GENERAL MORTALITY AMONG SELECTED CENSUS BUREAU SAMPLE COHORTS 1979-1981

NORMAN J. JOHNSON, BUREAU OF THE CENSUS
EUGENE ROGOT, NATIONAL HEART, LUNG, AND BLOOD INSTITUTE
CLAUDIA SANCHEZ GLOVER, BUREAU OF THE CENSUS
PAUL SORLIE, NATIONAL HEART, LUNG, AND BLOOD INSTITUTE
MARILYN MCMILLEN, NATIONAL CENTER FOR HEALTH STATISTICS

INTRODUCTION

The National Longitudinal Mortality Study (NLMS) is a long term collaborative study being conducted by the National Heart, Lung, and Blood Institute (NHLBI), the National Center for Health Statistics (NCHS) and the Bureau of the Census [Makuc et al, 1984 (1)]. Primary objectives of the NLMS are to analyze socioeconomic, demographic and occupational differentials in mortality within the United States. The study population consists of eight cohorts of selected Census samples. Deaths in this population are identified through periodic matching to the National Death Index (NDI), a file maintained by NCHS containing death certificate information for all deaths in the U.S. starting in 1979. The NLMS data base containing socioeconomic, demographic and mortality information is being created by linking Census records to the death certificate information using carefully controlled procedures to maintain the confidentiality of individual records.

In this paper we compare the mortality experience observed in the National Longitudinal Mortality Study with the U.S. experience for the period 1979-1981. Each of the cohorts consists of either one month of a Current Population Survey or a sample of the 1980 Decennial Census. Current Population Surveys are random samples of the noninstitutional population of the U.S. The Decennial Census sample is considered to be a random sample of the U.S. population as collected in the Census. Our purpose here is to determine how consistent our findings are with comparable results established for the U.S. and to determine the effects of any biases which may exist in our results due to the nature of the cohorts selected. The project is ongoing; therefore, some results presented here will be derived only for a subset of the eight cohorts.

BACKGROUND

Data for measuring socioeconomic differentials in mortality are not routinely available in the United States because death certificates, the primary source of national mortality data, do not generally contain information on educational attainment or income. Although the usual occupation of the decedent is collected on the death certificate, this item is not yet coded by all states. Further, Census population data needed as the denominator for computing occupation specific death rates are based on current occupation rather than usual occupation. Studies using record linkage, such as the NLMS, provide an approach to overcoming these problems in analyzing national socioeconomic differentials in mortality.

THE NATIONAL DEATH INDEX

A major impetus in the development of the NLMS was the relatively recent creation of the National Death Index (NDI). This index was designed to provide a simplified method of identifying deaths among participants in prospective studies. It is a centralized computerized index of death records in the U.S. beginning with deaths in 1979. The NCHS maintains the NDI using data supplied by state vital statistics offices. Users of the NDI provide a file with up to 14 identifier items on each person. Records are matched when they agree on any one of seven criteria established by the NCHS. The output from the NDI match is a file containing possible matches. Users must then decide which of these matches are true positive and which are false positive.

THE CENSUS SAMPLES

The eight cohorts which comprise the National Longitudinal Mortality Study consist of seven Current Population Surveys (CPS) from the period from 1973 to 1981, and the April 1980 Census Enumeration sample (E). The cohorts are identified in Table 1.

The CPS is a survey conducted by the Bureau of the Census consisting of a sample of households from the civilian noninstitutional population of the United States. The Current Population Survey was selected because it is an ongoing national survey that provides data on a variety of socioeconomic and demographic factors and it contains data necessary for record linkage with the NDI. The primary purpose of the CPS is to provide estimates of monthly labor force participation. As many as 64,000 households are interviewed in some months. Certain geographic areas are oversampled to provide reliable estimates for the nation, each state and certain large SMSA's. The CPS includes persons of all ages, but most questions are asked of persons 14 years of age and older.

