
Estimating Deaths for the United States in 1900 by Cause, Age, and Sex

AVIVA BERK, MA
LYNN C. PARINGER, MA
THEODORE D. WOOLSEY

MOST HISTORICAL COMPILATIONS of mortality statistics for the United States begin with 1900, the year for which the Bureau of the Census began the annual collection of mortality statistics. (Responsibility for national vital statistics was transferred to the Public Health Service in 1946.) In 1900, however, the registration of deaths by State governments was still imperfect in many parts of the country. Hence, the Census Bureau limited tabulations of the data to those areas in which registration was believed to be reasonably complete and accurate. These areas became known as the Death-Registration Area (DRA), which originally consisted of 10 States and the District of Columbia—called Death-Registration States (DRS)—and a number of cities in nonregistration States. As registration improved, more States were added to the DRA. In 1933, the DRA finally included the whole of the continental United States (1).

Two limitations have been, at the least, an inconvenience for researchers: (a) the published death rates for the DRS in 1900 represented only 20 million of

the country's 76 million population and (b) that population was solely from the northeast and north-central regions. The resulting bias in historical series purporting to represent the nation is one which researchers have had to pass over with little more than a reference. The best possible estimate of mortality in the rest of the country would require intensive study of local data and a large investment, and the need did not appear to justify these.

In recent months, however, the Georgetown University Public Services Laboratory needed an estimate of national mortality in 1900 for a study it was conducting for the National Institutes of Health. This study, briefly, included an attempt to determine the costs of illness and disease to the national economy in 1900. Some of these costs were due to premature death, as determined by use of the model made practical by Rice (2). The estimate was needed for disease categories based on the Eighth Revision, International Classification of Diseases, Adapted for Use in the United States (ICDA-8). Hence, an intermediate step was the estimation of the numbers of deaths by cause, age, and sex in the continental United States in 1900. Because much of the other data available for estimating direct and indirect costs in that year were of uncertain reliability, according to the Rice model, we believed that a fairly crude set of estimates of the numbers of deaths would suffice. Yet, the obvious strategy of projecting death rates

Ms. Berk and Ms. Paringer are research associates, Public Services Laboratory, Georgetown University. Mr. Woolsey is a health statistics consultant. The research was carried out by the Public Services Laboratory under contract No. NO1-OD-5-2121 with the Division of Program Analysis of the National Institutes of Health; project manager, Dr. Herbert Woolley, and principal investigator, Selma J. Mushkin.

Tearsheet requests to Selma J. Mushkin, Director, Public Services Laboratory, Georgetown University, Washington, D.C. 20057.

in the DRS, or the DRA, onto the total resident population enumerated in 1900 seemed too crude.

Examination of statistics for eight major causes of death in 1930, when the death registration area was nearly complete, had revealed that the ratio of the death rate in all registration States in 1930 to the death rate in the original States in that same year ranged from 1.49 for diarrhea and enteritis to 0.81 for diseases of the heart. This range suggested some important differences in the mortality experience of the original States, 10 States and the District of Columbia, and that of the country as a whole. Therefore, we made estimates taking into account at least some of the differences without spending a great deal of time and effort.

These estimates may be helpful to other researchers. They are presented here with a description and discussion of the admittedly unsophisticated methods used. To our knowledge, no such estimates have been published hitherto.

Methods

Two sources of mortality data for 1900 were used: the vital statistics rates reported by Linder and Grove (1) and the Census Bureau volume covering mortality for 1900-04 (3). The former, presented in a book familiarly called the "rate volume" by vital statistics workers, include death rates by age in the original 1900 DRS area for a selected list of nine major causes of death. The Census Bureau volume contains numbers of deaths by age, sex, and cause of death aggregated for the 1900 DRA, that is, including the cities in nonregistration States.

The first step was to start with data from the rate volume for the following seven cause-of-death groups for which rates by cause and age in 1900 were shown; these groups accounted for 46.8 percent of all deaths in the DRS in 1900:

- Tuberculosis (all forms)
- Cancer and other malignant tumors
- Diabetes mellitus
- Intracranial lesions of vascular origin
- Heart disease (all forms)
- Pneumonia (all forms) and influenza
- Nephritis

The age-specific death rates for these groups and for the same area were also computed for 1940, the first census year in which the death registration area covered the entire United States. The following relationship was then assumed:

$$M_{ca} (1900 \text{ U.S.}) = M_{ca} (1900 \text{ DRS}) \times M_{ca} (1940 \text{ U.S.}) + M_{ca} (1940 \text{ DRS})$$

in which M is the death rate; the subscript ca represents cause of death c in age group a ; and the parenthetical qualifier signifies the year and area.

In other words, it was assumed that within an age group for a particular cause of death, the ratio of the death rate in the entire country to the death rate in the DRS was the same in 1900 as it was later found to be in 1940.

