

QUALITY OF MORTALITY REPORTING IN SSA LINKED DATA:
SOME PRELIMINARY DESCRIPTIVE RESULTS

Wendy Alvey and Faye Aziz, Social Security Administration

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 partially in response to this need that the Social Security Administration has joined with the National Center for Health Statistics and the National Cancer Institute in an effort to restudy the usefulness of its death reporting system for addressing such questions.

In the time allotted today, we will present some preliminary analyses of the pilot work undertaken as part of that effort. We will begin by telling you something about the overall study. This will be followed by a brief description of the pilot effort itself: what design was used to obtain the sample and what matching conditions were imposed. Next, some of the coverage differences encountered between the linked components will be examined; finally, our expectations for further developments will also be mentioned.

Background and Purpose.--In order to study industrial mortality rates, one can employ SSA's Continuous Work History Sample (CWHHS), which records, on a longitudinal basis, the industry, age, race, and sex of 1% of the covered U.S. labor force [1]. Mortality information for workers can also be obtained from the CWHHS when it is available in the summary earnings records Social Security maintains for administrative purposes. The question this paper attempts to answer in a preliminary way is, "How good is the mortality reporting on our earnings files?" The suitability of the CWHHS for studying industrial mortality will, of course, depend on this, among other things.

Perhaps some background on the earnings files, themselves, might be in order before we describe the study in detail. The Social Security Administration has a summary earnings record on file for each individual to whom a social security number has been issued since 1936. Virtually all adult residents of the United States are represented in this file. It is to this and other SSA record systems that the fact of death is posted when it is reported to Social Security. Furthermore, the existence of survivor's benefits and lump sum death payments (often available to defray funeral costs) are strong incentives to report deaths to SSA on a routine basis [2]. What we are interested in determining, therefore, is to what extent are these incentives effective and for what groups are deaths underreported?

In order to "get a handle" on the amount of undercoverage existing in our own administrative system, we approached the National Center for Health Statistics (NCHS) about obtaining a sample of decedents which could be matched to social security earnings and benefit records on a one-to-one basis [3]. With NCHS' help, each State and the District of Columbia were contacted to inquire

about obtaining copies of death certificates. All, except New Mexico, which does not maintain its documents in sort by certificate number, agreed to cooperate.

Social Security then set about to design the sample, with the help of the National Cancer Institute. About 23,000 decedents, age 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. Oversampling was carried out for blacks and younger decedents, in an effort to compensate for expected coverage problems. Individuals who died in smaller States were oversampled so as to allow us to examine differential mortality rates geographically; and men were over proportionately selected to meet the needs of related research projects [4].

Study Implementation and Pilot Subsample.--Once the design was agreed upon and the National Cancer Institute had drawn the sample, Social Security recontacted each jurisdiction, sending them a list of sampled death certificate numbers and a few additional variables to confirm the identity of each decedent. This was done last summer. By late fall, death certificates had already been received from nearly all of the States and coding and keying of the sampled documents had begun.

Because of the massive amount of work before us, it was decided to "minimize our disasters" by selecting a 10% pilot subsample of all in-scope jurisdictions which were already in-house. As shown below, of the 22,884 records sampled, about 85 percent had been received by the cut-off date for this pilot study. (Documents from the remaining five jurisdictions were not completely received until spring of this year.) The 10% subsample of those certificates eligible for inclusion represents about 9 percent of the overall sample or 2,055 decedents. It is this group to which we will confine ourselves for most of the rest of this presentation.

	Sample Totals	
	Number	Percent
Total sample.....	22,884	100.0
Received by cut-off.....	19,411	84.8
Included in subsample.....	2,055	9.0

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 asks for the social security number of the decedent, that number was selected as the prime matching key.

In order to access the social security earnings records, the social security numbers of the pilot cases were sent to our Bureau of Data Processing

in Baltimore for machine searching. The left side of figure 1 illustrates the results of that effort for the pilot subsample. As you can see, about 11 percent of the sampled death certificates had the social security question left blank. In another 1.0 percent of the cases, the responses provided were not valid numbers. All of these cases will require further--manual--searching. The remaining (shaded) portion, or 1,804 of the 2,055 death certificates in the subsample, had a potentially usable social security number. For this 88 percent, our next step was to match together the information from our earnings data with the death certificates.

