CHANGES IN THE IMPACT OF EDUCATION AND INCOME ON MORTALITY IN THE U.S.

Sherwin Rosen, University of Chicago, NBER
and
Paul Taubman, University of Pennsylvania, NBER

Social scientists and physicians wish to analyze the differential life expectancy and age-specific death rates of various socioeconomic groups to learn the causes of differential longevity and to project the future development of programs such as social security. Yet there is relatively little empirical information on the relationship between age-specific death rates and various socioeconomic characteristics, because it has proved difficult and costly to obtain a large random sample of the U.S. population with interesting sociodemographic and mortality information [2]. Indeed, the only major study that we are aware of that groups individuals by characteristics such as education and income [2] and does not restrict itself to a few types of death is by Kitagawa and Hauser [3].

Kitagawa and Hauser (KH) matched "340,000 death certificates of a total of 535,000 deaths which occurred in the United States during the months May-August, 1960, to the 25% 1960 Census Records" to obtain the requisite socioeconomic information [4].

In this paper we propose an alternative method to obtain data on death rates. We use a new sample from the 1970's to reexamine the KH results. Our detailed results differ in many ways from KH's. For example, while they find no differences by education for white males over 65, we find marked differences.

1. THE GENERAL METHOD

It is possible to replicate the KH methodology using any other random sample. We propose, however, to use another method which is described in this section.

In the KH method, a sample of death certificates is collected and matched back to an appropriate random sample of the population. Our methodology is the reverse, i.e., socioeconomic data is collected from a random sample of the population and it is subsequently determined who in the sample died within a particular time interval. Subsequent mortality can be ascertained in two types of samples.

1. A longitudinal sample where death is recorded as a reason for not obtaining a re-interview. For example, Social Security's Retirement History Survey [5] includes data on some 1,000 respondents who died between the first wave in 1969 and the third wave in 1973. (We are analyzing this sample but do not report on it here.)

2. A cross-section sample where death is obtained by subsequently matching the interviews to Social Security Administration (SSA) files. The present paper employs one such data base, the March 1973 Current Population Survey [6].

An important advantage of the KH method is that it is possible to obtain a large sample of deaths even at fairly young ages, say, under 45. Moreover, there is no reason to suspect more accurate reporting of death for certain groups which is a difficulty with the SSA files [7]. This latter advantage may be offset, however, by bias arising from higher match rates for some groups, e.g., nontransients, since KH only matched about 80% of the certificates.

Using CPS-SSA linked data has several advantages over the time-consuming and expensive process of matching death certificates and census documents. The March Current Population Survey also contains more demographic and economic information than did the 25% census sample used by KH in 1960. Furthermore, the SSA portion contains pre- and post-sample information on earnings and labor force participation and some health and disability data.

One disadvantage of using CPS-SSA linked data for mortality research is that in order to obtain an adequate sample, deaths must be measured over a several year interval during which marital status, etc., may change. Another disadvantage, not already mentioned, is that the CPS data used in this paper was limited to the noninstitutionalized population in 1973, although it will include data on a spouse who is temporarily institutionalized. Those in hospitals and nursing homes have a higher death rate than other people of the same age, education, etc. (see KH). If use of these institutions is related to education, income, etc., some of our estimates will be biased. However, the institutionalized population is not large, and data on death rates for this population in KH indicate that the bias should not be great.

2. THE MARCH 1973 CPS-SSA SAMPLE

The March 1973 Current Population Survey was a stratified multi-stage cluster sample of about 45,000 interviewed households containing some 136,000 persons. To be eligible for the CPS, an individual had to be living in one of the 50 States or the District of Columbia. The institutional population was excluded as were all Armed Forces members except those living off-post or on-post with their families.

As part of the 1973 Exact Match Study, an attempt was made by the Census Bureau to obtain the social security numbers of the 101,000 sampled persons 14 years or older. In about 80,000 cases, potentially usable numbers were recorded by the interviewers. An additional 10,000 adults were matched to SSA files using CPS data on name, age,

sex, and race. Available evidence suggests that the matching was done accurately [8], and that inability to match everyone has not caused the matched subset to differ greatly from the whole CPS [9].

