MORTALITY COVERAGE IN SOCIAL SECURITY'S EARNINGS AND BENEFIT RECORD SYSTEMS

Linda DelBene and Faye Aziz, Social Security Administration

Social Security, because of its responsibility in administering the national old-age, survivors, disability, and health insurance programs, maintains an extensive and unique collection of data on individuals in this country. From the initial application for a social security number, a record is created with basic identifying and demographic data. Throughout an individual's lifetime, earnings and employment information is accumulated for all work experience acquired in covered employment. Finally, for each person who becomes eligible for benefits in one of Social Security's programs, a further record is kept of the type and status of these payments.

Clearly, these earnings and benefit record systems provide a tempting source of statistical data for researchers both within and outside Social Security. One particular area in which these data have been and will continue to be used is the study of mortality. For this reason, we have become increasingly aware of the need to assess the quality and usefulness of death statistics from Social Security's major administrative record systems. The purpose of this paper is to describe the work that has been done in this regard, and to share the preliminary results of our efforts.

The paper will begin with a brief discussion of the study sample which was selected for our project, and some background on each of the data sources which we sought to link together for the sample cases. The next step will be to describe in some detail the actual methods used to retrieve social security earnings and benefit data for the decedents in our sample, and to provide an assessment of the quality of this data linkage. Finally we will look at the mortality rates on both of these administrative files, and see to what extent deaths are reported--or underreported--to Social Security.

THE STUDY SAMPLE

In 1976, we approached the National Center for Health Statistics and the National Cancer Institute to secure their support and participation in the project. With their help, a study sample was selected which consisted of approximately 23,000 decedents for 1975, from each state except New Mexico, plus Washington, D.C. Copies of the death certificates for each of the sampled cases were then purchased from the states' vital records offices. Upon receipt of the certificates, specified data items were extracted from each; these included the decedent's name, sex, race, dates of birth and death, social security number (if it was present), occupation, industry, place of birth and death, and the death certificate number. Because there were certain items that would be too costly, in terms of both money and manpower, for us to extract and code directly from the certificate, we also sought each state's permission to use the decedent files which are maintained by NCHS. By matching our death certificate file to the NCHS file, we were able to obtain cause of death data for the sampled individuals, coded according to the Eighth Revision International

Classification of Diseases, Adapted for Use in the United States.

SSA ADMINISTRATIVE DATA

The next step was to attempt to match, to each decedent in the sample, data from several of Social Security's administrative record systems. Specifically, by using the social security number obtained from the death certificate as the primary identifier, a search was made of the longitudinal earnings and benefit record files maintained by Social Security in its central offices in Baltimore. We also requested certain demographic items from the SS-5, Application for a Social Security Number. More will be said later about the methods used and the completeness of this matching effort, but it may be useful at this point to provide some background on the administrative files being studied. 1/

The Summary Earnings Record (SER)--Social Security maintains a summary earnings record for each person with "covered" earnings who is assigned a social security number. Most types of employment in the United States are covered by Social Security. In fact, in 1975, 90.0 percent of those workers in paid employment were considered covered under the Old-Age, Survivors, Disability and Health Insurance (OASDHI) program. These earnings records are kept on file so that eligibility for benefits can later be established.

For the purpose of our project, the summary earnings record provided us with a variety of demographic and earnings information for our decedents. And, of particular interest to us for this study, there are certain indicators of death which appear on the SER. First, there may be an actual date of death posted on the record. Second, there are several death indications included as part of the type-of-claim data, The third item which may be present is a code designating the source of death information on Social Security's records.

The Master Beneficiary Record (MBR)--Monthly cash benefits are paid by Social Security to workers who gain insured status on the basis of their own earnings records, and to eligible dependents and survivors of such workers. Social Security's Master Beneficiary Record file contains a record for each person who has been entitled to receive retirement, survivors or disability insurance benefits since 1962.

The beneficiary records which we received for our sample cases have quite a variety of data on them. In addition to basic demographic data, there is detailed information on beneficiary status and payments. We also were able to identify two different indications of death. The most obvious is a date of death. Secondly, there is a code indicating the termination of a ledger account due to death.