Of course, just because an account number was valid and an earnings record was retrieved, does not necessarily mean that the social security record we found was that of the sampled individual. In order to confirm that a proper match had been made, several other variables were compared: the first six letters of the surname, sex, race, and month and year of birth. Upon comparing the agreement of the SSA data for these variables to their corresponding NCHS death certificate information, we were able to separate the cases with usable account numbers into three groups. (These are depicted in the histogram on the right side of figure 1.) About 80 percent were considered "probably good" matches, for which no further linkage efforts are required. Around 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 [5].

Some Results.--Now let us go on to describe some findings with regard to coverage for these probably good matches which comprise that 80 percent of the pilot subsample. The table below presents the coverage rates for that group. As it shows, only 8.8 percent of the good matched cases had deaths which were not reported on their social security earnings record. Another 4.4 percent had a death reported, but no date of death was available. The balance of the cases (87 percent) had dates of death on their SSA file: for 3.4 percent of these good matches the month of death did not agree with that on the corresponding death certificate and for 83.4 percent, month of death did agree.

	Sample Number	Totals Percent
"Probably good" matches.....	1,643	100.0
Death reported on earnings record:		
No SSA date of death.....	117	4.4
SSA date of death available:		
SSA and NCHS dates of death disagree.....	76	3.4
SSA and NCHS dates of death agree.....	1,293	83.4
Death not reported on earnings record.....	157	8.8

Examination of the coverage rates for good matches by SSA age is shown in figure 2. It reveals that, as expected, death reporting was virtually complete for those 65 or older. No doubt, eligibility for Medicare is a major contributing factor. In fact,

