

MORTALITY

Joel C. Kleinman, Ph.D.^a

INTRODUCTION

The oldest measure of the health status of a population is the death rate. Historically it has been quite useful in documenting progress (or lack thereof) in public health and medicine. For example, the death rate from tuberculosis decreased from 194 per 100,000 population in 1900 to 46 per 100,000 in 1940 to 3 per 100,000 in 1970. On the other hand, the death rate from breast cancer has actually increased from 48 per 100,000 in 1940 to 56 per 100,000 in 1970.

It must be remembered, however, that some conditions which kill cause relatively little disability before death while other conditions which seldom cause death (such as arthritis) are responsible for a great deal of disability. Obviously both kinds of data are needed to obtain a picture of the population's health.

However, morbidity data for small areas are obtainable only through special population surveys. Because mortality is relatively easy to define and nearly every death in the U.S. is officially registered, mortality data are extremely useful for comparing different areas. Although death rates provide only a limited assessment of the health status of an area's population, such data can be useful in identifying communities or population subgroups with special needs for certain types of preventive or curative services.

A wide variety of factors influence mortality: age, sex, race, socioeconomic status, occupation, environment, heredity, smoking and other personal habits, and place of resi-

dence. Some of these determinants of mortality are discussed in detail by Kitagawa and Hauser¹ and Erhardt and Berlin.²

This note will discuss overall mortality. Future notes will concentrate on cause-specific mortality data and their uses.

DEFINITIONS

In order to interpret mortality data meaningfully, information about the population at risk is needed. The *number* of deaths is seldom useful for comparative purposes; the number needs to be related to the population at risk. This is done by calculating a rate. The crude annual death rate per 1,000 population for an area is defined as the total number of deaths of area residents for the year divided by the midyear population of the area multiplied by one thousand:

$$\bullet \text{ CDR} = \frac{\text{number of deaths during the year}}{\text{population of the area at the midpoint of the year}} \times 1,000$$

For example, there were 1,934,388 resident deaths in the U.S. in 1974 and the estimated population was 211,389,000. Thus the crude death rate was

$$\frac{1,934,388}{211,389,000} \times 1,000 = 9.2 \text{ per } 1,000.$$

The crude death rate is of limited use by itself since, as will be demonstrated below, it is influenced by the population composition. Comparison of areas or time periods using crude death rates will almost certainly be misleading.

^aDivision of Analysis, National Center for Health Statistics.

Figure A. Relationship between crude and age-specific death rates

Consider two communities with identical age-specific death rates but different age distributions:

Age	Community A			Community B		
	Population	Deaths	Death rate per 1,000 ¹	Population	Deaths	Death rate per 1,000 ¹
Total	100,000	1,252	12.5	100,000	2,064	20.6
Under 1 years	1,989	35	17.6	1,648	29	17.6
1-4 years	5,714	4	0.7	4,286	3	0.7
5-14 years	17,525	7	0.4	14,989	6	0.4
15-24 years	16,667	20	1.2	14,167	17	1.2
25-34 years	12,000	18	1.5	9,333	14	1.5
35-44 years	9,286	26	2.8	7,857	22	2.8
45-54 years	8,824	60	6.8	7,353	50	6.8
55-64 years	13,263	205	15.5	10,903	169	15.5
65-74 years	8,829	294	33.3	17,658	588	33.3
75-84 years	4,421	338	76.5	8,842	676	76.5
85 years and over	4,482	245	165.3	2,964	490	165.3

¹ Although the communities are hypothetical, these are actual U.S. age-specific death rates for 1974.

Thus even though the age-specific death rates are identical, community B's crude death rate is 65 percent larger than A's. The reason is simple: community B has a much older population. For example, 15 percent of A's population is 65 or older compared to 19 percent of B's.

The CDR is a weighted average of the age-specific death rates:

Community A:

$$12.5 = 17.6 \times \frac{1,989}{100,000} + 0.7 \times \frac{5,714}{100,000} + 0.4 \times \frac{17,525}{100,000} + \dots + 76.5 \times \frac{4,421}{100,000} + 165.3 \times \frac{4,482}{100,000}$$

Community B:

$$20.6 = 17.6 \times \frac{1,648}{100,000} + 0.7 \times \frac{4,286}{100,000} + 0.4 \times \frac{14,989}{100,000} + \dots + 76.5 \times \frac{8,842}{100,000} + 165.3 \times \frac{2,964}{100,000}$$

The higher rates for the older age groups are given more weight in community B.

The methods for age adjustment presented in the appendix are ways of comparing rates which avoid the different weights used in the CDR.

It is therefore useful to limit the numerator and denominator of the crude death rate to specific subgroups of the population, usually by age, race, or sex (or combinations thereof). The specific death rates thus defined have the same form as the crude death rate but are limited to the deaths for the specific subgroup

and population at risk in that subgroup. The following are examples of specific death rates:

- Age-specific death rate = ASDR =
$$\frac{\text{number of deaths in the age group during the period}}{\text{population of the age group at the midpoint of the period}} \times 1,000$$

- Race-specific death rate = RSDR =
$$\frac{\text{number of deaths in the race group during the period}}{\text{population of the race group at the midpoint of the period}} \times 1,000$$
- Age-sex-race-specific death rate = ASSDR =
$$\frac{\text{number of deaths in the age-sex-race-group during the period}}{\text{population of the age-sex-race-group at the midpoint of the period}} \times 1,000$$
- Cause-specific death rate = CSDR =
$$\frac{\text{number of deaths attributed to cause in population at risk during period}}{\text{population at risk at midpoint of period}} \times 1,000$$

The limitations of the crude death rate (CDR) can now be illustrated more clearly. Figure A shows that the CDR is a weighted average of the specific rates. The weights depend upon the composition of the area's population. Thus if two areas have different age distributions, their CDR's will differ even if they have identical age-specific rates. The use of the age-race-sex-specific death rates avoids this problem.

