Mortality Reporting in SSA Linked Data: Preliminary Results

by Wendy Alvey and Faye Aziz*

The earnings and benefit data of the Social Security Administration represent a fairly large and balanced sample of present and past wage earners nationwide. Since the incentives for reporting deaths to the Social Security Administration mean that deaths are reported as a matter of course, this data base serves as an interesting prospect for examining problems of differential mortality. Variables available include age, race, sex, industry, and place of employment. To check on the coverage and content differences between social security and death certificate data, a sample study is being undertaken that links the two sources for 1975 decedents. This article provides a preliminary examination of the differences between the presumably complete frame of death certificates and the social security record data. Related efforts now in progress to improve available information for use in further mortality research are also discussed briefly.

As researchers attempt to examine possible trends in industrial and occupational mortality, it has become increasingly apparent that improved data sources are needed. It is partly in response to this need that the Social Security Administration has joined with the National Center for Health Statistics (NCHS) and the National Cancer Institute in an effort to restudy the usefulness of its death-reporting system in addressing such questions.

Some preliminary analyses of the pilot work undertaken as part of that effort are presented here, together with the background of the overall study. Briefly described are the pilot effort itself, the sampling design, and the matching conditions imposed. Some of the coverage differences between the linked components are examined, and expectations for further developments are outlined.

Background and Purpose of Study

To study industrial mortality rates, one can employ the Continuous Work History Sample (CWHS) of the Social Security Administration, which records on a longitudinal basis the industry, age, race, and sex of 1 percent of the covered U. S. labor force.(1) Mortality information for workers can also be obtaind from the CWHS, when it is available in the summary earnings record maintained for administrative purposes. The question raised here, in a preliminary way, is "How good is the mortality reporting in the earnings file?" The suitability of the CWHS for studying industrial mortality will depend on the answer to that question, among other things.

Some background on the earnings files is necessary before the study is described in detail. The Social Security Administration has a summary earnings record on file for each individual to whom a social security number (SSN) has been issued since 1936. Virtually all adult residents of the United States are represented in the file. When the fact of death is reported to the Social Security Administration, it is posted to this file and other records. Furthermore, the existence of survivor benefits and lump-sum death payments (often available to help defray funeral costs) are strong incentives to report deaths on a routine basis.[1] The study is interested in determining, therefore, the extent to which these incentives are effective and the groups for whom deaths are underreported.

To get a handle on the amount of undercoverage in the administrative system of the Social Security Administration the NCHS was approached about obtaining a sample of

[1] See Wendy Alvey, "SSA Procedures for Reporting Death," in **Collaborative NCHS-SSA Efforts To Study the Feasibility of Creating Linked Statistical Samples for Use in Mortality Research**, Linked Administrative Statistical Sample (LASS) Working Notes, No. 1, pages 144–158, 1979.

decedents that could be matched to social security earnings and benefit records on a one-to-one basis.[2] *(2)* With the help of NCHS, every State and the District of Columbia was reached to inquire about obtaining copies of death certificates. All agreed to cooperate (except New Mexico, which does not maintain its documents in sort by certificate number).

The Social Security Administration then set about designing the sample, with the help of the National Cancer Institute. About 23,000 decedents aged 35 or older were to be selected. All death certificates chosen were for 1975 deaths. Stratification was based on age, race, sex, State, and cause of death. Over-sampling was carried out for blacks and younger decedents in an effort to compensate for expected coverage problems. Individuals who died in smaller States were over-sampled to permit examination of differential mortality rates geographically. Men were over-proportionately selected to meet the needs of related research projects.*(3)*

Study Implementation and Pilot Subsample

The design was agreed upon, and the National Cancer Institute drew the sample. In the summer of 1978, every jurisdiction was reached and sent a list of sampled death certificate numbers with a few additional variables to confirm the identity of each decedent. By late fall, death certificates had been received from nearly all the States and coding and keying of the sampled documents had begun.

