

### 3. Studying populations – basic demography

*Some basic concepts and techniques from demography - population growth, population characteristics, measures of mortality and fertility, life tables, cohort effects.*

#### **The “demi” in epidemiology**

Since the primary subject matter of epidemiology is people (except for veterinary epidemiologists, who apply the same concepts and methods to studying other animal populations), a logical place to begin the study of epidemiology is with some basic concepts of demography.

#### **Population growth – an epidemic of homo sapiens\***

For its first few million years, the species that we refer to as *homo sapiens* numbered probably fewer than 10 million, due to high mortality. In about 8000 B.C., with the beginning of agriculture, significant population growth began, bringing world population to about 500 million over a 6000-year period. At that point (1650 AD), growth accelerated sharply, so that world population doubled in 150 years (1 billion in 1800), doubled again in 130 years (1930), and doubled yet again in 45 years (4 billion in 1975). Every decade the world’s population increases by about 1 billion, mostly in the developing countries. The population will reach 6 billion in early 1999. It is projected to reach 9.5 billion by 2030 and 12.6 billion by 2100.

#### **World Population in mid-1997 (millions)**

<u>Region</u>	<u>Population</u>
Asia	3,552
Africa	743
Europe	729
Latin America & Caribbean	490
North America	298
Oceania (Australia, NZ, and Pacific)	29
World	5,840

(does not add due to rounding)

\* *Note about sources:* Much of the following has been drawn from publications by the Population Reference Bureau (PRB), especially “Population: A lively introduction” and “The future of world population” (see bibliography). This table comes from their 1997 World population data sheet. The PRB web site ([www.prb.org](http://www.prb.org)) has a wealth of data and links to sources of information on population- and health-related topics.

In 1997, 86 million more people lived on planet Earth than the previous year, for an estimated annual world population growth rate of 1.47%. At that rate, world population would double in 47 years. The world population growth rate is the difference between the birth rate of 24 per 1,000 people and the death rate of 9.

Over the time, differing growth rates can dramatically alter the age, geographic, racial, and affluence distribution of the world's population. In 1950, two thirds of the world's population lived in what is usually referred to as the developing world. The proportion was three-quarters in 1990 and is projected to grow to 85% by 2025 and 90% by 2100. Thus, whatever improvements in health take place in the industrialized world, world demographic and health indicators will be primarily influenced by the situation in the developing world.

### ***The Demographic Transition***

A fundamental model developed to describe population dynamics is the Demographic Transition model. The model posits four stages in the evolution of the population in a society.

1. High fertility, high mortality (pre-industrial)
2. High fertility, declining mortality (industrializing)
3. Declining fertility, low mortality
4. Low fertility, low mortality (stable population)

The first stage (pre-industrial) prevailed throughout the world prior to the past few centuries. Rapid population growth takes place in Stages 2 and 3, because high birth rates, necessary for population survival in Stage 1, are embedded in the cultural, religious, economic, and political fabric of pre-modern societies. As economic and public health advances decrease mortality rates, rapid population growth occurs until the society adjusts to the new realities and fertility decline.

The Demographic Transition Model was constructed from the European experience, in which the decline in death rates was gradual. It remains to be seen how this model will play out in the developing world of today, in which the decline in death rates has occurred much more rapidly and in which social change takes place against a backdrop of and in interaction with the post-industrial world of electronic communications, multi-national production and marketing, and international travel. There is some evidence that the model will also apply to the developing world of today. But the timetable for completion of the demographic transition in the developing world will determine the ultimate size of the world's population.

### ***Demographic balancing equation***

If birth and death are the two most fundamental demographic processes, migration is probably the third. The size of the world's population is (at least at present) completely determined by birth and death rates, but the population in any particular region or locale is also determined by net migration. These three processes are expressed in the demographic balancing equation—the increase (or decrease) in a population as the algebraic sum of births, deaths, immigration, and emigration. The following table gives the equation for the world and for the U.S. for 1991.