However, the amount of information generated by such specific rates can be overwhelming; a summary measure is needed. Thus a useful strategy is to calculate a few overall mortality "indexes" which can summarize the area's mortality experience without being affected by different population compositions. Three such indexes are presented in detail in the appendix. They will allow the planner to identify areas with an excess of mortality. These areas can then be studied to isolate the particular causes of death or age groups which are responsible for the overall excess of mortality.

SOURCES OF DATA

Mortality data are obtained from standard death certificates (modified slightly in some States). A complete description of the sources and classification of mortality data is given in references 3 and 4. Mortality data in varying levels of geographic detail can be obtained on magnetic tape from the National Center for Health Statistics (NCHS).⁵ NCHS also publishes annual data for counties and urban

places within counties.⁶ These publications give only the total numbers of deaths. More recent and detailed data for county or for subcity units (e.g., census tracts) may be available from State or local health departments.

In order to obtain population data for the denominators of the death rates, it is necessary to use sources from the U.S. Bureau of the Census.³ The basic problem in the use of census data is the lack of detailed population breakdowns (by age, race, and sex) for small areas (county level or below) during the period between censuses. Using census data for 2 or 3 years on either side of the census year (e.g., using 1970 data as a denominator for rates between 1968-72 or 1967-73) is probably reasonable in most areas. However, as one moves away from the census year the justification for using decennial census data decreases. Thus the planner should become familiar with local programs which provide updated population estimates (e.g., the Federal-State Cooperative Program,³ manual for local population projections⁷).

USES OF DATA IN PLANNING

The importance of general mortality data to the health planner is primarily as a clue to other health problems. A large portion of mortality is caused by illness for which there is little effective medical care intervention. Factors such as socioeconomic status which affect mortality may be beyond the ability of the health planner to modify, at least in the short run. On the other hand, as the end result of a series of events which adversely affect health, high mortality will usually accompany other health problems which may be more amenable to planning intervention.

Given the limited resources available for local health planning, mortality data can provide a method for directing attention to communities at high risk of significant health problems. The first step in such a method is to calculate overall mortality indexes for each community to determine those with excessive deaths (see appendix). Further analysis of this group of communities could then proceed in a number of ways. Specific causes of death could be examined to determine more precisely the reasons for the excessive deaths. Areas

with excessive deaths for many causes might be the subject of further studies using hospital discharge surveys or even limited population surveys to determine the prevalence of other nonfatal health problems. One of the values of mortality data is to limit the focus of such investigations from an entire HSA to the high-risk communities within the HSA. Since the collection of additional health statistics (e.g., hospital discharge or population surveys) is expensive, limiting the size of the target area can result in a substantial reduction in cost. Any combination of these approaches will help the HSA in setting priorities.

Relationship to Incidence and Case-Fatality

The planner needs to recognize that death rates are directly influenced by two components—incidence of disease or event (e.g., hypertension, motor vehicle accidents) and mortality from that cause (called case-fatality). These components are important because they show that unusually high death rates may be explained by phenomena which have different planning implications. If the disease is preventable the death rate can be reduced by reducing incidence via preventive measures (e.g., immunization). On the other hand, more effective medical care intervention (increased availability and quality of appropriate health services, health education to promote early detection) might also reduce mortality by decreasing case-fatality rates. The optimal mix of these strategies requires more detailed analysis to determine whether it is the incidence of the condition or the case-fatality rate that is unusually high.

In general, the preferred strategy is to reduce incidence since this will usually reduce disability and morbidity as well as mortality. However, technological innovations in medical care are almost always directed at decreasing case-fatality. Such a decrease frequently results in increased morbidity and disability since many survivors will suffer some form of impairment.

Motor vehicle accident mortality provides a good example of the two strategies. If the health planner intervenes at the preventive level by attempting to reduce speed limits, install traffic lights at hazardous intersections,

reduce the number of intoxicated drivers, etc., the total motor vehicle accident rate may be reduced. This will decrease morbidity and disability due to motor vehicle accidents as well as mortality. On the other hand, if the approach is to increase the capability of emergency medical services to respond to motor vehicle accidents, this may also contribute to lower mortality but will not necessarily reduce the number of disabled persons in the population. Of course these approaches are not mutually exclusive and the appropriate mix of strategies will vary from place to place.

Mortality as an Indicator of Service Needs

Finally, the age-specific death rates (or even the crude death rate) for an area can serve as an indicator of the demand for hospital and nursing home care. Hospitalization during the last year of life is extensive so that areas with high death rates will probably need more facilities than others. An explanation of excess mortality in certain communities may be that people migrate there when they are very ill. For example, individuals suffering from severe respiratory allergies may migrate to the Southwest at an exceptionally high rate. Individuals with substantially reduced incomes resulting from disabling chronic conditions may move in disproportionate numbers to low-income neighborhoods in order to find housing that they can afford. Some individuals may move to an area near a source of treatment that they require on a regular basis. Thus a high mortality rate for an area does not necessarily mean that living there, rather than in some other area, is likely to cause a reduction in longevity. Nevertheless, the infirm who have congregated there have above average requirements for services and this should be taken into account in the allocation of resources.

LIMITATIONS OF DATA

Registration and Classification

Since the original sources of the data are the persons completing the death certificates, there is potential for variations in complete-

ness of registration of deaths, completeness and validity of certain items on the certificate, and information on cause of death. The latter will be considered in future notes. As far as completeness of registration of deaths is concerned, it is generally believed that this currently presents no problems.⁴ However, incomplete coverage of certain population subgroups (especially the poor and minorities) by the U.S. Bureau of the Census may provide a denominator which is smaller than it should be. The result is that the death rate for that group would be artificially high. If an area exhibits an excess of mortality for every cause, it may be due to underenumeration of the population at risk. Further, if populations are changing or migrating within an area, the rates may be inaccurate if based on old census counts. This is especially a problem in inner city areas and small areas.

