VARIATION IN LIFE EXPECTANCY DURING THE TWENTIETH CENTURY IN THE UNITED STATES*

DAVID W. SMITH AND BENJAMIN S. BRADSHAW

The National Center for Health Statistics (NCHS) reports life expectancy at birth (LE) for each year in the United States. Censal year estimates of LE use complete life tables. From 1900 through 1947, LEs for intercensal years were interpolated from decennial life tables and annual crude death rates. Since 1948, estimates have been computed from annual life tables. A substantial drop in variation in LE occurred in the 1940s. To evaluate these methods and examine variation without artifacts of different methods, we estimated a consistent series of both annual abridged life tables and LEs from official NCHS age-specific death rates and also LEs using the interpolation method for 1900–1998. Interpolated LEs are several times as variable as life table estimates, about 2 times as variable before 1940 and about 6.5 times as variable after 1950. Estimates of LE from annual life tables are better measures than those based on the mixed methods detailed in NCHS reports. Estimates from life tables show that the impact of the 1918 influenza pandemic on LE was much smaller than indicated by official statistics. We conclude that NCHS should report official estimates of intercensal LE for 1900–1948 computed from life tables in place of the existing LEs that were computed by interpolation.

Period life expectancy summarizes the current age-specific death rates for a population (Swanson and Siegel 2002: chap. 13). Life expectancies are primary public health measures that can be compared across nations, large subpopulations, or years; life expectancies have been used as indicators of both economic success and effectiveness of medical care (Bunker 2001; Bunker, Frazier, and Mosteller 1994; McKeown 1979; McKinlay and McKinlay 1977; Neison 1844, 1845; Sen 1998). Trends to longer life expectancy reflect the long-term transition from infectious and parasitic conditions to chronic and degenerative conditions as predominant causes of death (Omran 1971; Salomon and Murray 2002).

We examine patterns in annual variation in life expectancy at birth (hereafter, LE) over the twentieth century (1900–1998) reported by the National Center for Health Statistics (NCHS) for the U.S. Registration Area. The LEs have been published many different times; for an accessible copy, see Arias (2004:34, table 12). We measured variation for these LEs as the absolute values of the differences between two successive years (e.g., 1971 minus 1970). The differences for 1917–1918 and 1918–1919 were not included because of the unusually high death rates in 1918 associated with the influenza pandemic. The magnitude of the absolute differences dropped sharply and quickly in the 1940s and has remained low ever since. The medians for two eras, 1901–1939 and 1951–1998 show this change. For males the median was 1.60 years in the early era and 0.20 in the later era. For females the median dropped from 1.80 to 0.20 years. (The means, which are more sensitive to outliers, were somewhat larger but had a similar pattern.) Median absolute annual change was about 8 or 9 times as great before 1940 than after 1950. Such a large, rapid change from high variation to low variation calls for explanation.

Different methods of computing LE were used over the twentieth century by NCHS and its predecessor organizations, the National Office of Vital Statistics and the Census Bureau (Hetzel 1997): (1) life tables and LEs for decennial census years, (2) interpolated

*David W. Smith, University of Texas School of Public Health–San Antonio Campus, 8550 Datapoint, Suite 200, San Antonio, TX 78229-3440; E-Mail: smithd2@uthscsa.edu. Benjamin Bradshaw, Division of Management, Policy, and Community Health, University of Texas School of Public Health–San Antonio Campus. We appreciate the comments of Stephen Blanchard and of the reviewers and editors on our manuscript.

LEs for intercensal years before 1948 (Greville and Carlson 1951), and (3) life tables and LEs for intercensal years since 1948, with three variations (Anderson 1999; Greville and Carlson 1953; Sirken 1966). We describe these methods further later.

We compare two methods of estimating annual LE: from annual life tables and by interpolation from decennial tables. We compare results of these methods using annual variation in LEs in the U.S. registration area over the twentieth century. We also comment on the difference in estimates for 1918, when LEs were shortened by the influenza pandemic that year.

METHODS OF ESTIMATING OFFICIAL LIFE EXPECTANCY

The LEs published by the NCHS for census years were computed from complete life tables that used deaths in a three-year interval as numerators of age-specific death rates. Deaths were counted in the year before, the year of, and the year after a decennial census, except for the 1900 tables, which used deaths in 1900–1902.

