights do not incorporate adjustments to compensate for nonmatches.

- 2) However, the matched estimates for persons aged 65 or older fall considerably short of the number of deaths occurring in the U.S. resident population as indicated by NCHS data, especially in the first and second full years of the panel reference period. The observed shortfalls also tend to be larger for women than men for any given age group and year of death, and in particular for women age 80 or older at death as compared to those below age 80.
- 3) It will be argued that these shortfalls with respect to NCHS estimates of total deaths in the U.S. resident population plausibly stem from four factors: 1) deaths of persons institutionalized at the time the survey sample was drawn, 2) deaths occurring in January – April of the initial reference year of each panel that are missed because of interviewing rules used in the initial panel interview, 3) nonmatches, and 4) deaths of elderly individuals who are not receiving cash benefits from either the OASDI or SSI programs.

A brief description of the nature of each of these effects is given below:

*The hypothesized institutionalization effect.*—Since the SIPP is restricted to the noninstitutional population, persons living in institutions at time of initial interview ( $t_1$ ), are excluded. Therefore, the great majority of deaths stemming from the institutionalized population as defined at the time  $t_1$ , which occur subsequent to the initial interview, will not be observed among members of the original SIPP sample, i.e., 100-level individuals<sup>16</sup>. However, with the passage of time members of the institutional population as of ( $t_1$ ) who die are replaced by persons originally belonging to the survey population. Eventually, the institutional population will come to consist principally of persons who belonged to the survey population at time ( $t_1$ ), and, at that point, in principle, all deaths occurring in the institutional population may be observed from the sample identified at initial interview as representing the noninstitutional population.

---

<sup>16</sup> Of course, some elderly persons residing in institutions at the time of sample selection will subsequently leave the institutional setting and return to the survey population. If they die after returning to the survey population, nominally their deaths could be identified via match to SSA benefit records. Presumably such deaths represent but a small minority of deaths stemming from institutional population existing at the time of sample selection.

*The initial interview effect.*—Given SIPP interviewing rules, information is not obtained for persons who lived in the civilian noninstitutional population during the initial interview reference period but did not live at a sampled address at the time of interview (Bureau of Census 1991, p. 2-11). Consequently, members of the noninstitutional population who die during the months covered by initial interview reference period will not be represented by the SIPP sample and thus their deaths cannot be identified based on the match with SSA benefit records.

Given that the standard SIPP interview covers a four-month period, other things being equal one would expect that about 1/3 of deaths occurring during the initial 12 reference months of each panel would be missed simply because of the initial interview criteria. However, because of the way that the staggered SIPP interviewing pattern interacts with calendar time, only about 3/5 of the individuals living at sampled addresses during the first four months of the year and who exit the survey population prior to the initial interview due to death are missed by the initial interview; the balance are included in the reference period of the second interview and would nominally be identified in that interview. Working this through on a calendar year basis indicates that about 21 percent of deaths occurring during the initial calendar year of the 1990 and 1991 panels would not be observable from the matched sample because of the initial interview effect. The details of this mechanism are described in Appendix B. In the second through final interviews this problem does not arise, because deaths of persons who exit the sample subsequent to a successful initial interview will be identified by means of the match conditioned only on the availability of a valid SSN and the completeness of mortality information in SSA benefit records.

*Nonmatches.*—Obviously a sample member's death cannot be identified via match to SSA benefit records absent an SSN. As noted, SSN's considered to be operationally valid were not available for about 7 percent of persons aged 65 or older in the context of the 1990 panel and about 10 percent of the elderly for the 1991 panel. Consequently, if the SSN availability is independent of subsequent mortality, we would expect to miss, respectively 7 and 10 percent, of the deaths for elderly sample members for this reason.

*Deaths of nonbeneficiaries.*—Essentially only deaths of beneficiaries are identified in the Master Beneficiary and Supplemental Security Income record systems that were accessed to develop year of death information for public release. For the years covered by the 1990 and 1991 panel matches (1990 – 1996) approximately 94 percent of persons aged 65 or older were receiving cash benefits under the OASDI or SSI programs. Since death postings to the benefit record systems are restricted to beneficiaries, to the extent that probability of death is independent of beneficiary status, at most the match could be expected to

identify about 94 percent deaths of persons aged 65 or older; about six percent of such deaths could not be identified.<sup>17</sup>

*Comparing the estimates for the initial two calendar years of each panel.*-- The potential of these four factors to account for differences between the matched survey and NCHS estimates of the total number of deaths during the first two calendar years covered by each of the two SIPP panels under consideration is assessed in table E.<sup>18</sup> The total number of deaths as estimated from the 1990 panel for 1990 and 1991 is approximately 3.1 million. After making allowances for institutionalized deaths not observable from the matched sample, initial interview effects, and nonmatches, the expected number of deaths observable from the survey in the initial two calendar years is about 2 million. At approximately 1.9 million, the observed estimate based on the survey match represents 90 percent ( $\pm 9$  percent based on a 90-percent confidence interval) of the expected estimate.

While the reconciliation procedure incorporates uncertainties of its own, if taken at face value it suggests that the matched estimates likely represent approximately 80-100 percent of the deaths potentially observable from the 1990 panel during its first two calendar years (see bottom panel of table E).

*Comparing the estimates for the third and subsequent panel calendar years.*-- The information required to carry out the reconciliation is not readily available for balance of

---

<sup>17</sup> A significant fraction of such deaths could likely be identified the basis of the variable representing reason for leaving the household. In the context of the 1990 panel, it appears to identify additional deaths of persons aged 65 that older that occurred during the life of the panel but were not identified by the survey match. Such deaths could add as much as an additional 6 percent to those identified as occurring during the survey reference period solely on the basis of the survey match.

<sup>18</sup> Given that the initial interview effect is restricted to the first full calendar year for each panel, a more convincing reconciliation would be based on just the first full calendar year of the panel. However, it was carried out on the basis of two years because information used to approximate the number of deaths occurring in the institutional population subsequent to the initial interview was only available for a two-year period (Manton 1988).



calendar years observable via the match. However, consideration of the relationship between the survey match estimates and NCHS mortality data by year, age and gender is instructive nonetheless. Detailed comparisons from the two sources for the full year, gender age matrix are given in appendix tables B-8 and B-9, but are recast in summarized form in table F to facilitate discussion. (Corresponding information from the 1991 panel is given in table G.)

Before discussing the estimates a word about the information provide in tables F and G will be helpful. Estimates of the number of deaths by year, and sex by year, and age, and sex by age from the NCHS and the survey match appear in the first two columns of the tables, followed by the standard error and 90-percent confidence intervals for the respective survey estimates in columns 3-5. The survey estimates are expressed as a percentage of their corresponding NCHS estimates in column 6, together with the corresponding standard error in column 7. Selected subgroup comparisons are denoted in the last column of the table. The statistics required to evaluate these comparisons appear in columns 8, 9, and 10.

Turning first to estimates from the 1990 panel (table F), it is evident that the matched estimates for males nominally begin to approach the total number of deaths as reported to the NCHS (reaching or exceeding 90 percent) by the third full calendar year following the beginning of the survey. Thereafter, i.e. for the calendar years 1992-1996, the matched estimates cannot be said to differ from the NCHS estimates at the .10 level given that the value of the upper bound of the 90-percent confidence interval for the survey match estimate exceeds the corresponding number of deaths reported to NCHS.

The calendar year estimate for the number of female deaths based on the survey match as a percentage of NCHS deaths also eventually reaches approximately the same level (88 percent) as for males, but not until 1996, seven full years after the initiation of the survey. This interpretation of the nominal pattern of differences between estimates of female deaths from the two sources is sustained by the finding that the null hypothesis of no difference between the matched survey and NCHS estimates may be rejected at the .10 level for 1991 through 1995, but not for 1996 (again, as can be seen from a comparison of columns 1 and 5 of the table, the NCHS estimate exceeds the value of the upper bound of the 90-percent confidence interval for the survey estimate for 1991-1995, but not 1996). Comparisons of the relative completeness of male and female estimates of the number of deaths by calendar year identified by the survey match are generally consistent with this pattern as well, with estimates for females for all calendar years but two (1991 and 1996) below those for males as assessed at the .10 significance level. Taken together, this evidence of the growing completeness in the matched survey estimates with respect to the NCHS benchmarks over time, and which appears to be more marked and rapid for males than females, is consistent with the institutionalization effect noted earlier.

Variations in completeness of the matched survey estimates by age and gender also support the institutionalization hypothesis. For males, the null hypothesis of no difference between the survey match estimates by age and the corresponding NCHS estimates can only be rejected at the .10 level for males age 80 and over. However, the discrepancy between the matched survey estimate and the corresponding NCHS estimate is significant at the .10 level for the oldest age group. For females, survey-match

estimates are nominally below the NCHS estimates with all the differences significant at the .10 level. (Note again that the NCHS estimate exceeds the upper bound of the 90-percent confidence interval for the survey estimate in each instance.) Review of the match estimates expressed as a percentage of the NCHS estimates also indicates that survey match estimate for females age 80 and over at death is less complete than for the younger age groups (58 percent vs. 84-86 percent). All these observed differences are statistically significant at the .10 level. Furthermore, when the completeness of the survey match estimates is considered by gender within the four age groups, only the estimates for females age 80 or over at death may be said to differ from those for males.<sup>19</sup> These patterns by age and gender are also consistent with the hypothesized institutional effect, in that matched estimates are generally less complete for females than males with gender differences in completeness increasing with age and that are particularly marked for the oldest age group.

*Review of 1991 panel estimates.*--Elements of some of these patterns also appear in the matched death estimates stemming from the 1991 panel. While the reconciliation of the NCHS and matched survey estimates for the first two years of the survey indicates that the latter accounted for about 96 percent ( $\pm 12$  percent based on a 90-percent confidence interval) of the expected number of deaths identifiable from the survey via the match (table E, last two columns on the right), the patterns of variation in the relationship between the NCHS and matched survey estimates by calendar, year, gender, and age evident for the 1990 panel are not as clearly present in the 1991 panel estimates (table G).

---

<sup>19</sup> The test statistics that support this finding are based on the standard errors for the matched survey estimates for males and females by age expressed as a percentage of NCHS estimates. These percentages and their corresponding standard errors are given in columns 6 and 7 of table G. However, the test statistics *per se* do not appear in the table.

For example, a consistent pattern of increasing completeness of the survey match estimates following the initial year or two of the survey is not clearly evident. While matched estimates of female deaths are clearly less complete with respect to the NCHS data than those for males (67 vs. 80 percent), the consistent and contrasting variations in completeness of the matched estimates with respect to NCHS estimates by age for males and females that are apparent in the 1990 panel are not evident for the 1991 panel. If age patterns are considered without distinction to gender, the matched estimates for the 80 and over age group as compared to the three younger age groups are nominally less complete (67 percent vs. 78-82 percent), but there is no consistent pattern after taking into account of likely effects of sampling error.

In sum, the reconciliation between the matched and NCHS mortality estimates for the initial two years of the panel provides evidence that 1991 panel match may identify at least four-fifths or more of the deaths occurring in the survey population during that period subject to the limitations arising from the restriction of identifiable deaths to SSA beneficiaries and the lack of usable SSN's for some sample members. And while there is some evidence consistent with the hypothesis that an institutional effect could account for some of the short-fall in survey estimates with respect to those stemming from NCHS sources subsequent to the first two survey calendar years, the evidence is considerably weaker than for survey match estimates based on the 1990 panel.

*General assessment.*--Considering the findings for the two panels taken together, it seems that a reasonable case can be made that the survey matches identifies a high proportion (perhaps as much as 90-95 percent) of deaths attributable to the survey population age 65 and over, after making allowance for missing SSN's and the

limitations of the record data employed in the match (that is, only beneficiary deaths are observable). Furthermore, for the 1990 panel, the evidence is consistent with the view that whatever shortfall remains with respect to NCHS estimates after the initial two years of the panel for the balance of the period covered by the match is attributable, by-and-large, to the exclusion of institutionalized persons from the survey population rather than from a failure of the match to identify deaths occurring in the survey population. This case cannot be made as convincingly for the 1991 panel match estimates, but given the essentially identical design of the two panels, the use of the same match procedures with very similar outcomes, and reliance on the same beneficiary record systems to identify deaths among panel members, it seems to represent a plausible view, nonetheless.

## **CONCLUSIONS**

Public release of information characterizing type of OASDI benefit for social beneficiaries provides significant expansion information of use to social scientists and policy researchers interested in the contribution that the Social Security program makes to the economic well-being of the disabled, retired and survivor populations in the United States. By increasing the reach of the SIPP for research on the OASDI program, release of the type of benefit code contributes directly to the mission of a survey whose central purpose is to document participation in the Nation's transfer programs. The release is also advantageous to the Social Security Administration, since it brings with it the potential to generate additional research on its principal program. The release of information on year of death also increases the research potential of the SIPP, especially to the extent that the data on year of death contributes to a better understanding of the

effects of socio-economic status on longevity among the aged. Apart from this general question, better understanding of differential mortality among the aged is of importance to the Social Security Administration given the central role that actuarial considerations play in the design of its programs. Release of the mortality information also offers an opportunity for research leading to a better understanding of the mortality experience of various beneficiary subpopulations *per se*.

These potential advantages for the SIPP program and for increasing the relevance of the SIPP for social-security related research would clearly be strengthened were such releases to become routine and if the time-horizon for the release of mortality information were to be extended. Sample size limitations, which are potentially a crucial limiting factor for studies based on even the two panels involved in the current release, could be more adequately addressed, if matched data on type of benefit were to become available for the balance of extant panels. Sample size will necessarily be a consideration for any mortality research undertaken with these data as well. Finally, if the time horizon for the release of year of death data were to be expanded sufficiently, in principle, virtually the entire mortality experience of the elderly cohort defined as of the beginning of any SIPP panel could eventually be observed.

If future releases are undertaken, there are a few important steps of a technical nature that could arguably improve their usefulness for research purposes. For example, providing a flag that identified whether a usable SSN was available to support the survey-benefit record match would better enable users to assess possible biases associated with the presence or absence of an SSN, permit better definition of the appropriate universe for the matched characteristics, and support reweighting to deal with the effects of

nonmatch bias. Finally, the potential effects of mismatches on the estimates intended for release could also be investigated and reported on.