Since health planners are primarily interested in determining rates for particular communities, the validity of the residence classification on the death certificate is especially important. First it should be pointed out that mortality data obtained from local sources may have different residence classification than NCHS.⁸

The most recent results concerning the validity of the NCHS residence classification are from a 1960 study which matched death certificates to census records.⁸ In general the classification by Census and NCHS correspond. However, "an inverse relationship exists between the size of the geographic area and the degree of difference found between census assignments and those by NCHS. Differences diminish as the population size of the area increases. Compared with census assignments, NCHS somewhat overstates the numbers of deaths for individual urban places and understates those for rural areas. A comparison of NCHS annual data with figures tabulated by selected State offices of vital statistics shows the same pattern."⁸ Thus the discrepancy was also greater in nonmetropolitan than metropolitan counties. The implication is that death rates will be overstated for larger geographic areas and understated for smaller areas. However, the degree of underestimation or overestimation within a particular State or Health Service Area (HSA) should not be so great as to distort the ranking of areas in relation to each other.

Another problem with residency classification is that for persons residing in long-term institutions the death certificate asks for the usual place of residence prior to admission to the institution, while the census allocates such persons to the location of the institution. Thus in small counties with large institutional populations the death rates may be too low. Indeed, Sauer et al.⁹ found four counties in North Carolina and three in Georgia for which the death rates for white males aged 45-64 were underestimated by 3 percent to 50 percent because of institutional populations.

Finally, there are a few States which use last mailing address of the decedent as the basis for classification of residence. Death rates calculated for rural areas in these States may be misleading since rural mailing addresses often are the post office location.

Stability

In the previous note¹⁰ it was emphasized that the observed infant mortality rate should be considered an estimate subject to chance fluctuation. This is just as true for all death rates, especially for children and young adults where death rates are very low. In fact not until age 60 does the age-specific death rate exceed the infant death rate. The chance variation is relatively higher for small rates than for larger ones. Thus it is useful to combine more than 1 year of mortality data. When this is done the sum of the numbers of deaths for each year is divided by the population at the middle of the period and the number of years combined to obtain an average annual death rate. For example, consider an area with 52 deaths in 1969, 63 deaths in 1970, and 39 deaths in 1971. If the 1970 population was 5,000, the average annual death rate per 1,000 is

$$\frac{52 + 63 + 39}{3 \times 5,000} \times 1,000 = \frac{154}{15,000} \times 1,000 = 10.27$$

Confidence intervals should be used with death rates as they were with infant mortality rates.¹⁰ Figure B illustrates the calculations. It is also suggested that for most comparative purposes ratios of rates be used rather than absolute differences. The rationale is the same as it was for infant mortality.¹⁰ However, it is especially important when dealing with indexes (appendix).

The formula below is in a format which is easier to compute than that presented in the previous Note.⁹

Let

d = number of deaths upon which rate is based

n = denominator of rate (e.g., if rate is an average annual rate for 3 years, n is 3 times the midpoint population)

The 95 percent confidence limits are

$$\text{upper limit: } \frac{1,000}{n} (d + 1.96 \sqrt{d})$$

$$\text{lower limit: } \frac{1,000}{n} (d - 1.96 \sqrt{d})$$

If the death rate is higher than 100 per 1,000 (as it would be for ages over 65) replace \sqrt{d} in the above formulas by $\sqrt{d \left(1 - \frac{d}{n}\right)}$

If the rate is expressed per 10,000 instead of per 1,000 replace the factor 1,000/ n by 10,000/ n .

Confidence intervals for the ratio of a difference between two independent death rates follow the same formulas as those for infant mortality rates. The easiest method is to use the separate confidence limits. Thus if the 95-percent confidence intervals for each rate are

$$r_1 \pm CL_1$$

$$r_2 \pm CL_2$$

then the confidence interval for $R = \frac{r_1}{r_2}$ is

$$R \pm R \sqrt{\left(\frac{CL_1}{r_1}\right)^2 + \left(\frac{CL_2}{r_2}\right)^2}$$

and the confidence interval for $D = r_1 - r_2$ is

$$D \pm \sqrt{CL_1^2 + CL_2^2}$$

Remember that the formula for the confidence interval for R is valid only when the rate in the denominator (r_2) is based on 100 or more deaths.

REFERENCES

¹Kitagawa, E. M., and Hauser, P. M.: *Differential Mortality in the U.S.*, Cambridge, Mass. Harvard University Press, 1973.

²Erhardt, C. L., and Berlin, J. E., eds.: *Mortality and Morbidity in the U.S.*, Cambridge, Mass. Harvard University Press, 1973.

³Bureau of Health Planning and Resources Development: *Guide to Data for Health Systems Planners*. DHEW Document No. HRA 76-14502. Health Resources Administration, Apr. 1976.

⁴National Center for Health Statistics: *Vital Statistics of the United States, 1972*, Vol. II, Part A,

Section 6. DHEW Pub. No. (HRA) 75-1114. Health Resources Administration. Washington. U.S. Government Printing Office, 1975.

⁵National Center for Health Statistics: *Standardized Micro-Data Tape Transcripts*. DHEW Pub. No. (HRA) 75-1115. Health Resources Administration. Washington. U.S. Government Printing Office, 1974.