LEs reported by the NCHS for years between censuses before 1948 were derived from an interpolation method developed by Greville and Carlson (1951) in which the crude death rate was multiplied by the LE for each decennial life table and values for years in between were interpolated using a straight line. They computed an interpolated LE for each year as the interpolated value for the year divided by the crude death rate for the year. They assigned the LE using the 1900 census and deaths in 1900–1902, extrapolating to 1901, and interpolating for 1902–1909. Although Greville and Carlson did not explicitly state how they generated LEs for 1946 and 1947, they appear to have extrapolated these values using life table estimates for 1940 and 1945.

This procedure was based on a property of a stationary population that the LE is the inverse of the crude death rate and their product is 1.0. More crucially, Greville and Carlson (1951:166) assumed that "While both the crude death rate and the reciprocal of the average length of life may be expected to fluctuate from year to year . . . , nevertheless, the ratio of the two quantities may reasonably be expected to change slowly and fairly smoothly, ..."

From 1948 through 1996, intercensal LEs were computed from abridged life tables using two similar methods (Greville 1953; Sirken 1966). Both methods used the number of deaths in the year as numerators of death rates and the estimated population at midyear as denominators for age intervals of five years (except less than 1 year and from 2 through 4 years). Both methods also used the most recent decennial table to estimate the lifetimes of those dying within each age interval. Since 1997, annual LEs have been computed without relying on the most recent decennial table (Anderson 1999).

METHODS

We obtained annual age- and sex-specific death rates for the United States for 1900–1998 from the NCHS (National Center for Heath Statistics n.d.). We obtained crude death rates for 1900–1990 from the same source and for 1991–2001 from Kochanek et al. (2004). Both sources use the age intervals of under 1, 1–4, 5–84 by 10-year intervals, and 85 and over.

We computed life tables and LEs for each calendar year. We used the same survival times within age intervals for those dying within the interval for each year: 0.1 year for those under age 1 year, 2 years for those aged 1 to 4 years, and 5 years for ages 5 through 84 years. To estimate remaining lifetime at age 85, we used the number living at age 85 divided by the death rate among those over 85. Because several of these decisions differ from the methods used by the NCHS, our estimates of LE will differ somewhat from the published values.

We used the LEs from decennial life tables published by NCHS and crude death rates from NCHS to compute LEs using the method of interpolation described by Greville and Carlson (1951) for each intercensal year from 1901 to 1999. Because decennial LEs have not been reported for 2000, we used the average LEs and average crude death rates reported

for 1999–2001 instead (from Kochanek et al. 2004) to interpolate values for 1991–1999. We assigned the LE for the 1900 census to 1900 to interpolate values for 1901–1909, though Greville and Carlson (1951) assigned the LE for the 1900 census to 1901.

We computed LEs separately for all males and for all females. Separate race-specific numbers or those for the total population were not necessary for our purpose.

We evaluated several methods to measure annual variation in LEs, including first differences, residuals from moving means, and residuals from lowess estimates as statistical models for LEs over time (Cleveland 1979). For equally spaced data, lowess estimates are equivalent to moving means, and we modeled LEs with moving means of seven observations using the following coefficients: –0.059, 0.057, 0.304, 0.397, 0.304, 0.057, and –0.059. (Other methods produced quite similar substantive results.) These values were derived from a weighted quadratic lowess regression, as were the coefficients for the following exceptions (Chambers et al. 1983:121–23). Because of the well-known increase in deaths with the influenza pandemic in 1918, we excluded this year from estimated moving averages. For 1915–1921, we modified the formulas to remove dependence on the observations for 1918 and used six years for moving means for each year. We computed estimates for three beginning and three ending years using asymmetric moving means. For example, estimates for the first and last years in each series used four observations, and estimates for the second and next-to-last years used five.

We computed residuals (observed values minus moving means) for each year except 1918. We report the absolute values of residuals graphically and their medians for two eras: 1903–1939 and 1951–1995. Computation of medians excluded three end points when the moving means were less reliable because they used fewer observations. Both eras exclude 1940–1950, the period when the apparent change in variation took place. The earlier era includes 3 decennial estimates and 34 for other years. The later era includes 4 decennial estimates and 41 for other years.

