## REFERENCES

# Identifying Deaths before 1979 Using the Social Security Administration Death Master File

*Teresa M. Schnorr and Kyle Steenland*

For cohort studies, the Social Security Administration (SSA) traditionally has been the principal source of deaths that occurred before 1979. In 1988, the SSA abolished a system that provided a relatively complete accounting of deaths and replaced it with the Death Master File. We examined the completeness of the SSA Death Master File by comparing it with the U.S. Vital Statistics records and by searching the SSA Death Master File for known decedents from seven cohorts. Overall, only 53% of reported U.S. deaths and 75% of known deaths in our seven cohorts were included in the SSA Death Master File. Ascertainment was better after 1975 (89–95%). A re-analysis of two cohorts that excluded deaths before 1979 not found in the SSA Death Master File resulted in 20–35% decreases in both standardized mortality ratios and dose-response trends. Although the SSA system before 1988 provided relatively complete vital status information, the SSA Death Master File is inadequate for vital status determination. New cohorts with a substantial number of deaths before the inception of the National Death Index in 1979 will be most seriously affected. (Epidemiology 1997;8:321–323)

**Keywords: mortality studies, Social Security Administration, vital status.**

For cohort studies, the Social Security Administration (SSA) traditionally has been the principal source of deaths that occurred before 1979. Before 1988, the Social Security Administration provided researchers with complete vital status information on study subjects (alive, dead, or no match), using records on earnings and benefits for both alive and dead subjects. In 1988, the SSA abolished this system, with its relatively complete accounting of deaths, and replaced it with the Death Master File (DMF), which is compiled using more limited data on all U.S. decedents whose beneficiaries received death benefits from the SSA. The SSA DMF and the National Death Index (NDI), which began in 1979, are now the primary sources for identifying deaths in cohort studies in the United States. Several papers have demonstrated that the NDI is very effective at identifying known decedents (between 93% and 98%) and that vital status data provided by the SSA before 1988 are nearly as complete (between 83% and 93%).[1–3] A similar evaluation of the SSA DMF has not been conducted.

From the [1]National Institute for Occupational Safety and Health, Division of Surveillance, Hazard Evaluations and Field Studies, Cincinnati, OH.

Address correspondence to: Teresa Schnorr, Industrywide Studies Branch, National Institute for Occupational Safety and Health, Division of Surveillance, Hazard Evaluations and Field Studies, 4676 Columbia Parkway, Cincinnati, OH 45226.

Submitted March 22, 1995; final version accepted December 16, 1996.



## Methods

To examine the completeness of the SSA DMF, we first compared the number of deaths contained in the SSA DMF with the U.S. Vital Statistics data, the most complete source for U.S. deaths. Although the SSA DMF would not be expected to include all U.S. deaths because before the 1970s, a substantial proportion of the population did not pay into the Social Security system, it is nonetheless instructive to determine what percentage of U.S. deaths have been included by the SSA DMF over time.

We then selected seven cohorts that had been followed for vital status before the change in the SSA procedures for providing death information, each with follow-up thought to be over 90% complete.[4–10] Sources of follow-up included the information on vital status supplied by the SSA before 1988, information provided by the Internal Revenue Service, the Post Office, company records, the SSA DMF, and the NDI.

To determine the proportion of known deaths found by the SSA DMF, we searched the SSA DMF for known decedents in these seven cohorts, using the Social Security Number and last name to identify matches. Deaths that were not identified by an exact Social Security Number match were submitted to Epidemiologic Resources Inc., a commercial vital status service, which also searched the SSA DMF, using several search algorithms based on name, date of birth, and partial Social Security Number.

Currently, investigators outside the government rely principally on the SSA DMF and the NDI (NDI reports only deaths after 1978) to identify decedents. One might assume that study subjects not found to be dead via these

**TABLE 1. Number and Percentage of Deaths Identified on SSA Death Master File Compared with U.S. Vital Records, by 5-Year Intervals**

| Year | Number of Deaths* | | Percentage of U.S. Deaths Included in SSA Death Master File |
|---|---|---|---|
| | SSA Death Master File | U.S. Vital Statistics† | |
| 1940–1944 | 28,738 | 7,070,940 | 0.4 |
| 1945–1949 | 58,127 | 7,130,297 | 0.8 |
| 1950–1954 | 132,222 | 7,429,878 | 1.7 |
| 1955–1959 | 297,268 | 8,031,054 | 3.7 |
| 1960–1964 | 1,802,057 | 8,781,824 | 20.6 |
| 1965–1969 | 4,986,613 | 9,394,680 | 53.1 |
| 1970–1974 | 7,078,151 | 9,730,046 | 72.7 |
| 1975–1979 | 8,541,048 | 9,556,551 | 89.4 |
| 1980–1984 | 8,947,415 | 10,017,909 | 89.3 |
| 1985–1989 | 8,625,517 | 10,646,493 | 81.0 |
| 1990–1991 | 3,330,674 | 4,317,981 | 77.1 |
| Total | 48,992,966 | 92,107,653 | 53.2 |