Cohort E is the April 1980 Enumeration sample, a national sample of housing units enumerated in the 1980 Census of Population and Housing. The Enumeration sample (E-sample) is part of the Post Enumeration Program (PEP) to measure the

TABLE 1. Cohort Identification and Overall Mortality Probability

COHORT	SAMPLE DATE	SAMPLE	STARTING DATE	YEARS FOLLOWUP ‡ 79	80	81	FACTOR	N	D	q	SPECIAL CHARACTERISTICS
A	03/73	CPS	MONDAY JAN 1,1979	1	1	1	3	130186	3000	7.68	AGES 6+
B	02/78	CPS	MONDAY JAN 1,1979	1	1	1	3	92799	2540	9.12	AGES 14+
C	03/79	CPS	SUNDAY MAR 18,1979	0.79	1	1	2.79	41605	712	6.13	
D	04/80	CPS	SUNDAY APR 13,1980		0.72	1	1.72	182632	2102	6.69	
E	04/80	E	TUESDAY APR 1,1980		0.75	1	1.75	123965	1506	6.94	INCLUDES INST, PERSONS W SSN'S
F	08/80	CPS	SUNDAY AUG 17,1980		0.37	1	1.37	181039	1653	6.66	
G	12/80	CPS	SUNDAY DEC 14,1980		0.05	1	1.05	174886	1165	6.34	
H	03/81	CPS	SUNDAY MAR 15,1981			0.8	0.8	60139	297	6.17	
								987251	12975		

N = Cohort Size
D = Number of Observed Deaths
q = crude mortality rate per 1000/year
‡ Fraction of year containing mortality follow up

undercount of the 1980 Decennial Census. The E-sample differs from CPS samples in that its sampling frame is the resident population of the U.S. at the time of the Census. It is not limited to the civilian noninstitutionalized population. In addition, data items available for the E-sample differ from those of the CPS, and only those cases with Social Security Number were included in the NLMS sample (roughly half of the file).

LINKING CENSUS AND MORTALITY DATA

Census records and mortality data were matched using a two step procedure. The first step, performed at the National Center for Health Statistics, consisted of matching NLMS records to the National Death Index using their seven criteria matching algorithm. The results of this match were reviewed in a second step at the Census Bureau using a modified Newcombe model approach to record linkage in order to screen out false positive matches occurring in the NCHS match. Documentation of the probabilistic method used for this study is given in reference (2), with related material found in (3) and (4).

OBJECTIVES AND METHODOLOGY

The objectives of this paper are basically twofold. First, we wish to compare the NLMS results to those for the U.S., and second, we wish to present new data on characteristics of the decedents. At this time we are able to present only limited data, since the eight Census base files, needed as denominators for rates, have not been completely assembled and the deaths occurring in the first follow-up period, needed for numerators of rates, have not been linked with appropriate records on the base files. For this reason some of our results will be for cohorts D and F and not all cohorts combined.

Some material, to be presented, will be given in terms of crude death rates. Other material will be presented using age-standardized mortality ratios (SMR's). SMR's were derived by an indirect age-adjustment using the U.S. 1980 age-specific rates as standard rates. Five age groups were used: < 15, 15-44, 45-64, 65-74 and 75+. The standard rates were applied to the population at risk in each age group to obtain expected numbers of deaths. The SMR was then calculated as the number of observed deaths divided by the expected number. The SMR for the U.S. is set at 100.

RESULTS

A brief summary of results of the matching procedure for the NLMS is outlined in Table 1. As indicated in Table 1, out of approximately 1 million records in the eight cohorts under study, 12,975 deaths were identified in the matching process. Special characteristics of three of the cohorts have been identified in the table. Cohorts A and B have age restrictions. Cohort E includes persons in institutions and consists only of the subset of persons having social security numbers from the complete PEP sample. The other cohorts include persons of all ages not in institutions. The two cohorts with age restrictions differ further from the other files in that deaths occurring before January 1, 1979 can not be identified from the National Death Index. The observed death rate for the three cohorts identified with special characteristics is higher than that of the other five cohorts due in part to these attributes. The death rates for the remaining five cohorts are remarkably consistent reflecting the fact that they are random samples of essentially the same population. In the remainder of this paper we compare the mortality experience in the NLMS to the U.S. experience and discuss some characteristics of the decedents in our study.