FIGURE 1.—DERIVATION OF STUDY UNIVERSE

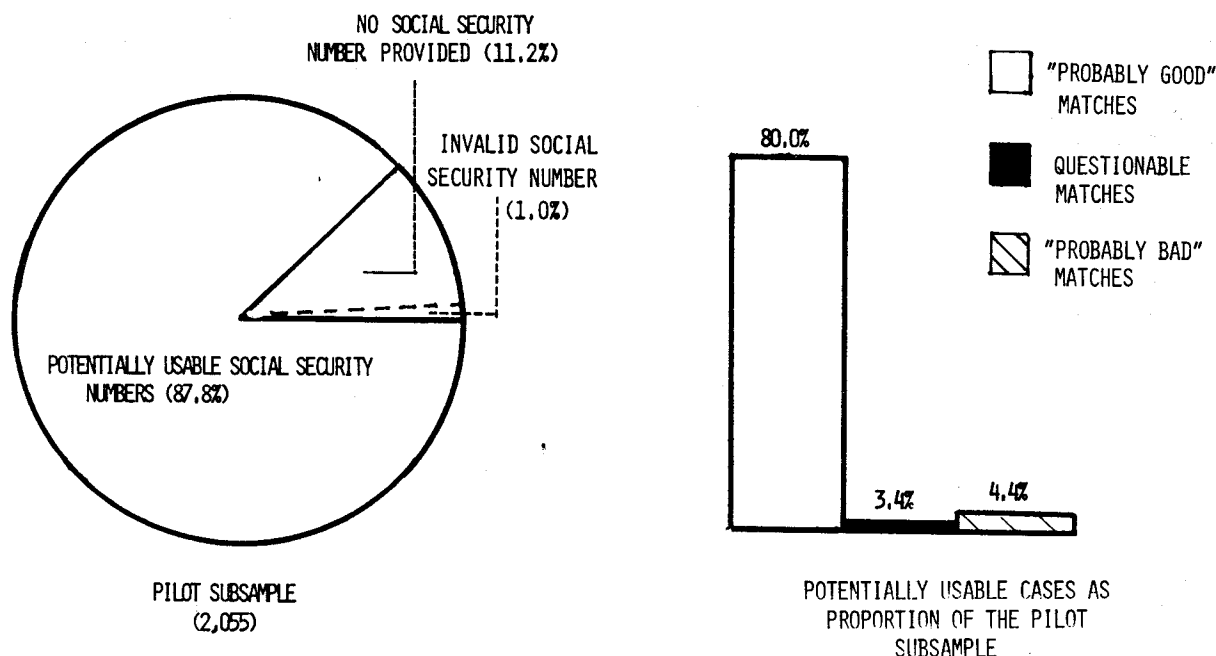


Figure 2.--Percent Distribution of Good Matches by SER Age Attained in 1975 (in years)

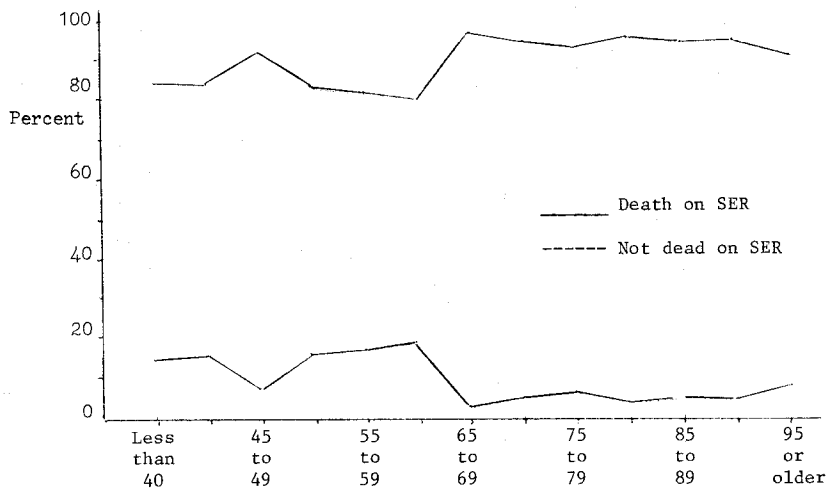


Figure 3.--Percent Distribution of Good Matches with Death on SER by Region

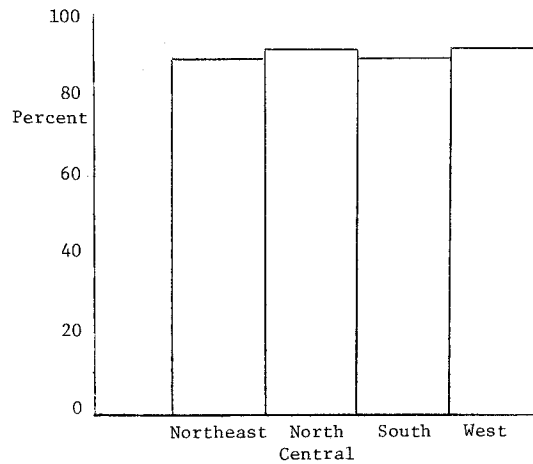
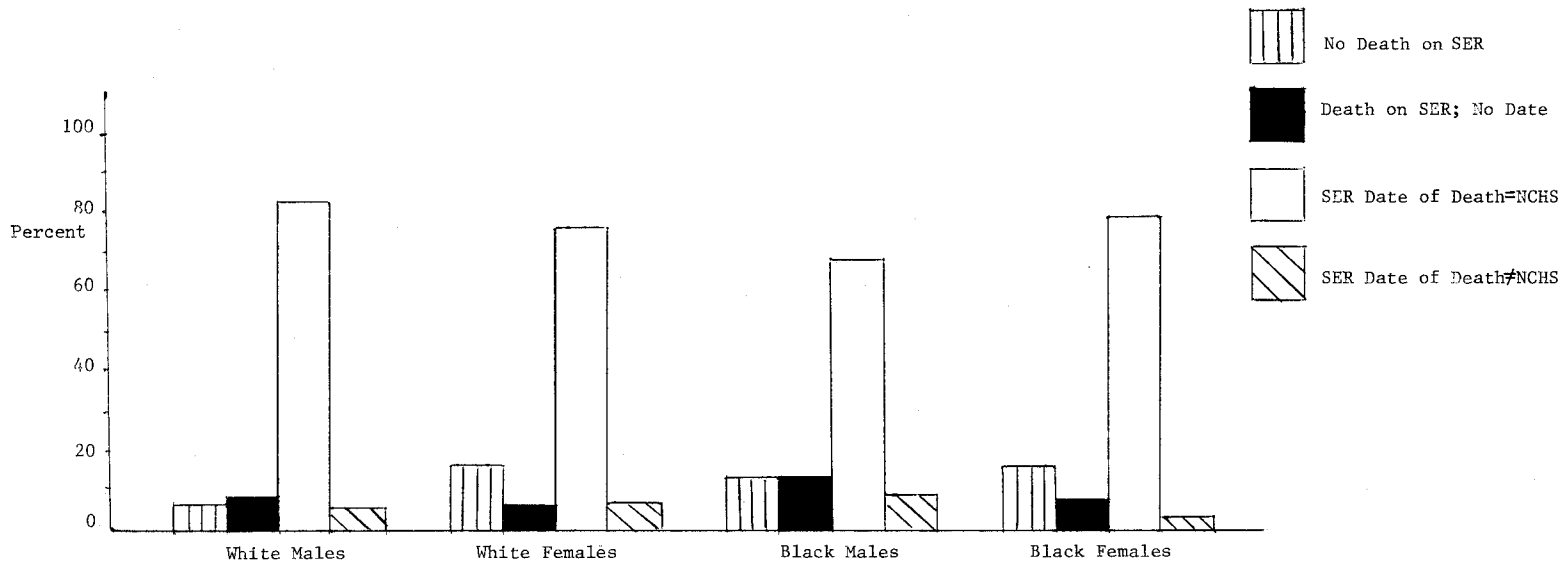