⁶National Center for Health Statistics: *Vital Statistics of the United States, 1973*, Vol. II, Part B. DHEW Pub. No. (HRA) 76-1102. Health Resources Administration. Washington. U.S. Government Printing Office, 1975.

intervals for death rates

The following data are based on deaths of white males aged 45-54 years and population for two U.S. counties (see table I in appendix):

	<i>Santa Cruz, Calif.</i>	<i>DeKalb, Ga.</i>
Population (n)	6,051	20,201
Number of deaths (d)	63	155
Death rate $\left(\frac{d}{n} \times 1,000\right)$	10.41	7.67

The 95-percent confidence intervals for the counties are

$$\text{Santa Cruz: } \frac{1,000}{6,051} (63 \pm 1.96 \sqrt{63}) = .1653 (63 \pm 15.5570) = 10.41 \pm 2.57$$

$$\text{Interval: } (7.84, 12.98)$$

$$\text{DeKalb: } \frac{1,000}{20,201} (155 \pm 1.96 \sqrt{155}) = .0495 (155 \pm 24.4018) = 7.67 \pm 1.21$$

$$\text{Interval: } (6.46, 8.88)$$

The confidence intervals for the ratio and difference are

$$\text{Ratio: } R = \frac{10.41}{7.67} = 1.36$$

$$1.36 \pm 1.36 \sqrt{\left(\frac{2.57}{10.41}\right)^2 + \left(\frac{1.21}{7.67}\right)^2} = 1.36 \pm 1.36 \sqrt{.0609 + .0249} =$$

$$1.36 \pm 1.36 \sqrt{.0858} = 1.36 \pm 1.36 (.2930) = 1.36 \pm .40$$

$$\text{Interval: } (0.96, 1.76)$$

$$\text{Difference } D = 10.41 - 7.67 = 2.74$$

$$2.74 \pm \sqrt{2.57^2 + 1.21^2} = 2.74 \pm \sqrt{8.0690} = 2.74 \pm 2.84$$

$$\text{Interval: } (-0.10, 5.58)$$

Since the interval for D includes zero (or the interval for R includes 1) the rates for the two counties are not significantly different.

⁷Bureau of Health Planning and Resources Development, Health Resources Administration, and U.S. Bureau of the Census, *Guide to Local Area Population Projection*. (In press.)

⁸National Center for Health Statistics: Comparison of place of residence on death certificates and matching census records, United States, May-Aug. 1960. *Vital and Health Statistics*. PHS Pub. No. 1000-Series 2-No. 30. Public Health Service. Washington. U.S. Government Printing Office, Jan. 1969.

⁹Sauer, H. I., et al.: Cardiovascular disease mortality patterns in Georgia and North Carolina. *Pub. Health Rep.* 81:455-465, May 1966.

¹⁰National Center for Health Statistics: Infant mortality, by J. C. Kleinman. *Statistical Notes for Health Planners*, No. 2. Health Resources Administration. Washington, D.C., July 1976.

¹¹Kitagawa, E. M.: Theoretical consideration in the selection of a mortality index, and some empirical comparisons. *Hum. Biol.* 38:293-308, Sept. 1966.

¹²Woolsey, T. D.: Adjusted death rates and other indices of mortality. *Vital Statistics Rates in the U.S., 1900-1940*, Chapter IV. U.S. Bureau of the Census. Washington. U.S. Government Printing Office, 1943.

¹³Kleinman, J. C.: A new look at mortality indexes with emphasis on small area estimation. *Proceedings of the American Statistical Association, Social Statistics Section. 1976*, (In press.)

¹⁴Haenszel, W.: A standardized rate for mortality in units of lost years of life. *Am. J. Public Health.* 40:17-26, Jan. 1950.

APPENDIX

AGE-ADJUSTED MORTALITY INDEXES^b

We have seen (figure A) that the crude death rate is not a useful measure for comparing mortality among areas or monitoring changes over time because it depends upon the population composition as well as the specific death rates. Since there are such a large number of specific death rates which can be computed it is useful to summarize those rates in an overall mortality index which takes into account different population compositions. These indexes are usually called age-adjusted or standardized death rates. Three indexes will be presented in this note. For a more complete discussion see Kitagawa¹ and Woolsey.^{1,2} The indexes will be explained in reference to age-adjustment. A future note will discuss more general adjusted rates.

Table I shows the age-specific death rates for white males in two U.S. counties. These counties will be used to illustrate the adjustment procedures. Note that the death rate for Santa Cruz males (15.62) is over twice as high as the rate for DeKalb males (6.58). The difference in overall death rates is greatly influenced by the age distributions—nearly 15 percent of Santa Cruz's white males are 65 or over compared to only 4 percent of DeKalb's white males.

Examination of the age-specific death rates shows that DeKalb has substantially lower rates in the 15-54 age groups and slightly higher rates in the older age groups (figure I). If a single measure is used to summarize the age-specific rates, it will not reflect this "crossover" in the curves (see figure I). Whenever such a crossover is present, the indexes presented here should be used with caution. In a later section we discuss the use of the indexes over narrower age ranges. When plotting death rates which vary widely (as they do over the entire age range) it is convenient to use a logarithmic scale.

The following notation will be used in the discussion (for a particular race-sex group):

^bThis is a more indepth discussion of age-adjusted rates than that presented in the Census Guide³ (which included only the direct method of adjustment).

p_a = area population in age group a

p = $\sum p_a$ = total population

d_a = number of deaths in area for age a

d = $\sum d_a$ = total number of deaths in area

$m_a = \frac{d_a}{p_a} \times 1,000$ = area's age-specific death rate per 1,000

Note that the race-sex specific death rate can be written as

$$\text{RSDR} = \frac{d}{p} \times 1,000 = \sum m_a \frac{p_a}{p} \quad (1)$$

This was illustrated in figure 1; the area's death rate is a weighted average of age-specific rates using the area's age distribution as weights.