PRELIMINARY ANALYSIS

The Pilot Study 2/--Because of the nature of the project, we sought a way to test our plans on a relatively small scale. Therefore, before we attempted the various data extractions and link-

ages for the entire sample of 23,000 decedents, a 10-percent "pilot" subsample was drawn from the 46 jurisdictions for whom we had received death certificates by November 1978. This subsample of 2,052 cases represented 9.0 percent of the complete decedent sample. 3/

For these pilot cases an attempt was made to link the death certificate and NCHS data to the corresponding social security summary earnings records. (At a later time we also retrieved benefit record data for the pilot cases. Our preliminary analysis, however, centered on the earnings files.) The key to the record linkage was the social security number (SSN). In 87.8 percent of the pilot cases there was a potentially usable SSN present on the death certificate. On 11.3 percent of the certificates there was no SSN at all, and on the remaining .9 percent of the certificates the SSN's provided were not valid account numbers. The potentially usable SSN's were sent to Social Security in Baltimore for machine searching of the earnings files.

In order to judge the quality of the matches which were then created between death certificate data and SSA earnings records, several variables from both sources were compared: the first six letters of the surname, sex, race and month and year of birth. As a result of these comparisons, we determined that 79.9 percent of the pilot subsample were probably good matches, 3.4 percent were questionable, and 4.4 percent were probably bad matches. In order to locate valid account numbers for nonmatches, questionable matches and bad matches, a manual search of Social Security's National Employee Index file would later be carried out.

Some Pilot Results--The preliminary analysis of death reporting on SSA earnings records was limited to the cases which we felt were "probably good" matches (a total of 1,640 decedents). A majority of the good matches (83.4 percent) actually had dates of death on their earnings record. Another 7.1 percent had some indication of death, but without a date of death available. In fact, only 9.6 percent of the good matched cases had no death reported at all on their social security earnings record.

The Manual Search--We were very pleased with the high coverage rates which we found for the original portion of the pilot subsample of decedents. However, because these rates reflected death reporting only for the "good" matched cases, we anticipated that they might be somewhat overstated. More specifically, we expected that if we were able to include coverage rates for the 20 percent of the subsample which were not originally considered good matches, our coverage rates would fall closer to other independent estimates. 4/

In order to do this, it was necessary to carry out a manual search of SSA's microfilm National Employee Index files to attempt to locate accurate social security numbers for 414 decedents. This number includes those decedents who had no SSN recorded on their death certificates, or for whom the SSN was not a usable number. It also includes the decedents who had potentially usable SSN's on their death certificates, but were later determined to produce questionable or bad matches to the earnings file.

Of the 414 records sent for searching, 252 new social security numbers were tentatively assigned. After retrieving the earnings records for these SSN's, we linked them to their respective death certificate records and sought to assess the quality of these new matches. The quality of the total matching effort had been somewhat improved by conducting the manual search. Ninety-six percent of the subsampled decedents were now matched to SSA earnings records, as opposed to 87.8 percent previously. More significant, of course, is that our "good" matches now made up almost 88 percent of the subsample.

Revised Coverage Rates--When we had taken into account the outcome of the manual searching, we were ready to see what effect, if any, this additional information would have on our earlier coverage rates. Figure 1 compares coverage rates before and after this search. After manual searching we had a total of 1,803 probably good matches in the pilot subsample. Of these decedents, 81.3 percent now had dates of death on their SSA earnings record and 6.9 percent more had some indication of death (without a date). The degree of undercoverage (those records with no death report at all) had risen from 9.5 percent to 11.8 percent.

Figure 1.--Mortality Coverage Rates on the Pilot Subsample for "Probably Good" SER Matches Before and After Manual Searching

"Probably Good" Matches	Before Manual Search		After Manual Search	
	Number	Percent	Number	Percent
Total.............................	1,640	100.0	1,803	100.0
Death reported on SSA earnings record:				
No date of death on SSA record...	116	7.1	125	6.9
Date of death present on SSA record:				
Disagrees with NCHS date*......	21	1.3	28	1.6
Agrees with NCHS date*.........	1,346	82.1	1,438	79.8
Death not reported on SSA earnings record.............................	157	9.6	212	11.8

*The NCHS year of death was always 1975. If the SSA record showed a year of death from 1974-1976, we placed it in the category "Agrees with NCHS date."

