

# A MATCHED-RECORDS STUDY OF AGE AT DEATH INFORMATION

Bert Kestenbaum, Social Security Administration

The accuracy of age at death information is essential to the correct calculation of mortality rates and life table functions; the identification of important demographic phenomena such as the "race crossover"; the tracking of gains in the reduction of mortality due to advances in medicine, public health measures, changes in life styles, and other factors; and the determination of the mortality experience of various study groups. This paper reports the very high extent of consistency in age at death information for the elderly between the death certificate in two States in 1987 and the social security/medicare administrative record.<sup>1</sup> Among matched records, 92½ percent agree exactly on year of age; the discrepancy is but one year in another 4½ percent of matched records. Also, 97 percent of the matched records fall in the same 5-year age interval in the two record systems.

Demographers have occasionally expressed skepticism towards the accuracy of age at death information for the extreme aged, those age 85 and over.<sup>2</sup> Yet high agreement rates are obtained even for this subgroup. About 89½ percent agree exactly on year of age, with another 6½ percent discrepant by only one year. About 96 percent are in the same 5-year age grouping in the two files.

Although the consistent reporting of an item of information does not guarantee its accuracy, we believe that these high rates of agreement can be taken as evidence of the credibility of the age information in both files. We view the vital registration system as the big winner here. The credibility of the age information in the social security/medicare file was always less suspect, given that an applicant for social security benefits or medicare coverage is generally required to prove his/her age exactly. Admittedly, the generality of this assertion is far from perfect, particularly for the now extreme aged, some of whom were awarded medicare eligibility at that program's inception in 1965 without proof of exact age.

Rates of agreement are substantially lower for blacks than for whites. In most discrepant cases, the black decedent is older at death according to program records than according to vital records. In particular, among matched records the numbers of black decedents age 85 and over at death in the respective record systems are 2,400 and 2,178 (table 1). This finding is significant for the correct placement of the race crossover phenomenon: if program records are the more accurate, the crossover,

if it occurs, does so at a later age than that shown in official publications.

The very high match rate achieved in this study obviates the need for caveats about the possibility that the results would differ substantially were all records successfully matched. In fact, less than 3 percent of the 1987 death certificates were not matched to the social security/medicare administrative record. Nonetheless, we undertook an investigation of this potential bias, the results of which suggest that it is unlikely that the ages on the unmatched death certificates are of particularly questionable quality.

## A PROTOTYPE STUDY

The only matched-records study of age at death information in the United States of comparable scope was executed by the National Center for Health Statistics (NCHS) as a byproduct of the famous Kitagawa-Hauser project.<sup>3</sup> A sample of death certificates from all States for May-August 1960 was matched manually, using name and address, to census schedules. Included in the sample were approximately 113 thousand certificates for decedents age 65 and over, among which 80.8 percent were successfully matched (figure 1).

The NCHS study found that the death certificate age and the census age were in the same 5-year age interval for just 84.9 percent of elderly decedents. The percent agreement on single year of age is not shown separately for the elderly, but for the group 45 and over it was only 68.1 percent. The report contains speculation that the true rates of agreement are even lower, considering the effects of match bias. One aspect of match bias is that people whose age is reported inconsistently may be more likely to be people for whom the identifiers used in forming matches are also reported inconsistently. Another aspect of match bias in that study stems from the occasional practice of the clerical staff to use the consistency of age information in forming a match.

The study found a marked difference in agreement rates between whites and blacks: only 54.6 percent of elderly blacks were reported in the same 5-year age grouping in the two data systems! From its analysis of the data, the NCHS concluded that the race crossover in mortality, which, at that time, occurs according to

Table 1.--Matched records: certificate age group and Social Security/Medicare record age group, by race