The national 1900 death rates obtained from this equation were then multiplied by the appropriate population counts from the 1900 census to produce estimates of the total numbers of deaths in the United States, by age for each of the seven cause-of-death groups.

The next step was to subdivide the deaths in each of the age-by-cause-of-death cells by sex. We did this by assuming that the male-female distribution of the deaths within each cell in the country as a whole was the same as it was in the deaths in the DRA (see Discussion). The DRA data for 1900 (3) were used instead of the DRS data because the rate volume did not show cause-of-death statistics crossed with age and sex.

It was now necessary to make the estimate for the remaining disease groups. To do this, an estimate was first made of deaths in the entire United States by age and sex for all cases combined. A relationship was assumed that is analogous to that used for the seven major diseases:

$$M_{as} (1900 \text{ U.S.}) = M_{as} (1900 \text{ DRS}) \times M_{as} (1940 \text{ U.S.}) + M_{as} (1940 \text{ DRS})$$

in which the subscript now represents an age-sex grouping.

Again, the death rates obtained were multiplied by the appropriate 1900 census counts to produce estimates of the numbers of deaths by age and sex in 1900. From these aggregates, totals were subtracted for the corresponding age-sex groups resulting from the combination of the figures for the seven major disease groups. This procedure provided totals for the country by age and sex for all other causes.

The next steps were directed to distributing the all-other-causes totals into the other cause-of-death groups being used in the study. We again used the DRA figures from the 1900-04 mortality volume. It was first necessary to reorganize the titles of the causes of death as published in that volume to correspond as closely as possible to the desired categories of ICDA-8. This required assumptions about where in ICDA-8 the various cause-of-death titles in the early mortality tables belonged. These assump-

tions were made with the advice of experts in disease terminology. In a few instances, the process involved splitting up age-sex-cause of death totals from the early volume.

Percentage distributions by cause of death were then calculated within each all-other-causes total from the 1900 DRA, and these were applied to the corresponding all-other-causes totals that had been estimated for the United States. Thus, for these causes—53.2 percent of all deaths in the DRS in 1900—the distribution was taken to be the same in the country as a whole as it was within the DRA, within an age-sex total.

The results of all the steps taken were then consolidated to form tables 1 and 2.

Discussion

The particular form of the estimates was dictated more by the availability of data than by any evident logic. Nevertheless, the underlying assumption is that for general mortality and for seven major causes of death, the ratio within each age group of the death rates in the DRS to the corresponding death rates in the continental United States remained unchanged from 1900 to 1940. In fact, this certainly did not occur, but we think it is reasonable to suppose that this set of ratios did not change very much over the 40-year period. The change in the ratios from 1930 to 1940 can be examined under the minor assumption that rates for the 1930 DRS accurately represented the United States since by that time it included all areas except Texas. For major causes of death, the ratios changed very little in that decade. An exception was diarrhea and enteritis mortality, for which the rates fell rapidly during those years (*la*).

At the very least, it seems that the basic assumption made should give a more accurate picture of mortality by age in the United States at the beginning of the century than would be obtained by assuming that the age- and cause-specific rates in the population of the DRS were the same as those in the entire country.

The estimates could have been improved had it been possible to separate the deaths by cause and age for the entire list of causes of death in the DRS for 1900, or to compute death rates in 1900 and 1940 for the DRA, instead of the DRS, in the same cause groupings that were available in the first census report for the DRA.

The most serious shortcoming of the estimates, however, may be in the way the deaths by sex were

determined. The populations of the southern and western States had distinctly different sex ratios from those in the northeast and north-central areas. Hence, the assumption that the ratios of male to female deaths were the same biased the estimates somewhat toward too high a number of deaths of females.

There are alternative possibilities for making identical estimates, but they were not explored because the ones produced were considered adequate for our purpose. We hope that other persons will study the matter in more depth and prepare estimates by an independent method. A comparison of those results with ours would help to determine the extent of reliability of the ones we have presented.

References

1. Linder, F. E., and Grove, R. D.: Vital statistics rates in the United States, 1900–1940. Bureau of Census. U.S. Government Printing Office, Washington, D.C., 1943; (a) table 14.
2. Rice, D. P.: Estimating the cost of illness. Health Economics Series, No. 6. Public Health Service Publication No. 947-6. U.S. Government Printing Office, Washington, D.C., May 1966.
3. U.S. Bureau of the Census: Mortality Statistics, 1900–1904. U.S. Government Printing Office, Washington, D.C., 1906, pp. ccii–cciii.