RESULTS

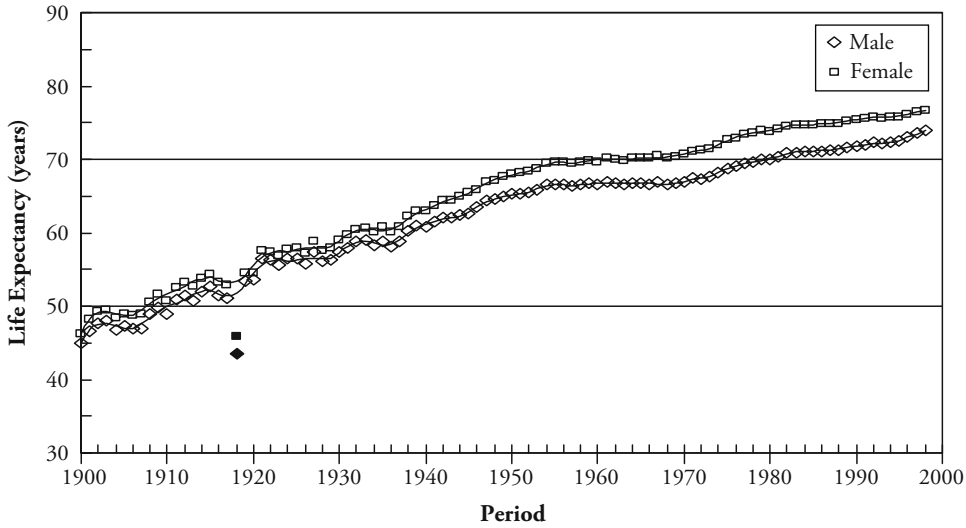
Annual LEs computed from life tables with seven-year moving mean estimates are shown in Figure 1 and in Appendix Table A1. LEs computed using interpolation are shown in Figure 2. The absolute residuals (absolute values of annual LE minus the estimate) for the former are shown in Figure 3, and those for the latter are shown in Figure 4. Medians of the absolute values of the residuals for each era (1903–1939 and 1951–1995) computed both ways and for estimates published by NCHS are shown in Table 1.

Interpolated estimates of LE are more variable than life table estimates in each era, as shown by the medians of the absolute residuals. Variation is greater in the earlier era than in the later era, by either method of estimation. In the earlier era, the medians of interpolation residuals are about 2.5 times those of life table residuals; during the later era, they are about 7 times as great. For estimates using life tables, the medians for the earlier era are about 6 times those of the later era. For interpolation estimates, the medians of the earlier era are 2 to 3 times as large as those of the later era.

The medians of the absolute residuals of the NCHS estimates for the earlier era are 11 to 12 times those of the later era. These ratios include a methodological component as well as temporal change.

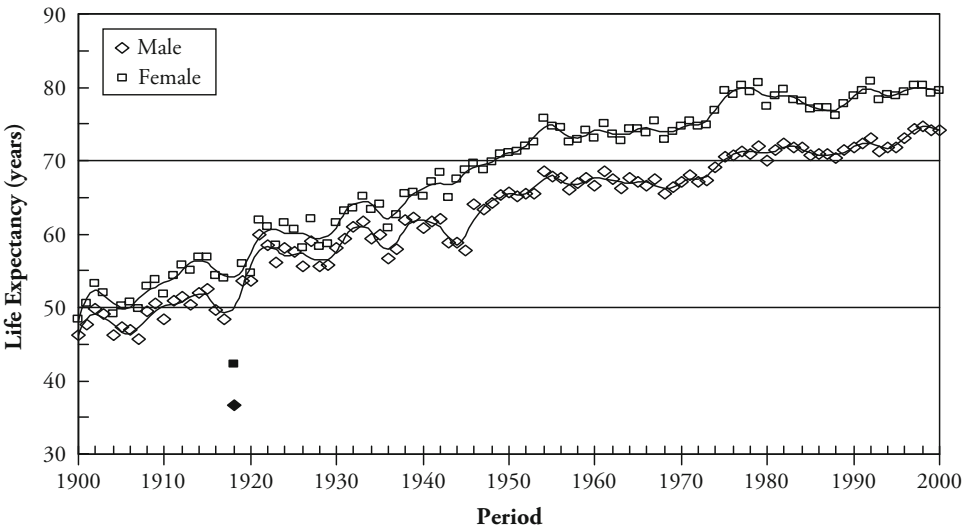
The most unusual year of the twentieth century for death rates and LE was 1918 due to the influenza pandemic. A “close-up” of the LEs from 1915 to 1920 is shown in Figure 5. Our life table estimates of LE that year are 48.1 years for females and 43.2 years for males, 5.9 years and 6.6 years shorter, respectively, than the official estimates. Our estimates for 1918 are shorter than our estimates for 1917 by 6.8 years for females and 7.5 years for males. The decreases from 1917 to 1918 in the official estimates are the same, 11.8 years, for females and males. Our estimates of these changes are 58% of the NCHS estimates for females and 64% for males.