THE FULL SAMPLE

After conducting the preliminary work outlined above, the next step was to complete the data linkage for the balance of the decedent sample. Since the November 1978 cut-off date for inclusion of decedents in the pilot subsample, we had received death records from the remaining 5 states. We therefore were ready to extract and edit data from all death certificates not in the 10% subsample. 5/

Matching Death Records to Administrative Records--Data were extracted from a total of 20,829 new death certificates. We then requested earnings and benefit record searches for each valid, potentially usable social security number we had available on these certificates. We found that there were substantially more matches to the summary earnings record file (17,323) than to the master beneficiary record file (13,596). There are two explanations for the lower MBR match rate. First of all, an SER is established for everyone with a social security number; an MBR is established only for those entitled to receive social security benefits. 6/ Therefore, we could not expect all decedents to have an MBR. Secondly, we had some difficulty in matching the benefit data, even after it was located, back to the death certificate extract.

Again it was necessary to judge the quality of the death certificate-SSA record matches by determining the level of agreement between the linked sources on certain key variables (surname, sex, race, month and year of birth). This inspection yielded 73.5 percent of all decedents with a "probably good" match to the SER, and 60.6 percent to the MBR. Approximately 17 percent of the sample had no SER match at all, while almost 35 percent were not matched to the MBR.

Creation of Matched File for 100% Sample of Decedents--Several steps in our data linkage plans remained. When we attempted to link the balance of the death certificate-SER-MBR records to the basic NCHS decedent files, we found 682 nonmatches. These are apparently cases in which the states' vital records offices supplied us with incorrect certificates for requested decedents. For the present analysis, therefore, these 682 records have only NCHS data.

Finally, we combined the 10% pilot subsample described earlier with the 90 percent just mentioned. This resulted in a "complete" linked dataset of 22,881 death certificate-NCHS-SER-MBR records. The analysis which follows is derived from this complete file, with percentages based on weighted numbers.

Demographic Characteristics of the 100% Sample--Before looking at coverage rates on our decedent file, it may be useful to describe the demographic make-up of the entire sample according to the quality of the death certificate-SSA record match.

Sex and Race: White decedents consistently have a higher proportion of good matches to both the SER and MBR than blacks, and males higher than females. Furthermore, about 20 percent of the females and 27 percent of blacks in the sample have no match at all to the SER. Nonmatch rates to the MBR are even higher.

Age: As figure 2 shows, the proportion of good SER matches by age is relatively stable across the entire spectrum. The highest match rate appears to be for the 50-59 year age group. The SER nonmatch rates are also fairly level; the only noticeable deviation occurs for those decedents in the sample less than 45 years old, whose nonmatch rate rises to an average 20.4 percent.

On the other hand, MBR match rates show more of an effect from age differences. The proportion of good MBR matches is fairly consistent for those 65 years and older (about 72 percent), but falls quite drastically to an average 47.5 percent for those less than 65 years. The MBR nonmatch rate is also very high for this group.

Death Coverage Rates on SSA Files--In order to find out the level of mortality reporting on Social Security's record systems, we have chosen to limit our analysis to the 17,118 decedents with good death certificate-SER matches. Figure 3 breaks this group into categories according to the presence or absence of some type of death indication on the earnings record. Eighty-six percent of the good matches actually have a date of death on their SER, and another 5.2 percent have a death indicator without a date of death. In fact, by looking only at the Social Security earnings files we have already identified 91.1 percent of the sample deaths.

Before going any further, we will examine mortality coverage on the earnings records by several demographic breakdowns in order to identify those deaths which are most--and least-- likely to be reported to Social Security.

Sex: Approximately 88 percent of the men under consideration had a date of death on their earnings record, and another 6.3 percent had some indication of death without a date. Undercoverage of death reporting then is quite low for males--approximately 5.7 percent. Females are less well covered: 13.9 percent have no indication of death whatsoever. Only 82.7 percent of the females actually had a date of death on their SER.

Race: There is slightly less difference in undercoverage rates by race. While 8.4 percent of the white decedents had no death report, blacks and other races averaged 13.5 percent undercoverage. As anticipated, a greater percentage of whites actually had dates of death on their earnings record.

Age: Clearly, death reporting is best for those 65 years and older. This could be due to several factors. Dependents of a decedent over the age of 65 (retirement age) are likely to report a death in order to receive survivors' benefits. It is also probable that participation in Medicare is another reason for the high reporting of deaths after age 65.

Obviously, the high degree of death coverage on the earnings file is very encouraging. It is important to remember, however, that if we were able to locate valid social security numbers for the portion of the sample not originally considered "good" matches, we could expect our coverage rates to fall slightly. At the present time, however, we do not have the resources available to conduct a manual search of the National Employee Index for the approximately 3,000 decedents involved.