AGE, SEX AND RACE COMPARISONS

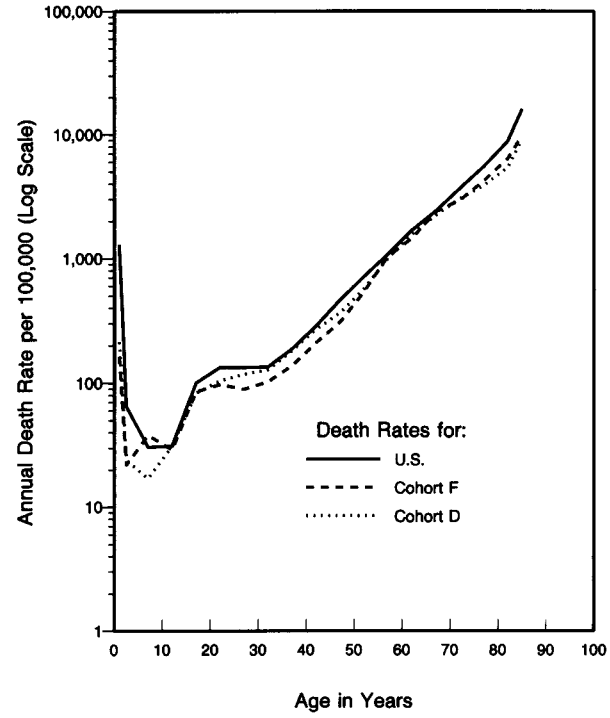
A comparison of the age, sex and race

specific mortality experience of the National Longitudinal Mortality Study with the U.S. experience will be made on two of the cohorts, cohorts D and F. In Figure 1, the U.S. annual death rate per 100,000 is plotted on a log scale for all sex and race groups combined by 5-year age categories up to 85+. The curve is compared to similar curves for the D and F cohorts. The U.S. rates have been taken from published NCHS data (5). The plots show the similarity in the death rates for the two NLMS cohorts. The greatest discrepancy between the U.S. rates and the NLMS rates occurs in the early ages particularly in the less-than-one age group. The CPS does not include children who die in the first day of life. According to national records, about 40 percent of deaths in the first year of life occur within the first 24 hours after birth. In addition, children may not have all variables needed for matching to the NDI, such as social security number. There is a noticeably lower rate for the two NLMS cohorts compared to the U.S. for ages between 20 and 35 and for ages greater than 65. The latter occurs mainly because the CPS does not include institutionalized persons, such as those in nursing homes. Many older persons are institutionalized before they die. However, persons who enter an institution after the CPS survey and subsequently die there are included in the NLMS. Also, there is some ascertainment loss, of perhaps 5 percent, occurring in the matching process due to recording errors which occur in the files being matched. In spite of these noted exceptions, the graphs show that the mortality rates of the NLMS cohorts are quite similar to the U.S. rates.

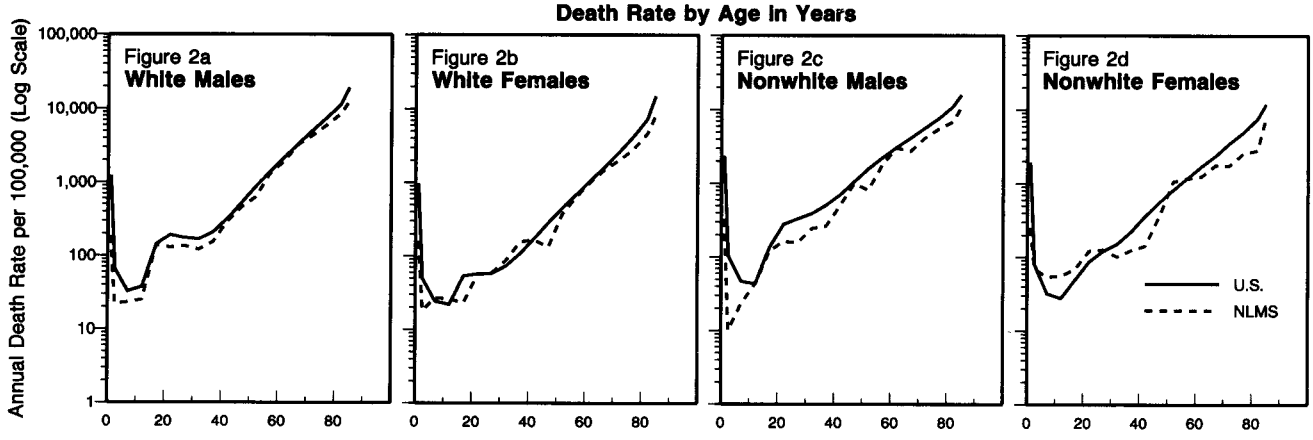
Similar graphs to that in Figure 1 are given in Figure 2 for each sex-race category. Because of their similar nature and close beginning date cohorts D and F have been combined in these graphs and the estimated death rate computed by averaging the two corresponding rates.