Indirect Method

The indirect method of adjustment compares the area's death rate with an "expected" death rate based on the area's population composition and a standard set of rates. The method will be illustrated for white males using U.S. rates as the standard. If the area's age-specific rates for white males were exactly the same as the U.S. age-specific rates for white males, the expected number of white male deaths in age group a would be the product of the U.S. rate and the number of white males in age group a in the area divided by 1,000:

$$\frac{1}{1,000} M_a p_a$$

where the M_a is the U.S. age-race-sex specific rate per 1,000. Thus the total number of expected deaths for white males is just the sum of these products over all age groups:

$$\frac{1}{1,000} \sum M_a p_a$$

The standardized mortality ratio (SMR) is

Table I. White male death rates for two U.S. counties, 1970

Santa Cruz, California						
Age	Population	Number of deaths	Death rates	Years of life lost	Expected deaths	Expected years of life lost
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total	56,478	882	15.62	6,480	884.18	5,893.0
Under 1 year	847	16	18.89	1,112	17.90	1,244.1
1-4 years	3,274	4	1.22	268	2.75	184.3
5-14 years	10,279	3	0.29	180	4.93	295.8
15-24 years	10,201	24	2.35	1,200	17.44	872.0
25-34 years	6,413	14	2.18	560	11.35	454.0
35-44 years	5,661	26	4.59	780	19.47	584.1
45-54 years	6,051	63	10.41	1,260	53.43	1,068.6
55-64 years	5,402	112	20.73	1,120	119.01	1,190.1
65-74 years	4,920	225	45.73	---	236.65	---
75-84 years	2,781	275	98.89	---	280.85	---
85 years and over	649	120	184.90	---	120.40	---

DeKalb, Georgia						
Age	Population	Number of deaths	Death rates	Years of life lost	Expected deaths	Expected years of life lost
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total	173,051	1,139	6.58	16,712	1,190.66	19,635.4
Under 1 year	3,325	64	19.25	4,448	70.26	4,883.1
1-4 years	12,669	12	0.95	804	10.64	712.9
5-14 years	37,883	13	0.34	780	18.18	1,090.8
15-24 years	29,136	31	1.06	1,550	49.82	2,491.0
25-34 years	27,383	41	1.50	1,640	48.47	1,938.8
35-44 years	24,223	68	2.81	2,040	83.33	2,499.9
45-54 years	20,201	155	7.67	3,100	178.37	3,567.4
55-64 years	11,128	235	21.12	2,350	245.15	2,451.5
65-74 years	4,901	252	51.42	---	235.74	---
75-84 years	1,867	199	106.59	---	188.55	---
85 years and over	335	69	205.97	---	62.15	---

then defined as a ratio of observed to expected number of deaths:

$$SMR = \frac{d}{\frac{1}{1,000} \sum M_a p_a} = \frac{\sum d_a}{\frac{1}{1,000} \sum M_a p_a}$$

By using some algebra we can express the SMR as a ratio of rates. Since

$$d_a = \frac{1}{1,000} m_a p_a$$

the SMR can be written as

$$SMR = \frac{\sum m_a p_a}{\sum M_a p_a} = \frac{\sum m_a \frac{p_a}{p}}{\sum M_a \frac{p_a}{p}} \quad (2)$$

Thus the SMR is also a ratio of the area's observed death rate to its expected death rate.