## REFERENCES

- Anderson, Robert N., Kenneth D. Kochanek and Sherry L. Murphy (1997).  
*Report of Final Mortality Statistics, 1995, Monthly Vital Statistics Report*, vol. 45, no. 11, Supplement 2, June 12.
- Aziz, Faye and Warren Buckler (1992).  
*The Status of Death Information in Social Security Administration Files*, American **Statistical Association Proceedings, Social Statistics Section, 1992**, Alexandria, Va., American Statistical Association, pp. 262-267.
- Bureau of the Census (1991).  
*Survey of Income and Program Participation Users Guide (Supplement to the Technical Documentation)*, Washington, D.C.
- Bureau of the Census (1991).  
*Source and Accuracy Statement for the 1990 Public Use Files from the Survey of Income and Program Participation, Survey of Income and Program Participation (SIPP) 1990 Panel, Wave 1 Retangular Microdata File Technical Documentation*, prepared by the Data User Services Division, Bureau of the Census, Washington: The Bureau, 1991, pp 8-1 to 8-39.
- Bye, Barry V. and Salvatore J. Gallicchio (1993).  
*Sampling Variance Estimates for SSA Program Recipients from the 1990 Survey of Income and Program Participation*, **Social Security Bulletin**, vol. 53, no. 3, pp. 75-87.
- Bye, Barry V. and Salvatore J. Gallicchio (1988).  
*A Note on Sampling Variance Estimates for Social Security Program Participants from the Survey of Income and Program Participation*, **Social Security Bulletin**, vol. 55, no. 10, pp. 4-21.
- Grad, Susan (1989).  
*Income and assets of Social Security beneficiaries by type of benefit*, **Social Security Bulletin**. vol. 52, no. 1, pp. 2-10.
- Jabine, Thomas B. (1990)  
**Survey of Income and Program Participation (SIPP): Quality Profile**, U.S. Department of Commerce, Bureau of the Census, May.
- Manton, Kenneth G. (1988).  
*A longitudinal Study of Functional Change and Mortality in the United States*, **Journal of Gerontology: SOCIAL SCIENCES**, vol. 43, no. 5, S153-61.
- Martini, Alberto and Daniel Dowhan (1997).  
*Why SIPP and CPS Produce Different Measures Among the Elderly*, **Social Security Bulletin**, vol. 60, no. 4, pp. 50-55.
- Scheuren, Frederick and H. Loch Oh (1975).  
*Fiddling Around with Mismatches and Nonmatches*, **American Statistical Association Proceedings, Social Statistics Section, 1975**, pp. 673-678.
- Social Security Administration (1987-1992, 1999).  
*Annual Statistical Supplement to the Social Security Bulletin*, Washington. D.C., U.S. Government Printing Office.



Vaughan, Denton R. (1992).

*Description and evaluation of [the] type of benefit code (TOB) for the May 1990 Survey of Income and Program Participation file*, note to the record, Social Security Administration, Office of Research and Statistics, Division of Economic Research, November 17.

Vaughan, Denton R. (1989).

*Development and Evaluation of a Survey-Based Type of Benefit Classification for the Social Security Program. **Social Security Bulletin** vol. 52, no. 1, pp. 12-26.*

Vaughan, Denton R. (1978).

*Errors in Reporting Supplemental Security Income Reciprocity in a Pilot Household Survey, **American Statistical Association Proceedings, Section on Survey Research Methods, 1978.*** American Statistical Association, Washington, D.C., pp. 288-293.

Table A.--Number of Social Security beneficiaries in current payment status, December 1990, by gender and selected type of benefit based on independent and matched survey estimates, 1990 Sipp panel

[Number in thousands]

Code value	Code category description(type of benefit)	Independent estimate <sup>1</sup>	Matched survey	Program minus survey †	DIFF/ S.E. diff †	Survey as percent of program <sup>2</sup>
		Total				
	Total	36,367	34,671	1,696 *	1.93	95.3
	Subtotal, age 18 or over.....	34,024	32,894	1,130		96.7
	Subtotal, excluding all other.....	32,790	32,046	743	0.9	97.7
	All retired workers excluding					
1	dually entitled women.....	18,834	18,375	459	0.9	97.6
2	Disabled worker.....	2,807	2,624	183	1.3	93.5
3,4	Aged wife.....	4,844	4,659	185	1.0	96.2
3	Not dually entitled <sup>3</sup> .....	2,884	2,642	241 *	1.7	91.6
4	Dually entitled .....	1,961	2,017	(56)	(0.5)	102.9
5,6	Aged widow.....	6,304	6,388	(84)	(0.4)	101.3
5	Not dually entitled <sup>5</sup> .....	4,166	4,172	(6)	(0.0)	100.1
6	Dually entitled <sup>6</sup> .....	2,139	2,216	(78)	(0.6)	103.6
7	All other .....	3,577	2,624	953 **	6.9	73.4
	Under age 18 .....	2,343	1,777	566 **	4.2	75.8
		Men <sup>8</sup>				
	Subtotal.....	...	...	...	...	...
	Subtotal, excluding all other.....	13,963	13,689	274	0.7	98.0
1	All retired worker men.....	12,124	12,102	22	0.1	99.8
2	Disabled worker.....	1,839	1,587	253 **	2.5	86.3
3,4	Aged husband.....	(4)	(4)	...	...	...
3	Not dually entitled .....	(4)	(4)	...	...	...
4	Dually entitled .....	(4)	(4)	...	...	...
5,6	Aged widower.....	(4)	(4)	...	...	...
5	Not dually entitled .....	(4)	(4)	...	...	...
6	Dually entitled.....	(4)	(4)	...	...	...
7	All other .....	(7)	1,199	...	...	...
	Under age 18.....	(7)	9 964	...	...	...

See notes at end of table

Table A.--Number of Social Security beneficiaries in current payment status, December 1990, by gender and selected type of benefit based on independent and matched survey estimates, 1990 Sipp panel, Continued

[Number in thousands]

Code value	Code category description(type of benefit)	Independent estimate <sup>1</sup>	Matched survey	Program minus survey †	DIFF/ S.E. diff †	Survey as percent of program <sup>2</sup>
		Women <sup>8</sup>				
	Subtotal.....	...	...	...	...	...
	Subtotal, excluding all other.....	18,820	18,357	462	0.9	97.5
1	Retired worker, not dually entitled.....	6,708	6,273	435 *	1.84	93.5
2	Disabled worker.....	968	1,037	(69)	(0.85)	107.1
3,4	Aged wife.....	4,847	4,659	188	1.0	96.1
3	Not dually entitled <sup>3</sup> .....	2,884	2,642	241 *	1.7	91.6
4	Dually entitled .....	1,964	2,017	(53)	(0.5)	102.7
5,6	Aged widow.....	6,304	6,388	(84)	(0.4)	101.3
5	Not dually entitled <sup>5</sup> .....	4,166	4,172	(6)	(0.0)	100.1
6	Dually entitled <sup>6</sup> .....	2,139	2,216	(78)	(0.62)	103.6
7	All other .....	(7)	1,425	...	...	...
	Under age 18.....	(7)	9 813	...	...	...

(. . .) - Not applicable or not available.

\*\* -- Difference significant at the 0.05 level using a two-tailed test.

\* -- Difference significant at the 0.10 level using a two-tailed test.

†† Tests of the statistical significance of a difference between matched survey estimates and program data are based on generalized variance parameters developed by Bye and Gallicchio (SSB, Fall 1993, vol. 56, pp. 75-87) for survey estimates of social security beneficiaries, and for program data, table 10.A1, p. 315, SSB, 1991 ASS.

<sup>1</sup> Derived from program estimates generally based on 10 percent sample data, with nominal adjustments to account for the differences between the program and survey universes.

<sup>2</sup> Survey estimate divided by program estimate times 100.

<sup>3</sup> Includes only wives with entitlement based on age.

<sup>4</sup> Adult males receiving other than retired or disabled worker benefits are included in program data "total" and subtotals for "age 18 or over" and "excluding all other". The corresponding survey estimates are included in "all other".

<sup>5</sup> Includes only widows with entitlement based on age.

<sup>6</sup> Excludes dually entitled disabled widows.

<sup>7</sup> Distribution by gender not available from program data. Included in the appropriate row of the total panel of the table.

<sup>8</sup> Gender classification based on the Master Beneficiary Record (MBR) data unless noted otherwise.

<sup>9</sup> Gender for minor beneficiaries under age 18 based on survey report from 4th wave.

Source: Social Security Bulletin, Annual Statistical Supplement 1991, tables 5.A1 and 5.G2 and the 1990 SIPP panel matched to SSA program records. Matched estimates employ January 1991 cross-sectional weight with no correction for nonmatches.

Run stream TOB\_DOD\_COMB2 of 2/9/00 (Job 10 @ 12:15)

Data file: sv790mrg of 1/6/00.

I:\TONY\[tobd90cmb\_expB\_2\_10.xls]Sheet1

11/22/99; r rev. 12/05/2000

Table B.--Number of Social Security beneficiaries in current payment status, December 1991, by gender and selected type of benefit based on independent and matched survey estimates, 1991 SIPP panel

[Number in thousands]

Code value	Code category description (benefit type)	Program data <sup>1</sup>	Matched survey	Program minus survey †	DIFF/ S.E. dif †	Survey as percent of program <sup>2</sup>
		Total				
	Total.....	37,054	34,462	2,592 **	2.4	93.0
	Subtotal, age 18 or over.....	34,656	32,550	2,106 **	2.0	93.9
	Subtotal, excluding all other.....	33,398	31,623	1,775 *	1.8	94.7
	Women with only a retired worker benefit and all retired worker men .....					
1	Disabled worker.....	19,094	18,412	682	1.1	96.4
2	Aged wife .....	2,981	2,669	312 *	1.8	89.5
3,4	Wife only <sup>3</sup> .....	4,923	4,511	412 *	1.7	91.6
3	Dual wife .....	2,886	2,464	421 **	2.6	85.4
4	Aged widow.....	2,038	2,047	(10)	(0.1)	100.5
5,6	Widow only <sup>5</sup> .....	6,400	6,031	369	1.3	94.2
5	Dual widow <sup>6</sup> .....	4,184	3,862	322	1.5	92.3
6	All other .....	2,216	2,169	47	0.3	97.9
7	Under age 18.....	3,656	2,839	817 **	4.6	77.7
		2,399	<sup>9</sup> 1,912	486 **	3.4	79.7
		Men <sup>8</sup>				
	Subtotal.....	...	...	...	...	...
	Subtotal, excluding all other.....	14,283	13,746	536	1.0	96.2
1	All retired worker men.....	12,349	12,035	314	0.7	97.5
2	Disabled worker.....	1,934	1,711	222 *	1.7	88.5
3,4	Aged husband.....	(4)	(4)			
3	Husband only.....	(4)	(4)	...	...	...
4	Dual husband.....	(4)	(4)	...	...	...
5,6	Aged widower.....	(4)	(4)			
5	Widower only .....	(4)	(4)	...	...	...
6	Dual widower .....	(4)	(4)	...	...	...
7	All other .....	(7)	1,219	...	...	...
	Under age 18.....	(7)	<sup>9</sup> 1,045	...	...	...
	See notes at end of table.....					

Table B.--Number of Social Security beneficiaries in current payment status, December 1991, by gender and selected type of benefit based on independent and matched survey estimates, 1991 SIPP panel, Continued

[Number in thousands]

Code value	Code category description (benefit type)	Program data <sup>1</sup>	Matched survey	Program minus survey †	DIFF/ S.E. dif †	Survey as percent of program <sup>2</sup>
		Female <sup>8</sup>				
	Subtotal.....	...	...	...	...	...
	Subtotal, excluding all other.....	30,447	28,419	2,027 **	2.20	93.3
1	Retired worker, not dually entitled.....	6,745	6,376	368	1.24	94.5
2	Disabled worker.....	1,048	958	90	0.92	91.4
3,4	Aged wife .....	4,923	4,511	412 *	1.7	91.6
3	Wife only <sup>3</sup> .....	2,886	2,464	421 **	2.55	85.4
4	Dual wife .....	2,038	2,047	(10)	(0.06)	100.5
5,6	Aged widow.....	6,404	6,031	373	1.3	94.2
5	Widow only <sup>5</sup> .....	4,188	3,862	326	1.51	92.2
6	Dual widow <sup>6</sup> .....	2,216	2,169	47	0.30	97.9
7	All other .....	(7)	1,620	...	...	...
	Under age 18 .....	(7)	9 867	...	...	...

(. . .) - Not applicable or not available.

\*\* -- Difference significant at the 0.05 level using a two-tailed test.

\* -- Difference significant at the 0.10 level using a two-tailed test.

† Tests of the statistical significance of a difference between matched survey estimates and program data are based on generalized variance parameters developed by Bye and Gallicchio (SSB, Fall 1993, vol. 56, pp. 75-87) for survey estimates of social security beneficiaries as adjusted for the smaller 1991 panel sample size, and for program data, table 10.A1, SSB, 1991 ASS, p. 315.

<sup>1</sup> Program estimates generally based on 10 percent sample data with nominal adjustments to account for the differences between the program and survey universes.

<sup>2</sup> Survey estimate divided by program estimate times 100.

<sup>3</sup> Includes only wives with entitlement based on age.

<sup>4</sup> Adult males receiving other than retired or disabled worker benefits are included in program data "total" and subtotals for "age 18 or over" and "excluding all other". The corresponding survey estimates are included in "all other".

<sup>5</sup> Includes only widows with entitlement based on age.

<sup>6</sup> Excludes dually entitled disabled widows.

<sup>7</sup> Distribution by gender not available from program data. Included in the appropriate row of the total panel of the table.

<sup>8</sup> Gender classification based on the Master Beneficiary Record (MBR) data unless noted otherwise

<sup>9</sup> Gender of beneficiaries under age 18 based on survey report 4th wave.

Source: Social Security Bulletin, Annual Statistical Supplement 1992, tables 5.A1 and 5.G2 and the 1990 SIPP panel matched to SSA program records. Matched estimates employ January 1992 cross-sectional weight with no correction for nonmatches.

Run stream: TOBj92\_db1 of 2/1/00 (run time 16:13) Data file: sv691mrg of 1/6/00.