Figure 1. Life Expectancy for Males and Females in the U.S. Registration Area Estimated From Life Tables With Moving Mean Averages for 1900–2000[a]



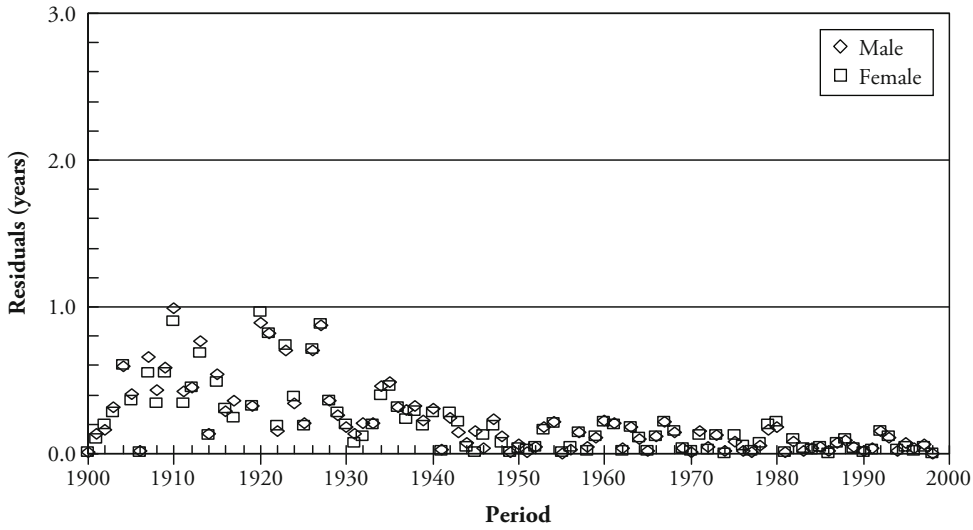
[a]The values for 1918 are shown as solid objects.

Figure 2. Life Expectancy for Males and Females in the U.S. Registration Area Estimated Using Interpolation With Moving Mean Averages for 1900–2000[a]



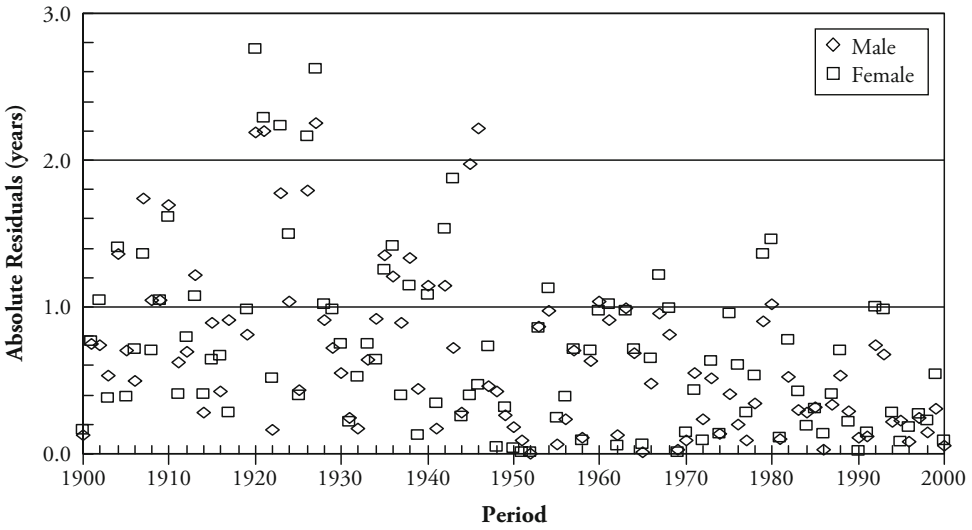
[a]The values for 1918 are shown as solid objects.