Figure I. Age-specific death rates for white males in two U.S. counties, 1970

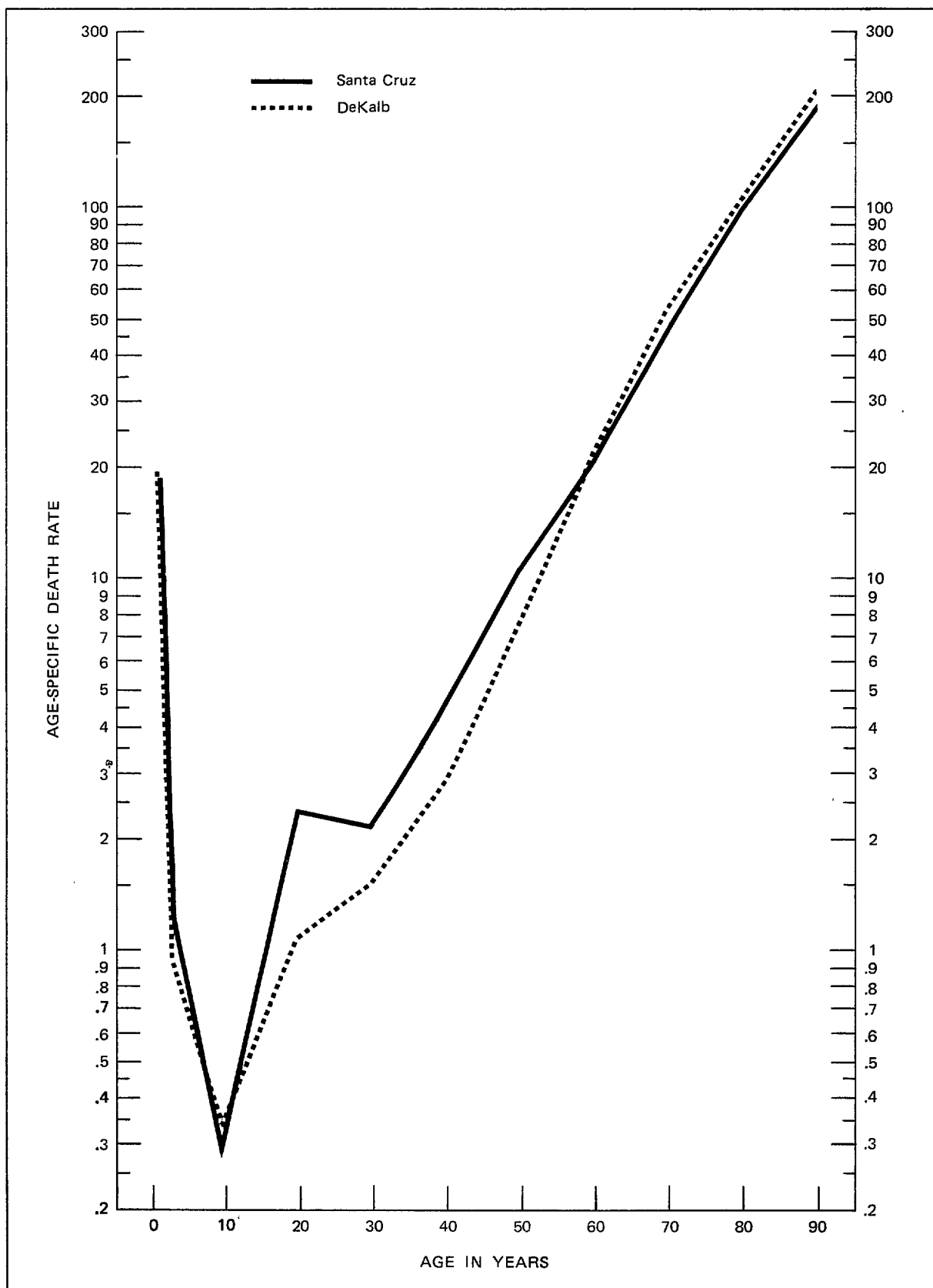


Table II. U.S. population distribution and age-specific death rates, 1970

Age	Proportion (P_a/P) of United States population	Death rates per 1,000			
		White male	White female	Other male	Other female
Under 1 year0172	21.13	16.15	40.20	31.69
1-4 years.....	.0673	.84	.66	1.45	1.23
5-14 years.....	.2005	.48	.30	.65	.42
15-24 years.....	.1744	1.71	.62	3.05	1.08
25-34 years.....	.1226	1.77	.84	5.04	2.16
35-44 years.....	.1136	3.44	1.93	8.74	4.91
45-54 years.....	.1143	8.83	4.63	16.46	9.79
55-64 years.....	.0915	22.03	10.15	30.47	18.87
65-74 years.....	.0612	48.10	24.71	54.74	36.76
75-84 years.....	.0301	100.99	66.99	89.81	63.93
85 years and over0074	185.52	159.80	114.05	102.89

The absolute value of the SMR is influenced by the standard set of rates M_a . However, since the purpose of the SMR is to compare areas, it is the relative values of the SMR's which are important. Choosing the standard as the set of U.S. rates for the period under study is recommended since that will allow comparison of areas between and within different States or HSA's. These rates are shown in table II for 1970.

For the two counties in table I, the expected numbers of deaths are shown in column (6).

Thus the SMR's are

$$\text{Santa Cruz: } SMR_S = \frac{882}{884.18} = .998$$

$$\text{DeKalb: } SMR_D = \frac{1,139}{1,190.66} = .957$$

Note that these indexes are much closer than the crude rates.

A combined SMR for all races and both sexes is often used instead of the individual race and sex-specific SMR's. In this case, the observed and expected number of deaths are each summed over all race-sex groups and a single ratio obtained. The problem with doing this is that there may be problems specific to a race-sex group which will not show up in the combined index. Indeed, for U.S. counties in 1969-71 there were substantial differences among the race-sex specific SMR's.¹³

Direct Method

Another procedure for age adjustment is the direct method. In the direct method a standard age distribution is chosen and the area's age-specific death rates are weighted according to the standard. A reasonable choice for the standard is the U.S. total population (all races, both sexes) for the year under study or the average population of the areas under study. If time trends are being analyzed, the population for the most recent decennial census year can be used.

The direct age-adjusted rate is

$$DAR = \sum m_a \frac{P_a}{P}$$

where P_a = standard population in age group a and

$$P = \sum P_a$$

(The values of P_a/P for the U.S. 1970 population are shown in table II.) By comparing this expression with (1), it is seen that the only difference between the race-specific death rate (RSDR) and the direct age-adjusted rate (DAR) is the weights which are applied to the age-specific rates.

For the two counties in figure I

$$\text{Santa Cruz: } DAR_S = 18.89 \times .0172 + 1.22 \times .0673 + \dots + 184.90 \times .0074 = 11.90$$

$$\text{DeKalb: } DAR_D = 19.25 \times .0673 + \dots + 205.97 \times .0074 = 11.85$$

Thus the two direct age-adjusted rates are very similar, like the SMR's.

It is important to remember that the direct age-adjusted rate is a death rate for a hypothetical population. Its absolute value has little meaning. To keep this point in mind it is best to express the DAR as a ratio to the U.S. death rate for the appropriate race-sex group. Thus we define the *comparative mortality figure* CMF as

$$CMF = \frac{\sum m_a \frac{P_a}{P}}{\sum M_a \frac{P_a}{P}} \quad (3)$$

where M_a is the race-sex-specific death rate for age group a . For example, for white males the denominator is

$$21.13 \times .0172 + .84 \times .0673 + \dots + 185.52 \times .0074 = 11.80$$

Thus the CMF's for the two counties are

$$\text{Santa Cruz: } CMF_S = \frac{11.90}{11.80} = 1.008$$

$$\text{DeKalb: } CMF_D = \frac{11.85}{11.80} = 1.004$$

The difference between the counties in CMF's is smaller than the difference in SMR's, since the CMF uses the same set of weights for each county while the SMR uses each county's population. Thus if the counties had exactly the same set of age-specific death rates their CMF's would be equal (unlike the situation with respect to the SMR). Comparing formulas (2) and (3) reveals that the only difference is the weights applied.

Comparison of the Indexes

It has already been pointed out that the different age distributions used in the SMR makes two SMR's less comparable than two CMF's. However, there are two reasons for using the SMR as it stands:

- (1) *Data availability.*—The SMR requires only the total number of deaths in an area for each race-sex group with no need to calculate all the age-specific rates. This is a special advantage when the only data available are published county or area totals.