The rates for the combined NLMS cohort approximates the corresponding curve for the U.S. quite well for white males and females with the same exceptions as noted for the data in Figure 1. Greater differences between the combined cohorts D and F and the U.S. curves can be seen for nonwhite males and nonwhite females. The large differences for nonwhites may be due to sampling error because of the smaller number of deaths observed for nonwhites than for whites. The noticeable difference in the death rates between the U.S. and the combined NLMS cohort for the 1-4 age group is not fully understood.

Figure 1
Annual Death Rates for U.S. and Two NLMS Cohorts



Death Rate by Age in Years



SEASONAL VARIATION

The graph in Figure 3 shows the crude death rate each month on an annual basis for the U.S. and for seven of the eight NLMS cohorts combined. The death rate for the NLMS was based on combining all persons alive at the beginning of the month in all cohorts except cohort E. Cohort E was not included in this graph because of the potential biases which might exist due to the requirement of a Social Security Number to be in the cohort. It was not always possible to collect this information for persons dying between April 1, 1980 and late 1980 or early 1981, especially for persons living alone. Therefore, the observed death rate for persons in Cohort E is unrealistically low. The graph covers the three years 1979-1981. Adjustments have been made to cohort A to account for the number of unobserved deaths which occurred before the beginning of followup by the National Death Index. As shown in Figure 3, the seasonal pattern for the NLMS resembles the U.S. pattern. We expect that the NLMS would be consistentlv below the U.S. rate due to the exclusion of the institutional populations. We believe this difference to be about 20 percent. In addition, there may be a loss of perhaps 5 percent in ascertainment of death resulting from the match to the National Death Index. This difference is observed in the graph except for a three month period in early 1979 and a five month period in early 1980. The high rate observed for the NLMS in the first three month period may occur because only cohorts A and B are used. These two cohorts consist of older persons then the other cohorts.

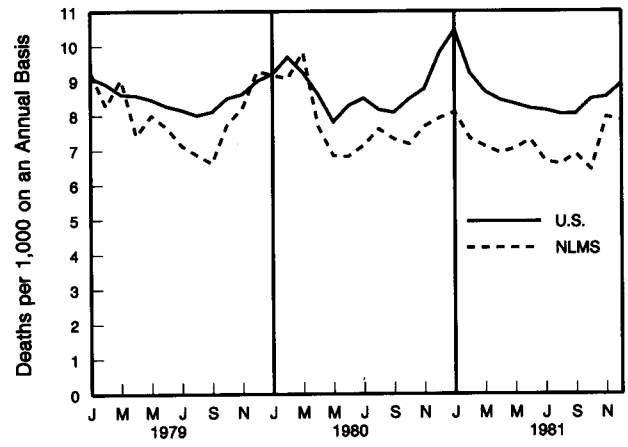
GEOGRAPHICAL COMPARISONS

Figure 4a and 4b are maps of the U.S. showing the distribution of SMR's, shaded by quartiles, for the U.S. and for cohorts D and F combined. The correlation across states between the U.S. rates and the combined cohort D&F is .37, significantly different from 0 at the .05 level. A possible explanation for this low correlation might be that state of residence for the U.S. is taken from the death certificate where as in the NLMS it is taken from the CPS record at the starting date for the cohort. Thus, moving that has taken place before death would lower the expected correlation computed here. The followup period is, however, a short one.