**The demographic balancing equation for the United States  
(from McFalls, 1991) (numbers in thousands)**

	Starting population	+	( Births – Deaths )	+	( Immigration – Emigration )	=	Ending population
	Starting population	+	(Natural increase)	+	(Net migration)	=	Ending population
World	= 5,245,071	+	(142,959 – 50,418)				
	= 5,245,071	+	92,541			=	5,337,612
U.S.	= 248,168	+	(4,179 – 2,162)	+	(853 – 160)		
	= 248,168	+	2,107	+	693	=	250,878

In recent decades, on a world basis, the migration has perhaps had its greatest impact on urbanization. In the forty years from 1950 to 1990, the urban population in the countries of the Third World increased over five-fold, from 286 million to about 1.5 billion. About 40 percent of this growth resulted from rural to urban migration. The U.N. predicts that by the year 2000 there will be 19 Third World cities with populations over 10 million. In contrast to Tokyo, Los Angeles, New York, London, and other glamorous metropolises, overcrowded urban areas in poor countries are characterized by inadequate housing, sanitation, transportation, employment opportunities, and other essentials of healthy living, the ingredients for misery and the spread of microorganisms.

***Population age structure and the population pyramid***

For every 10 people in the world:

- 3 are younger than 15 years of age
- 4 live in an urban area
- 6 live in Asia (2 in China, 1 in India)
- 8 live in developing countries

An important dynamic in population growth is the reciprocal relationship between the rate of natural increase (births - deaths) and the age structure of the population. The latter is one of the strongest influences on the growth rate of a population, since both fertility and mortality vary greatly by age. A younger population has a higher rate of natural increase; a high rate of natural increase in turn lowers the median age of a population.

In Africa, which has the highest birth (40/1,000) and growth (2.6%) rates, only 56% of the population are older than 15 years. In contrast, in Europe, where average birth rates have been close to replacement level for many years, four-fifths of the population (81%) are older than 15 years. In fact, Europe as a whole experienced overall negative growth in 1997, due to birth and death rates of 10 and 14, respectively, in Eastern Europe (including Russia). Since nearly all (96%)

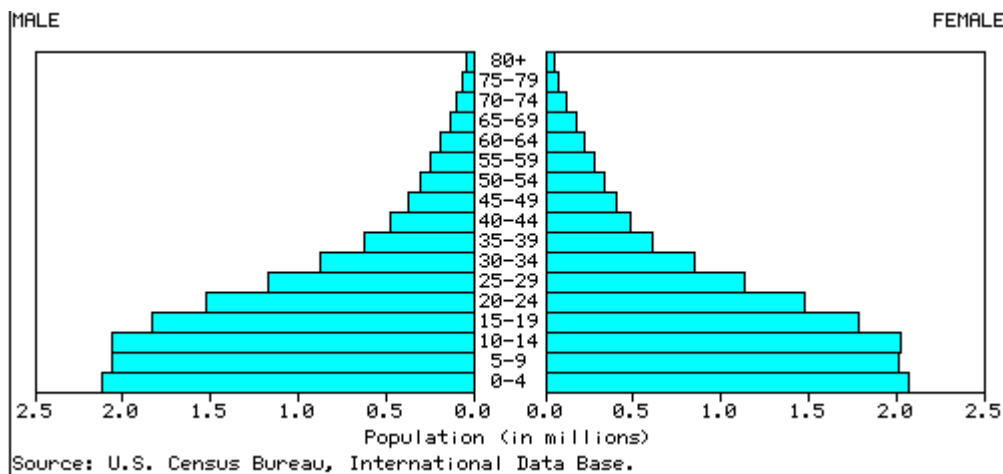
of the increase in the world's population takes place in the developing world, the developing countries are becoming younger while the wealthier countries are becoming older.

Nevertheless, fertility control is increasing in the developing world. As it does, the age structure of the population shifts upwards, since the larger birth cohorts of previous years are followed by relatively smaller birth cohorts. The average age of the world's population, around 28 years, is projected to increase to 31-35 years, so that the proportion of persons 60 years and older will grow from about 9% to 13-17% (Lutz, 1994). This proportion will range from as low as 5% in sub-Saharan Africa to much as 30% in Western Europe. In China, where fertility has been successfully regulated for decades, the proportion of the population age 60 and older will rise to about 20% in 2030 (Lutz, 1994).