Figure 3. Absolute Values of Residuals[a] of Life Expectancy Estimated From Life Tables for Males and Females in the U.S. Registration Area for 1900–2000, Excluding 1918[a]



[a]A Residual is an estimated life expectancy minus its moving mean.

Figure 4. Absolute Values of Residuals[a] of Life Expectancy Estimated Using Interpolation for Males and Females in the U.S. Registration Area for 1900–2000

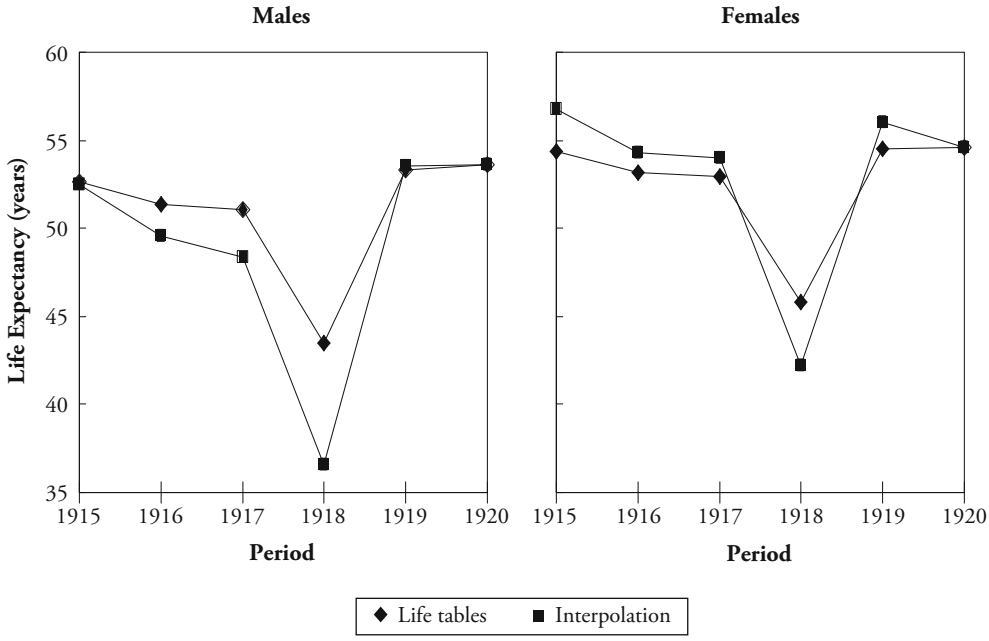


[a]A residual is an estimated life expectancy minus its moving mean.

Table 1. Medians of Residuals of Life Expectancies for Males and Females by Different Methods for Two Periods: 1903–1939 (except 1918) and 1951–1995

	Male		Female	
Source	1903–1939	1951–1995	1903–1939	1951–1995
Life Table Method	0.36	0.05	0.34	0.06
Interpolation Method	0.90	0.32	0.77	0.42
NCHS Report	0.89	0.07	0.79	0.07

Figure 5. Life Expectancy in the U.S. Registration Area for Males and Females Estimated From Life Tables and by Interpolation for 1915–1920



DISCUSSION

Interpolated estimates and life table estimates of annual LE differ substantially for the United States over the twentieth century. Much of the change in variation in NCHS-reported LEs from the earlier era to the later era is an artifact of the different methods. Estimates of LE using life tables are more appropriate for comparisons because they more accurately reflect the current age-specific death rates.

When Greville and Carlson (1951) prepared their intercensal estimates of LE for 1948 and earlier, computation was labor intensive. The interpolation method that they developed was clever and insightful, made excellent use of the available data, and required only limited computing effort. Today, computation is quite cheap and statistical methods have been

developed. We conclude that it is appropriate for NCHS to prepare a new series of official estimates of annual life expectancies for the twentieth century, consistently computed and using the best data and methods available now. Without such estimates, we advise others to prepare their own life tables and estimates of LE from the official, published estimates of age-specific death rates.