- (2) *Public Health implications.*—By weighting according to the area's age distribution, the SMR emphasizes the rates as they apply to the area's population. Thus a high rate combined with high population concentration is emphasized. This property makes the indirect method more appropriate for planning applications.

A disadvantage which is shared by the CMF and the SMR is an emphasis on death rates in the elderly. For example, suppose the death rate in the 85 or over group doubles. The number of deaths in this group also doubles and the two indexes for Santa Cruz become

$$SMR_S = 1.133, \text{ a 13.5-percent increase over original } SMR_S$$

$$CMF_S = 1.124, \text{ a 11.6-percent increase over original } CMF_S$$

However, if the death rate for the 15-24 group doubles the indexes become

$$SMR_S = 1.025, \text{ a 2.7-percent increase over original } SMR_S$$

$$CMF_S = 1.043, \text{ a 3.5-percent increase over original } CMF_S$$

Thus an increase in rate for the young age group has little effect on the CMF or SMR since the number of deaths involved is small while the opposite is true for an increase in rate among the elderly. This is why the SMR and CMF were nearly equal for the two counties despite the differences in the death rates for ages 15-44. For use in planning, the emphasis on the elderly is unfortunate since *mortality* in this group is probably least amenable to planning intervention (the opposite might well be true for morbidity or disability).

Years of Life Lost

One method for emphasizing mortality in the younger age groups is to express deaths in terms of years of life lost rather than numbers of deaths.¹⁴ An easy method for doing this is to assume that each individual has 70 "productive" years of life and so a death at age a results in $70-a$ years of life lost when $a < 70$. When using the 11 age groups, all deaths are assumed to occur at the midpoint of the in-

terval. Since the midpoint of the 65-74 interval is 70, all deaths in this group and the two older groups are omitted from the computation of years of life lost. For example, the numbers of years of life lost in the two counties (column 5 of table I) are:

$$\text{Santa Cruz: } 16 \times 69.5 + 4 \times 67 + 3 \times 60 + \dots + 112 \times 10 = 6480$$

$$\text{DeKalb: } 64 \times 69.5 + 12 \times 67 + 13 \times 60 + \dots + 235 \times 10 = 16712$$

In order to use years of life lost in an index we need to adjust for different age distributions as we did previously. We can use either a direct or indirect approach which have the same advantages and disadvantages discussed above. For ease of calculation and public health implications we choose the indirect method. As we did for the SMR, we calculate an expected number of deaths for each age group. But instead of adding these up we first translate expected deaths into expected years of life lost by multiplying expected deaths by 70 minus the midpoint of the age interval. The expected years of life lost for the two counties are given in column 7 of table I:

$$\text{Santa Cruz: } 17.90 \times 69.5 + 2.75 \times 67 + 4.93 \times 60 + \dots + 119.01 \times 10 = 5,893.0$$

$$\text{DeKalb: } 70.26 \times 69.5 + 10.64 \times 67 + 18.18 \times 60 + \dots + 245.15 \times 10 = 19,635.4$$

The years of life lost (*YLL*) index for each county is the ratio of observed to expected years of life lost:

$$\text{Santa Cruz: } YLL_S = \frac{6,480}{5,893.0} = 1.100$$

$$\text{DeKalb: } YLL_D = \frac{16,712}{19,635.4} = 0.851$$

In this case, Santa Cruz, with its higher death rates in the young age groups, does show a moderately higher *YLL* index than DeKalb.

The formula for *YLL* is:

$$YLL = \frac{\sum d_a (70 - y_a)}{1,000 \sum M_a p_a (70 - y_a)} = \frac{\sum m_a p_a (70 - y_a)}{\sum M_a p_a (70 - y_a)}$$

where y_a is the midpoint of the age interval and the sum extends only to the 55-64 age group:

The *YLL* index, unlike the SMR and CMF, emphasizes the differences in age-specific mortality at younger ages. This can be illustrated by the effect of doubling the death rate for the 15-24 age group in Santa Cruz. The number of deaths in this group will increase from 24 to 48 and the corresponding years of life lost will increase to 2400. The total observed years of life lost becomes 7680 and the *YLL* index is

$$YLL = \frac{7,680}{5,893.0} = 1.303$$

This represents an 18.5 percent increase over the original *YLL* (compare this to the 3-4 percent increases in SMR and CMF given earlier).

A detailed comparison of these three indexes is presented in the reference.¹³ For health planners, the years of life lost index seems most appropriate.

Confidence Intervals for Mortality Indexes

The confidence intervals for the indexes discussed previously are obtained by the formulas in figure II. These formulas are similar to the ones used to compute the indexes and so the additional computations are not time-consuming. Note that the confidence interval for *YLL* is wider than the intervals for the CMF and SMR. This is because the *YLL* emphasizes death rates for younger age groups and these death rates generally have larger random error than those based on large number of deaths (as in the older age groups). Thus aggregation over years will usually be required even when dealing with these summary indexes. If 3 years of data are combined, the *YLL* is reasonably stable for populations of 10,000 or more.¹³

When comparing the same index for different areas, the ratio of the two indexes should be used. Figure III shows the formulas for computing 95-percent confidence intervals for the ratio. If the interval does not include a ratio of one, the indexes are significantly different at the 5-percent level. In order for the formula to be used, the indexes must be independent, i.e., no death included in one index can be included in the other. In addition, the index in the denominator should be based on at least 100 deaths.