### ***The population pyramid***

Demographers display the age structure of a population by constructing a graph in which the population size in each age band is depicted by a horizontal bar that extends from a centerline to the left for one gender and to the right for the other, with the age bands arranged from lowest (at the horizontal axis) to highest. A population pyramid for a population that is growing rapidly, e.g. Kenya, resembles a pillar that is very broad at the base (age 0-1 years) and tapers continuously to a point at the top. In contrast, the population pyramid for a zero-growth population, e.g. Denmark, resembles a bowling pin, with a broader bottom and middle, and narrower base and top.

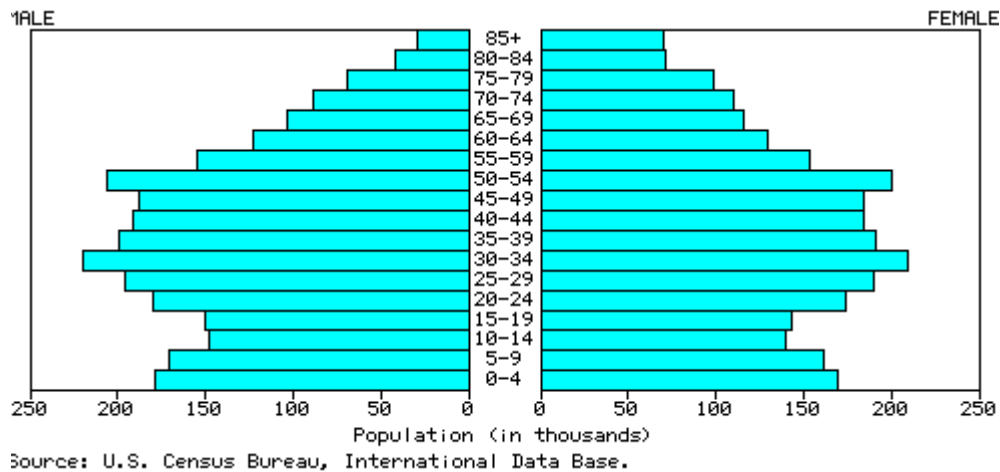
**Kenya, 1998**



The population pyramid for a country shows the pattern of birth and death rates over the past decades, since apart from immigration and emigration, the maximum size of any age group is set by the birth cohort that it began as, and its actual size shows its subsequent mortality experience. For example, the 1989 population pyramid for Germany shows the deficit of older males resulting from losses in World Wars I and II and narrowings corresponding to the markedly lower wartime birth

rates. Similarly, bulges in the reproductive years often produce bulges at the bottom, since more women of reproductive age usually translates into more births.

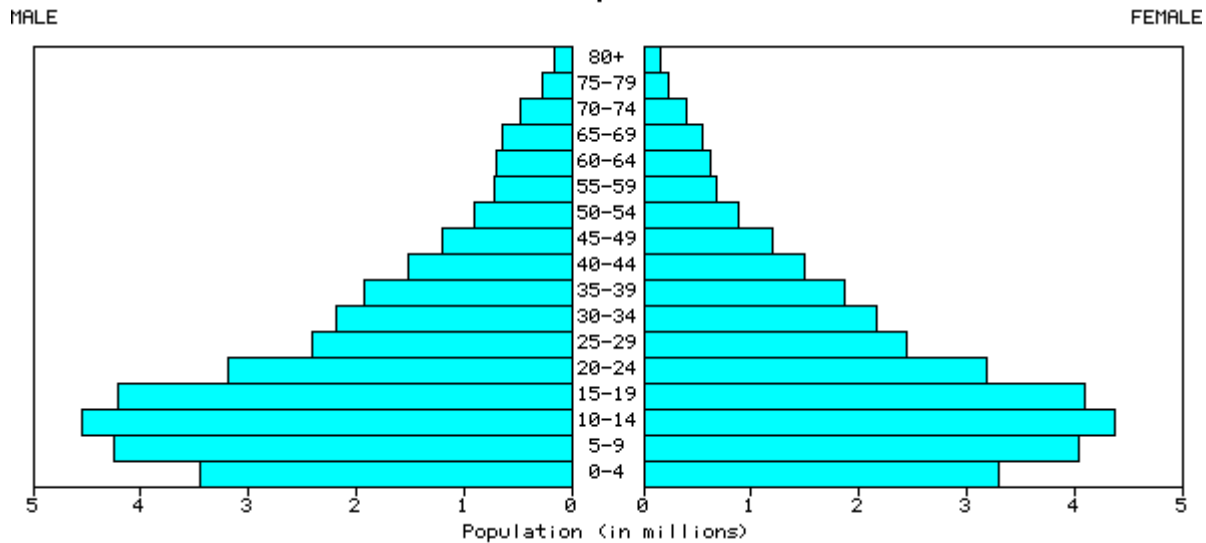
### Denmark, 1998 (note change in scale)