Because of the massive amount of work, it was decided to "minimize disasters" by selecting a 10-percent pilot subsample of all jurisdictions covered in the study that were already in-house. As the tabulation below shows, of the

| Records sampled	Number	Percent
Total	22,884	100.0
Received by cut-off	19,411	84.8
Included in subsample	2,055	9.0

22,884 records sampled, 84.8 percent had been received by the cut-off date for this pilot study.. (Documents from the five remaining jurisdictions were not completely received until spring of this year.) The 10-percent subsample of those certificates eligible for inclusion represents 9.0 percent of the overall sample. It is this group on which the remainder of the article will focus.

Quality of SSN Information

Once the death certificate information had been processed, steps were initiated to match the decedents' records to their own social security earnings data. Since each State

[2] For other research based on these methods, refer to reports in the series, **Studies From Interagency Data Linkages,** Office of Research and Statistics, Social Security Administration.

Chart 1.—Derivation of study universe

PILOT SUBSAMPLE
(2,055)

No SSN provided (11.2%)

Invalid SSN (1.0%)

Potentially usable SSN (87.8%)
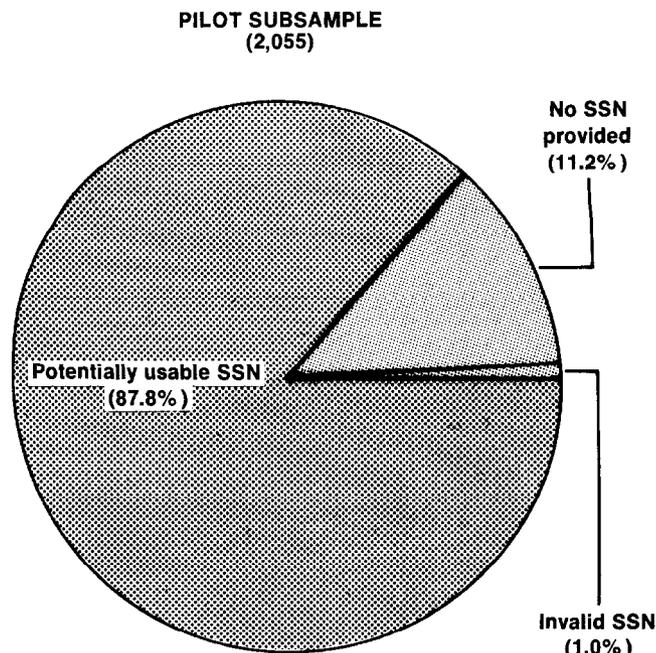
POTENTIALLY USABLE CASES AS
PERCENTAGE OF PILOT SUBSAMPLE

80.0%

3.4%

4.4%

"Probably good" matches   Questionable matches   "Probably bad" matches
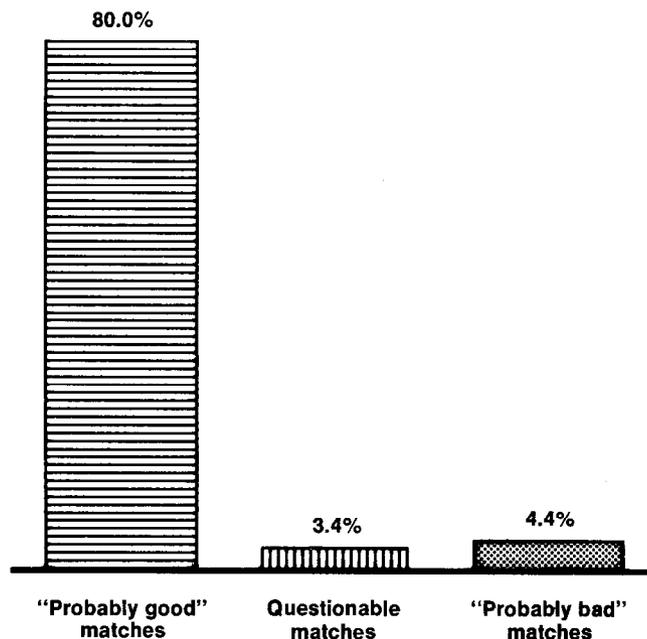
asks for the SSN of the decedent, that number was selected as the prime matching key.