The total number of deaths per year in the CPS-SSA Exact Match sample is given in Table 1. It should be noted that about 400 people in the SSA files were recorded as having died prior to March 1973, though CPS information is recorded for March 1973. This anomaly occurs due to mismatches, CPS and SSA punching and transcription errors, or some women's (legal) use of their deceased husband's social security number. We exclude these people from our analysis as well as the few individuals whose date of death is unknown. The remaining 3,030 physically matched cases will constitute our basic study group [12].

Table 1.--Annual Number of Deaths in the CPS-SSA Sample through June 1977

Year of Death	Physical Matches	"Good" Matches
Total.....	3,445	2,786
Year Unknown.....	8	6
Before 1973.....	386	67
1973	534*	479
1974.....	700	626
1975.....	815	736
1976.....	801	697
1977 (part year).....	201	181

Note: In the Exact Match Study, users are provided all the matches obtained plus rules for determining which matches might be considered accurate. The rule employed here to determine a "good" match is the same as that described in [8].

*There were 20 physical matches for persons who were recorded as dying in the first three months of 1973.

Since the CPS survey was undertaken in March, we would expect recorded 1973 deaths to be about 75% of the number in all of 1974. This compares favorably with the 73% found in the sample. The total number of deaths in 1975 is a bit higher than in 1976 and more than 10% higher than the number in 1974. While the 1976 deviation may be due to sampling variability, we believe that further updating of information will probably raise the 1976 figure relative to the 1975 figure. We suspect that the 1974 figure is more affected than the one for 1975 by exclusion of the institutionalized from the CPS, since some who were not institutionalized in March, 1973 became so subsequently. The 1977 figures are obviously incomplete.

Table 2 contains the breakdown by race, sex, and age of post-March 1973 mortality rates. As should be expected: blacks have a higher age-specific death rate than whites for each sex except for males 65+ and females 35-44; death rates rise with age; and white males have a lower life expectancy than white females. Given the small numbers of nonwhites in the sample who died, we concentrate on whites in this study to facilitate further comparisons. Also, to provide comparability with KH, we exclude those less than 25 years old.

To obtain the comparable Vital Statistics rates applicable to the sample, we calculate the expected death rate for 1973 through 1977 for each person of a given age level in 1973. We then calculate an average for each age group. However, since Vital Statistics does not publish single year death rates for those over 65, we fit and extrapolate the equation $\log Y = a+b(\text{age})$ where Y is the Vital Statistics death rate for those between 60 and 65.

In the last two columns, we present the ratio of the sample death rate to that of the weighted Vital Statistics. For white males and females under 65, the ratios are about 70% and 50% respectively. In the 65+ group, 74% of white females and 94% of white males are recorded although this comparison is less informative since we estimate the Vital Statistics rate for this group.

The comparisons in Table 2 suggest that attention be focussed only on white males. For them, there seemed little reason to suspect that this small SSA undercount is correlated with education and most other independent variables to be used in the present study.

3. RESULTS

We have calculated tables, available upon request, which classify death rates by a variety of sociodemographic and economic variables for different age groups. The results are very similar to those obtained from ordinary least squares regressions whose results are summarized in Table 3 in the 1973-1977 columns. The least squares regressions are also very similar to logit estimates when these are evaluated at the cell means. (The 1960 columns are discussed later.)

We have included the expected death rate variables described above. We have also included its square because the extrapolated variable is not appropriate above age 85 and because the expected death rate variable does not take account of correlations with education and other independent variables.

Consider first the results by education level in the top panel of Table 3. While the pattern is not without anomalies, death rates tend to decrease with education and the least educated have a 40 to 60% higher death rate than the most educated. These differences are statistically significant at the 5% level [10].