I:\TONY\[tobd91cmb\_expA\_1\_28.xls]Sheet1

3/2/2000; rev.: 12/05/2000

Table C.--Percent of people in current pay as of December 1990 and 1991 for whom a survey social security amount is present for the subsequent January, 1990 and 1991 SIPP panels

Benefit category (code value)	[Number in thousands]			
	December 1990 current payment status according to SSA benefit records		December 1991 current payment status according to SSA benefit records	
	Total <sup>1</sup>	Social security benefit reported in survey for 1/91, as percent of total	Total <sup>1</sup>	Social security benefit reported in survey for 1/92, as percent of total
Total.....	248,688	14.2	251,270	14.2
Subtotals.....				
Matched and in current pay.....	34,671	92.2	34,462	91.2
Excluding other.....	32,046	97.5	31,623	96.8
Age 18 or over, only.....	32,894	96.9	32,550	96.3
Retired worker (1).....	18,375	97.8	18,412	97.6
Disabled worker (2).....	2,624	90.2	2,669	87.9
Aged wife.....	4,659	98.2	4,511	97.4
Not dually entitled (3).....	2,642	97.8	2,464	97.1
Dually entitled (4).....	2,017	98.7	2,047	97.8
Aged widow.....	6,388	99.0	6,031	97.9
Not dually entitled (5).....	4,172	99.1	3,862	97.0
Dually entitled (6).....	2,216	98.8	2,169	99.6
All other (7).....	2,624	27.2	2,839	29.3
Age 18 and over <sup>2</sup> .....	847	75.6	926	78.7
Under age 18 <sup>2</sup> .....	1,777	4.1	1,912	5.44
Matched and not in current pay, matched but no MBR record located, or not matched (9) <sup>3</sup> .....	214,017	1.63	216,808	2.05

<sup>1</sup>Using public use core file weight for the corresponding January.

<sup>2</sup>Age based on Master Beneficiary Record.

<sup>3</sup>Those reporting reciprocity in the survey in this subgroup represent, respectively, 11.4 and 14.3 percent of the number of individuals reported to be receiving a social security benefit in the survey for January 1991 and January 1992.

Source: 1990 and 1991 SIPP panel public use files matched to SSA beneficiary records.

Data files: sv790mrg.sd2 and sv691mrg.sd2 of 1/6/00.

Run streams: TOB\_DOD\_COMB1 & TOBDODCOMB2 (Jobs2 & 3 of 2/23/00).

I:\TONY\[dol\_con9091.xls]Sheet1

3/2/2000; rev.: 12/05/2000

Table D.-- Reconciliation of current pay status from the record match with survey reports of benefit receipt based on the sample for whom benefit records were located

[Numbers in thousands]

	1990 panel		1991 panel	
	Number	Percent	Number	Percent
Benefit record located, benefit in current pay.....	32,894	100.0	32,550	100.0
Benefit reported in survey.....	31,882	96.9	31,338	96.3
Benefit not reported in survey.....	1,012	3.1	1,212	3.7
Benefit record located, no benefit in current pay, but benefit reported in survey.....	623	1.9	731	2.2
Net bias in survey reciprocity reports among those with benefit record.....	390	1.2	481	1.5
-----				
Report benefit in survey, no evidence of a current pay benefit from record.....	3,473	100.0	4,436	100.0
No record located.....	2,851	82.1	3,705	83.5
Record located, no benefit in current pay.....	623	17.9	731	16.5

Source: 1990 and 1991 SIPP panel public use files matched to SSA beneficiary records.

Table E.--Reconciliation of estimates of the number of deaths of people aged 65 and over as identified from the SIPP-SSA benefit record match and the U.S. death registration system, initial two calendar years of the 1990 and 1991 SIPP panels -31-

[Number in thousands]				
Element	1990 panel, calendar years '90-'91		1991 panel, calendar years '91-'92	
	Number	As % of total	Number	As % of total
Total reported deaths, U.S. resident population <sup>1</sup> .....	3,105.6	100.0	3,138.7	100.0
<b>Less:</b>				
Estimated deaths among nonbeneficiaries <sup>2</sup> .....	173.9	5.6	174.5	5.6
<b>Equals:</b>				
Estimated number of deaths of U.S. residents observable in SSA benefit records...	2,931.7	94.4	2,964.2	94.4
<b>Less:</b>				
Estimated deaths occurring in 1990-91 or 1991-92 among beneficiaries institutionalized at time the survey sample was drawn <sup>3</sup> .....	439.8	14.2	444.6	14.2
<b>Equals:</b>				
Estimated number of beneficiary deaths occurring in the survey population.....	2,491.9	80.2	2,519.6	80.3
<b>Less:</b>				
Estimated beneficiary deaths occurring in the survey population during Jan.- April of the initial panel year, but not observable from survey <sup>4</sup> .....	282.8	9.1	274.0	8.7
<b>Equals:</b>				
Estimated number beneficiary deaths occurring in the survey population, less those occurring prior to the initial interview of the first calendar year of the panel.....	2,209.2	70.4	2,245.5	71.5
<b>Less:</b>				
Estimated number of deaths occurring in the survey population as adjusted above and attributable to nonmatched beneficiaries <sup>5</sup> .....	154.6	5.0	224.6	7.2
<b>Equals:</b>				
Expected number of deaths of U.S. residents observable based on the survey-benefit record match.....	2,054.5	66.2	2,021.0	64.4
<b>Survey estimate</b>				
Number.....	1,856.6	59.8	1,932.1	61.6
Standard error.....	112.4	3.6	143.4	4.6
Survey estimate as percent of expected estimate.....	90.4	...	95.6	...
90 percent confidence interval				
Upper bound (number and as % of expected estimate).....	2,044.3	99.5	2,171.5	107.4
Lower bound (number and as % of expected estimate).....	1,668.9	81.2	1,692.7	83.8

<sup>1</sup> Centers for Disease Control, see tables 4 and 7 for sources.

<sup>2</sup> Estimated as the product of the total number of reported deaths and the proportion of the population aged 65 and over not receiving social security or supplemental security income benefits in 1990 and 1991 (.056).

<sup>3</sup> Estimated as 15 percent of deaths among beneficiaries aged 65 and over occurring in the first two calendar years covered by each panel based on findings of the 1982 and 1984 National Long Term Care Surveys.

<sup>4</sup> Estimated as the cumulative product of the estimated number of deaths of persons aged 65 and over that occurred in January through April of the initial year of each panel, the proportion of the elderly population receiving benefits (.944), the estimated proportion of deaths attributable to the noninstitutionalized (.85) population, and the proportion of January - April calendar time (.625) attributable to an initial interview estimate.

<sup>5</sup> Estimated as the product of the proportion of the aged 65 and over sample without a valid SSN in the 1990 and 1991 panels (.07 and .10 respectively) and the number of aged beneficiary deaths occurring in the survey population, net of initial interview and other effects previously accounted for.

Source: 1990 and 1991 SIPP panel public use files matched to SSA beneficiary records.



Table F.--Number of deaths 1990-1996, by year, sex,<sup>1</sup> and age<sup>2</sup> as reported to the NCHS and as identified from the 1990 SIPP panel based on a match to Social Security Administration benefit records, people age 65 and over in year of death

[Numbers in thousands]<sup>3</sup>

Characteristic	Number of deaths		Survey estimate					Subgroup differences			
			Standard error	90 % confidence interval		As a percent of the NCHS estimate		S.E. of difference	Difference		Subgroup
	NCHS	Survey		Lower bound	Upper bound	Percent	S.E.		Difference	± S.E.	
<b>Year and sex by year</b>											
Total	11,413	8,889	300	8,388	9,390 -	77.9	2.63	7.65	-19.37	-2.53	Female vs. male total
1990.....	1,542	746	68	632	861 -	48.4	4.44	7.00	-22.64	-3.23	90 vs. 91
1991.....	1,564	1,110	85	969	1,252 -	71.0	5.41	7.90	-8.84	-1.12	91 vs. 92
1992.....	1,575	1,258	91	1,107	1,409 -	79.9	5.75	8.16	-4.16	-0.51	92 vs. 93
1993.....	1,654	1,390	96	1,230	1,550 -	84.0	5.78	8.20	-1.37	-0.17	93 vs. 94
1994.....	1,670	1,426	97	1,264	1,588 -	85.4	5.81	8.06	4.91	0.61	94 vs. 95
1995.....	1,694	1,364	95	1,206	1,522 -	80.5	5.59	8.22	-12.59	-1.53	95 vs. 96
1996.....	1,714	1,595	103	1,423	1,767 +	93.1	6.02	...	...	...	Null
Subtotal	5,263	4,648	194	4,324	4,972 -	88.3	7.04	...	...	...	Males
1990.....	703	437	52	351	524 -	62.1	7.36	10.70	-9.25	-0.86	90 vs. 91
1991.....	732	522	57	427	617 -	71.4	7.76	11.75	-20.19	-1.72	91 vs. 92
1992.....	735	673	65	565	782 +	91.6	8.82	12.81	-13.28	-1.04	92 vs. 93
1993.....	766	803	71	684	922 +	104.9	9.29	12.80	10.06	0.79	93 vs. 94
1994.....	768	728	68	615	841 +	94.8	8.80	12.31	3.00	0.24	94 vs. 95
1995.....	776	713	67	601	824 +	91.8	8.61	12.39	-6.84	-0.55	95 vs. 96
1996.....	782	771	70	655	888 +	98.6	8.91	...	...	...	Null
Subtotal	6,151	4,241	183	3,935	4,547 -	69.0	2.98	...	...	...	Females
1990.....	839	309	43	237	381 -	36.8	5.16	8.91	-33.85	-3.80	90 vs. 91
1991.....	832	588	60	487	689 -	70.7	7.26	10.21	1.08	0.11	91 vs. 92
1992.....	840	585	60	484	685 -	69.6	7.17	9.88	3.55	0.36	92 vs. 93
1993.....	888	587	60	486	688 -	66.1	6.79	9.99	-11.33	-1.13	93 vs. 94
1994.....	902	698	66	587	808 -	77.4	7.33	10.10	6.46	0.64	94 vs. 95
1995.....	918	651	64	545	758 -	70.9	6.94	10.40	-17.48	-1.68	95 vs. 96
1996.....	932	824	72	703	944 +	88.4	7.74	...	...	...	Null
<b>Age and age by sex</b>											
Total	11,413	8,889	300	8,388	9,390 -	77.9	2.63	8.39	-1.81	-0.23	Total
65 and over.....	1,470	1,281	91	1,128	1,434 -	87.1	6.22	7.70	0.84	0.14	65-69 vs. 70-74
70-74.....	1,892	1,683	106	1,505	1,860 -	88.9	5.62	6.07	19.87	3.12	70-74 vs. 75-79
75-79.....	2,168	1,910	114	1,719	2,101 -	88.1	5.27	6.38	20.70	3.41	75-79 vs. 80 plus
80 and over.....	5,884	4,015	177	3,719	4,311 -	68.2	3.01	6.91	18.90	2.73	80 plus vs. 85 plus
Males	5,263	4,648	194	4,324	4,972 -	88.3	3.69	10.94	-1.66	-0.16	Males
65 and over.....	863	771	70	655	888 +	89.3	8.07	10.30	0.12	0.01	65-69 vs. 70-74
70-74.....	1,063	967	79	836	1,098 +	91.0	7.39	8.82	5.55	0.62	70-74 vs. 75-79
75-79.....	1,133	1,029	81	893	1,165 +	90.9	7.17	9.00	22.75	0.64	75-79 vs. 80 plus
80 and over.....	2,204	1,880	113	1,691	2,070 -	85.3	5.14	9.57	21.09	0.42	80 plus vs. 85 plus
Females	6,151	4,241	183	3,935	4,547 -	69.0	2.98	12.28	-2.30	-0.21	Females
65 and over.....	606	510	56	416	603 -	84.0	9.25	10.84	1.24	0.16	65-69 vs. 70-74
70-74.....	829	716	67	604	828 -	86.3	8.08	7.94	27.08	3.10	70-74 vs. 75-79
75-79.....	1,035	881	75	756	1,006 -	85.1	7.22	8.73	18.08	3.56	75-79 vs. 80 plus
80 and over.....	3,680	2,135	122	1,932	2,338 -	58.0	3.31	9.82	15.78	2.65	80 plus vs. 85 plus

See also notes for table B-8.

(...) - Not applicable

"+" - Value of upper bound on the 90-percent confidence interval exceeds corresponding NCHS estimate.

"-" - Value of upper bound on the 90-percent confidence interval is less than the corresponding NCHS estimate.

<sup>1</sup>Sex as represented on the MBR.

<sup>2</sup>Survey age as of January of 1990 adjusted to January of year of death as identified in SSA administrative records.

<sup>3</sup>Survey deaths weighted by the January 1990 cross-sectional weight with no adjustment for nonmatches.