Age on death certificate	Age on Social Security/Medicare record									
	Total, all ages	Under 65	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100 and over
<b>All races</b>										
Total, all ages	121,127	162	17,171	20,685	23,948	23,555	18,835	11,800	4,176	795
65-69	17,403	68	17,069	198	38	14	6	8	2	0
70-74	20,805	20	77	20,363	265	51	16	9	1	3
75-79	24,087	25	12	109	23,459	378	79	18	5	2
80-84	23,622	22	8	5	163	22,927	419	58	14	6
85-89	18,725	13	3	5	21	177	18,110	341	45	10
90-94	11,525	9	2	4	1	5	173	11,155	161	15
95-99	4,135	4	0	1	1	2	27	193	3,879	28
100 & over	825	1	0	0	0	1	5	18	69	731
<b>White, not Hispanic</b>										
Total, all ages	104,288	101	14,342	17,625	20,540	20,320	16,491	10,499	3,708	662
65-69	14,446	39	14,290	89	15	8	2	2	1	0
70-74	17,664	13	42	17,466	116	15	6	4	0	2
75-79	20,583	15	4	62	20,291	178	28	2	1	2
80-84	20,374	13	3	2	108	19,999	229	15	3	2
85-89	16,480	12	2	4	8	115	16,086	226	24	3
90-94	10,356	6	1	2	1	2	125	10,104	107	8
95-99	3,701	3	0	0	1	2	14	139	3,529	13
100 & over	684	0	0	0	0	1	1	7	43	632
<b>Black</b>										
Total, all ages	9,288	39	1,545	1,711	1,829	1,764	1,258	754	297	91
65-69	1,652	22	1,505	91	22	4	1	6	1	0
70-74	1,787	5	29	1,585	119	33	10	4	1	1
75-79	1,909	4	7	33	1,638	164	45	15	3	0
80-84	1,762	5	3	1	40	1,514	146	40	10	3
85-89	1,181	0	0	0	10	48	1,017	84	17	5
90-94	643	1	1	1	0	1	29	560	45	5
95-99	262	1	0	0	0	0	9	37	202	13
100 & over	92	1	0	0	0	0	1	8	18	64
<b>Hispanic</b>										
Total, all ages	7,551	22	1,284	1,349	1,579	1,471	1,086	547	171	42
65-69	1,305	7	1,274	18	1	2	3	0	0	0
70-74	1,354	2	6	1,312	30	3	0	1	0	0
75-79	1,595	6	1	14	1,530	36	6	1	1	0
80-84	1,486	4	2	2	15	1,414	44	3	1	1
85-89	1,064	1	1	1	3	14	1,007	31	4	2
90-94	526	2	0	1	0	2	19	491	9	2
95-99	172	0	0	1	0	0	4	17	148	2
100 & over	49	0	0	0	0	0	3	3	8	35

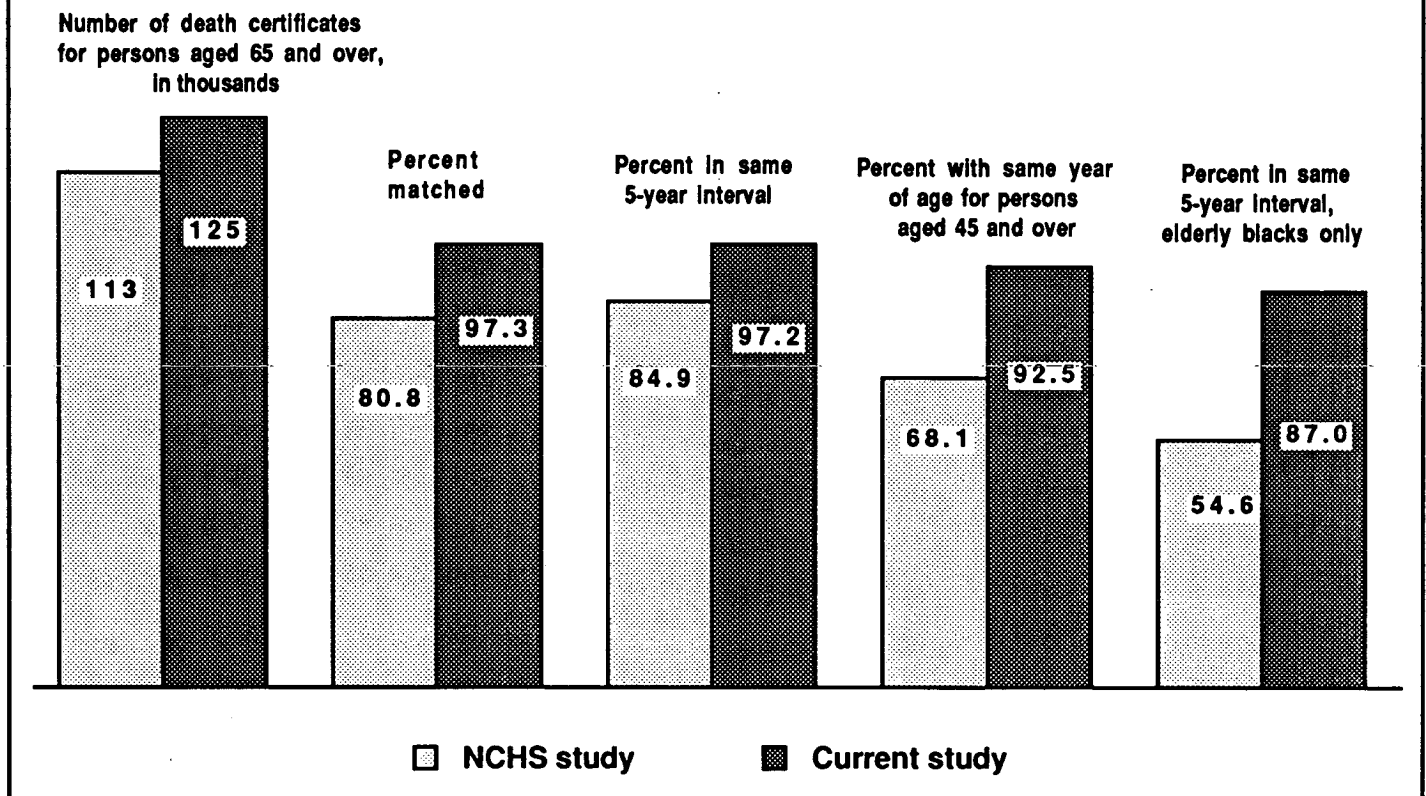
official data in the 75-84 age range, in fact occurs after age 85, if at all.