Figure 4.--Percent Distribution of Good Matches by Presence of Death on SER and SER Race-Sex



our chief interest with regard to these decedents was to examine content differences, rather than coverage problems, as we knew that they would be well-covered.

Similarly, coverage by region, illustrated in figure 3, showed that identified deaths were distributed fairly evenly nationwide. There was, however, a good bit more variation when these data were looked at State-by-State [6]. At least some of this, we feel, is undoubtedly due to sampling fluctuations.

Finally, we looked at death reporting by SSA race and sex. (See figure 4.) In this chart, the striped columns represent those not covered; the shaded areas present persons for whom no date of death is available on the earnings record; the clear portions illustrate the distribution of decedents whose SSA month of death is the same as their NCHS month of death; and the cross-hatched areas show those cases where the month of death disagrees.

As you can see, white males have considerably less undercoverage (5.9 percent) than white females (12.6 percent). Correspondingly, agreement on month of death for these males is also somewhat greater. Furthermore, it is interesting to note that while the black females shown in this chart behave essentially like the white females, the amount of agreement for reported deaths is substantially lower for black males than for white males (72 percent as opposed to 86 percent). This reflects the fact that the extent of reporting for women 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.

Limitations.--We are, of course, pleased with the high coverage rate which we obtained for this portion of the sample. However, it is important to emphasize here that we are only looking at the unweighted counts of the "probably good" matches to only one of SSA's administrative systems--the summary earnings record files.

Other data suggest that our figures overstate the extent of death reporting present on the earnings records for 1975 decedents. However, when coverage rates are examined for the 20 percent of the sample which we did not designate as good matches, our pilot results will, no doubt, come closer into line with other currently available estimates. Then, we will have more reliable data to look at coverage rates by variables which are not now available on SSA's administrative record files, such as place of death, cause of death, etc.

This Leads to Our Future Plans.--As you can well imagine, the material presented here today is just the beginning of a major analytic effort. These results are not only preliminary in nature, but incomplete, as well.

First of all, data for the five jurisdictions omitted here must be added to the subsample. Then, weighting of the entire sample is necessary. In fact, this has already been attempted. Unfortunately, computer problems and other unexpected delays prevented us from having these "corrected" figures available for this presentation.