Table G.--Number of deaths 1991-1996, by year, sex,<sup>1</sup> and age<sup>2</sup> as reported to the NCHS and as identified from the 1991 SIPP panel based on a match to Social Security Administration benefit records, people age 65 and over in year of death

[Numbers in thousands]<sup>3</sup>

Characteristic	Number of deaths		Survey estimate					Subgroup differences				
			Standard error	90 % confidence interval		As a percent of the NCHS estimate		S.E. of difference	Difference	Difference ÷ S.E.	Subgroup	
	NCHS	Survey		Lower bound	Upper bound	Percent	S.E.					
Year and year by sex												
Total	9,871	7,190	323	6,651	7,729 -	72.8	3.27	5.98	13.3	2.22	Female vs. male, total	
Total by year	1991.....	1,564	705	83	567	844 -	45.1	5.30	8.84	-32.8	-3.71	91 vs. 92
	1992.....	1,575	1,227	111	1,041	1,413 -	77.9	7.08	9.91	-0.3	-0.03	92 vs. 93
	1993.....	1,654	1,294	115	1,102	1,485 -	78.2	6.93	9.90	-3.5	-0.35	93 vs. 94
	1994.....	1,670	1,363	118	1,166	1,561 -	81.7	7.07	9.80	4.9	0.50	94 vs. 95
	1995.....	1,694	1,300	115	1,108	1,492 -	76.7	6.79	9.55	0.8	0.08	95 vs. 96
	1996.....	1,714	1,302	115	1,109	1,494 -	76.0	6.72	...	...	...	Null
Subtotal		4,559	3,647	208	3,299	3,995 -	80.0	4.57	...	...	...	Males
Males by year	1991.....	732	417	63	311	522 -	56.9	8.61	13.68	-28.9	-2.11	91 vs. 92
	1992.....	735	631	78	501	762 +	85.8	10.63	14.51	8.6	0.60	92 vs. 93
	1993.....	766	591	76	465	717 -	77.2	9.87	14.72	-16.7	-1.14	93 vs. 94
	1994.....	768	721	84	581	861 +	93.9	10.92	15.21	4.5	0.30	94 vs. 95
	1995.....	776	694	82	556	831 +	89.3	10.59	14.35	13.5	0.94	95 vs. 96
	1996.....	782	594	76	467	720 -	75.9	9.68	...	...	...	Null
Subtotal		5,311	3,543	205	3,201	3,885 -	66.7	3.85	...	...	...	Females
Females by year	1991.....	832	289	52	201	376 -	34.7	6.27	11.00	-36.2	-3.29	91 vs. 92
	1992.....	840	596	76	469	722 -	70.9	9.03	12.97	-8.1	-0.63	92 vs. 93
	1993.....	888	702	83	564	841 -	79.1	9.31	12.78	7.8	0.61	93 vs. 94
	1994.....	902	642	79	510	774 -	71.2	8.75	12.09	5.2	0.43	94 vs. 95
	1995.....	918	606	77	478	734 -	66.0	8.34	12.21	-10.0	-0.82	95 vs. 96
	1996.....	932	708	83	569	847 -	76.0	8.92	...	...	...	Null
Age and age by sex												
Total by age	65 and over.....	9,871	7,190	323	6,651	7,729 -	72.8	3.27	10.53	0.8	0.08	Total
	65-69.....	1,253	985	99	820	1,150 -	78.6	7.90	9.69	-4.2	-0.43	65-69 vs. 70-74
	70-74.....	1,631	1,270	114	1,080	1,459 -	77.8	6.96	7.79	15.5	1.99	70-74 vs 75-79
	75-79.....	1,867	1,531	126	1,320	1,741 -	82.0	6.74	7.98	11.3	1.42	75-79 vs. 80 plus
	80 and over.....	5,120	3,405	200	3,072	3,739 -	66.5	3.90	8.81	12.1	1.38	70-74 vs 80 plus
Males by age	65 and over.....	4,559	3,647	208	3,299	3,995 -	80.0	4.57	13.82	-1.7	-0.12	Males
	65-69.....	735	584	75	459	709 -	79.4	10.21	13.26	-7.3	-0.55	65-69 vs. 70-74
	70-74.....	915	742	85	600	884 -	81.1	9.30	11.38	12.9	1.13	70-74 vs 75-79
	75-79.....	975	861	92	707	1,015 +	88.4	9.45	11.26	5.6	0.50	75-79 vs. 80 plus
	80 and over.....	1,935	1,460	123	1,255	1,665 -	75.5	6.34	12.02	4.0	0.33	70-74 vs 80 plus
Females by age	65 and over.....	5,311	3,543	205	3,201	3,885 -	66.7	3.85	15.54	3.9	0.25	Females
	65-69.....	517	401	62	298	504 -	77.5	11.94	13.44	-1.4	-0.10	65-69 vs. 70-74
	70-74.....	716	528	71	409	647 -	73.7	9.94	10.11	14.0	1.38	70-74 vs 75-79
	75-79.....	892	669	81	535	804 -	75.0	9.04	10.92	12.6	1.15	75-79 vs. 80 plus
	80 and over.....	3,186	1,945	144	1,705	2,186 -	61.1	4.52	12.77	4.0	1.29	70-74 vs. 80 plus

See also notes for table B-9.

(...) - Not applicable

"+" - Value of upper bound on the 90-percent confidence interval exceeds corresponding NCHS estimate.

"-" - Value of upper bound on the 90-percent confidence interval is less than the corresponding NCHS estimate.

<sup>1</sup>Sex as represented on the MBR.

<sup>2</sup>Survey age as of January of 1990 adjusted to January of year of death as identified in SSA administrative records.

<sup>3</sup>Survey deaths weighted by the January 1990 crosssectional weight with no adjustment for nonmatches.

## Appendix A. Definitions and Explanations

**Social security benefits, overview.** Benefits paid under the Old-Age, Survivors, and Disability Insurance (OASDI) program, popularly known as social security, may be classified in two broad categories: (1) worker (primary) benefits and (2) auxiliary (secondary) benefits. Primary (worker benefits) are payable to workers who meet the insured status, and age or disability criteria of the social security program. Auxiliary benefits are payable to the dependents (generally children and spouses) of a retired or disabled worker, or of an appropriately insured deceased worker.

Some individuals are simultaneously entitled to a primary and auxiliary benefit. If the auxiliary benefit exceeds the benefit payable on the basis of the individual's own work record, the individual is considered to be dually entitled.

Additional details concerning the characteristics defining benefit receipt and type of benefit that are relevant to the benefit types identified by the type of benefit code for the 1990 and 1991 panels of the Survey of Income and Program Participation (SIPP) are provided below.

**In current-payment status.** Eligible for a monthly payment, provided applicable deductions (excluding the monthly Medicare Supplementary Medical Insurance (SMI) premium) are less than a full month's benefit. A benefit in current-payment status at the end of a month is usually payable in the first week of the following month.

The type of benefit code developed for the 1990 and 1991 SIPP panels is intended to identify sample members in current payment status as of December 1990 and 1991 respectively. Such individuals would have received their benefits via the mail or electronic deposit in following month, that is January 1991 for the 1990 panel and January 1992 for the 1991 panel.

**Disabled.** The inability to engage in substantial gainful activity by reason of any medically determinable physical or mental impairment that can be expected to result in death or to last for a continuous period of not less than 12 months. This means for a disabled worker, a disabled adult child, or a disabled widow, widower or surviving divorced spouse, the inability to engage in substantial gainful activity. A person must not only be unable to do his or her previous work but cannot, considering age, education, and work experience, engage in any other kind of substantial gainful work which exists in the national economy. Somewhat different criteria apply to blind beneficiaries.

**Substantial gainful activity (SGA).** Remunerative work that is substantial, as determined from consideration of the amount of money earned, and/or the number of hours worked, and the nature of the work. In 1990 and 1991 for nonblind beneficiaries, earnings above \$500 per month would ordinarily demonstrate substantial gainful activity. Earnings below \$300 per month would indicate that SGA has not occurred. If earnings are between \$300 and \$500 per month other factors are considered. Self-employment activity is generally evaluated in terms of the time spent and degree of effort, as

compared to that of non-disabled self-employed individuals. Somewhat different criteria apply to blind beneficiaries.

### **Primary (worker) benefits**

**Retired-worker (old-age) benefit.** Monthly benefit payable to a retired worker aged 62 or older who meets the insured status requirements of the social security program. This type of benefit is identified as (1) under the type of benefit code. Women entitled to a retired worker benefit and a larger auxiliary benefit are classified under the appropriate dual entitlement category (see below). In contrast, all men with a retired-worker benefit in current payment status are assigned to category 1 regardless of dual entitlement status.

**Disabled-worker benefit.** The monthly benefit payable to a worker under age 65 who meets the Social Security Administration's definition of disability and the requirements of disability insured status. All persons in current payment status for a disabled-worker benefit, regardless of dual entitlement status, are classified under type of benefit code category (2).

### **Auxiliary (secondary) benefits<sup>1</sup>**

**Aged wife only benefit.** The monthly benefit payable to a wife or a divorced wife of a retired or disabled worker if the wife or divorced wife is at least age 62 and is not entitled to a smaller primary benefit based on her own work record. A divorced wife must have been married to the worker for at least 10 years before the divorce became final. This type of benefit is classified under type of benefit code category (3).

**Dually entitled aged wife benefit (dual wife).** The monthly benefit payable to an individual as an aged wife, as defined above, when that wife is also entitled, simultaneously, to a smaller retired-worker benefit based on her own work record. This type of benefit is classified under type of benefit code category (4).

**Aged widow only benefit.** The monthly benefit payable to a widow or surviving divorced wife of a worker who met the appropriate insured status requirements at time of death. The individual must be aged 60 or older and not entitled to a smaller primary benefit based on her own work record. A surviving divorced wife's marriage to a worker must have lasted 10 years before the divorce became final. Widows and surviving divorced wives may remarry at age 60 or older and continue receiving widow benefits based on the prior marriage. This type of benefit is classified under type of benefit code category (5).

---

<sup>1</sup> Men entitled to auxiliary benefits, except those who are dually entitled, are included in type of benefit category 7. Men may receive benefits as the husband or surviving spouse of a retired, disabled, or deceased worker in the same manner as women.

**Dually entitled aged widow benefit (dual widow).** The monthly benefit payable to an individual as an aged widow, as defined above, when that widow is also entitled, simultaneously, to a smaller retired-worker benefit based on her own work record. This type of benefit is classified under type of benefit code category (6).

**Benefit types not identified separately (combined in code category 7) <sup>1</sup>**

**Wife benefit based on entitled child in her care.** The monthly benefit payable to the wife or divorced wife of a retired or disabled worker who has an entitled child of the worker in her care. The child must be under age 16 or disabled.

**Widowed mother benefit.** The monthly benefit payable to the widow or surviving divorced wife of a retired or disabled worker, or an appropriately insured deceased worker who has the entitled child of the worker in her care. The child must be under age 16 or disabled according to the definition employed by SSA.

**Minor child benefit.** The monthly benefit for a minor child of a retired or disabled worker, or an appropriately insured deceased worker. The child must be unmarried and under age 18. Under certain circumstances, a minor child's benefit may be payable to the stepchild or grandchild of an insured worker.

**Student child benefit.** The monthly benefit payable to a full-time unmarried elementary or secondary school student aged 18-19 who is the child of a retired or disabled worker, or an appropriately insured deceased worker. Student's benefits end at age 19 or at the end of the current semester or quarter, whichever is later.

**Disabled adult child benefit.** The monthly benefit for the child of a retired or disabled worker, or an appropriately insured deceased worker. The child must be aged 18 or older, disabled according to the definition employed by the Social Security Administration, and have a disability that began before age 22. There is no upper age limit for disabled adult children's benefits. With few exceptions, a disabled child must also be unmarried.

**Disabled widow benefit.** The monthly benefit payable to a widow of an appropriately insured worker who is (1) aged 50-64, and (2) meets the Social Security Administrations definition of disability as applied to widows.

## **Appendix B. Development of the OASDI type of benefit code and year of death variable**

### **GENERAL DESCRIPTION OF THE RECORD MATCH**

Both items included in this public release were developed from an exact match between individual members of the survey sample and agency benefit records systems based on the sample member's Social Security Number (SSN). SSN's were obtained as part of the regular survey interview. Reported numbers were subsequently validated based on standard SSA automated and manual procedures in effect during the early and mid 1990's. The validation involved comparison of comparison of name, date of birth, sex, and race as obtained from the survey with the same variables from SSA administrative records.

An attempt was also made to locate SSN's for sample members who did not provide them in the interview<sup>1</sup> as well as to locate valid SSN's for sample members whose reported SSN's were initially flagged as invalid. The search and validation procedures yielded SSN's considered to be operationally valid for approximately 93 percent of original sample members aged 65 or older from the 1990 panel and about 90 percent of original sample members of that age group from the 1991 panel. Post validation SSN availability rates were slightly lower for adults aged 18-64 (about 88-90 percent) and somewhat lower still for persons entering the sample after the initial interview and for persons under age 18 (about 80 percent for the 1990 panel<sup>2</sup>).

---

<sup>1</sup> However, SSN searches were not undertaken for persons who refused to provide an SSN.

<sup>2</sup> Though collected in the interview, SSN's for persons entering the sample after wave 4 were not validated for the 1990 panel.

Among persons age 18 and over and considered to have a valid SSN, all five confirmatory items employed in the validation process were in agreement more than four-fifths of the time for both the 1990 and 1991 panels. For virtually all cases considered as having a valid SSN, there was agreement on at least 4 of the five items. The characteristic with the highest level of disagreement was race (about 10 percent), followed by name (below 5 percent), date of birth (also less than 5 percent), and sex (less than 1 percent).

No explicit attempt is made here to assess the possible impact of mismatches on the representation of type of benefit or year of death. Clearly, mismatches are likely to be present, but there has been no assessment of the extent of mismatches in the context of linkages of SIPP panels and SSA administrative data. On the one hand the high level of consistency between current payment status as indicated for the cases taken to be matched and reports of benefit receipt in the survey suggest that the prevalence of mismatches is quite low; on the other hand the presence of a handful of cases for which the “matched” information yielded a year of death prior to the survey reference period, indicates that some mismatches are very likely present, as well (see discussion on year of death, below and appendix table B-7 ).<sup>3</sup>

#### **CREATION OF THE OASDI TYPE OF BENEFIT CODE.**

With the exception of presence in sample,<sup>4</sup> all information used in the development of the OASDI type of benefit code was taken from the Master Beneficiary Record (MBR) for the matched sample individuals from each panel. The basic unit of observation in the

---

<sup>3</sup> Those interested in obtaining a better sense of the potential impact of various matching errors in an environment somewhat similar to the one represented by the SIPP/SSA administrative record match may consult Scheuren and Oh (1975, pp. 627-633).

<sup>4</sup> As defined on the basis of a final weight of greater than zero for the month of January.

MBR system is the benefit record. Individuals frequently have multiple records present in the MBR corresponding to current pay benefits, prior current pay benefits, benefits applied for and denied, and so forth. Two different sets of MBR benefit record extracts were employed in this work. Individual variables in each extract were defined in a similar manner. Each contained selected information on benefit characteristics for three different time periods: as of December of the first and second calendar years of the respective panel, and as of the date the record system was accessed. The two extracts differed principally with respect to when the MBR was accessed<sup>5</sup>, in the procedures employed to locate primary and auxiliary records for SIPP sample members with usable SSN's, and in how information for primary and auxiliary benefits was consolidated.