Why wasn't the interpolation method more successful? The U.S. population was not stationary at any time, so the assumption that life expectancy equals the inverse of the crude death rate (1/CDR) cannot be theoretically justified, but might have been a reasonable approximation. Although we know of no theoretical argument for it, such as constant population growth together with regularly changing death rates, we found that the relationship of LE and 1/CDR is quite close to a straight line with strong correlations for both males (0.964) and females (0.958) from 1900 to 1998. This strong relationship suggests that LE could be estimated quite accurately as a linear function of 1/CDR for a few years, and the estimated relationship could be used for interpolation. The intercepts and coefficients of 1/CDR are, respectively, 17.21 and 0.51 for males and 20.68 and 0.44 for females, where LE is measured in years and CDR is a proportion between 0 and 1. Because these estimates are for the United States, the use of a linear relationship for other countries probably requires estimates using data from those countries.

Our life table methods differed in several minor respects from life table methods used for statistics reported by NCHS. Principally, we used 10-year age intervals for death rates, as published by NCHS, whereas NCHS used 1-year or 5-year intervals. In addition, we used a single set of survival times within age intervals for every year for both males and females, whereas NCHS used different values after each census. Where estimates could be compared, our life table estimates of LE are close to those of NCHS.

A focus on the estimates for 1918 is apropos to concerns about both bioterrorism and a potential pandemic of natural influenza (Gensheimer et al. 2003; Meltzer, Cox, and Fukuda 1999; World Health Organization 2005). The NCHS-reported estimates of LE in 1918 may be misleading in painting a darker picture of the effect of another pandemic. We estimate that LE in 1918 decreased by about 60% as much as the decrease computed from official LEs. The effect of a new pandemic on life expectancy may be much less than has been speculated, depending on which age groups are most affected. Although the impact of influenza in 1918 was great by any measure, we need accurate measures to plan for another such occurrence.

The reduction in annual variation in LE occurred sometime in the 1940s. It is well-known that death rates dropped and LEs increased steadily from well before 1900 until about 1950; the change is associated with the epidemiologic transition, with dramatic reductions in parasitic and infectious diseases as causes of death. We have shown that the end of the transition is contemporaneous with substantially less variable life expectancy.

Why has the dramatic reduction in annual variation in life expectancy gone unnoticed despite the long history of analysis by many government statisticians and university scholars? The reduction in variation is obvious regardless of which method of computation of residuals we use, and others could have computed these or similar residuals. Figures 3 and 4 and Table 1 tell the same story about this change. The principal reason we advance is that our predecessors focused solely on the levels of death rates and LEs, such as those shown in Figures 1 and 2. Little attention has been given to residual variation.

We suggest that the change in variation in LEs between the earlier era and the later era reflects the degree of control that could be applied to infectious disease in the two eras. Although the epidemiologic transition was ongoing throughout the twentieth century, something happened during the 1940s that caused a striking break in annual variation in life expectancy. Our ongoing research suggests that the introduction of antibiotics in the 1940s might have contributed to this change from high variation to low variation.

Appendix Table A1. Life Expectancies for Males and Females for Each Year, 1900–1998

	Male		Female	
Period	Computed From Life Tables	Published by NCHS	Computed From Life Tables	Published by NCHS
1900	44.88	46.3	47.76	48.3
1901	46.55	47.6	49.82	50.6
1902	47.70	49.8	51.05	53.4
1903	47.96	49.1	51.14	52.0
1904	46.76	46.2	50.05	49.1
1905	47.33	47.3	50.78	50.2
1906	46.89	46.9	50.84	50.8
1907	46.94	45.6	51.12	49.9
1908	48.88	49.5	52.52	52.8
1909	49.84	50.5	53.44	53.8
1910	48.96	48.4	52.77	51.8
1911	50.85	50.9	54.37	54.4
1912	51.41	51.5	55.19	55.9
1913	50.78	50.3	54.74	55.0
1914	52.04	52.0	55.76	56.8
1915	52.68	52.5	56.23	56.8
1916	51.37	49.6	55.16	54.3
1917	51.05	48.4	55.11	54.0
1918	43.47	36.6	48.34	42.2
1919	53.35	53.5	55.81	56.0
1920	53.59	53.6	55.58	54.6
1921	56.47	60.0	58.67	61.8
1922	56.27	58.4	58.76	61.0
1923	55.53	56.1	58.13	58.5
1924	56.42	58.1	59.36	61.5
1925	56.51	57.6	59.40	60.6
1926	55.82	55.5	58.77	58.0
1927	57.31	59.0	60.40	62.1
1928	56.08	55.6	59.31	58.3
1929	56.31	55.8	59.64	58.7
1930	57.30	58.1	60.83	61.6
1931	57.97	59.4	61.58	63.1
1932	58.87	61.0	62.16	63.5
1933	59.01	61.7	62.62	65.1
1934	58.20	59.3	62.21	63.3
1935	58.84	59.9	62.92	63.9
1936	58.14	56.6	62.40	60.6
1937	58.77	58.0	63.16	62.4
1938	60.27	61.9	64.36	65.3