Beyond that, our 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. As for the longer run, we will be continuing work in the area of mortality research with our colleagues at NCHS and the National Cancer Institute [7].

ACKNOWLEDGEMENTS

The authors would like to thank Thomas J. Mason at the National Cancer Institute for his help in drawing the sample for this project. At SSA, appreciation is also extended to Pearl MacLin and her staff for their extraordinary efforts in coding the death certificate data; Richard Wehrly and Joseph Manno, for their aid in matching the sampled decedents; Joan Reynolds for her typing assistance; and numerous members of the Statistical Analysis and Survey Implementation Branch in the Division of Economic Research for their extensive contributions to this project.

NOTES AND REFERENCES

[1] The following are examples of industrial mortality research using Social Security CWHS data:

"Mortality and Industrial Employment (I)," by John Goldsmith and David Hirschberg (in Journal of Occupational Medicine, vol. 18, no. 3, pp. 161-164, March 1976);

Letters to the Editor by Pierre Decoufle and John Goldsmith (in Journal of Occupational Medicine, vol. 19, pp. 582-586, 1977);

"Mortality and Industrial Employment (II)," by John R. Goldsmith (in Journal of Occupational Medicine, vol. 19, no. 4, pp. 249-254, April 1977); and

"Mortality and Industrial Employment (III)," by John R. Goldsmith and Madeline Thresh (from American Journal of Epidemiology, vol. 106, no. 2, pp. 109-124, 1977).

Other examples of epidemiological studies employing SSA data include:

"Relation of Duration of Employment and Prior Respiratory Illness to Respiratory Cancer among Beryllium Workers," by Thomas Mancuso (in Environmental Research, vol. 3, no. 3, pp. 251-275, July 1970).

[2] See "SSA Procedures for Reporting Death," by Wendy Alvey (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, pp. 144-158, February 22, 1979).

- [3] Social Security has done extensive work in the area of data linkages using exact matching techniques. For other research based on these methods, refer to reports in the series on Studies from Interagency Data Linkages, Social Security Administration. See, also, "The 1973 CPS-IRS-SSA Exact Match Study," by Beth Kilss and Frederick J. Scheuren (in Social Security Bulletin, vol. 41, no. 10, October 1978).
- [4] Our interest in mortality data grew out of a proposal by Fritz Scheuren at the NBER Workshop on Estate Multiplier Estimates of Wealth (May 2, 1975). Scheuren suggested that by using SSA mortality data, the standard estate multiplier procedure could be considerably upgraded. The following papers describe his proposals:

"Historical Perspectives on IRS Wealth Estimates with Prospectives for Improvements," by Fritz Scheuren (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, pp. 77-90, February 22, 1979); and

"1962 Federal Estate Tax Wealth Estimates for Top Wealthholders," by Fritz Scheuren (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, pp. 91-114, February 22, 1979).

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:

"Some Preliminary Results from a Validation Study of the Estate Multiplier Procedure," by H. Lock Oh and Fritz Scheuren (in 1976 American Statistical Association Proceedings, Social Statistics Section, part II, pp. 650-654).

- [5] For more information on the matching rules employed and the subsequent demographic make-up of the pilot subsample, some basic tables were provided in the handout which can be obtained from the authors upon request. Also, for those seeking more general information about the comparability of linked administrative and survey or census data, see, for example:
- "Comparability of Age on the Death Certificate and Matching Census Record, United States, May-August 1960," by Thea Hambright (in Vital and Health Statistics Data Evaluation and Methods Research, series 2, no. 29, June 1968); and
- "Validating the Social Security Number," by Cynthia Cobleigh and Wendy Alvey (in Studies from Interagency Data Linkages, Report No. 4, Social Security Administration, July 1975, pp. 89-124).
- [6] The basic tables mentioned in [5] provide this detailed information.
- [7] For documentation of the planning stages for this research, refer to

Early 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, January 30, 1979; and

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. 3, January 30, 1979.