The initial set of MBR extract records was developed in a two-stage process. First, records corresponding to retired and disabled-worker benefits for SIPP sample members with usable SSN's were identified. Second, a search for auxiliary benefit records was conducted, but it was restricted to two groups of sample members: 1) those whose retired or disabled worker record indicated the presence of dual entitlement to an auxiliary benefit, and 2) those for whom no retired or disabled-worker benefit had been initially located.

The resulting collection of benefit records was consolidated into a single record for each sample member for whom benefit records were located. In the majority of cases no dual entitlement exists, and the resulting record simply represents the individual's primary (worker) or auxiliary benefit. In instances of dual entitlement, the information for the dual auxiliary record was simply appended at the end of the individual's primary

---

<sup>5</sup> The initial extracts were obtained in March and April 1995 and the second in February and April 1997, respectively for the 1990 and 1991 panels.



(worker) benefit. For individuals dually entitled, the *front half* of the consolidated record contained information on the primary (worker) benefit and the “back half” contained information corresponding to the auxiliary benefit stemming from dual entitlement. For persons not subject to dual entitlement, the *front half* of the record corresponded to the primary (worker) or auxiliary benefit; the *back half* was blank.

This MBR information was combined with relevant information from the survey and from the Supplemental Security Income Record that could be used to identify death of an individual receiving SSI benefits.

As subsequently discovered by other SSA researchers, the sequential procedure for locating primary and auxiliary benefit records failed to uncover auxiliary benefits in instances where a primary (worker) benefit record existed but was not in current pay. Consequently, a second set of MBR records was developed that did not access primary and worker benefits in a sequential fashion and so was not subject to the problem of “missing” auxiliary records. The resulting file contained multiple benefit records for many individuals. These records had to be unduplicated and consolidated before they could be conveniently used for defining type of benefit and identifying year of death. The full collection of records was separated into two extract files which included only information required to define type of benefit and year of death, one corresponding to primary (worker) records and the other to auxiliary records. Records from each of the extract files corresponding to benefits in current pay as of December of the first full year of each panel were identified and matched.<sup>6</sup> The resulting consolidated record had two sections: the first corresponded to the records for primary (worker) benefits; the second to

---

<sup>6</sup> Subsequently, records for persons not in current pay as of the first part but showing evidence of death as of the date of extraction were added to the working file.

all records for auxiliary benefits, regardless of dual entitlement status. This second version of the MBR data was then combined with the existing file containing the survey, SSR and original MBR information.

### **Development of the type of benefit code**

There were four basic dimensions considered in defining the individual's type of OASDI benefit: 1) identification of type of benefit; 2) identification of current payment status; 3) identification of dual entitlement status; and 4) presence in sample in the month of benefit receipt.

*Identification of the type of benefit.*—Identification of type of benefit was based on a summary recode of the detailed Beneficiary Identification Code (BIC) known as the Type of Benefit Code (TOB). The relationship between the summary TOB recode and the BIC code is provided in table B-1.

Both MBR extracts employed in development of the code provided a representation of the TOB at three points in time (at extraction and as of December of the first and second full calendar years of each panel). The public release code was based on the TOB corresponding to the December of the first full calendar year of each panel, i.e., 12/91 and 12/92 respectively for the 1990 and 1991 panels. Auxiliary benefits for wives and widows were distinguished from the same type of benefit for husbands and widowers based on gender as given in the MBR.<sup>7</sup>

*Identification of current payment status.*—The payment status of benefits identified on the basis of TOB was confirmed by checking the so-called Ledger Account File (LAF) field. This field identifies the payment status of benefits covered by the MBR system. The LAF code was available for the same three periods as the TOB code. The version of the LAF corresponding to the December of the first full calendar year of each panel was used to define current payment status. Persons for whom no MBR record was located or for whom an MBR record was found but the record gave no indication of a benefit in current pay for the month of interest were assigned to the category "other," i.e., as (9).

*Identification of dual entitlement status.*—Women who were determined to have both a retired worker and an auxiliary benefit as an aged spouse or aged widow in current pay based on the TOB and LAF fields were designated as dually entitled. The small number of women dually entitled on the basis of disabled-worker benefits and all dually entitled males were simply classified according to their primary benefit.

*Presence in sample.*—The type of benefit code was only defined for sample persons for whom the final public use file weight was greater than zero for January of 1991 or January 1992, depending on which panel was involved.

---

<sup>7</sup> Users will note a few instances in which type of benefit is inconsistent with gender as taken from the survey.

Type of benefit assignment criteria are summarized in table B-2. Sample counts and weighted estimates for the code are given in table B-3.

*Final assignment of type of benefit.*--A separate type of benefit code was developed on the basis of each of the two sets of MBR records according to the previously described criteria. Comparison of the two codes indicated a high level of consistency with three significant exceptions involving a larger number of benefits in current pay for disabled workers, aged wives, and aged widows according to the second extract. The additional benefits for aged wives and widows occurred only among those without dual entitlement (see Table B-4).

Given that the second set of MBR records was obtained about two years after the first, yet was thought to offer a more complete account of benefits in current pay for auxiliary benefits, two concerns needed to be addressed before deciding how to use information from the two sources to define the final version of the code for public release. On the one hand, the evidence of additional aged wife and widow benefits that did not involve dual entitlement was consistent with the view that the second set of MBR records offered a more complete accounting of current pay benefits for certain types of auxiliary benefits than the former. On the other hand, the fact that the second set of records was accessed two years after the first raised the issue that SSA might have changed its view of benefit entitlement for the survey period of interest, particularly since the second version of MBR information indicated an increase for the type of benefit (disabled worker) thought to be most subject to the effects of retroactive payment.

Representation of social security receipt in the survey was reviewed to help resolve these concerns. As shown in table B-4, the level of social security receipt reported in the survey for the additional aged wife and aged widow benefits was very

high and indistinguishable from that of current pay benefits identified by means of the initial set of records. On the other hand, the additional disabled-worker benefits associated with the MBR as of 1997 tended to be unconfirmed by corresponding survey reports. Furthermore, the 9 percent of additional current pay disabled-worker benefits reports stemming from the later version of the MBR, though nominally “confirmed” by survey reports, could plausibly represent persons who were actually receiving SSI benefits during the five-month waiting period that precedes award of social security disabled-worker benefits.<sup>8</sup> In sum, the relationship between the survey and MBR representations of reciprocity shown in the table strongly suggests that the additional aged wife and aged widow benefits stem from benefit records that were missed the first time the MBR was accessed but located the second time the MBR was accessed in 1997, while the additional disabled-worker benefits identified by the second MBR extract more likely represent retroactive awards. Based on these results, the final type of benefit characterization was made on the basis of the 1995 MBR extract for all but aged wife and widow benefits not involving dual entitlement.

For a discussion of the nature of the resulting estimates see the body of this report.

## **DEVELOPMENT OF YEAR OF DEATH INFORMATION**

SSA benefit records may be used to establish two basic aspects of mortality: fact and date of death. Fact of death in this context is taken to mean the occurrence of death without reference to the timing of occurrence. Obviously timing is a function of calendar date. Fact of death in SSA benefit records may be established based on the reason for

---

<sup>8</sup> Essentially the same result was obtained for the three benefit types in the context of the 1991 panel (data not shown).

OASDI benefit termination as given by the Ledger Account File (LAF) field of the Master Beneficiary Record (MBR) or the reason for SSI benefit termination given in the Payment Status field (PSTAT) of the Supplemental Security Record (SSR). Beneficiary date of death is also available from each record system. As noted above, both record systems were accessed to obtain this information for SIPP sample members for whom usable SSN's were available. The variables used to define the fact and year of death are given in table B-5. In instances of dual entitlement, the relevant fields on both the primary and auxiliary benefit records were consulted.

*Source of year of death data.*—Nearly all deaths identified by the survey match were identified on the basis of the MBR. In fact for both panels at least 95 percent of deaths were observable only from the MBR; and less than 1 percent were found solely to the SSR. The balance, about 2-5 percent, were identifiable from either source, but in such instances information from the MBR rather than the SSR was employed (see table B-6).

The original set of MBR records was obtained in March and April 1995 and the second February and April of 1997. A year of death variable was developed from both sets of records. Deaths reported for the year in which the MBR was accessed were not considered for public release because they were seriously incomplete and judged not likely to be representative even of deaths occurring in the calendar months preceding the point at which the benefit records were accessed. Comparison of year of death based on the two MBR extracts for the years prior to 1994 showed a very modest tendency for more deaths to be identified for 1994 on the basis of the second extract. In addition, it contained arguably complete, or nearly complete, information for deaths occurring in 1995 and 1996. As a result, the public release year of death information is based

principally on the second MBR extract. Weighted and unweighted counts by year of death for the public release version of the year of death variable are given in table B-7. The number of deaths by year, age, and sex for persons age 65 and over at death as estimated on the basis of the survey match and as reported to the NCHS is given in tables B-8 and B-9. An interpretation based on a comparison of these two sets of estimates is provided in the body of this report.

*Consistency of information on fact and year of death.*—It was possible to establish a year of death for each instance when benefit termination was attributed to death.<sup>9</sup> Likewise, presence of a year of death information was always accompanied by an indication of benefit termination due to death of the beneficiary.

*Use of the January cross-sectional weight.*—On conceptual grounds the longitudinal full panel weight would seem to be the appropriate weight for an event such as death, that arises within the population as defined at the beginning of a given panel. In addition, it is intended that the full panel weight be defined for sample persons who leave the sample due to death.<sup>10</sup> However, it became apparent that a significant minority of those identified as dying by means of the survey match did not in fact have a positive full panel weight. As shown in table B-10, a full panel weight was not available for 17 and 11 percent, respectively, of sample members who died between 1990/91 and 1996 according to the survey match. Presumably some of these individuals left the sample due

---

<sup>9</sup> Two sources of inconsistency were identified and eliminated. On the second MBR extract obtained for both panels, an indication of benefit termination due to death was not accompanied by year of death for a significant fraction of dually entitled women. However, year of death was present on the corresponding primary benefit record. A second smaller inconsistency occurred principally for 1990 panel and arose in those few instances in which the LAF code indicated death of a Medicare covered individual not receiving a monthly OASDI benefit at time of death. However, for these cases year of death was present in the field reserved for a primary beneficiary's year of death (DODP-YY).

to death during the survey reference period but were not identified as having died, or otherwise exited both the sample and the survey population. Others may have left the sample, but not the survey population, for a variety of reasons and died subsequently. This problem with the full panel weight led to the decision to employ the cross-section core weight for January of the initial full calendar year for each panel as the basis of population estimates of year of death. It was present for more than 95 percent of identified deaths. The balance of identified deaths, those with neither a full panel nor January cross-sectional weight, correspond principally to the deaths of individuals who entered the sample in the second and subsequent waves. Plausibly, this group is a bit smaller for the 1990 panel as compared to the 1991 panel because, as noted earlier, SSN's were not developed for 1990 panel entrants subsequent to the fourth interview.

*Explanation of the initial interview effect on deaths observable the from the first calendar year of the 1990 and 1991 panels.*-In order to properly reconcile the NCHS and survey match estimates, it is important to understand how a particular aspect of SIPP panel design, characterized in the body of this report as the *initial interview effect*, prevents observation of a sizeable proportion of the deaths which occur in the noninstitutional population during the initial year of each panel.

The SIPP sample is divided into 4 equally sized subsamples or rotation groups. Each rotation group is interviewed sequentially over four successive months in a repeating pattern as many times as is called for by the design of a given panel (8 times in the case of the 1990 and 1991 panels). Thus the calendar period covered by the standard SIPP interview reference period of 4 months may be characterized as consisting of 16

---

<sup>10</sup> Nominally, a longitudinal weight should be present for all persons who left the survey population by moving to an ineligible address, such as an institution or outside the country. Clearly some of these

rotation/month time-segments (4 rotations each covering 4 calendar months). The way the staggered interviewing pattern by rotation interacts with calendar year means that for 10 of the 16 rotation/month time segments corresponding to the first 4 months of the year, persons living at a selected sample address during the reference period but not living at that address at interview will be excluded from the interviewed sample. As a result, approximately 10/16 or .625 of the deaths occurring in the first four months of the initial calendar year of panels with an interviewing pattern of the sort employed by the 1990 and 1991 panels will not be observable from the SIPP sample.

Table B-11 illustrates how this comes about. The top half of the table represents the interaction between rotation group and calendar month for the first four months of 1990 in the context of the 1990 panel. The bottom half provides the same information for the 1991 panel. The numbered boxes represent the sixteen rotation (by month) time-segments of the initial interview (appearing to the left of the thick stepped line); those to the right of the line represent the 16 rotation time-segments of the second interview. Boxes 1, 2, 3, 5, 6, and 9, on the far left, represent the six initial interview rotation time-segments pertaining to 1989; 4, 7, 8, and 10-16 correspond to the 10 rotation time-segments pertaining the first four calendar months of 1990 that are included in the initial interview. Exits from the survey population during these rotation time-segments will not be identified because such individuals are not included on the household roster as members of the household during the reference period. Consequently, deaths generating such exits will not be observable via match to SSA benefit records. These deaths represent 10/16 or .625 of the deaths during the first four months of the initial calendar year of the panel, and, assuming an approximately even distribution of deaths across the

---

individuals may subsequently have died also.



year, about 21 percent ( $.625 \times .333 = .208$ ) of all deaths occurring during the initial calendar year of the panel. The balance of the 16 rotation time-segments covering the first four months of the year lie to the right of the thick stepped line (the six boxes 1, 2, 3, 5, 6 and 9) and are included in the second interview reference period. Nominally, deaths occurring in these rotation time-segments may be observed by means of a match to SSA benefit records because such individuals had been listed as household members at the point of initial interview and, thus, prior to death. They would account for about  $6/16$  or  $.375$  of January-April deaths.

This initial interview effect operates in the same manner in the context of the 1991 panel as may be seen from the lower half of table B-11.