FOREIGN BORN

The data in Table 2. compares the SMR's for the U.S. with the SMR's for the combined cohort D and F for native born versus foreign born by sex. The 1980 U.S. age-specific rates are used to calculate expected values. The most noticeable difference is the lower ratio for foreign born males compared to native born. There

Figure 3.
Seasonal Variation of U.S. and NLMS Death Rate



is no similar effect for females. We have no explanation for this occurrence and we hope that this effect will be more fully understood with further study. When all files have been completed, we expect to be able to study the foreign born data by specific country of birth for four of the cohorts.

Figure 4a.
SMR Quartiles by State: U.S., 1980

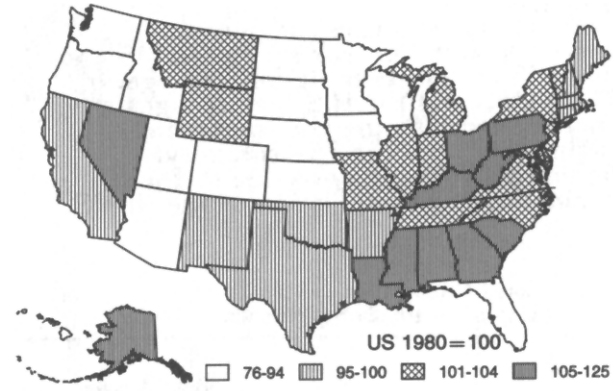


Figure 4b.
SMR Quartiles by State: NLMS Combined D and F, 1980-1981

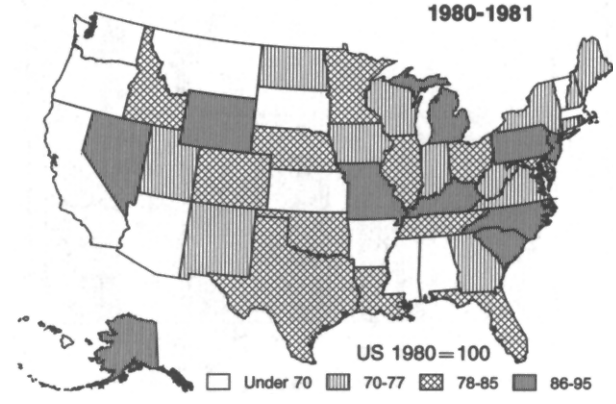


TABLE 2. Observed (O) and Expected (E) Deaths by Sex and Nativity: NLMS Cohorts D and F combined, 1980-1981

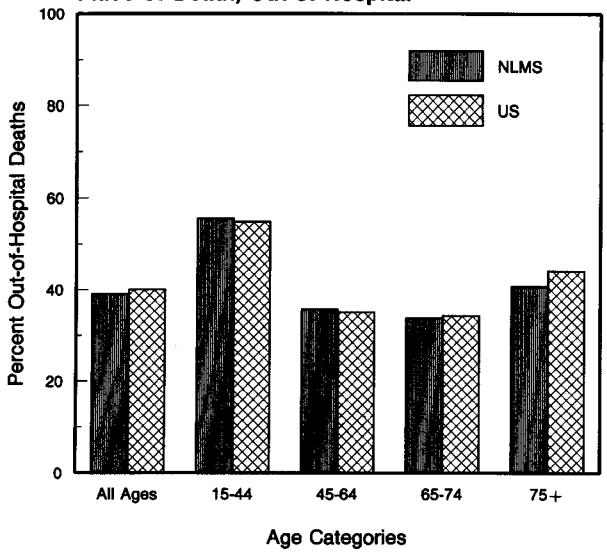
	NATIVE BORN	FOREIGN BORN	UNKNOWN	D&F TOTAL	U.S. 1980
MALES					
O	2017	167	3	2187	
E	1812.22	191.91	67.44	2071.57	
O/E	111	87		106	130
FEMALES					
O	1404	160	4	1568	
E	2423.63	278.76	93.77	2796.16	
O/E	58	57		56	79
TOTAL					
O	3421	327	7	3755	
E	4235.85	470.67	161.21	4867.73	
O/E	81	69		77	100