Figure 2.--Percent of 100% sample with "good" matches to SSA earnings
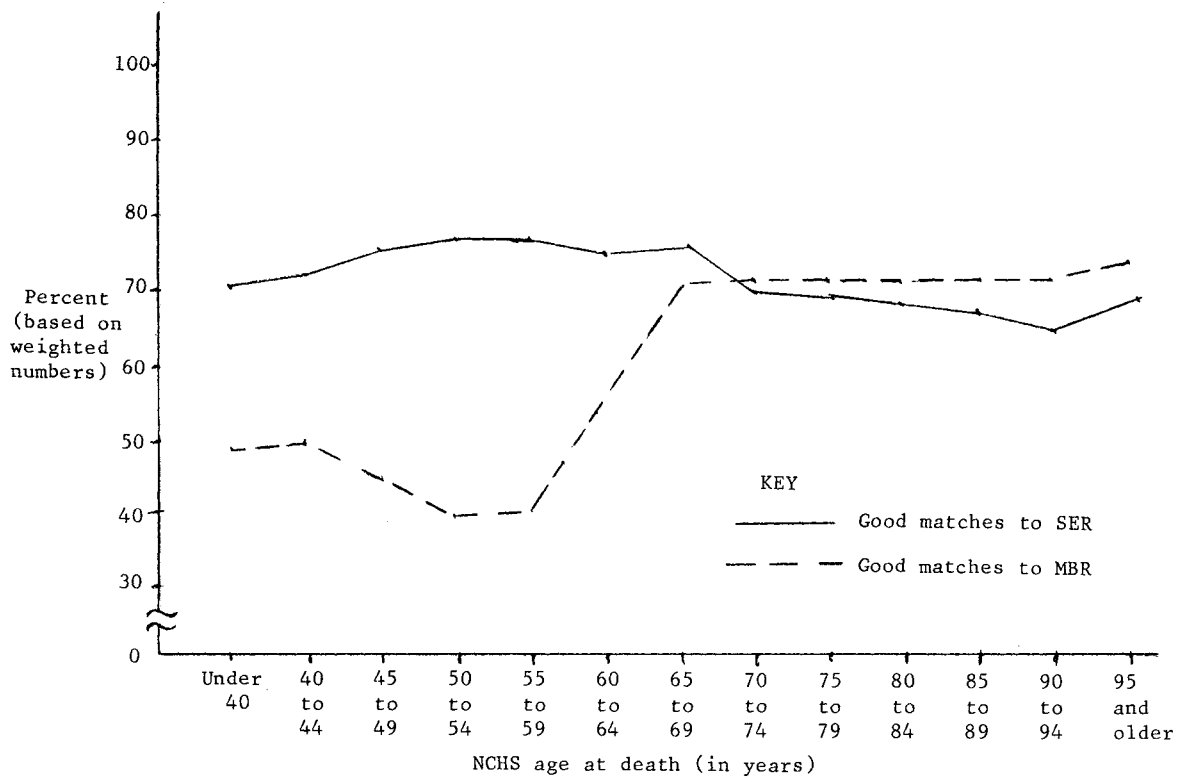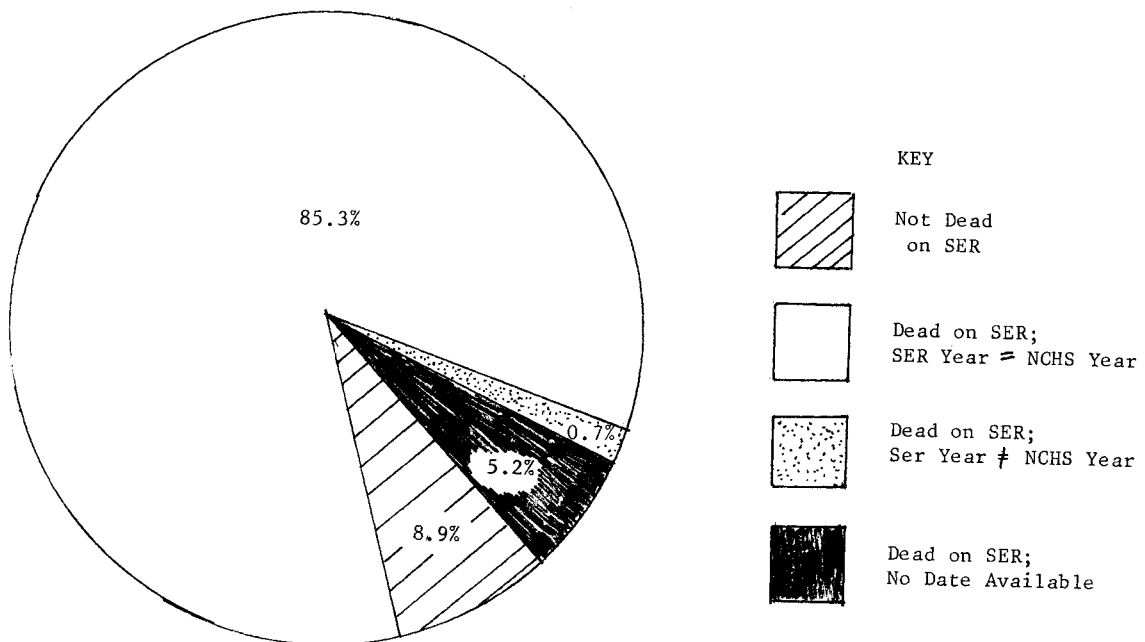and benefit data, separately, by age