Table B-1.--Relationship between summary Type of Benefit Code (TOB) and the full Beneficiary Identification Code (BIC) from the Master Beneficiary Record

Type of benefit	TOB code value	BIC values <sup>1</sup>	Type of claim (TOC) values <sup>1</sup>
Retired worker	1	A	1,2,3,4
Disabled worker	2	A	5,6
Aged spouse	3	B,B1,B3,B4,B6,B8,B9, BA,BD,BG,BH,BJ,BN,BP-BR,BT	...
Spouse caring for minor children	4	B2,B5,B7,BK,BL,BW,BY	...
Aged widow(er)	5	D,D1-D9,DA,DC,DD,DG,DH, DJ-DN,DP-DT,DV-DZ	...
Widow(er) caring for minor children	6	E,E1-E9,EA-ED,EF-EH,EJ, EK,EM	...
Disabled widow(er)	7	W,W1-W9,WB,WC,WF,WG, WJ,WR,WT	...
Adult disabled in childhood	8	C1-C9,CA-CK <sup>2</sup>	3,4,7
Student child	9	C1-C9,CA-CK <sup>2</sup>	R,8,9
Minor child	10	C1-C9,CA-CK <sup>2</sup>	0,1,5
Other, specified <sup>3</sup>	11	F01-F08,J01-J04,K01-K9, KA-KH,KJ,KL,KM	...
Other, n.e.c.	12	Other nonblank values not specified elsewhere	...
Denied	13	Ledger Account File (LAF) code = N, ND	...
Medicare covered, but not an OASDI recipient	99	M1,T2-T9,TA-TH,TJ-TN,TP-TZ	...
On MBR but not in current pay during period	00	LAF is blank	...
No MBR record found for this sample person	Blank	Blank	...

(. . .) - Not applicable.

<sup>1</sup> Master Beneficiary Record (MBR) 7080 Version Spread Record. Social Security Administration, Office of Software Improvement and Engineering, Division of Data Administration, specification number 02-003, January 25, 1989.

<sup>2</sup> Note that the 360 version spread allows for 99 values.

<sup>3</sup> Special Age-72 beneficiaries and their spouses, dependent parent of deceased workers, and uninsured beneficiaries.

Table B-2.--Summary of type of benefit code assignment criteria

Code category	Public-use code value	TOB value	LAF value	MBR gender	January survey weight <sup>1</sup>
Not in sample.....	0	...	...	...	not > zero
Disabled worker <sup>2</sup> .....	1	1			
Retired worker.....	2	2	C,E	...	> zero
Aged wife, not dually entitled.....	3	3	C,E	F	> zero
Aged wife, dually entitled.....	4	3	C,E	F	> zero
Aged widow, not dually entitled <sup>3</sup> ..	5	5	C,AD	F	> zero
Aged widow, dually entitled <sup>3</sup> .....	6	5	C,E	F	> zero
All other benefits in current pay.....	7	4,6,7,8,9 10,11,12	C,E,AD Blank or	...	> zero
Other <sup>4</sup> .....	9	Blank or 00,12,99	(nonblank & not C,E, or AD)	...	> zero

... - Not applicable.

<sup>1</sup> Final weight (FNLWGT) for 1/91 or 1/92, wave 4 public use files, respectively 1990 and 1991 panels.

<sup>2</sup>Also, year of adjudication for disabled worker benefit must be prior to the second full calendar year of the respective panel (DSD<=1990;1991, conditioned on panel). Records failing this test were assigned to "other" i.e., as (9).

<sup>3</sup> Also requires presence of a current pay retired-worker benefit.

<sup>4</sup> No SSN available for this sample member, SSN available but no MBR record found, or MBR record found, with no indication of benefit in current pay for 12/90 or 12/91 respectively, 1990 and 1991 panels.

Table B-3. Selected types of OASDI benefits identified by the SIPP - SSA benefit record match, 1990 and 1991 panels: Sample counts and weighted estimates

Category description	Code value	1990 panel		1991 panel	
		Sample count	Weighted estimate <sup>1</sup>	Sample count	Weighted estimate <sup>1</sup>
Total.....	...	62,211	248,688	40,295	251,270
Not in sample in January of the second full calendar year of the panel <sup>2</sup>					
Subtotal.....	0	6,750	-	4,989	-
In sample in January of the second full calendar year of the panel <sup>2</sup>					
Subtotal.....	...	55,461	248,688	35,306	251,270
Retired worker benefit.....	1	3,889	18,375	2,477	18,412
Disabled worker benefit.....	2	561	2,624	353	2,669
Aged wife only benefit.....	3	541	2,642	340	2,464
Dually entitled aged wife benefit.....	4	419	2,017	274	2,047
Aged widow only benefit.....	5	924	4,172	534	3,862
Dually entitled aged widow benefit.....	6	486	2,216	303	2,169
Other benefits in current pay <sup>3</sup>	7	654	2,624	398	2,839
Other <sup>4</sup> .....	9	47,987	214,017	30,627	216,808

... - Not applicable.

<sup>1</sup>Number in thousands based on the cross-sectional public use file weight for January of 1991 (1990 panel) or January of 1992 (1991 panel).

<sup>2</sup>As inferred by presence or absence of a positive cross-sectional weight for the appropriate month (see also note 1 this table).

<sup>3</sup>Including principally dependent and survivor benefits received by minor children or adult children disabled in childhood, benefits received by spouses of retired, disabled or deceased workers with minor children in their care, benefits received as a disabled widow, and all benefits received by men as a husband or widower of a retired, disabled or deceased worker.

<sup>4</sup>MBR record located, but not in current pay as of 12/91 (1990 panel) or 12/92 (1991 panel), operationally valid SSN available, no MBR record located, or operationally valid SSN not available.

#### Sources

Data files: sv790mrg.sd2 of and sv691mrg.sd2 both of 1/6/2000

Run streams: tob\_dod\_comb1.sas & tob\_dod\_comb2.sas

Jobs: Jobs 8 and 10 of 3/8/00.

I:\TONY\[code\_cnts\_tappnx.xls]Sheet1

Range: TOB

03/30/2000

Table B-4.--Principal instances of discrepant current pay status, by timing of MBR extract and representation of social security receipt in the survey, 1990 panel

Type of benefit by MBR version	Number of sample cases	Percent receiving a social security benefit in the January 1991 according to the survey
Disabled worker		
As represented by 1995 extract	561	89.3
Increment as of 1997extract	56	8.9
Aged wife (not dually entitled)		
As represented by 1995 extract	437	97.3
Increment as of 1997extract	100	96.0
Aged widow (not dually entitled)		
As represented by 1995 extract	802	98.8
Increment as of 1997extract	113	98.2

Source 1990 SIPP panel exact matched to SSA benefit records.

Table B-5.--Variables used to establish the fact and year of beneficiary death

Project mnemonic	Source mnemonic	Source	Description & relevant code values
<u>Variables used to establish fact of death</u>			
LAF	LAF	MBR	Ledger Account Field, as of date of extraction t1 - Benefits terminated because of death of beneficiary receiving OASDI benefits x1 - Benefits terminated because of death of Medicare covered individual not receiving a monthly OASDI benefit
RPSTAT	PSTAT	SSR	Edited final payment status field (most recent payment spell) t01 - Benefits terminated because of death of beneficiary s01 - Benefits suspended because of report of beneficiary death
<u>Variables used to identify year of death</u>			
BDOD-YY	BDOD	MBR	Beneficiary year of death, date of extraction field
DODP-YY	DODP	MBR	Primary account holder year of death where BDOD-YY was not present and the LAF code at extraction was "x1"
YOD	DOD	SSR	Year of death as extracted from the SSR variable date of death (DOD)

I:\TONY\ [DTHVARS.XLS]Sheet1

09/07/1997

rev.: 3/30/00

Table B-6.--Source of year of death by panel  
 [sample counts]

Source of data	Panel			
	1990		1991	
	Sample count	Percent distribution	Sample count	Percent distribution
Total.....	2,435	100.0	1,316	100.0
MBR only.....	2,364	97.1	1,245	94.6
Both.....	54	2.2	62	4.7
SSR only.....	17	0.7	9	0.7

Data files: sv790mrg.sd2 of and sv691mrg.sd2 both of 1/6/2000

Run streams: tob\_dod\_comb1.sas & tob\_dod\_comb2.sas

Jobs: Jobs 1 and 2 of 3/15/00.

I:\TONY\[dthcomb2\_wgtst1.xls]Sheet1

Range: dthcomb2

03/15/2000 rev.: 3/30/00

Table B-7. Year of death as identified by the SIPP - SSA benefit record match, 1990 and 1991 panels: Weighted and unweighted counts

Category description	Code value	1990 panel		1991 panel	
		Sample count <sup>1</sup>	Weighted estimate <sup>2</sup>	Sample count <sup>1</sup>	Weighted estimate <sup>2</sup>
Total.....	...	62,211	246,007	40,295	248,885
No year of death identified or identified year of death is 1997.....	0	59,776	235,413	38,979	240,248
Year of death identified <sup>3</sup> .....	...	2,435	10,594	1,316	8,638
Before 1990.....	< 1990	<sup>4</sup> 29	135	<sup>4</sup> 28	97
1990.....	1990	202	899	6	33
1991.....	1991	311	1,380	131	929
1992.....	1992	363	1,515	210	1,451
1993.....	1993	369	1,584	237	1,533
1994.....	1994	375	1,644	246	1,602
1995.....	1995	358	1,618	229	1,479
1996.....	1996	428	1,820	229	1,514

... - Not applicable.

<sup>1</sup> Includes cases for which January cross-sectional weight is not available.

<sup>2</sup> Number in thousands based on the cross-sectional public use file weight for January of 1990 (1990 panel) or January of 1991 (1991 panel).

<sup>3</sup> Excluding deaths occurring in 1997.

<sup>4</sup> The earliest year of death identified for the 1990 panel is 1975 and for the 1991 panel, 1970.

**Sources**

**Data files:** sv790mrg.sd2 of and sv691mrg.sd2 both of 1/6/2000

**Run streams:** tob\_dod\_comb1.sas & tob\_dod\_comb2.sas

**Jobs:** Jobs 8 and 10 of 3/8/00.

I:\TONY\[code\_cnts\_tappnx.xls]Sheet1

**Range:** YOD

03/30/2000



**Table B-8. Number of deaths of people age 65 and over by age<sup>1</sup> and sex<sup>2</sup> in the United States, 1990-1996 as reported to the NCHS and as identified from the 1990 SIPP Panel based on a match to Social Security Administration benefit records**

[Number of deaths in thousands]																					
Age at death <sup>1</sup>	Both sexes by year of death							Males by year of death							Females by year of death						
	1990	1991	1992	1993	1994	1995	1996	1990	1991	1992	1993	1994	1995	1996	1990	1991	1992	1993	1994	1995	1996
Reported deaths, U.S. resident population <sup>3</sup>																					
65 and over.....	1,542.5	1,563.5	1,575.2	1,654.3	1,669.8	1,694.3	1,713.7	703.4	731.6	735.3	765.9	768.2	776.2	782.2	839.1	831.9	839.9	888.4	901.6	918.1	931.6
65-69.....	217.3	214.5	211.1	213.4	209.3	204.3	200.0	128.4	126.4	124.2	125.4	123.0	119.4	116.7	89.0	88.1	86.8	87.9	86.3	84.9	83.4
70-74.....	260.6	264.2	266.8	274.5	275.5	276.5	273.8	147.7	149.5	149.9	154.2	154.5	154.6	152.5	112.9	114.7	116.9	120.3	121.1	122.0	121.3
75-79.....	301.1	301.8	301.7	312.8	313.6	315.5	321.2	157.9	158.3	158.3	163.7	162.7	164.4	167.4	143.2	143.6	143.5	149.1	150.9	151.1	153.8
80 and over.....	763.4	783.1	795.6	853.7	871.4	897.9	918.6	269.4	297.5	302.9	322.7	328.1	337.9	345.6	494.0	485.6	492.7	531.0	543.3	560.1	573.0
80-84.....	300.3	305.7	308.1	325.3	332.1	336.7	342.1	137.5	140.7	141.6	149.9	152.3	155.0	157.9	162.8	165.0	166.5	175.4	179.8	181.6	184.2
85 and over.....	463.1	477.4	487.5	528.4	539.3	561.3	576.5	151.9	156.8	161.3	172.8	175.8	182.8	187.7	311.2	320.6	326.2	355.7	363.5	378.4	388.8
Number of deaths estimated from the survey based on SSA match <sup>4</sup>																					
65 and over.....	746.2	1,110.4	1,258.0	1,389.9	1,425.8	1,363.7	1,595.0	437.1	522.3	673.4	803.1	728.2	712.5	771.4	309.1	588.1	584.6	586.8	697.6	651.2	823.5
65-69.....	169.2	178.8	199.8	180.8	244.7	145.0	162.5	138.5	105.9	108.1	121.7	137.1	71.8	88.3	30.7	72.9	91.7	59.2	107.6	73.2	74.3
70-74.....	184.1	217.3	261.4	293.9	236.7	236.8	252.8	104.2	120.3	140.2	180.7	160.4	132.1	129.2	79.9	97.0	121.1	113.2	76.3	104.6	123.6
75-79.....	191.7	241.2	335.9	285.1	270.0	295.8	290.3	116.4	78.6	192.0	164.0	149.5	165.5	163.2	75.4	162.7	144.0	121.1	120.4	130.2	127.1
80 and over.....	201.2	473.1	460.9	630.1	674.5	686.1	889.4	78.0	217.6	233.1	336.8	281.2	343.1	390.7	123.1	255.5	227.8	293.2	393.3	343.0	498.7
Survey deaths identified based on match as a percent of reported deaths among U.S. residents																					
65 and over.....	48.4	71.0	79.9	84.0	85.4	80.5	93.1	62.1	71.4	91.6	104.9	94.8	91.8	98.6	36.8	70.7	69.6	66.1	77.4	70.9	88.4
65-69.....	77.8	83.4	94.7	84.7	116.9	71.0	81.3	107.9	83.8	87.0	97.0	111.5	60.1	75.7	34.5	82.7	105.6	67.3	124.7	86.2	89.1
70-74.....	70.7	82.2	97.9	107.1	85.9	85.6	92.3	70.5	80.5	93.5	117.2	103.8	85.5	84.7	70.8	84.6	103.6	94.1	63.0	85.8	101.8
75-79.....	63.7	79.9	111.3	91.2	86.1	93.7	90.4	73.7	49.6	121.3	100.2	91.9	100.7	97.5	52.6	113.3	100.3	81.2	79.8	86.2	82.6
80 and over.....	26.3	60.4	57.9	73.8	77.4	76.4	96.8	29.0	73.1	76.9	104.4	85.7	101.5	113.0	24.9	52.6	46.2	55.2	72.4	61.3	87.0

<sup>1</sup> Age in January of year of death as derived from survey age for January 1990 and year of death as determined from match.