To access the social security earnings records, the SSN's of the pilot cases were sent to the Office of Systems in Baltimore for machine searching. The top part of chart 1

illustrates the results of that effort for the pilot subsample. As is evident, about 11 percent of the sampled death certificates had the social security questions left blank. In 1 percent of the cases, the responses provided were not valid numbers. All these cases required further (manual) searching. The shaded portion of the chart (1,804 of the 2,055 subsample death certificates) represents potentially usable SSN's. For this 88 percent, the next step was to match the information from the earnings data with the death certificates.

Of course, just because the SSN was valid and an earnings record was retrieved did not necessarily mean that the social security record found was that of the sample individual. To confirm that a proper match was made, several other variables were compared: The first six letters of the surname, sex, race, and month and year of birth. On comparing the agreement of the social security data for these variables with the corresponding NCHS death certificate information, the cases with usable SSN's could be put in three groups (shown on the lower part of chart 1). About 80 percent were considered "probably good" matches, for which no further linkage efforts are required. About 3.4 percent were called "questionable" matches, and the remaining 4.4 percent of the subsample were "probably bad" matches. For these latter two categories, manual searching may be necessary to obtain a more reliable data linkage.[3](4)

Some Study Results

Some findings on coverage for these "probably good" matches (that 80 percent of the pilot subsample) may be described now. As shown by the tabulation below, which

"Probably good" matches	Number	Percent
Total	1,643	100.0
Death reported on earnings record:		
No date of death on SSA record	117	4.4
Month of death on SSA and NCHS record:		
Disagree.......................	76	3.4
Agree	1,293	83.4
Death not reported on earnings record	157	8.8

presents the coverage rates for that group, only 8.8 percent of the good matched cases had deaths not reported on their social security earnings record; 4.4 percent had a death reported but with no date of death available. The balance of the cases (86.8 percent) had dates of death on their social security files: For 3.4 percentof these matches the month of death did not agree with that on the corresponding death certificate; the month of death was the same for 83.4 percent.

---

[3] Information on the matching rules employed and basic tables on the subsequent demographic makeup of the pilot subsample are available from the authors on request.

Chart 2.—Percentage distribution of good matches by age attained in 1975 (from summary earnings record)
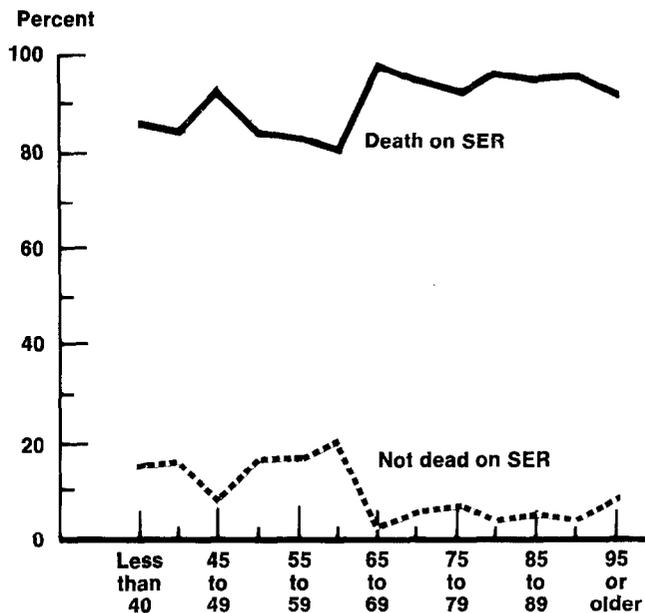

Chart 2 examines the coverage rates for good matches by age on the earnings record. It reveals that, as expected, death reporting was virtually complete for those aged 65 or older. No doubt, eligibility for Medicare is a contributing factor. In fact, with respect to these decedents the chief interest here was to examine content differences, rather than coverage problems, since coverage was expected to be high. Similarly, coverage by region showed that identified deaths were distributed fairly evenly nationwide. A good bit more variation was apparent, however, when these data were looked at State by State. At least some of this variation undoubtedly reflects sampling fluctuations.