#### DESIGN OF CURRENT STUDY

From among several populous States which we contacted about the availability of a death certificate file on magnetic media at nominal cost for research

purposes, positive replies were received from Massachusetts and Texas. The 54,268 death certificates in Massachusetts and the 123,033 certificates in Texas for 1987 included a combined total of 124,559 records for decedents age 65 and over. The removal of 52 duplicate records left us with a file of 124,507 records to attempt to match. Given the importance of the social security number in the matching effort, it was indeed fortunate

## Figure 1.--Comparison with NCHS study



that relatively few--about 2.3 percent--of the death certificates for the elderly were either missing a social security number entirely (1.8%) or had an impossible number recorded (0.5%).

These death certificates were matched to the Master Beneficiary Record (MBR) maintained by the Social Security Administration, the master file of persons either entitled to or denied social security benefits, or enrolled in the medicare program.<sup>4</sup> This file, features a date of birth which was established for purposes of program eligibility after evidentiary requirements were satisfied.

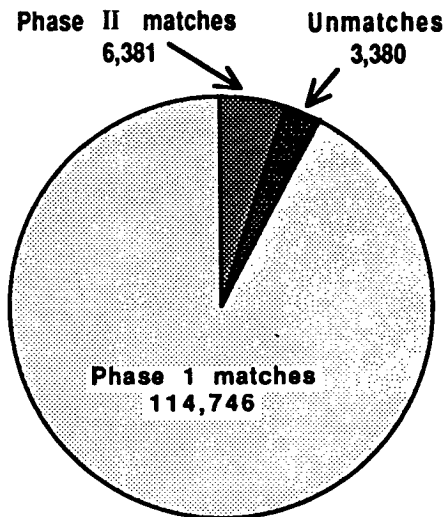
The matching was carried out in two phases. In the first phase, which netted almost 95 percent of the ultimate number of matches, records were selected from the MBR which matched exactly the social security number on the death certificate. From experience, we knew that the social security number entered on a death certificate or similar document sometimes is not the person's own number, but the claim number under which the person receives social security benefits or is enrolled in the medicare program. These two numbers will differ when the person is a secondary beneficiary, that is, when the person's sole basis for eligibility is the circumstance of being the spouse or other relative of a primary beneficiary. Accordingly, when the set of

social security numbers appearing on death certificates was applied to the MBR, any record with one of these numbers as either a claims number or an own number was selected. We are here taking advantage of another feature of the MBR which is missing in other administrative files that are available to us, namely, that the claim number, as well as the own number, is present in the record of an individual who is a secondary beneficiary.

Almost 115 thousand death certificates -- 114,746 to be exact -- were matched in this first phase (figure 2), including about 7 thousand death certificates where the social security number turned out to be not the decedent's own number, but the number of a primary beneficiary to whom the decedent was related.

Agreement on date of death was quite good. In fact, 96 thousand of the total of 115 thousand matched records matched exactly on month, day, and year of death; and another 17 thousand matched on month and year of death, with the failure to match on day of death attributable in the great majority of these cases to the absence of day of death information in that MBR record. Only 549 matched pairs matched on year of death and had discrepant month of death information, and most of the time the discrepancy was but one month. Only 700 matched MBR records did not have 1987 as

**Figure 2.--Success in matching**



the year of death; most of these had no year of death, the fact of death having never been reported to the Social Security Administration.