<sup>2</sup> Sex defined on basis of Master Beneficiary record.

<sup>3</sup> Includes deaths in Puerto Rico, Guam and the U.S. Virgin Islands which are excluded from the survey population.

<sup>4</sup> Survey deaths weighted using January 1990 public use file cross-sectional weight with no adjustment for nonmatches.

Source: U.S. Center for Disease Control, Monthly Vital Statistics Report, 41:7(s):2,15; 42: 2(s):2,16; 43: 6(s):2,16; 44: 7(s):2,7; 43:13:4,14; 45:11:2,17;47:9:2,18, and the 1990 panel of the Survey of Income and Program Participation based on a match to Social Security Administration benefit records.

Table B-9.--Number of deaths of people age 65 or over by age<sup>1</sup> and sex<sup>2</sup> in the United States, 1991-1996 as reported to the NCHS and as identified from the 1991 SIPP Panel based on a match to Social Security Administration benefit records

[Number of deaths in thousands]

Age at death <sup>1</sup>	Both sexes by year of death						Males by year of death						Females by year of death					
	1991	1992	1993	1994	1995	1996	1991	1992	1993	1994	1995	1996	1991	1992	1993	1994	1995	1996
	Reported deaths, U.S. resident population <sup>3</sup>																	
65 and over.....	1,563.5	1,575.2	1,654.3	1,669.8	1,694.3	1,713.7	731.6	735.3	765.9	768.2	776.2	782.2	831.9	839.9	888.4	901.6	918.1	931.6
65-69.....	214.5	211.1	213.4	209.3	204.3	200.0	126.4	124.2	125.4	123.0	119.4	116.7	88.1	86.8	87.9	86.3	84.9	83.4
70-74.....	264.2	266.8	274.5	275.5	276.5	273.8	149.5	149.9	154.2	154.5	154.6	152.5	114.7	116.9	120.3	121.1	122.0	121.3
75-79.....	301.8	301.7	312.8	313.6	315.5	321.2	158.3	158.3	163.7	162.7	164.4	167.4	143.6	143.5	149.1	150.9	151.1	153.8
80 and over.....	783.1	795.6	853.7	871.4	897.9	918.6	297.5	302.9	322.7	328.1	337.9	345.6	485.6	492.7	531.0	543.3	560.1	573.0
80-84.....	305.7	308.1	325.3	332.1	336.7	342.1	140.7	141.6	149.9	152.3	155.0	157.9	165.0	166.5	175.4	179.8	181.6	184.2
85 and over.....	477.4	487.5	528.4	539.3	561.3	576.5	156.8	161.3	172.8	175.8	182.8	187.7	320.6	326.2	355.7	363.5	378.4	388.8
	Number of deaths estimated from the survey based on SSA match <sup>4</sup>																	
65 and over.....	705.2	1,226.9	1,293.6	1,363.5	1,299.7	1,301.6	416.5	631.1	591.2	721.3	693.5	593.6	288.6	595.8	702.4	642.2	606.1	708.0
65-69.....	127.9	221.9	169.6	187.6	154.8	123.2	85.3	108.4	108.9	102.3	110.1	68.9	42.7	113.5	60.7	85.3	44.7	54.3
70-74.....	110.8	206.0	224.2	297.7	205.7	225.3	73.7	139.4	145.7	162.6	123.5	97.1	37.0	66.6	78.5	135.1	82.2	128.2
75-79.....	240.2	249.0	284.5	232.3	286.6	238.1	168.5	98.1	165.6	106.1	190.9	132.1	71.7	150.9	118.9	126.2	95.6	106.0
80 and over.....	226.3	550.0	615.3	645.9	652.6	715.1	89.1	285.2	171.0	350.2	268.9	295.6	137.2	264.8	444.3	295.7	383.7	419.5
	Survey deaths identified based on match as a percent of reported deaths among U.S. residents																	
65 and over.....	45.1	77.9	78.2	81.7	76.7	76.0	56.9	85.8	77.2	93.9	89.3	75.9	34.7	70.9	79.1	71.2	66.0	76.0
65-69.....	59.7	105.1	79.5	89.6	75.8	61.6	67.5	87.3	86.8	83.2	92.2	59.0	48.4	130.7	69.0	98.8	52.6	65.1
70-74.....	41.9	77.2	81.7	108.1	74.4	82.3	49.3	93.0	94.5	105.3	79.9	63.7	32.3	57.0	65.3	111.6	67.4	105.6
75-79.....	79.6	82.5	91.0	74.1	90.8	74.1	106.4	62.0	101.2	65.2	116.2	78.9	50.0	105.2	79.7	83.6	63.3	68.9
80 and over.....	28.9	69.1	72.1	74.1	72.7	77.8	29.9	94.1	53.0	106.8	79.6	85.5	28.3	53.7	83.7	54.4	68.5	73.2

<sup>1</sup> Age in January of year of death as derived from survey age for January 1990 and year of death as determined from match.

<sup>2</sup> Sex defined on basis of Master Beneficiary record.

<sup>3</sup> Includes deaths in Puerto Rico, Guam and the U.S. Virgin Islands which are excluded from the survey population.

<sup>4</sup> Survey deaths weighted using January 1990 public use file cross-sectional weight with no adjustment for nonmatches.

Source: U.S. Center for Disease Control, Monthly Vital Statistics Report, 41:7(s):2,15; 42: 2(s):2,16; 43: 6(s):2,17; 44: 7(s):2,7; 43:13:4,14; 45:11:2,17;47:9:2,18, and the 1991 panel of the Survey of Income and Program Participation based on a match to Social Security Administration benefit records.

I:\TONY\dtthent91c.xls]Sheet1

Range: survdths

1/20/2000; rev. 12/07/2000;

Data file: d:\mbrdata\sv691mrg of 1/6/00;

disk -- (ser.no. 2CA2-A55E).

Run stream: tob\_dod\_comb1.xls of 2/11/00 (Job 2 @ 10:16 hrs).

Table B-10.--Presence of positive survey weights with for cases with a year of death identified based on the survey match, by panel

[sample counts]

Presence of positive survey weights	Panel			
	1990		1991	
	Sample count <sup>1</sup>	Percent distribution	Sample count	Percent distribution
Total.....	2,488	100.0	1,316	100.0
Core <sup>2</sup> .....	2,427	97.5	1,238	94.1
Both.....	2,007	80.7	1,094	83.1
Core only.....	420	16.9	144	10.9
Neither.....	61	2.5	78	5.9

<sup>1</sup> Includes deaths identified as occurring in 1997.

<sup>2</sup> For January of the initial full calendar year of each panel.

Data files: sv790mrg.sd2 of and sv691mrg.sd2 both of 1/6/2000

Run streams: tob\_dod\_comb1.sas & tob\_dod\_comb2.sas

Jobs: Jobs 1 and 2 of 3/15/00.

I:\TONY\[dthcomb2\_wgtst1.xls]Sheet1

Range: wgtst1 03/16/2000 rev.: 3/30/00

## Appendix C. Sampling Errors and Inference

In principle, the information on type of OASDI benefit and year of death involved in this public release may be linked to any of the several public use files available for the 1990 and 1991 SIPP panels. All resulting estimates will be subject to the same sorts of sampling and nonsampling errors as other estimates based on data from these two panels in addition to errors arising from nonmatches and mismatches and those errors from the administrative record system *per se*.

The issue of nonsampling error as it may affect estimates from the SIPP has been discussed elsewhere (Bureau of the Census 1991b) and will not be dealt with here. While information available on the SIPP public use files will support direct calculation of sampling errors, not all users will be in a position to do so. On the other hand, generalized estimates of sampling error have been developed specifically for the SSA beneficiary population (Bye and Gallicchio 1988, 1993; Jabine 1990) using the half-sample replication method. Users of the information provided in this public release may wish to consider them when working with estimates involving type of OASDI benefit and year of death.

The discussion of sampling variability and standard errors given below generally follows the usual presentation, in summarized form, provided in Census Bureau technical documentation available for SIPP public use files and SIPP-based technical reports. However, estimates of sampling error presented are those estimated by Bye and Gallicchio for SSA beneficiaries. The reader may wish to consult the Census Bureau

sources (Bureau of the Census 1991b, Jabine 1990), as well as the material provided by Bye and Gallicchio for additional information.

*Use and calculation of standard errors.*-- Standard errors, as described later in this section, provide a measure of sampling error, that is variations that arise by chance because a sample rather than the entire population was surveyed. While they also reflect nonsampling errors to some degree, they do not account for systematic biases in the survey estimates such as might arise from underreporting of income receipt or, in the present context, from nonmatches.

The sample estimate and its standard error may be used to construct confidence intervals and conduct hypothesis testing. The confidence interval represents a range around the estimate that would include the average result of all possible samples with a known probability. If an estimate and its standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.645 standard error below the estimate to 1.645 standard error above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from 1.96 standard errors below the estimate to 1.96 standard errors above the estimate would include the average result of all possible samples.

Though not directly observable from a given sample, the average estimate derived from all possible samples is or is not contained in any particular computed interval. On the other hand, for a particular sample, one can say that the average estimate derived from all possible samples lies in a specified confidence interval with its associated level of confidence, e.g., 68, 90 or 95 percent in the examples given above.

In hypothesis testing, standard errors are employed to make inferences about population estimates using sample estimates. The most common hypothesis tested is whether observed differences between sample estimates reflect actual differences in the population. Testing may be performed at various levels of significance, where the significance level indicates the probability of concluding that population characteristics are different when, in fact, they are identical.

The test for a difference between two estimates may be performed by dividing the observed difference by its standard error as described below. The larger the resulting ratio, the smaller the chance of concluding the estimates differ when they actually do not. If the difference falls in the range  $\pm 1.645$  times its standard error, a conclusion of a population difference between the characteristics is not justified at the .10 level. However if the difference is greater than  $\pm 1.645$  its standard error there is only a ten percent chance that given the observed difference in sample estimates, the population estimates are the same. Conventionally such an observed difference is taken to reflect the existence of an actual difference in the population. Obviously, however, for a given test assessed at the .10 significance level there is a ten percent chance that the underlying population characteristics are in fact the same and the more tests performed, the greater the likelihood of inferring the presence of a population difference where none exists. For

example, 100 independent tests of observed differences between sample estimates in which there are no underlying population differences, would likely give rise to about 10 erroneous inferences of a difference.

*Standard error parameters and tables and their use.*--Bye and Gallicchio developed two parameters (denoted “a” and ”b”) that may be used to approximate the standard error behavior for a broad range of SSA beneficiary subgroups<sup>1</sup> in the 1990 panel context. Values for the “a” and “b” parameters values corresponding to both panels are given in table C-1. Since “a” and “b” parameter values for SSA beneficiaries based on direct estimation are not available for the 1991 panel, the values given in table C-1 for the 1991 panel reflect the values estimated from the 1990 panel that have been adjusted to reflect the smaller 1991 sample.<sup>2</sup>

The Bureau of the Census typically provides generalized standard errors such as those given in tables C-2 through C-5 for users who wish further simplification. (Tables C-2 and C-3 provide generalized standard errors for estimated numbers and percentages, respectively, for the 1990 panel. Corresponding information for the 1991 panel is provided in tables C-4 and C-5.) The standard errors in these tables are based on the generalized variance parameters for the respective panels given in table C-1.

---

<sup>1</sup> Bye and Gallicchio computed sampling errors for 148 beneficiary subgroups defined on the basis of age, sex, and marital status. They fit a curve to the estimated variances using the 126 cells with unweighted counts of 25 or more. They do not recommend use of the resulting generalized standard errors for child beneficiaries because analysis of the 1984 panel data showed that standard errors for this group were strongly related to family size. However, since taken as whole, the standard error child beneficiaries was similar to other groups of the same size (Bye and Gallicchio 1988, table I) and since we are interested in the effect of sampling error on the estimated for the total number of child beneficiaries, we felt it was appropriate to employ the generalized variance parameters estimated for beneficiaries age 18 and older to the total number child beneficiaries identified in the survey.

<sup>2</sup> As compared to the 1990 panel, the number of interviewed households and the number of sample members reporting social security benefits, was about 1/3 less in the 1991 panel.

*Standard errors of estimated numbers and percentages.*—There are two ways to obtain an approximate standard error,  $s_x$  of an estimated number of social security beneficiaries based on this public release in conjunction with other data available for these panels in public release form. The estimate may be obtained from tables C-2, C-3, C-4 or C-5 directly or by interpolation. Alternatively  $s_x$  may be approximated using the formula

$$S_x = \sqrt{ax^2 + bx} \quad (1)$$

Formula (1) will provide the approximate standard error ( $S_x$ ) of a population total or subtotal where  $x$  is the estimated size of the sub-population in thousands and  $a$  and  $b$  are the generalized variance parameters pertaining to the respective panels.

For example, from table A (p. 19 in the body of this report) we see that the estimate for the number of persons with disabled-worker benefits in current payment status as of December 1990, based on the 1990 panel match, is 2,624 thousand. Using formula 1 the approximate standard error of 2,624 thousand is:

$$S_x = \sqrt{(0.00047)(2,624)^2 + (5.9315)(2,624)} = 137.1$$

Thus the approximate 90-percent confidence interval for an estimate of 2,624 thousand (1.645 standard errors) is from about 2,398 thousand to about 2,850 thousand, while the approximate 95-percent confidence interval (1.96 standard errors) is from about 2,355 thousand to 2,893 thousand. One could conclude then that the average estimate derived from all possible samples lies in the ranges computed for roughly 90 and 95 percent of all samples, respectively.