There are several possible reasons why education can be negatively related to death rates. First, the better educated have higher income and the income elasticity of health care undoubtedly is positive. Second, the better educated work in different types of occupations which may expose

Table 2.--Post-March 1973 Death Rates for Persons 17 Years of Age or Older by Age, Race and Sex

Age Groups (in years)		CPS-SSA Exact Match				Weighted Vital Statistics		Ratio SSA to Vital Statistics	
		White Males	White Females	Black Males	Black Females	White Males	White Females	White Males	White Females
Persons 17-24 years old	# Died	29	6	4	1				
	# in group	6730	6949	814	981				
	Death Rate/100	.43	.086	.49	.10	.67	.22	.64	.39
Persons 25-34 years old	# Died	37	12	7	3				
	# in group	7132	7410	627	863				
	Death Rate/100	.52	.16	1.1	.35	.60	.50	.87	.32
Persons 35-44 years old	# Died	50	19	12	2				
	# in group	5718	5995	517	687				
	Death Rate/100	.87	.32	2.3	.29	1.2	.675	.72	.47
Persons 45-54 years old	# Died	127	65	17	9				
	# in group	6243	6453	521	680				
	Death Rate/100	2.03	1.0	3.26	1.32	3.14	1.7	.65	.59
Persons 55-64 years old	# Died	299	130	35	16				
	# in group	5099	5413	410	516				
	Death Rate/100	5.86	2.4	8.54	3.1	7.67	3.7	.76	.65
Persons 65+ years old	# Died	1140	724	87	67				
	# in group	5031	6373	401	527				
	Death Rate/100	22.66	11.36	21.69	12.71	24.01	15.40	.94	.74

Excludes 132 decedents under age 17 or of races other than white or black.

them to lower risk of accidents and of health hazards whose effects accumulate over time. Third, the better educated may be better able to follow their doctors' advice or may be more efficient in producing healthiness from given resources.

Turning to the next panel, we find that for white males age-specific death rates initially decrease sharply with 1972 family income and at some point begin to rise. (The earnings levels are chosen to correspond to those used by KH.) Several of the coefficients are statistically significant at the 5% level. Two explanations for the negative relationship are available: those with higher earnings can buy more and better health care and those who are seriously ill are forced to reduce hours of work.

Death rates by marital status are presented in the bottom panel. For males less than 65, those married have the lowest rate. Single persons have a 13% greater rate, even though there is less reason for SSA to learn about their deaths. Widowers have a rate more than twice as high as married men, while for divorced men the death rate is 90% larger. In the 65+ age group there is the same rank ordering as for the younger group except that single men do a bit better than married men; however, the qualification about reporting bias still applies. Most of the coefficients are statistically significant at the 5% level.

The possible explanations for the above results, which have been found by others, include the following: (1) married persons benefit from health care and emotional stability provided by the spouse; (2) married persons lead a less dissolute life; (3) those who remain single, widowed, or divorced contain a disproportionately large number of seriously ill people who do not remarry because of the illness; (4) widows and widowers suffer from grief; (5) widows and widowers shared a common environment.

4. COMPARISONS WITH KITAGAWA AND HAUSER

At this point, we wish to make a comparison with the results KH obtained from the 1960 Census. It should be noted that KH use a slightly different method to control for age. However, our results are not changed if we use their method.

The 1960 columns in Table 3 contain KH's results. Consider first the results for education. For white males, KH find that there is a strong negative effect of schooling on mortality for those less than 65. Those who only went to elementary school have a relative death rate 40 percentage points higher than those who completed college. For men older than 64, however, they find no effects of education. Our ratios are roughly comparable to theirs. For those over 64, however, we find that the effects of education are about as strong as for the younger men. This is in

marked contrast to KH.

The second panel indicates the effects of family income where the divisions in income are based upon 1959 constant dollar levels, adjusted for nationwide productivity growth between 1959 and 1972. For younger men, in both years there is a strong negative relationship between family income and death rates. For the older men, both studies hint at a common U-shape pattern and again the death rate is highest for the lowest income group in 1973.

Table 3.--Relative, Age Adjusted, Death Rates by Education, Earnings and Marital Status: 1960, 1973-1976, White Males

Item	25-64		65+	
	1960	1973-77	1960	1973-77
Education[a/]				
0-8 years	1.11	1.24	1.01	1.06
9-11	1.03	1.34	.99	.93
12	.91	.99		1.06
13-15	.85	1.13	.98	.77
16+	.70	.84		.77
Family income in 1959 $ [d/]				
<2,000	1.51	2.65	1.10	1.22
2,000-3,999	1.20	1.3	.99	.95
4,000-5,999	.99	1.05	.92	.89
6,000-7,999	.88	.9		1.09
8,000-9,999	.93	.8	.96	.93
10,000+	.84	.95		1.90
Marital status[b/]				
Single	1.75	1.13	1.44	.95
Married	1.00	1.00	1.00	1.00
Widower	1.78	2.84	1.33	1.44
Other [c/]	2.30	1.86	1.33	1.49

a/ Relative to average death rate.
b/ Relative to average death rate for married.
c/ Mostly divorced.
d/ Inflated by the CPI and by the growth in real per capita income from 1959-1972.