Figure II. Confidence intervals for mortality indexes

The confidence interval for one of the indexes discussed is approximated

$$\text{by } I \pm 1.96 s_1 \quad (1)$$

where I = index

s_1 = standard error of the index

The standard error of each index is approximately:

$$s_{SMR} = \frac{\sqrt{d}}{\frac{1}{1,000} \sum M_a p_a}$$

$$s_{CMF} = \frac{\sqrt{\sum \frac{1,000 m_a}{p_a} \left(\frac{P_a}{P}\right)^2}}{\sum M_a \frac{P_a}{P}}$$

$$s_{YLL} = \frac{\sqrt{\sum d_a (70 - y_a)^2}}{\frac{1}{1,000} \sum M_a p_a (70 - y_a)}$$

Using the data from table I, the following standard errors are obtained:

	<i>Santa Cruz, Calif.</i>	<i>DeKalb, Ga.</i>
s_{SMR}0336	.0283
s_{CMF}0349	.0330
s_{YLL}0845	.0426

Confidence intervals are obtained using formula (1):

	<i>Santa Cruz, Calif.</i>	<i>DeKalb, Ga.</i>
<i>SMR</i>	(.932, 1.064)	(.902, 1.012)
<i>CMF</i>	(.940, 1.076)	(.939, 1.069)
<i>YLL</i>	(.934, 1.266)	(.768, .934)

Use of Indexes for Portions of the Age Range

We have seen in figure I that there is a crossover in the age-specific death rates between Santa Cruz and DeKalb, i.e., for the younger age groups DeKalb has lower death rates, but for the older age groups DeKalb has higher rates. A single measure of mortality cannot reflect such differences. Yet examination of all the age-specific rates for more than two areas is very difficult and is subject to the vagaries of random error in the rates (see section in text, "Stability"). A compromise between a single measure and all eleven age-specific rates is to compute indexes for portions of the age range. A general strategy might be to examine the infant mortality rate separately (using live births as the denomina-

tor⁹), then use SMR's for the age ranges 1-34, 35-64, and 65 and over. The SMR's for the narrow age groups are obtained as before except that only observed and expected deaths for the relevant age groups are used. For example, the SMR for aged 1-34 in Santa Cruz is

$$SMR_{1-34} = \frac{4 + 3 + 24 + 14}{2.75 + 4.93 + 17.44 + 11.35}$$

$$= \frac{45}{36.47} = 1.234$$

The *YLL* index could also be used for the age groups under 65 (or even in the group 65 and over by modifying the "expected" years from 70 to, say, 95). However, over the narrower age ranges the differences between the *YLL* and the SMR should be small (e.g., *YLL*₁₋₃₄ for Santa Cruz is 1.223).

Figure III. Confidence interval for the ratio of independent indexes

Let I_1 = mortality index for area 1
 s_{I_1} = standard error of I_1
 I_2 = mortality index for area 2
 s_{I_2} = standard error of I_2
 $R = \frac{I_1}{I_2}$

The confidence limits for R are

$$\text{upper limit: } R + 1.96 R \sqrt{\left(\frac{s_{I_1}}{I_1}\right)^2 + \left(\frac{s_{I_2}}{I_2}\right)^2}$$

$$\text{lower limit: } R - 1.96 R \sqrt{\left(\frac{s_{I_1}}{I_1}\right)^2 + \left(\frac{s_{I_2}}{I_2}\right)^2}$$

For example, using the YLL 's for Santa Cruz, Calif. and DeKalb, Ga.

$$YLL_S = 1.100 \qquad s_{YLL_S} = .0845$$

$$YLL_D = 0.851 \qquad s_{YLL_D} = .0426$$

$$R = 1.293$$

$$1.96 R \sqrt{\left(\frac{s_{I_1}}{I_1}\right)^2 + \left(\frac{s_{I_2}}{I_2}\right)^2} = 2.534 \times .0917 = .232$$

$$R + .232 = 1.525$$

$$R - .232 = 1.061$$

Thus the years of life lost index is from 6 percent to 53 percent higher in Santa Cruz, Calif. than in DeKalb, Ga. Since the interval does not include 1, the two indexes are significantly different at the 5-percent level.

The confidence interval for the ratio can also be obtained directly from the confidence interval for each index.

Suppose these intervals are

$$I_1 \pm CL_1$$

$$I_2 \pm CL_2$$

Then the confidence interval for $R = \frac{I_1}{I_2}$ is

$$R \pm R \sqrt{\left(\frac{CL_1}{I_1}\right)^2 + \left(\frac{CL_2}{I_2}\right)^2}$$

These formulas should only be used when the index in the denominator is based upon more than 100 deaths.

The SMR's for each county are given below:

Age	Santa Cruz	DeKalb
1-34 years	1.234	0.763
35-64 years	1.047	0.904
65 years and over	0.972	1.069

These SMR's give a clearer picture of the differences between the two counties: Santa Cruz has much higher death rates in the young age groups, somewhat higher rates in middle age, and somewhat lower rates among the elderly.

Standard errors and confidence intervals are computed as before (figure II) except that the summation extends over only those age groups included in the index. For example, the standard error of the SMR for aged 1-34 in Santa Cruz is

$$SE = \frac{\sqrt{45}}{36.47} = .184$$

Statistical Notes for Health Planners is a cooperative activity of the National Center for Health Statistics and the Bureau of Health Planning and Resources Development, Health Resources Administration.

Information, questions, and contributions should be directed to Mary Grace Kovar, Division of Analysis, NCHS, Room 8A-55, Parklawn Building, 5600 Fishers Lane, Rockville, Md. 20857.

(HRA) 77-1237

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
 Public Health Service
 Health Resources Administration
 National Center for Health Statistics
 5600 Fishers Lane
 Rockville, Maryland 20857
 OFFICIAL BUSINESS
 PENALTY FOR PRIVATE USE, \$300

POSTAGE AND FEES PAID
 U.S. DEPARTMENT OF HEW
 HEW 390
 FIRST CLASS MAIL