PLACE OF DEATH

Using information recorded on the death certificate, decedents can be characterized by place of death. Adequate information on place of death was available from 39 states in 1979 and from 40 states in 1980 and 1981. Thus, the total number of deaths characterized by place of death is 10109, which is less than the total of 12,975 deaths observed in the National Longitudinal Mortality Study.

Bar charts showing the percent out-of-hospital deaths for the NLMS and the U.S. are given in Figure 5 by age. The NLMS and the U.S. percentages are very similar with the U.S. percentage slightly larger in the 75+ category. Again this effect may be due to the noninstitutional population of the NLMS.

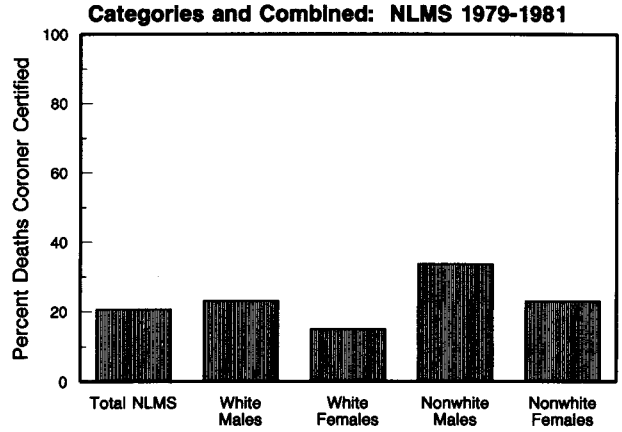
Figure 5.
Place of Death, Out-of-Hospital