* Deaths under age 15 years represented about 11% of all deaths in 1940, decreasing continuously to about 2% in 1992. Deaths under age 15 years are included in the data for U.S. Vital Statistics but might not have Social Security Numbers and therefore might not be expected to appear on the SSA DMF. The great majority of such deaths are infants. After 1988, SSA was required to collect information on all deaths. Before 1988, SSA collected death data only on beneficiaries.

† Sources: National Center for Health Statistics. Monthly Vital Statistics Report. Table 1. Deaths in the U.S. by Race and Sex. 1993:42(2S);15; National Center for Health Statistics, Division of Vital Statistics. Vital Status in the United States 1974. vol. 2. Mortality. Part A. Hyattsville, MD: National Center for Health Statistics, 1978.

two sources should be considered alive. To assess the hypothetical effect of this assumption, we reran analyses[11] for two of our cohorts (cadmium and uranium). Those previously identified as dead via sources other than the SSA DMF or NDI were considered alive until end-of-follow-up in these analyses. In essence, this assumption meant excluding deaths before 1979 not found in the SSA DMF. Results were then compared with previous analyses.

## Results

Table 1 and Figure 1 show the deaths contained in the SSA DMF compared with reported deaths in U.S. Vital Records. Overall, only 53% of U.S. deaths were included in the SSA DMF, with the percentage of deaths included increasing over time (over 89% after 1975).

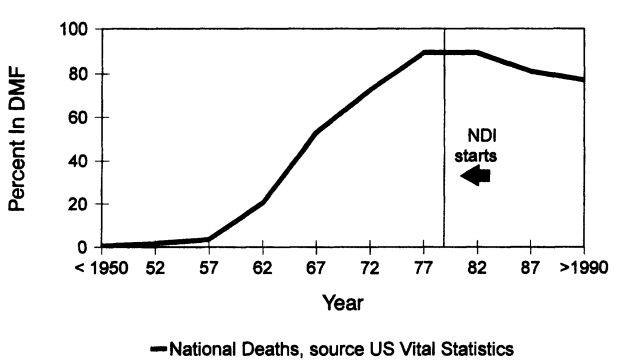


**FIGURE 1. Percentage of U.S. deaths in SSA Death Master File.**

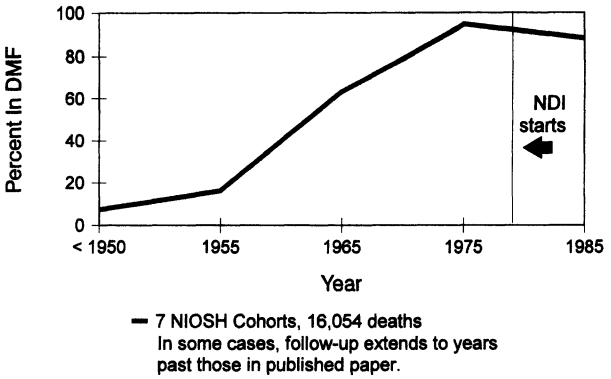


**FIGURE 2. Percentage of NIOSH cohort deaths found in SSA Death Master File.**

We then searched the SSA DMF for the known decedents (N = 16,054) in seven National Institute for Occupational Safety and Health (NIOSH) cohorts followed for different time periods between 1940 and 1993; we found 11,955 (75%). Figure 2 shows the percentage of known cohort deaths identified by the SSA DMF over time. Paralleling Figure 1, there is a clear trend of better ascertainment of deaths in more recent decades.

A search using a more extensive algorithm for the 4,099 deaths not found by the initial search of the SSA DMF found only 81 more potential matches. Hence, invalid Social Security Numbers could explain only a small proportion of the missed deaths in the SSA DMF.

We reran analyses for two cohorts that showed an excess of lung cancer deaths (cadmium and uranium), including as deceased only those cohort members identified as dead from the SSA DMF or the NDI. When deaths are not identified, these persons continue to accumulate person-years at risk. This accumulation has the effect of increasing the expected number of deaths as well as decreasing the observed number of deaths, resulting in potentially large decreases in SMRs. For cadmium, the original lung cancer SMR of 1.40 (25 lung cancers) dropped to 0.91 (20 lung cancers). For uranium, the original analysis yielded a lung cancer SMR of 5.29 (370 lung cancers), whereas follow-up relying only on SSA DMF and NDI data yielded a lung cancer SMR of 4.21 (341 lung cancers). Furthermore, the lung cancer dose response dropped 25% (from a rate increase of $1.69 \times 10^{-5}$ per working-level month (WLM) to $1.29 \times 10^{-5}$ per WLM), because the missing lung cancer deaths occurred in the 1950s through 1970s and were clustered in the higher exposure categories.