Using the pyramid it is easy to see how a growing population becomes younger and the transition to lower fertility makes it older. Widespread family planning makes new birth cohorts smaller, so that the pyramid consists of a broad middle (persons born before the adoption of family planning) being pushed upward by a narrower base. Initially this age distribution makes life easier for adults, especially women, since effort and resources for childrearing and support are proportionally lower. However, when the adults who first adopted family planning reach retirement age, there are fewer younger people available to support them. Unless productivity and savings have risen sufficiently, the society will be hard pressed to support its elderly members—an issue of concern in affluent societies today.

The population pyramid for Iran has a number of distinctive features. Iran embraced family planning in the 1960's, one of the first developing countries to do so. The Islamic revolution of 1979, however, regarded the family planning program as "pro-West" and dismantled it. Moreover, the war with Iraq made population growth seem advantageous. When the war ended and reconstruction became the priority, the government reversed its policy and inaugurated a new family planning program with an extensive information campaign and, in 1993, powerful economic disincentives for having more than three children. These measures reduced the total fertility rate (see below) from 5.2 children in 1989 to 2.6 children in 1997. (This account is taken from Farzaneh Roudi, *Population Today*, July/August 1999). The jump in the birth rate following the revolution can be seen in the large size of the 15-19 year-old band (born 1979-1983) compared to the next older one; the subsequent curtailment of births shows up as a relatively small number of children 4 years old and younger. (Note: these population pyramids come from the U.S. Bureau for the Census International Database and were downloaded from the Population Bureau web site.)

## Iran, 1998



Source: U.S. Census Bureau, International Data Base.

### ***Influence of population age composition***

Since rates of most diseases and medical conditions, injuries, and health-related phenomena such as violence vary greatly by age, a population's age structure affects much more than its growth rate. As the 76 million "Boomers" in the post-World War II baby boom cohort to which President Bill Clinton belongs have moved up through the population pyramid, as a pig which has been swallowed by a python, they expanded school and university enrollments, created an employment boom first in obstetrics, pediatrics, construction, urban planning, and diaper services, and subsequently increased demand for baby clothes, toys, appliances, teachers, school buildings, faculty, managers, automobile dealers, health professionals, and investment counselors.

But in their wake, the Boomers have faced the contraction of many of those employment opportunities as their larger numbers and the smaller job-creating needs of the following generation increased competition at every stage. On the horizon are substantial increases in the need for geriatricians and retirement facilities, providing more employment opportunity for the generations that follow the Boomers but also a heavier burden for taxes and elder-care. A baby "boomlet" is also moving up the pyramid, as the Boomers' children create an "echo" of the baby boom.

The baby boom is a key contributor to the projected shortfalls in funding for Social Security, Medicare, and pensions in the coming decades. The following projections were made several years ago but are still relevant:

### When the baby boom cohort retires

	1995	2030
Retired population (%)	12	20
Workers per retired person	3.4	2.0
Combined Social Security and Medicare tax rate per worker (including employer's share)	15%	28%

(Source: Who will pay for your retirement? The looming crisis. Center for Economic Development, NY, NY. Summarized in TIAA-CREF quarterly newsletter the Participant, November 1995: 3-5.)