Finally, death reporting by social security earnings record, race, and sex was examined (chart 3). In this chart each group of bars shows the data for (1) those not covered in the summary earnings record, (2) those for whom no date of death is available on the earnings record, (3) those whose month of death on the earnings record is the same as that shown by NCHS, and (4) the group for whom the SSA and the NCHS dates do not agree.

As the chart reveals, white men have considerably less undercoverage (6.8 percent) than white women (14.9 percent). Agreement on month of death is also greater for these men. It is interesting to note, furthermore, that, although the black women behave essentially like the white women, the amount of agreement for reported deaths is substantially lower for black men than it is for white men (67 percent, compared with 82 percent). This difference reflects the fact that, for women, the extent of reporting is affected by a failure to report their death, perhaps because no death benefit may be payable. Deaths for black men, on the other hand, are being reported, but the information provided is more likely to be either incomplete or inconsistent with the information supplied on the death certificate.

Chart 3.—Percentage distribution of good matches by presence of death on summary earnings record and race and sex from summary earnings record

Percent

☐ No death on SER
▓ Death on SER/No date
⊟ Date of death same SER/NCHS
■ Date of death not same SER/NCHS

WHITE   BLACK
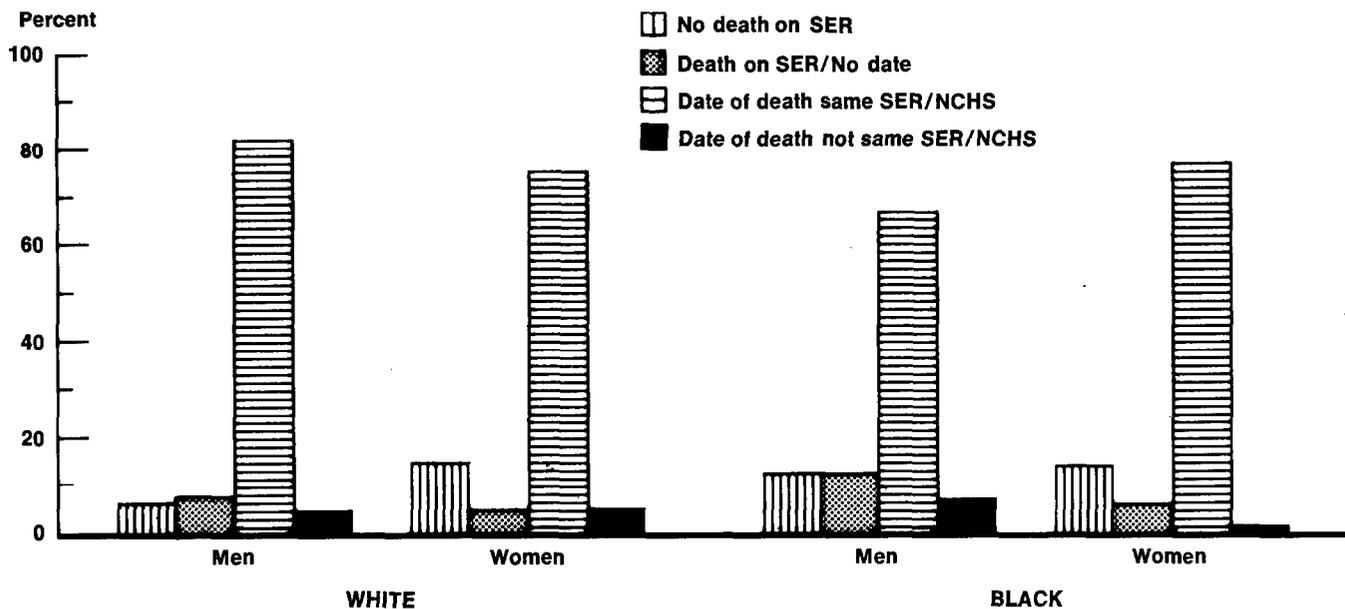
Limitations

The authors are pleased with the high coverage rate for this portion of the sample. It is important to emphasize, however, that what are being looked at here are the unweighted counts of the "probably good" matches to only one administrative system of the Social Security Administration—the summary earnings records files. Other data suggest that these figures overstate the extent of death reporting on the earnings records for 1975 decedents. When coverage rates are examined for the 20 percent of the sample not designated as good matches, the pilot results will undoubtedly come closer into line with other currently available estimates. Then, with the pilot sample it will be possible to reliably use such variables as place of death and cause of death—items not now in the administrative record files.