For the second phase, we began by selecting from the MBR all records showing both death in 1987 and last address in Massachusetts or Texas, and then removed from this file those records which were already matched in the first phase. To make the matching manageable we employed a simple technique technically known as "blocking." Both the file of yet-unmatched death certificates and the file of yet-unmatched MBR records were sorted according to the values of a key variable--the blocking variable--and only records with identical values of the blocking variable were eligible for matching. We used several blocking variables, one after another. Each time, before proceeding to the next round, the records matched in the current round were removed from the two files. The second phase yielded 6,381 matches.

The blocking variables we used, in the order of their use, are listed below, together with the number of certificates matched.

1. Exact date of death (4,509).
2. 8-digit number obtained by removing a single digit from the social security number (820).
3. Month of death, when day of death was missing from the MBR record (869).

4. Date of death after the certificate date of death was increased/ decreased by 1 (72).
5. Combined first and last names (95).
6. Last name, for uncommon names (16).

In both phases of the matching process, we required some level of name agreement between paired records to confirm a match. For pairs with identical social security numbers (even 8 digits) and exact dates of death, our requirement was that either the first names or the last names agree. For pairs with the same social security number but not the same exact date of death, or the same exact date of death but not the same social security number, the requirement was that both first and last names agree. When name was used as a blocking variable, we required that any discrepancy in date of death be within a tolerance.

Notice that age is not used in any way in establishing a match, thus avoiding one potential source of match bias.

Before attempting to ascertain name agreement, all names were passed through a computer program, which we wrote, that standardizes names by dealing with irregularities such as punctuation within names, and suffixes and titles.

For name agreement we required only that the first or last 'n' characters agree, where 'n' is the length of the shorter character string in the pair. Thus, 'TIM' and 'TIMOTHY' are considered to be in agreement, as are 'ELIZABETH' and 'BETH'. Furthermore, under certain circumstances, we resorted to a variety of devices such as ignoring one discrepant letter or one extra letter, SOUNDEX coding<sup>5</sup>, and a lookup table of common variations on first names, as well as a very limited amount of clerical review, to uncover name agreements. The leniency of the name-matching procedures results in more "false positives" than would be forthcoming from a more strict match rule. But the effect is to underestimate the extent of age agreement.

Reliance on the MBR, however, exposes us to a certain type of match bias. Consider a typical decedent at age 55 whose age is incorrectly entered on the certificate as '65': unless some survivor benefits are payable, this person will likely not appear on the MBR because of his young age, and his faulty certificate record will not be matched. Although the effects of this type of match bias must be small, given that relatively few certificates remain unmatched to the MBR, we sought evidence for its existence through another matching effort.

The certificates not matched to the MBR which bear a social security number were matched to a second administrative file, one which is not limited to the

Table 2.—Matched records: percent agreement, by selected characteristics

Characteristic	65 & over		85 & over	
	Same exact age	Same 5-year interval	Same exact age	Same 5-year interval
Total, all matched records	92.5%	97.2%	89.4%	96.2%
<u>Race</u>				
White, not Hispanic	94.6	98.2	91.7	97.2
Black	72.6	87.0	63.2	84.6
Hispanic	88.4	95.5	81.7	92.8
<u>Age</u>				
65-69	95.4	98.1	---	---
70-74	94.6	97.9	---	---
75-79	93.6	97.4	---	---
80-84	92.1	97.1	---	---
85-89	90.6	96.7	90.6	96.7
90-94	89.3	96.8	89.3	96.8
95-99	86.5	93.8	86.5	93.8
100 & over	80.1	88.6	80.1	88.6
<u>Sex</u>				
Male	94.2	98.0	91.0	96.9
Female	91.0	96.5	88.7	95.9
<u>Marital status</u>				
Married	95.2	98.2	91.9	97.2
Not married	90.7	96.5	88.9	96.0
<u>Place of death</u>				
Hospital or nursing home	92.6	97.1	88.9	95.7
At home or other	92.5	97.2	89.5	96.3
<u>Beneficiary status</u>				
Regular beneficiary	92.9	97.3	90.0	96.5
Not regular beneficiary	85.8	94.2	84.3	93.8

beneficiary population, the NUMIDENT file of applicants for a social security number. Among the 2,283 record pairs which were acceptably matched, only 21 had a NUMIDENT age of less than 65, leading to the conclusion that the type of match bias we were concerned about is not a factor. (Note that only 1,117 certificates -- including 573 with no social security number -- are not matched to either administrative file.)

The age agreement among these matched records is quite good: about 94 percent are in the same 5-year interval in both files. These results are not included in the analysis proper because the NUMIDENT age can be an allegation unsupported by documentation.