The U.S. is the fastest growing industrialized country, with a 1% growth rate (about 30% of which is due to immigration). Providing for the needs of senior citizens will be even more difficult in Europe, where most countries are now close to zero population growth and already 14% of the population are age 65 years or older. It has been projected that in 100 years there will be only half as many Europeans as today, which for many raises concerns about economic health, military strength, and cultural identity.

### **Sex composition**

Another fundamental demographic characteristic of a population is its sex ratio (generally expressed as the number of males per 100 females). A strongly unbalanced sex ratio affects the availability of marriage partners, family stability, and many aspects of the social, psychological, and economic structure of a society.

Sex ratios are affected by events such as major wars and large-scale migration, by cultural pressures that favor one sex, usually males, by unequal mortality rates in adulthood, and by changes in the birth rate. Because of higher male mortality rates, the sex ratio in the U.S. at birth falls from about 106 at birth, to about 100 by ages 25-29, and to 39 for ages 85 and above. Migration in search of employment is a frequent cause of a sex ratio far from 100. For example, oil field employment in the United Arab Emirates has brought that country's sex ratio as high as 218.

Although changes in birth rates do not alter sex ratios themselves, if women usually marry older men, a marked increase or decrease in the birth rate will produce an unbalanced sex ratio for potential mates. Girls born at and after a marked increase in the birth rate will encounter a deficit of mates in the cohort born before birth rates increased; boys born before a marked decrease will encounter a deficit of younger women. The substantial declines in birth rates in Eastern Europe following the collapse of Communism may lead to a difficult situation for men born before the collapse. In the U.S., casualties from urban poverty and the War on Drugs have created a deficit of marriageable men, particularly African American men. Because of assortive mating and the legacy of American apartheid, the effects of the deficit are concentrated in African American communities,

with many African American forced to choose between raising a family by themselves or remaining childless.

Women's status in society is a key factor in relation to the sex ratio and fertility in general. For example, women's opportunities for education and employment are strongly and reciprocally related to the birth rate. In China, where a "one-child" policy for urban families was adopted as a dramatic step toward curbing growth in its huge population, the sex ratio at birth is now 114 (a normal ratio is 105 boys for 100 girls). The approximately 12% shortfall of girls arises from families' desire for a male offspring and is believed to be due to a combination of sex-selective abortion, abandonment, infanticide, and underreporting (Nancy E. Riley, China's "Missing girls": prospects and policy. Population Today. February 1996;24:4-5).

### ***Racial, ethnic, and religious composition***

Race (a classification generally based on physical characteristics) and ethnicity (generally defined in relation to cultural characteristics), though very difficult to define scientifically, have been and continue to be very strong, even dominant factors, in many aspects of many societies. Thus, the racial, ethnic, and religious composition of a population is linked with many other population characteristics, as a function of the beliefs, values, and practices of the various groups and of the way societies regard and treat them. While people in the United States are most conscious of racial and ethnic issues in relation to African Americans, Latinos, Asian Americans, and Native Americans/American Indians, conflicts related to race, ethnicity, and religion are a major phenomenon throughout the world and throughout history, as the following VERY selective list recalls:

Balkans - Serbs, Croats, and Muslims (Bosnia), Serbs and Albanians (Kosovo)

Northern Ireland - Catholics and Protestants

Rwanda - Hutu's and Tutsi's

Middle East/Northern Africa - Jews, Christians, and Muslims

Iran's massacre of Bahai's

Kurds in northern Iran and Turkey

Indonesia - massacres of ethnic Chinese

East Timor

India/Pakistan - Hindus and Muslims

Europe - Christians and Jews (centuries of persecution climaxing though not ending with the Nazi's systematic extermination of over 6 million Jews, gypsies, and other peoples)

Germany - Catholics and Protestants (The Hundred Years War)

Americas - Europeans, white Americans, African Americans, and Native Americans/American Indians



The pervasiveness, strength, viciousness, and persistence of human reactions to differences in physical features, practices, beliefs, language, and other characteristics have had and will have powerful effects on public health.