## Discussion

The SSA system used to provide a relatively complete accounting of deaths using earnings and benefits data for both alive and dead individuals. In 1988, the SSA abolished this system and created the SSA DMF system, based on death benefits alone.

Our data show that the SSA DMF includes only 53% of U.S. deaths since 1940. In our seven cohort studies, the SSA DMF ascertained approximately 75% of known

deaths, and the greatest underascertainment occurred in the earlier decades. The previous SSA vital status database was the original source of 90% of the missed deaths occurring before 1979 (when NDI was created). After 1979, the SSA DMF identified no death that was not identified by NDI. This last finding indicates that the SSA DMF is superfluous for the post-1979 period if NDI is used.

Our analysis shows that the SSA DMF is inadequate for identification of deaths and cannot be used as the primary source of vital status in cohort studies. For deaths that occurred before 1979 (pre-NDI), however, there is no other large database of U.S. deaths that is publicly available. Other databases provide death information on special groups, such as veterans (Veterans Administration data) or elderly individuals (Health Care Finance Administration data, which also provides less than complete coverage of the total population before the 1970s). Occupational studies may have access to company death records, which can provide death information on pensioners but usually miss shorter-term workers. These sources may not cover the majority of subjects in a particular study. Although one can also determine whether a subject is alive via use of state drivers' license bureaus and retail credit agencies, finding no record of a subject in these databases cannot reliably be taken to mean the subject is dead.

An additional problem with the SSA DMF is that slightly different versions of the data have been released over time. Conversations with the SSA have failed to clarify the reason for such differences.

The problem identified here will affect primarily newly assembled cohorts in which a substantial number of deaths occurred before the availability of NDI data in 1979. The failure to identify pre-NDI deaths will result in an underestimation of mortality. In studies in which exposure increases mortality, the effect of exposure may be missed owing to this underestimation. Authors should be clear about the source of their vital status data so that potential underascertainment can be appreciated; sensitivity analyses that estimate the potential effect of undercounting deaths when using the SSA DMF might be useful.

## References


1. Wentworth DN, Neaton JD, Rasmussen WL. An evaluation of the Social Security Administration Master Beneficiary Record File and the National Death Index in the ascertainment of vital status. Am J Public Health 1983;73:1270–1274.
2. Boyle CA, Decouflé P. National sources of vital status information: extent of coverage and possible selectivity in reporting. Am J Epidemiol 1990;131: 160–168.
3. Curb JD, Ford CE, Pressel M, Palmer C, Babcock C, Hawkins CM. Ascertainment of vital status through the National Death Index and the Social Security Administration. Am J Epidemiol 1985;121:754–766.
4. Fingerhut M, Halperin W, Marlow D, Piacitelli L, Honchar P, Sweeney M, Greife A, Dill P, Steenland K, Suruda A. Cancer mortality in workers exposed to 2,3,7,8-TCDD. N Engl J Med 1991;324:212–218.
5. Steenland K, Selevan S, Landrigan P. The mortality of lead smelter workers: an update. Am J Public Health 1992;82:1641–1645.
6. Ward E, Okun A, Ruder A, Fingerhut M, Steenland K. A mortality study of workers at seven beryllium processing plants. Am J Ind Med 1992;22:885–904.
7. Rinsky RA, Ott G, Ward E, Greenberg H, Halperin W, Leet T. Study of Mortality Among Chemical Workers in the Kanawah Valley of West Virginia. Am J Ind Med 1988;13:429–438.
8. Hornung R, Meinhardt T. Quantitative risk assessment of lung cancer in U.S. uranium miners. Health Phys 1987;52:417–430.
9. Stayner LT, Smith R, Thun M, Schnorr T, Lemen R. A dose-response analysis and quantitative assessment of lung cancer risk and occupational cadmium exposure. Ann Epidemiol 1992;2:177–194.
10. Steenland K, Stayner L, Greife A, Halperin W, Hayes R, Hornung R, Nowlin S. Mortality among workers exposed to ethylene oxide. N Engl J Med 1991;324:1402–1407.
11. Steenland K, Beaumont J, Spaeth S, Brown D, Okun A, Jurcenko L, Ryan B, Phillips S, Roscoe R, Stayner L, Morris J. New developments in the life table analysis system of the National Institute for Occupational Safety and Health. J Occup Med 1990;32:1091–1098.