#### DETAILED FINDINGS

The high extent of age at death consistency among the elderly between the death certificate and the

social security/medicare administrative record, even among the extreme aged, which is the major finding of this investigation, does, of course, vary among population subgroups. Among the six ways we partitioned the population, the partition by race exhibits the largest differentials. Blacks have very much lower levels of agreement than whites, and hispanics have substantially lower levels than non-hispanic whites (table 2).

Agreement rates fall as age increases. One reason may be that the decedent has outlived those persons who were knowledgeable about his/her age. The individual may have suffered memory loss, or preferred not to divulge his/her real age. Perhaps an important factor is that earlier cohorts are less-educated or don't care as much about knowing their own precise ages.

Agreement is less for females than for males, for

unmarried persons than for married persons, and for persons not entitled to earnings-related social security benefits than for persons so entitled. In each of these three comparisons the lower-agreement group is the older of the two groups, but the differential remains, although it is smaller, after the differences in age are controlled for through a standardization procedure.

These differences were expected. It was hypothesized that women have a greater tendency than men to conceal their true age. Married persons should have higher agreement rates than persons not married because a surviving spouse would be knowledgeable about the decedent's age. It is known that persons not entitled to regular, earnings-related social security benefits (whether the earnings are their own or the earnings of a spouse or other relative serving as the basis for their entitlement as secondary beneficiaries) often were permitted to satisfy the proof-of-age requirements for medicare or certain special, non-earnings-related social security benefits with evidence of having achieved some minimum age, rather than of exact age. In fact, for this reason the mortality experience of persons not entitled to earnings-related benefits is excluded from the medicare administrative data used in computing the extreme aged section of the official decennial life tables.<sup>6</sup>

It was also expected that agreement would be better for persons dying in hospitals or nursing homes than for persons dying at home or elsewhere on the theory that the decedent's age would be available from records of the hospital or home. In fact, the differential was quite small, and, for the extreme aged, in the direction not expected.

#### IN CLOSING

This study, finding a high extent of age consistency among elderly decedents between the death certificate and the Master Beneficiary Record, is part of an effort to increase the usefulness and improve the accuracy of administrative files for mortality and population data. An earlier paper<sup>7</sup>, besides presenting an estimate of the centenarian population, advanced the thesis that the quality of mortality studies of the very old would be well-served by excluding the experience of persons whose record indicates entitlement to Part A of Medicare -- the (usually) free hospital insurance -- but not to Part B, the medical insurance which is purchased by monthly premiums, because in large part these persons are decedents of whose death we were not notified.

Our current work focuses on the mortality statistics for the very old which emerge from the joint consideration of the Social Security Administration's Master Beneficiary Record and the Health Care Financing Administration's Health Insurance Master files, and on the differences between mortality rates

computed from individual records and mortality rates computed, using standard assumptions and techniques, from counts of population and deaths.

#### NOTES

- [1] Throughout this paper, the social security/medicare age at death is the age (in completed years) computed by subtracting the social security/medicare record date of birth from the death certificate date of death, while the certificate age at death is the datum appearing in the "age" block of the certificate, rather than the result of a similar computation using the certificate date of birth. This is consistent with the practice followed by the National Center for Health Statistics in its publication program. Actually, in Texas the age item is edited to be consistent with date of birth/date of death, but Massachusetts tolerates discrepancies of one or two years and makes followback efforts to resolve discrepancies exceeding the tolerance. Other States follow assorted procedures.
- [2] For example, the Bureau of the Census report, *Demographic Aspects of Aging and the Older Population in the United States* (series P-23 No. 59, May 1976, p. 65) asserts: "Reporting of the age of decedents among the extreme aged (85 years and over) in the vital registration system is believed to be quite inaccurate. The exact or even approximate age of most decedents at these ages is not known to surviving relatives, friends, or neighbors, and their reports tend to be a guess, with a tendency towards exaggeration of age."
- [3] National Center for Health Statistics, *Comparability of Age on the Death Certificate and Matching Census Record: United States, May-August 1960, Series 2, No. 29, June 1968.*
- [4] A complete description of the Master Beneficiary Record appears in *Internal Revenue Service, Statistical Uses of Administrative Records: Recent Research and Present Prospects*, vol. 1, March 1984, pp. 33-35.
- [5] We modified the "modified NYSIIS name coding procedure" presented in *USDA Statistical Research Service, Selection of a Surname Coding Procedure for the SRS Record Linkage System*, February 1977.
- [6] National Center for Health Statistics, *U.S. Decennial Life Tables for 1979-81: Methodology of the National and State Life Tables*, vol. 1, no. 3, May 1987, pp. 1-7 and 1-8.

[7] Bert Kestenbaum, "Centenarians: An Administrative Records Perspective," paper presented at the annual meeting of the Population

Association of America, Baltimore, Maryland, March 1989.