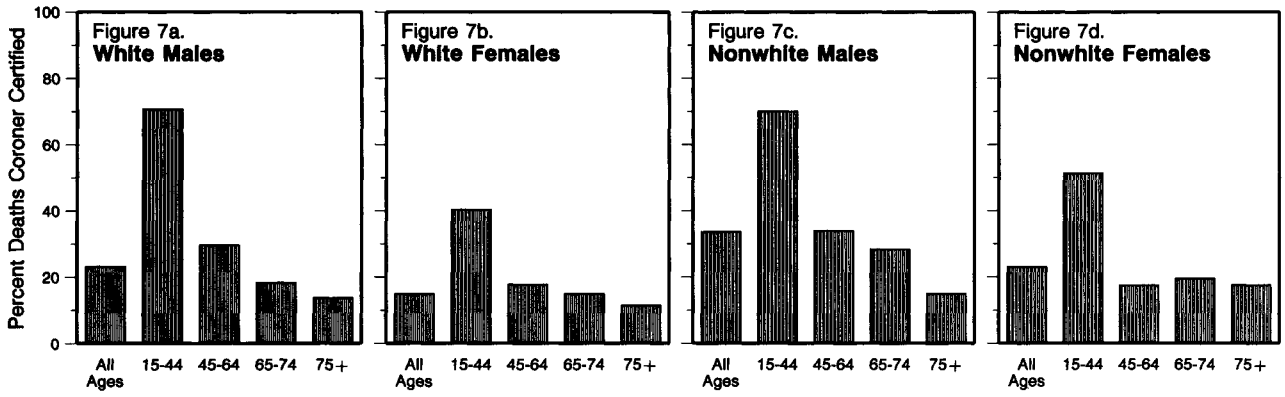
DEATHS CERTIFIED BY CORONERS OR MEDICAL EXAMINERS

Causes of death on certificates are provided by either a medical examiner, a coroner or an attending physician. Medico-legal deaths, those certified by a medical examiner or a coroner, comprise a unique group in terms of medical documentation and description of cause of death. While the medico-legal examiner's function is to rule out foul play, the coroner seldom performs the necessary procedures to determine the particular natural cause of death. Thus, natural causes of death with a high frequency of coroner certification may be subject to more uncertainties in diagnosis than those diseases that are rarely certified by coroners. For the purposes of this presentation, the medico-legal certifiers as a group will be designated as coroners. There are no published data on the percent of coroner certified deaths for the U.S. Bar charts showing the percent coroner certified deaths in the NLMS, are given in Figures 6-7d. Only a small percentage of death certificates were lacking this information, mainly those from Alaska. In our study, about 20 percent of all deaths were certified by a coroner and 80 percent were certified by an attending physician. A separate bar chart is given for all categories combined and for race-sex categories in Figure 6. The distribution for each

Figure 6.
Coroner Certification Sex-Race Categories and Combined: NLMS 1979-1981



race-sex group is further broken down by age categories in Figures 7a-7d. As a group, nonwhites have a greater percentage of coroner deaths than whites. From Figures 7a-7d we see that there is generally a decrease in the percentage of coroner deaths as age increases. The largest proportion of coroner deaths occurs in the 15-44 age group. For this group, the number of coroner certified deaths is more than twice that of non-coroner deaths and men in this age



group have a greater percentage of coroner certified deaths than non-coroner certified deaths. Nonwhite females have about 50 percent coroner certified deaths in this age group. Information on specific causes of death in these ages for the different sex-race groups may explain why the high percentage of coroner certified deaths is being observed. This information will be presented in a paper to follow (6).

SUMMARY

In this paper we have presented original findings from the National Longitudinal Mortality Study and compared our results to U.S. figures when these were available. We have attempted in this way to establish the results of the mortality followup for 1979-1981 for the National Longitudinal Mortality Study as reasonable and generally consistent with U.S. vital statistics. The findings attest to the quality of results obtained from the probability samples of the CPS and show that for a variety of demographic variables, NLMS results are quite representative of results for the entire U.S. We have observed the effect of the noninstitutional composition of NLMS cohorts compared to the U.S. population. We have also observed that due to the definition of eligibility for CPS interview and other reasons, rates for the less-than-one-year age group are much lower than expected and that rates for the age groups less than 15 are also low. Rates in this paper were frequently shown for two of the cohorts combined. Because of the similar nature of the cohorts, pooling of the results was reasonable. We intend to study the possibility of pooling additional cohorts in order to improve the reliability of our results. We expect, with the completion and pooling of other cohort data files, to be able to develop further the results presented here.

REFERENCES

1. Makuc, D., McMillen, M., Feinleib, M., McMillen, D., Schwartz, S. and Rogot, E. An Overview of the U.S. National Longitudinal Mortality Study. 1984 Proceedings of the Section on Social Statistics, American Statistical Association.

2. Rogot, E., Sorlie, P. and Johnson, N. Probabilistic Methods in Matching Census Samples to the National Death Index. Submitted to The Journal of Chronic Diseases.

3. Rogot, E., Feinleib, M., Ockay, K., Schwartz, S., Bilgrad, R. and Patterson, J. On the Feasibility of Linking Census Samples to the National Death Index for Epidemiologic Studies: A Progress Report. American Journal of Public Health. November, 1983, Vol. 73, No. 11.

4. Rogot, E., Schwartz, S., O'Conner, K. and Olsen, C.L. The Use of Probabilistic Methods in Matching Census Samples to the National Death Index. 1983 Proceedings of the Section on Survey Research Methods, American Statistical Association, 319-324.

5. National Center for Health Statistics: Advance Report, Final Monthly Report, 1980. Monthly Vital Statistics Report, Vol.32-No. 4, Supp. DHHS Pub No.(PHS) 83-1120. Public Health Service. Hyattsville, Md., August 1983.

6. Rogot, E., Sorlie, P., Johnson, N., Glover, C. and Makuc, D. Mortality by Cause of Death Among Selected Census Bureau Sample Cohorts for 1979-1981. 1985 Proceedings of the Section on Survey Research Methods, American Statistical Association.