* * *

Authors' note: Robert J. Myers, former Chief Actuary of the Social Security Administration and now professor of actuarial science at Temple University, who reviewed this paper, suggested a better method for estimating the sex groups within each age group. With this method, most of the shortcomings mentioned in the Discussion could be avoided. However, because the project for which these estimates were made is now aimed at other tasks and we are engaged in other work, we could not redo the estimates. The method he suggested is as follows:

1. Within each age group find the "expected" number of deaths of males and the "expected" number of deaths of females in the United States for a particular cause-of-death group by multiplying the male death rate in the DRA by the U.S. male population and the female death rate in the DRA by the U.S. female population.

2. Distribute the estimated U.S. deaths of both sexes in each age group proportionally into estimated U.S. male deaths and estimated U.S. female deaths in the same proportion as the expected numbers of deaths obtained in step 1.

We hope that some other workers in need of estimates of U.S. mortality for 1900 will carry through the arithmetic proposed by Myers and will publish their results.

Table 1. Estimated deaths, by age and sex, United States, 1900

Age group (years)	Total		Males		Females	
	Number	Percent	Number	Percent	Number	Percent
All ages	1,410,968	100.0	743,376	100.0	667,592	100.0
Under 1	355,590	25.2	196,379	26.4	159,211	23.8
1-4	197,535	14.0	100,695	13.5	96,840	14.5
5-14	76,633	5.4	34,894	4.7	41,739	6.3
15-24	115,367	8.2	55,106	7.4	60,261	9.0
25-34	122,991	8.7	63,987	8.6	59,004	8.8
35-44	105,426	7.5	57,842	7.8	47,584	7.1
45-54	97,695	6.9	54,343	7.3	43,352	6.5
55-64	103,125	7.3	56,426	7.6	46,699	7.0
65-74	114,164	8.1	61,543	8.3	52,621	7.9
75-84	90,808	6.4	47,518	6.4	43,290	6.5
85 and over	31,634	2.2	14,643	2.0	16,991	2.5

NOTE: Percentages may not total 100 because of rounding.

Table 2. Estimated number of deaths by diagnosis, age, and sex, United States, 1900

Disease category	Total males and females	Males					Females				
		All ages	Under 25	25-44	45-64	65 and over	All ages	Under 25	25-44	45-64	65 and over
All diseases	1,410,968	743,376	387,074	121,829	110,769	123,704	667,592	358,051	106,588	90,015	112,902
Infective and parasitic diseases	428,023	218,830	142,765	48,018	17,859	10,188	209,193	145,266	40,688	14,226	9,013
Neoplasms	35,917	12,840	435	1,521	5,701	5,183	23,077	554	4,626	10,993	6,904
Endocrine, nutritional, and metabolic diseases ...	6,880	3,537	905	777	1,011	844	3,343	834	599	1,135	775
Diseases of the blood and blood-forming organs .	3,625	1,450	584	273	344	249	2,175	1,108	567	319	181
Mental disorders	4,867	3,479	153	1,832	1,113	381	1,388	107	712	300	269
Diseases of the nervous system and sense organs	65,433	34,637	23,176	3,085	2,810	5,566	30,796	21,512	2,558	1,911	4,815
Diseases of the circulatory system	176,170	92,000	12,039	12,281	27,900	39,780	84,170	11,764	11,420	24,453	36,533
Cerebrovascular diseases ¹	(75,106)	(38,859)	(3,359)	(4,442)	(12,243)	(18,815)	(36,247)	(2,843)	(3,262)	(11,482)	(18,660)
Diseases of the respiratory system	274,911	144,816	81,539	19,407	19,403	24,467	130,095	71,517	13,064	15,575	29,939
Diseases of the digestive system, oral cavity, salivary glands, and jaws	54,870	27,952	13,336	5,182	4,825	4,609	26,918	13,307	6,578	3,128	3,905
Diseases of the genitourinary system .	116,846	64,223	7,787	12,450	20,772	23,214	52,623	9,867	14,473	14,958	13,325
Complications of pregnancy, childbirth, and puerperium	9,209	0	0	0	0	0	9,209	3,522	5,658	29	0
Diseases of the skin and subcutaneous tissue ..	2,159	1,220	744	164	182	130	939	585	136	97	121
Diseases of the muscu- loskeletal system and connective tissue	6,897	3,336	1,507	602	543	684	3,561	1,962	535	427	637
Congenital anomalies ...	45,470	25,786	25,786	0	0	0	19,684	19,684	0	0	0
Certain causes of perinatal morbidity and mortality	32,020	18,133	18,133	0	0	0	13,887	13,887	0	0	0
Symptoms and ill-defined conditions	86,794	46,236	37,561	1,553	1,942	5,180	40,558	32,432	2,240	1,474	4,412
Accidents, poisonings, and violence	60,877	44,901	20,624	14,684	6,364	3,229	15,976	10,143	2,734	1,026	2,073

¹ Cerebrovascular diseases are included in diseases of the circulatory system above.