Figure 3.--Mortality coverage rates on the SER for "probably good"
SER matches from the 100% sample
(Percents based on weighted numbers)

In the meantime, we were curious to see if we could identify any further deaths in our sample by examining Social Security's master beneficiary record file. Specifically, we selected the 1,643 good SER matches which had no indication of death on the earnings file. Almost 74 percent of these decedents did not have a good match to the MBR files, so we looked no further at them. As figure 4 illustrates, however, 173 of the remaining decedents did have a date of death on their MBR, and 100 more had some other indication of death. Our results must be considered tentative at this time. Efforts to improve the death certificate-MBR match rates may in turn yield higher mortality coverage rates on this administrative file.

Figure 4.--MBR Coverage Rates for Cases in the 100% Sample with No Death Indication on the SER

Description	Count	Weighted Number	Percent
Not dead on Ser:			
Total...................	1,643	111,061	100.0
Bad MBR match.........	1,321	81,691	73.6
Good MBR match:			
Not dead on MBR......	49	4,726	4.3
Dead on MBR, no date.	100	14,315	12.9
Date of death on MBR.	173	10,329	9.3

CONCLUSION

The work done on this mortality project is obviously still in its preliminary stages. Our current plans involve the creation of a matched dataset which can be made available to the research community for further analysis, probably late in 1981. In the meantime, we will continue to participate in the interagency effort currently underway to use administrative record systems in the area of mortality research.

ACKNOWLEDGEMENTS

The authors would like to thank their coworkers in the Division of Economic Research for their contributions to this project. In particular, we are grateful to Henry Ezell who provided valuable data processing assistance, and to Helen Kearney and Joan Reynolds for typing the various drafts of the paper.

NOTES AND REFERENCES

1/ For more information on the programs administered by Social Security, including the Federal retirement, survivors, disability, health insurance, black lung and supplemental security income programs, see: Social Security Handbook, 1978, U.S. Department of Health, Education, and Welfare. For statistics on program participation, see: Social Security Bulletin, Annual Statistical Supplement, 1976, U.S. Department of Health, Education, and Welfare.

2/ For background on this project, the sample design, and the pilot effort in particular, see: Some Background on the 1975 NCI-NCHS-SSA Mortality Study, LASS Working Notes No. 1, Office of Research and Statistics, Social Security Administration, February 22, 1979. See also: The 1975 NCI-NCHS-SSA Mortality Study: Initial Stages and Some Preliminary Results, LASS Working Notes No. 7, Social Security Administration, October 23, 1979.

3/ The original subsample actually included 3 more decedents, for a total of 2,055; however, we have omitted them from the current analysis because they have no NCHS data. They were apparently cases in which the states' vital records offices substituted another certificate for a requested decedent.

4/ Tabulations from the CWHS Mortality Study indicate that in 1970, mortality coverage on the SER was only 67.6%; by 1977 it had risen to 82.5%. For more information on this work, see: Aziz, F. and Buckler, W., "Mortality and the Continuous Work History Sample," a paper presented at the 1980 annual meetings of the American Statistical Association.

5/ It must be noted that although we drew the 10 percent subsample from the final 5 jurisdictions and coded data from these death certificates, time constraints have kept us from including this data in our "100%" sample yet. These decedents currently have only NCHS data on our matched file.

6/ Since 1975, MBR records are also established for persons who apply for retirement, survivors or disability insurance benefits, but are not entitled to receive them.