Source: 1960 is KH, pp. 12, 109. 1973 is CPS-SSA.

In both samples, married men have the lowest death rate for those less than 65. KH also find this for men 65+, but we do not. However, in our sample, death rates for single men who do not collect any social security benefits (including Medicare) may be underestimated. A comparison of our results with KH's results suggests that by 1973, the differential for widowed and divorced people has grown.

The major puzzle that emerges from these comparisons is the differential effects of education for older men. KH find no effects and we find marked effects. Our results may be biased because of the exclusion of the permanently institutionalized from the CPS. However, exclusion of the institutionalized from KH's data has little effect on their results and we do not think we missed enough deaths to affect our results greatly.

The differences in the two sets of findings may be due to a variety of substantive factors. First, between 1960 and 1973 there were marked advances in medicine that could prevent, cure, or retard the progress of fatal diseases. For example, polio vaccines and kidney dialysis machines were introduced. These innovations extend life expectancy in general but may have been more beneficial to people in the more educated group. Without more detailed knowledge of the medical advances and correlations of particular diseases and education, we cannot determine how important or valid this explanation is.

Second, Medicare and Medicaid were introduced midway between 1960 and 1973. Reducing the price and increasing the use of medical facilities can lead to longer life expectancies. However, these changes suggest that those with low education, who include a disproportionate share of the poor, must have received a major increase in the quality and quantity of medical resources.

Between 1960 and 1973, the food stamp program was introduced and social security benefits were increased at a rate faster than inflation. But, once again, it appears that these programs conferred more benefits on those with less education. (The changes in these programs may explain why single men 65+ do relatively better in the mid-1970's.)

Third, the passage of thirteen years also affected the sample composition. Those in the older age group in 1960 must have been born before 1890 and for the most part have finished their formal education by 1915. Most of their time in the labor force would have been prior to 1960. The group we studied would have been educated and begun their labor force experience some 13 years later. While the content of schooling changes over time, and while Taubman and Wales [11] have shown that the relationship of years of schooling to IQ changed noticeably over the relevant birth cohorts, it was true for both groups that, on average, the more educated were more intelligent. Thus, it is difficult to explain the KH results if the effect of education on death rates occurs because the more educated know how to take better care of themselves.

We have come up with a possible explanation though we do not have data with which to test it. The annual death rates for those under 65 are relatively small. Many of these deaths may be related to a person's occupation because of: accidents; exposure to chemicals and other dangerous elements; or stress. In a modern, industrialized economy, the more educated generally have different jobs and may be less exposed to health hazards and perhaps to stress. This type of explanation can explain the similarity in findings for those under 65.

If exposure to occupational hazards has a long-

term cumulative effect, it is also possible to explain the shift in results for those over 64. During the 13 year interval between the 1960 cohort of older men used by KH and the 1973 cohort used by us, the structure of the American economy switched, becoming more industrial and less agricultural. Education and occupation became more tightly tied together. Thus, those over 64 who had low education were much more likely to be exposed to industrial health hazards in 1973 than in 1960. Let us caution, however, that much work remains to be done before this hypothesis can be considered as proven.

Indeed, since writing the above paragraph, we have examined the relationship of death rates to and education for those 78 and older in 1973, the survivors of KH's 65+ group. For these men, we obtain the same coefficients on education as for our entire 65+ group. Thus, education is now important for a cohort for which it was presumably unimportant in 1960. This does not seem to support the health hazards from work idea.

We have calculated multiple regressions using all the variables described above, measures of health, labor force participation, region, veteran status, etc. (The tables are available upon request.) Briefly, we find that the coefficients on marital status are fairly robust to alternative specifications but that the education and family income effects are decreased in the multiple regressions. We believe that these results indicate that people who are very ill participate less and thus have lower income, and that the more educated tend to live and work in healthier environments.