Future Plans

The material presented here is just the beginning of a major analytical effort. These results are not only preliminary in nature but incomplete as well.

First, data for the five jurisdictions omitted here must be added to the sample. Then, weighting for the entire sample is necessary. In fact, this weighting has already been started. Beyond that, the study's short-term goals are to finish the analysis of the complete pilot subsample, including examination of the match and coverage rates for the death certificates requiring manual searches. Some analysis of the content differences for the matched subsample is contemplated as well. For the longer run, work is continuing in the area of mortality research with colleagues in the National Cancer Institute and NCHS. *(5)*

Addenda

(1) The following are examples of industrial mortality research using social security data from the CWHS:

John Goldsmith and David Hirschberg, "Mortality and Industrial Employment (I)," **Journal of Occupational Medicine,** March 1976, pp. 161–164.
Pierre Decoufle and John Goldsmith, "Letters to the Editor," **Journal of Occupational Medicine,** September 1977, pp. 582–586.
John R. Goldsmith, "Mortality and Industrial Employment (II)," **Journal of Occupational Medicine,** April 1977, pp. 249–254.

Among epidemiological studies using social security data is—

Thomas Mancuso, "Relation of Duration of Employment and Prior Respiratory Illness to Respiratory Cancer Among Beryllium Workers," **Environmental Research,** July 1970, pp. 251–275.

(2) The Social Security Administration has done extensive work in the area of data linkage using exact matching techniques. In addition to the work cited in footnote 2, see Beth Kilss and Frederick J. Scheuren, "The 1973 CPS-IRS-SSA Exact Match Study," **Social Security Bulletin,** October 1978.

(3) Interest in mortality data grew out of a proposal by Frederick Scheuren at the NBER Workshop on Estate Multiplier Estimates of Wealth (May 2, 1975). He suggested that by using Social Security Administration mortality data, the standard estate multiplier procedure could be considerably upgraded. The following papers describe his proposals:

Frederick Scheuren, "Historical Perspectives on IRS Wealth Estimates with Prospectives for Improvement," in **Background on Collaborative NCHS-IRS-SSA Efforts To Study the Feasibility of Creating Linked Statistical Samples for Use in Mortality Research** (LASS Working Notes, No. 1, 1979, pp. 77-90).

Frederick Scheuren, "1962 Federal Estate Tax Wealth Estimates for Top Wealthholders," in **Background on Collaborative NCHS-IRS-SSA Efforts. . . . .** (LASS Working Notes, No. 1, 1979, pp. 91-114).

Subsequently, a validation study of the estate multiplier procedure led to some of the sampling constraints imposed on the current study. For more information on that work, see

H. Lock Oh and Frederick Scheuren, "Some Preliminary Results From a Validation Study of the Estate Multiplier Procedure," in **1978 American Statistical Association Proceedings, Social Statistics Section,** part II, pp. 650-654.

(4) For general information about the comparability of linked administrative and survey or census data, see

Theta Hambright, "Comparability of Age on the Death Certificate and Matching Census Record, United States, May-August 1960," **Vital and Health Statistics Data Evaluation and Methods Research** (Series 2, No. 29), June 1968.

Cynthia Cobleigh and Wendy Alvey, "Validating the Social Security Number," in **Studies From Interagency Data Linkages** (Report No. 4) Office of Research and Statistics, Social Security Administration, July 1975, pp. 89-124.

(5) For documentation of the planning stages for this research, see

Further Developments in Collaborative NCHS-IRS-SSA Efforts To Study the Feasibility of Creating Linked Statistical Samples for Use in Mortality Research (LASS Working Notes, No. 2), 1979.

Further Development in Collaborative NCHS-IRS- SSA Efforts To Study the Feasibility of Creating Linked Statistical Samples for Use in Mortality Research (LASS Working Notes, No. 3), 1979.