The approximate standard error of a percentage may be derived on the basis of formula

(2):

$$S_{(x,p)} = \sqrt{\frac{b}{x}(p)(100-p)} \quad (2)$$

where x is the population total in thousands forming the base of the percentage, p is the percentage ( $0 \leq p \leq 100$ ), and b is the value of the appropriate “b” parameter from table C-1. For example, we observe, also in table A that about 35 percent of those in current payment status for disabled-worker benefits are women ( $1037/2624 \times 100 \approx 39.5$ ). From equation (2) the standard error for the percentage of disabled-worker beneficiaries who are women would be

$$S_{(x,p)} = \sqrt{\frac{5.9315}{2,624}(39.5)(100-39.5)} \approx 2.97$$

Thus the approximate 90-percent confidence interval for this percentage (1.645 standard errors) is from 34.6 percent to 44.4 percent, while the approximate 95-percent confidence interval (2 standard errors) would be from 33.7 to 45.3 percent.

The formula for deriving the standard error of the difference between two estimates x and y is:

$$S_{(x-y)} = \sqrt{S_x^2 + S_y^2 - 2rS_xS_y} \quad (3)$$

where  $S_x$  and  $S_y$  are the standard errors of the estimates x and y and r is the correlation coefficient between the characteristics estimated by x and y. The estimates may be

numbers, averages, percentages, ratios, etc. Underestimates or overestimates of the standard error of a difference result if the estimated correlation coefficient is less than or greater than zero. Generally estimates of  $r$  for the characteristics of SSA beneficiaries are not available.<sup>3</sup> However, in the present context, we will assume  $r$  is zero.

For example, from previously published data (Social Security Administration 1992, table 5.A12, p. 185, not shown here) we see that 3.6 percent of the 1,439 thousand women below age 70 and receiving aged wife benefits as of mid 1990 also received money income from earnings while 18.6 percent of the estimated 1,626 thousand women below age 70 and receiving benefits as aged widows were also receiving money income from earnings.<sup>4</sup> From formula (2) the standard errors of these percentages are approximately 1.21 and 2.59 percent, respectively. Assuming that the two estimates are not correlated, the standard error of the difference of 15 percentage points is:

$$S_{(x-y)} = \sqrt{1.21^2 + 2.59^2} \approx 2.86$$

To determine whether the two percentages differ significantly at the 10-percent significance level, multiply the standard error of the difference by 1.645 and compare the result (about 4.7 percentage points) to the estimated difference of 15 percentage points. Since the difference is clearly larger than 1.645 times the standard error of the difference, one may conclude that the estimates of 3.6 and 18.6 differ at indicated significance level. To be considered statistically significant at the 5-percent significance level, the standard

---

<sup>3</sup> However the covariances for mean and median benefit amounts for husbands and wives are available (see again Bye and Gallicchio 1993, tables 5 and 6).

<sup>4</sup> In each instance, the estimates pertain to benefits that do not involve dual entitlement.

error of the difference would have to be less than 1.96 times the estimated difference of 15 percentage points, which it is by a considerable margin, *i.e.*,  $(2.86 \times 1.96 \approx 5.6)$ .

*Small sample sizes.*—In the context of the SIPP panels of the early to mid 1990ies, the Census Bureau cautions that summary measures for population subgroups with a population base of less than 200 thousand are likely to reveal little useful information because of the large standard errors involved. This is suggested by the large coefficients of variation<sup>5</sup> for estimates below 200 thousand shown tables C-2 and C-4. Furthermore, in the context of the 1990 panel, an estimate of 200 thousand would typically be based on about 40 sample cases. Clearly for estimates based on fewer cases, nonsampling error affecting only a few of them could produce a large error in the estimate for the group. Given the smaller sample size of the 1991 panel, 40 sample cases would typically be associated with a population estimate of about 300,000. On the other hand, were estimates to be based on the combined samples from both panels, the typical population estimate based on 40 sample cases would be on the order of 100,000-150,000.

*Assessing comparisons between survey estimates and independent estimates.*--The foregoing discussion relates primarily to estimating the effect of sampling error on the precision of survey estimates of numbers and proportions of SSA beneficiaries with a given characteristic or on comparisons between such survey-based estimates. However, the evaluation of the information to be included in the public release relies heavily on comparisons between survey estimates and independent estimates based on sources that are either not affected by sampling error because they may be treated as 100 percent count data (NCHS estimates of the number of deaths by age year and sex and SSA

---

<sup>5</sup> The coefficient of variation is the ratio of the standard error to its respective estimate. Obviously, the larger the coefficient of variation, the less certain one can be of the underlying value of the estimate.

estimates of the number of women receiving benefits as dually entitled aged widows or dually entitled aged wives), or as is the case for other the independent estimates for other benefit categories, are subject to only minimum sampling error because of the large samples involved.<sup>6</sup>

Three approaches were taken to evaluating the effect of sampling error on comparisons between the survey and independent estimates.

- 1) For survey estimates of the number of beneficiaries by type, except for those involving dual entitlement, formula (3) was employed since both the survey and independent estimates were based on samples and the standard errors were available or could be easily derived. Standard errors for the independent estimates were taken from look-up tables provided in the Annual Statistical Supplement to the Social Security Bulletin (SSA, 1991, table 10.A.1), using interpolation as required.
- 2) For numbers of deaths by year, age and sex, standard errors for a given survey estimate were derived based on the generalized “a” and “b” parameters given in table C-1. The standard errors were then used to generate the corresponding 90 percent confidence intervals to assess the null hypothesis of no difference between the survey and independent estimate. If the value of the relevant independent estimate fell outside the confidence interval, the null hypothesis was rejected. As noted in the body of the report, when the appropriate independent estimate fell outside the confidence interval, it

---

<sup>6</sup> With the exception of benefits involving dual entitlement for women, which are based on universe counts, estimates of the number of beneficiaries by type of benefit employed in developing the independent estimates use to evaluate the completeness of the type of benefit match are based on ten percent samples. The approximate c.v. for the smallest beneficiary subgroup considered, disabled worker women, is only

invariably exceeded the upper bound, implying a short-fall in survey-match estimate.

This approach was also employed in comparing the independent and survey match estimates of the number of dually entitled wives and widows because the independent estimates for these benefit categories were based on universe counts, rather than ten-percent sample data.

- 3) While this approach is sufficient to assess whether or not an estimate based on the survey match is likely to differ from its corresponding independent estimate, it does little to inform the issue of relative completeness, that is whether the estimates for one sub-group are likely more or less complete than those of another.

In order to assess whether two or more survey estimates differed with respect to their corresponding independent estimates, comparisons were made between survey estimates expressed as a percentage of their respective independent estimates, i.e.,  $(x/a) \times 100$ , where  $x$  and  $a$  are respectively the survey and corresponding independent estimate. The standard error of this percentage is derived as

$$S_{(p)} = \sqrt{\text{Var}_x / a^2} \times 100 \quad (4)$$

and  $\text{Var}_x$  is derived by squaring the appropriate value given by formula (1). Having such percentages and their corresponding standard errors for two different decedent or beneficiary subpopulations, say female decedents age

65-69 (survey match estimate represents  $84.0 \pm 9.25$  percent of the NCHS figure) and 80 and over (survey match estimate represents  $58.0 \pm 3.31$  percent of the NCHS figure), the difference ( $S_D$ ) between the two with respect to their corresponding independent estimate is simply 84-58 or 26 percentage points. Assessment of the null hypothesis of no difference between the two estimates with respect to their corresponding independent estimate was then made by dividing the observed difference by the standard error of the difference derived on the basis of the formula (3), assuming  $r$  is zero. In the case of this example  $S_D \approx 9.82$  and  $D/S_D \approx 2.65$ . Unless otherwise noted, the hypothesis of no difference was evaluated at the .10 level, that is the difference between the two estimates had to be greater than 1.645 times its standard error.

Table C-1.--Generalized variance parameters for  
SSA beneficiaries, 1990 and 1991 SIPP panels

Parameter	Panel	
	1990	1991
"a".....	0.00047	0.00073
"b".....	5.9315	9.1591

C:\My Documents\DOCS\EXCEL\SIPPTOB\  
[SE\_90\_91 panels.xls]Sheet1  
04/04/2000

Table C-2.-- Standard errors for estimated population totals, SSA recipients, 1990 SIPP panel<sup>1</sup>

Estimate in 1000's	Standard error	Coefficient of variation
75 .....	21.2	0.282
100 .....	24.5	0.245
250 .....	38.9	0.156
500 .....	55.5	0.111
750 .....	68.7	0.092
1,000 .....	80.0	0.080
2,500 .....	133.3	0.053
5,000 .....	203.5	0.041
7,500 .....	266.3	0.036
10,000 .....	326.1	0.033
15,000 .....	441.3	0.029
20,000 .....	553.7	0.028
25,000 .....	664.9	0.027
35,000 .....	885.1	0.025

<sup>1</sup>After Bye and Gallicchio (SSB 1993, vol. 53, no. 3) .



Table C-3.--Standard errors for estimated percentages of SSA recipients, 1990 SIPP panel <sup>1</sup>

Base of percents in 1000's	Percent											
	1 or 99	2 or 98	5 or 95	8 or 92	10 or 90	15 or 85	20 or 80	25 or 75	30 or 70	35 or 65	40 or 60	50
75 .....	2.80	3.94	6.13	7.63	8.44	10.04	11.25	12.18	12.89	13.41	13.78	14.06
100 .....	2.42	3.41	5.31	6.61	7.31	8.70	9.74	10.55	11.16	11.62	11.93	12.18
250 .....	1.53	2.16	3.36	4.18	4.62	5.50	6.16	6.67	7.06	7.35	7.55	7.70
500 .....	1.08	1.52	2.37	2.95	3.27	3.89	4.36	4.72	4.99	5.20	5.34	5.45
750 .....	0.88	1.25	1.94	2.41	2.67	3.18	3.56	3.85	4.08	4.24	4.36	4.45
1,000 .....	0.77	1.08	1.68	2.09	2.31	2.75	3.08	3.33	3.53	3.67	3.77	3.85
2,500 .....	0.48	0.68	1.06	1.32	1.46	1.74	1.95	2.11	2.23	2.32	2.39	2.44
5,000 .....	0.34	0.48	0.75	0.93	1.03	1.23	1.38	1.49	1.58	1.64	1.69	1.72
7,500 .....	0.28	0.39	0.61	0.76	0.84	1.00	1.12	1.22	1.29	1.34	1.38	1.41
10,000 .....	0.24	0.34	0.53	0.66	0.73	0.87	0.97	1.05	1.12	1.16	1.19	1.22
15,000 .....	0.20	0.28	0.43	0.54	0.60	0.71	0.80	0.86	0.91	0.95	0.97	0.99
20,000 .....	0.17	0.24	0.38	0.47	0.52	0.61	0.69	0.75	0.79	0.82	0.84	0.86
25,000 .....	0.15	0.22	0.34	0.42	0.46	0.55	0.62	0.67	0.71	0.73	0.75	0.77
35,000 .....	0.13	0.18	0.28	0.35	0.39	0.46	0.52	0.56	0.60	0.62	0.64	0.65

d Gallicchio (SSB 1993, vol. 53, no. 3) .

.91 panels.xls]Sheet1

rev.: 3/23/00

Table C-4.-- Standard errors for estimated population totals,  
SSA recipients, 1991 SIPP panel<sup>1</sup>

Estimate in 1000's	Standard error	Coefficient of variation
75 .....	21.2	0.28
100 .....	24.5	0.25
250 .....	39.1	0.16
500 .....	56.1	0.11
750 .....	69.7	0.09
1,000 .....	81.6	0.08
2,500 .....	139.3	0.06
5,000 .....	218.9	0.04
7,500 .....	292.6	0.04
10,000 .....	363.9	0.04
15,000 .....	503.5	0.03
20,000 .....	641.2	0.03
25,000 .....	778.0	0.03
35,000 .....	1,050.4	0.03

<sup>1</sup>After Bye and Gallicchio (SSB 1993, vol. 53, no. 3) as adjusted to account for the smaller 1991 panel size.

I:\TONY\[SE\_90\_91 panels.xls]Sheet1

03/13/1998

03/23/2000

Table C-5.--Standard errors for estimated percentages of SSA recipients, 1991 SIPP panel <sup>1</sup>

Base of percents in 1000's	Percent											
	1 or 99	2 or 98	5 or 95	8 or 92	10 or 90	15 or 85	20 or 80	25 or 75	30 or 70	35 or 65	40 or 60	50
75 .....	3.48	4.89	7.62	9.48	10.48	12.48	13.98	15.13	16.01	16.67	17.12	17.47
100 .....	3.01	4.24	6.60	8.21	9.08	10.81	12.11	13.10	13.87	14.43	14.83	15.13
250 .....	1.90	2.68	4.17	5.19	5.74	6.83	7.66	8.29	8.77	9.13	9.38	9.57
500 .....	1.35	1.89	2.95	3.67	4.06	4.83	5.41	5.86	6.20	6.46	6.63	6.77
750 .....	1.10	1.55	2.41	3.00	3.32	3.95	4.42	4.79	5.06	5.27	5.41	5.53
1,000 .....	0.95	1.34	2.09	2.60	2.87	3.42	3.83	4.14	4.39	4.56	4.69	4.79
2,500 .....	0.60	0.85	1.32	1.64	1.82	2.16	2.42	2.62	2.77	2.89	2.97	3.03
5,000 .....	0.43	0.60	0.93	1.16	1.28	1.53	1.71	1.85	1.96	2.04	2.10	2.14
7,500 .....	0.35	0.49	0.76	0.95	1.05	1.25	1.40	1.51	1.60	1.67	1.71	1.75
10,000 .....	0.30	0.42	0.66	0.82	0.91	1.08	1.21	1.31	1.39	1.44	1.48	1.51
15,000 .....	0.25	0.35	0.54	0.67	0.74	0.88	0.99	1.07	1.13	1.18	1.21	1.24
20,000 .....	0.21	0.30	0.47	0.58	0.64	0.76	0.86	0.93	0.98	1.02	1.05	1.07
25,000 .....	0.19	0.27	0.42	0.52	0.57	0.68	0.77	0.83	0.88	0.91	0.94	0.96
35,000 .....	0.16	0.23	0.35	0.44	0.49	0.58	0.65	0.70	0.74	0.77	0.79	0.81

<sup>1</sup>After Bye and Gallicchio (SSB 1993, vol. 53, no. 3) as adjusted to account for the smaller 1991 panel size.

I:\TONY\SE\_90\_91 panels.xls]Sheet1

03/13/1998 rev.: 3/23/00