## **Demographic concepts, measures, and techniques**

The discussion above uses many demographic terms, concepts, and measures. We now give precise definitions.

The (crude) **birth rate** is the number of births during a stated period divided by population size.

The (crude) **death rate** is the number of deaths during a stated period divided by population size.

Population-based rates are usually expressed per 100, 1000, 10,000, or per 100,000 to reduce the need for decimal fractions. For example, 2,312,132 deaths were registered in the United States in 1995, yielding a (crude) death rate was 880 per 100,000 population. This rate represented a slight increase over the preceding year's rate of 874 (Source: Anderson et al., *Report of final mortality statistics*, 1995. Monthly Vital Statistics Report 45(11) suppl 2, June 12 1997, National Center for Health Statistics (CDC), <http://www.cdc.gov/nchswww/data/mv4511s2.pdf>). Birth rates are generally expressed per 1,000 per year. For example, the lowest birth rates in the world are about 10, in several European countries; the highest are about 50, in several African countries.

When the numerator (deaths or births) in a given calculation is small, data for several years may be averaged, so that the result is more precise (less susceptible to influence by random variability). For example, taking the average number of births over three years and dividing by the average population size during those years yields a 3-year average birth rate. The average population size may be the average of the estimated population size for the years in the interval or simply the estimated population for the middle of the period (e.g., the middle of the year for which the rate is being computed). Where the population is growing steadily (or declining steadily), the mid-year population provides a better estimate than the January 1<sup>st</sup> or December 31<sup>st</sup> population size, so the mid-year population is also used for rates computed for a single year. Typical birth and death rate formulas are:

$$\text{Birth rate} = \frac{\text{Births during year}}{\text{Mid-year population}} \times 1,000$$

$$\text{Death rate} = \frac{\text{Deaths during year}}{\text{Mid-year population}} \times 1,000$$

$$\text{5-year average death rate} = \frac{\text{Deaths during the year period}}{\text{Population estimate for the middle of the third year}} \times 1,000$$

## **Fertility and fecundity**

An obvious limitation of the birth rate is that its denominator includes the total population even though many members (e.g., young children) cannot themselves contribute to births - and only women give birth. Thus, a general fertility rate is defined by including in the denominator only women of reproductive age:

$$\text{General fertility rate} = \frac{\text{Births during year}}{\text{Women of reproductive age (mid-year estimate)}} \times 1,000$$

Note that in English, *fertility* refers to actual births. *Fecundity* refers to the biological ability to have children (the opposite of sterility). In Spanish, however, *fecundidad* refers to actual births, and *fertilidad* (opposite of *sterilidad*) refers to biological potential (Gil, 2001).

## **Disaggregating by age**

A key consideration in interpreting overall birth, death, fertility, and almost any other rates is that they are strongly influenced by the population's age and sex composition structure. That fact does not make these "crude" overall rates any less real or true or useful. But failure to take into account population composition can result in confusion in comparing crude rates across populations with very different composition.

For example, the death rate in Western Europe (10) is higher than in North Africa (8). In other words, deaths are numerically more prominent in Western Europe than in North Africa. It would be a serious error, though, to interpret these rates as indicating that conditions of life and/or health care services are worse in Western Europe than in North Africa. The reason is that Western Europe would be expected to have a higher (crude) death rate because its population is, on the average, older (15% age 65 or above) than the population of North Africa (4% age 65 and above).

To enable comparisons that take into account age structure, sex composition, and other population characteristics, demographers (and epidemiologists) use specific rates (i.e., rates computed for a specific age and/or other subgroup - demographers call these "refined" rates). These specific rates can then be averaged, with some appropriate weighting, to obtain a single overall rate for comparative or descriptive purposes. Such weighted averages are called adjusted or standardized rates (the two terms are often used interchangeably; however, standardization is only one method for deriving adjusted rates). The United States age-adjusted death rate for 1995 was 503.9 per 100,000, slightly lower than the 507.4 age-adjusted death rate for 1994 (NCHS data in Anderson et al., 1997, see above). The reason that the age-adjusted death rate declined from 1994 to 1995 while the crude death rate increased is that the latter reflects the aging of the U.S. population, whereas the

former is adjusted to the age distribution of a “standard” population (in this case, the U.S. population for 1940).