(continued)

(Appendix Table A1, continued)

Period	Male		Female	
	Computed From Life Tables	Published by NCHS	Computed From Life Tables	Published by NCHS
1939	60.93	62.1	65.05	65.4
1940	60.84	60.8	65.31	65.2
1941	61.45	63.1	66.14	66.8
1942	62.09	64.7	67.01	67.9
1943	61.99	62.4	66.91	64.4
1944	62.41	63.6	67.61	66.8
1945	62.65	63.6	68.26	67.9
1946	63.45	64.4	68.64	69.4
1947	64.39	64.4	69.59	69.7
1948	64.56	64.6	69.99	69.9
1949	64.98	65.2	70.50	70.7
1950	65.28	65.6	70.89	71.1
1951	65.39	65.6	71.22	71.4
1952	65.54	65.8	71.49	71.6
1953	65.77	66.0	71.88	72.0
1954	66.58	66.7	72.69	72.8
1955	66.58	66.7	72.77	72.8
1956	66.57	66.7	72.88	72.9
1957	66.34	66.4	72.68	72.7
1958	66.55	66.6	72.90	72.9
1959	66.72	66.8	73.24	73.2
1960	66.54	66.6	73.17	73.1
1961	67.02	67.1	73.65	73.6
1962	66.83	66.9	73.50	73.5
1963	66.56	66.6	73.39	73.4
1964	66.77	66.8	73.73	73.7
1965	66.74	66.8	73.85	73.8
1966	66.62	66.7	73.87	73.9
1967	66.87	67.0	74.28	74.3
1968	66.49	66.6	74.02	74.1
1969	66.70	66.8	74.35	74.4
1970	66.98	67.1	74.73	74.7
1971	67.39	67.4	75.07	75.0
1972	67.38	67.4	75.16	75.1
1973	67.57	67.6	75.37	75.3
1974	68.16	68.2	75.95	75.9
1975	68.76	68.8	76.70	76.6
1976	69.10	69.1	76.87	76.8
1977	69.45	69.5	77.33	77.2

(continued)

(Appendix Table A1, continued)

Period	Male		Female	
	Computed From Life Tables	Published by NCHS	Computed From Life Tables	Published by NCHS
1978	69.65	69.6	77.45	77.3
1979	70.05	70.0	77.91	77.8
1980	69.94	70.0	77.56	77.4
1981	70.42	70.4	77.97	77.8
1982	70.82	70.8	78.28	78.1
1983	70.98	71.0	78.23	78.1
1984	71.12	71.1	78.29	78.2
1985	71.08	71.1	78.27	78.2
1986	71.17	71.2	78.36	78.2
1987	71.35	71.4	78.44	78.3
1988	71.34	71.4	78.35	78.3
1989	71.64	71.7	78.65	78.5
1990	71.82	71.8	78.88	78.8
1991	72.03	72.0	79.04	78.9
1992	72.33	72.3	79.25	79.1
1993	72.14	72.2	78.95	78.8
1994	72.35	72.4	79.07	79.0
1995	72.54	72.5	79.05	78.9
1996	73.08	73.1	79.19	79.1
1997	73.56	73.6	79.34	79.4
1998	73.88	73.8	79.43	79.5

Note: Values were computed from life tables using death rates for each period.

Source: National Center for Health Statistics (Arias 2004).

REFERENCES

Anderson R.N. 1999. "Method for Constructing Complete Annual U.S. Life Tables." *Vital and Health Statistics* 2(129). Washington, DC: National Center for Health Statistics.