5. CONCLUSION

In this paper, we have proposed and used a new type of sample to examine socioeconomic determinants of age-specific mortality. The new methodology appears quite promising, especially for males. The decision to record date of death for the uninsured should improve the usefulness of the technique for females.

We have found noticeable effects on mortality of education, income, marital status, and proxies for health. The marital status effects are robust to changes in specification. The effects of education and current income are diminished sharply when health measures are included. We think this represents two different phenomena. In one case, low education causes poor health; in the other case, poor health causes a reduction in labor force participation and increases the chances of dying.

A comparison with Kitagawa and Hauser's study based on the 1960 Census shows some similarities and some changes on the 13-year interval. The most important difference is that they find no effects of education for white males 65+ in 1960, but we find strong effects in 1973-77. Some possible explanations are advanced to explain this shift.

ACKNOWLEDGEMENTS

This research was supported by NIA Grant No. 5 ROL AG00913-02 POP and the Henry J. Kaiser Family Foundation grant to the National Bureau of Economic Research (NBER).

NOTES AND REFERENCES

[1] There are some studies based on census tract or other geographic units, but these studies have relatively little data and analysis is complicated because of colinearity in the grouped data. For a recent survey of empirical results, see Brenner, H., Estimating the Social Choice of National Economic Policy: Implications for Mental and Physical Health, and Criminal Aggression, a study prepared for the JEC, Washington, 1976.

[2] There are studies for specific causes of death. There are studies that group people by family and marital status. There is also an interesting study by Grossman; however, his sample only includes white males 32-54 years old who are in the top half of the IQ distribution. (See Grossman, M., The Correlation Between Health and Schooling, N.E. Terleckyj (ed.) Household Production and Consumption, New York, Columbia University Press, 1975.)

[3] Kitagawa, E. and P. Hauser, Differential Mortality in the U.S., Harvard University Press, 1973.

[4] Kitagawa and Hauser, p. 318. They exclude a random four-fifths of whites aged 75 and over. They then use appropriate weights in their analyses. They were unable to match about 20% of the death certificates to the complete Census enumeration. Also, about 3% of the people in the complete enumeration, who should have been in the 25% sample, could not be located.

[5] Irelan, L.M., et al, Almost 65: Baseline Data from the Retirement History Study, Office of Research and Statistics, Social Security Administration, Research Report No. 49, 1976. See also, Rabin, William, Longitudinal Retirement History Study: Description of Data Systems, Unpublished SSA Working Note, November, 1977.

[6] Kilss, B. and F. Scheuren, The 1973 CPS-IRS-SSA Exact Match Study, in Social Security Bulletin, Vol. 41, No. 10, October, 1978.

[7] Alvey, W. and Aziz, F., Quality of Mortality Reporting in SSA Linked Data: Some Preliminary Descriptive Results, 1979 Proceedings American Statistical Association Section on Survey Research Methods.

[8] Scheuren, F. and H. L. Oh, Fiddling Around with Non-Matches and Mis-matches, 1975 Proceedings American Statistical Association, Social Statistics Section, pp. 627-633.

[9] Oh, H.L., Scheuren, F. and Vogel, L., Methods of Estimation for the 1973 Exact Match Study, Studies from Interagency Data Linkages, Report No. 10 (forthcoming). There is a slight tendency for those with higher incomes to be overrepresented. In the basic Exact Match File weighting adjustments were developed to "correct" this. For simplicity, however, the present analyses were unweighted and hence may be slightly biased because of this overrepresentation.

[10] This and all other significance statements in this paper refer to nominal levels only; they were calculated under the assumption that the matched dataset is a simple random sample. For details in the actual 1973 CPS sample design, see [9].

[11] Taubman, P. and T.J. Wales, Mental Ability and Higher Educational Attainment in the Twentieth Century, NBER and Carnegie Commission, McGraw Hill, 1972.

[12] The basic data bases employed in this paper are documented in--

Aziz, F., Kilss, B. and Scheuren, F., 1973 Current Population Survey--Administrative Record Exact Match File Codebook, Studies from Interagency Data Linkages, Report No. 8, 1978.

DelBene, L., 1937-1976 Social Security Longitudinal Earnings Exact Match File, Chapter 1, Studies from Interagency Data Linkages, Report No. 9, 1979.