### ***Total fertility rate (TFR)***

Standardization of rates and ratios will be explained later. But there is another important technique that is used to summarize age-specific rates. For fertility, the technique yields the ***total fertility rate*** (TFR) -- the average number of children a woman is expected to have during her reproductive life. The average number of children born to women who have passed their fecund years can, of course, be obtained simply by averaging the number of live births. In contrast, the TFR provides a projection into the future.

The TFR summarizes the fertility rate at each age by projecting the fertility experience of a cohort of women as they pass through each age band of their fecund years. For example, suppose that in a certain population in 1996 the average annual fertility rate for women age 15-19 was 110 per 1000 women, 180 for women age 20-29, and 80 for women 30 years and older. The TFR is simply the sum of the annual fertility rate for each single year of age during the fecund years. So 1,000 women who begin their reproductive career at age 15 and end it at age 45 would be expected to bear:

**Calculation of total fertility rate (TFR)**  
**For 1000 women from age 15 through age 45 years**

Age	Births	
15	110	
16	110	
17	110	(average annual fertility
18	110	from ages 15-19 = 110/1000)
19	110	
20	180	
21	180	
22	180	(average annual fertility
...		from ages 20-29 = 180/1000)
29	180	
30	80	
31	80	(average annual fertility
...		from ages 30-45 = 80/1000)
44	80	
45	80	
	3,630	

or about 3.6 children born to each woman.

(This TFR could also be calculated more compactly as  
 $110 \times 5 + 180 \times 10 + 80 \times 16 = 3,630$ )

Note that the TFR is a hypothetical measure based on the assumption that the age-specific fertility rates do not change until the cohort has aged beyond them. The TFR is a projection, not a prediction – essentially, a technique for summarizing a set of age-specific rates into an intuitively meaningful number.

## **Life expectancy**

The technique, of using current data for people across a range of ages to project what will happen to a person or population who will be passing through those ages, is also the basis for a widely-cited summary measure, life expectancy. Life expectancy is the average number of years still to be lived by a group of people at birth or at some specified age. Although it pretends to foretell the future, life expectancy is essentially a way of summarizing a set of age-specific death rates. It thus provides a convenient indicator of the level of public health in a population and also a basis for setting life insurance premiums and annuity payments.

In order to understand life expectancy and several other demographic summary measures, such as the TFR, it is important to appreciate the difference between these demographic summary measures and actual predictions. A prediction involves judgment about what will happen in the future. Life expectancy and TFR's are simply ways of presenting the current experience of a population. Thus, my prediction is that most of us will live beyond our life expectancy!

The explanation for this apparent paradox is that life expectancy is a representation of age-specific death rates as they are at a given time. If age-specific death rates remain constant, then life expectancy today will be an excellent estimate of the average number of years we will live. However, how likely are today's age-specific death rates to remain constant? First, we can anticipate improvements in knowledge about health, medical care technology, and conditions of living to bring about reductions in death rates.

Second, today's death rates for 40-90 year-olds represent the experience of people who were born during about 1900-1960. Today's over-forties Americans lived through some or all of the Great Depression of the 1930s, two world wars, the Korean War, the Vietnam War, atmospheric nuclear bomb testing, unrestrained DDT use, pre-vaccine levels of mumps, polio, measles, rubella, chicken pox, pre-antibiotic levels of mycobacterium tuberculosis, syphilis, and other now-treatable diseases, varying levels of exposure to noxious environmental and workplace substances, a system of legally-enforced apartheid in much of the nation, limited availability of family planning, and lower general knowledge about health promotive practices, to list just a smattering of the events and conditions that may have affected subsequent health and mortality. Although changes in living conditions are not always for the better (death rates in Russia and some other countries of the former Soviet Union have worsened considerably following the breakup of the Soviet Union), the United States, Western Europe, Japan, and many countries in the developing world can expect that tomorrow's elderly will be healthier and longer-lived than the elderly of the previous generation. For these reasons life expectancy, computed from today's age-specific death rates, most likely underestimates the average length of life remaining to those of us alive today.