Arias, E. 2004. "United States Life Tables, 2001." *National Vital Statistics Reports,* Vol. 52, No. 14. National Center for Health Statistics, Hyattsville, MD. Available online at http://www.cdc.gov/nchs/data/nvsr/nvsr52/nvsr52_14.pdf

Bunker, J.P. 2001. "The Role of Medical Care in Contributing to Health Improvements Within Societies." *International Journal of Epidemiology* 30:1260–63.

Bunker, J.P., H.S. Frazier, and F. Mosteller. 1994. "Improving Health: Measuring Effects of Medical Care." *Milbank Quarterly* 72:225–58.

Chambers, J.M., W.S. Cleveland, B. Kleiner, and P. Tukey. 1983. *Graphical Methods for Data Analysis*. Belmont, CA: Wadsworth Press.

Cleveland, W.S. 1979. "Robust Locally Weighted Regression and Smoothing Scatter Plots." *Journal of the American Statistical Association* 74:829–36.

Gensheimer, K.F., M.I. Meltzer, A.S. Postema, and R.A. Strikas. 2003. "Influenza Pandemic Preparedness." *Emerging Infectious Diseases* 9:1645–49.

Greville, T.N.E. 1953. "Method of Constructing the Abridged Life Tables for the United States, 1949." *Vital Statistics–Special Reports,* Vol. 33, No. 15. Washington, DC: National Center for Health Statistics.

Greville, T.N.E. and G.A. Carlson. 1951. "Estimated Average Length of Life in the Death Registration States." *Vital Statistics–Special Reports,* Vol. 33, No. 9. Washington, DC: National Center for Health Statistics.

Hetzel, A.M. 1997. *History and Organization of the Vital Statistics System*. Washington, DC: National Center for Health Statistics.

Kochanek K.D., S.L. Murphy, R.N. Anderson, and C. Scott. 2004. "Deaths: Final Data for 2002." *National Vital Statistics Reports,* Vol. 53, No 5. Hyattsville, MD: National Center for Health Statistics.

McKeown, T. 1979. *The Role of Medicine: Dream, Mirage or Nemesis?* Princeton, NJ: Princeton University Press.

McKinlay, J.B. and S.M. McKinlay. 1977. "The Questionable Contribution of Medical Measures to the Decline of Mortality in the United States in the Twentieth Century." *Milbank Fund Quarterly* 55:405–28.

Meltzer, M.I., N.J. Cox, and K. Fukuda. 1999. "The Economic Impact of Pandemic Influenza in the United States: Priorities for Intervention." *Emerging Infectious Diseases* 5:659–71.

National Center for Heath Statistics. n.d. "HIST290: Death Rates for Selected Causes by 10-Year Age Groups, Race, and Sex: Death Registration States, 1900–32, and United States, 1933–98." Available online at http://www.cdc.gov/nchs/datawh/statab/unpubd/mortabs/hist290.htm

Neison, F.G.P. 1844. "On a Method Recently Proposed for Conducting Inquiries Into the Comparative Sanatory Condition of Various Districts." *Journal of the Statistical Society of London* 7:40–68.

———. 1845. "Contributions to Vital Statistics, Especially Designed to Elucidate the Rate of Mortality, the Laws of Sickness, and the Influences of Trade and Locality on Health, Derived From an Extensive Collection of Original Data, Supplied by Friendly Societies, and Proving Their Too Frequent Instability." *Journal of the Statistical Society of London* 8:290–350.

Omran, A. 1971. "The Epidemiological Transition: a Theory of the Epidemiology of Population Change." *Milbank Memorial Fund Quarterly* 49:509–38.

Salomon, J.A. and C.J.L. Murray. 2002. "The Epidemiologic Transition Revisited: Compositional Models for Causes of Death by Age and Sex." *Population and Development Review* 28:205–28.

Sen, A. 1998. "Mortality As an Indicator of Economic Success and Failure." *The Economic Journal* 108:1–25.

Sirken, M.G. 1966. "Comparison of Two Methods of Constructing Abridged Life Tables by Reference to a 'Standard' Table." *Vital and Health Statistics,* Vol. 2, No. 4. Washington, DC: National Center for Health Statistics.

Swanson, D. and J. Siegel. 2002. *Methods and Materials of Demography,* condensed ed. New York: Academic Press.

World Health Organization. 2005. *WHO Global Influenza Preparedness Plan.* Geneva: World Health Organization.