Since it is a summary of a set of age-specific mortality rates, life expectancy can be computed from any particular age forward. Life expectancy at birth summarizes mortality rates across all ages. Life expectancy from age 65 summarizes mortality rates following the conventional age of retirement. Accordingly, life expectancy at birth can be greatly influenced by changes in infant mortality and child survival. The reason is that reductions in early life mortality typically add many more years of life than reductions in mortality rates for the elderly. The importance of knowing the age from which life expectancy is being computed is illustrated by the following excerpt from a column

prepared by the Social Security Administration and distributed by Knight Ridder / Tribune News Service (*Chapel Hill Herald*, June 28, 1998: 7):

*Q. I heard that the Social Security retirement age is increasing. Is this true and if so, why?*

A. Yes, it's true. When Social Security was just getting started back in 1935, the average American's life expectancy was just under age 60. Today it's more than 25 percent longer at just over 76. That means workers have more time for retirement, and more time to collect Social Security. And that's why Social Security's retirement age is gradually changing ... to keep pace with increases in longevity. A worker retiring today still needs to be age 65 to collect full benefits, but by 2027, workers will have to be age 67 for full retirement benefits.

It is certainly the case that longevity today is much greater than when the Social Security system was begun, so that it is now expected to provide support over a much larger fraction of a person's life. However, the life expectancies cited are life expectancies from birth. Although children who die obviously do not collect retirement benefits, neither do they make contributions to Social Security based on their earnings. For Social Security issues, the relevant change in life expectancy is that from age 62 or 65, when workers become eligible to receive Social Security retirement benefits. Every year's increase in life expectancy beyond retirement means an additional year of Social Security benefits. This life expectancy (now 15.7 and 18.9 years, respectively, for U.S. males and females age 65 years) has also increased greatly since 1935.

### ***Life expectancy computation and the current life table***

Life expectancy is computed by constructing a demographic life-table. A demographic life table depicts the mortality experience of a cohort (a defined group of people) over time, either as it occurs, has occurred, or would be expected to occur. Imagine a cohort of 100,000 newborns growing up and growing old. Eventually all will die, some as infants or children, but most as elderly persons. The demographic life table applies age-specific risks of death to the surviving members of the cohort as they pass through each age band. Thus, the demographic life table (also called a current life table) is a technique for showing the implications on cohort survival of a set of age-specific death rates.

**Excerpt from the U.S. 1993 abridged life table  
(total population)**

Age interval (years)	Risk Of Death	Number still alive	Deaths
$x-x+n$	${}_nQ_x$	$l_x$	${}_nD_x$
(A)	(B)	(C)	(D)
<= 1 yr	.00835	100,000	835
1-5	.00177	99,165	176
5-10	.00106	98,989	105
10-15	.00126	98,884	125
15-20	.00431	98,759	426
20-25	.00545	98,333	536
25-30	.00612	97,797	599
30-35	.00797	97,198	775
35-40	.01031	96,423	994
40-45	.01343	95,429	1,282
45-50	.01842	94,147	1,734
50-55	.02808	92,413	2,595
55-60	.04421	89,818	3,971
60-65	.06875	85,847	5,902
65-70	.10148	79,945	8,113
70-75	.14838	71,832	10,658
75-80	.21698	61,174	13,274
80-85	.32300	47,900	15,472
>= 85 yr	1.00000	32,428	32,428

(Source: National Center for Health Statistics)

(The algebraic symbols beneath the column headings show traditional life table notation; “x” refers to the age at the start of an interval, “n” to the number of years of the interval.)

For example, here are the first four columns of the U.S. 1993 abridged life table, from the National Center for Health Statistics web site (“abridged” means that ages are grouped rather than being listed for each individual year). The table begins with a cohort of 100,000 live births (first line of column C). For each age interval (column A), the cohort members who enter the interval (column C) are subjected to the risk of dying during that age interval (column B), producing the number of deaths shown in column D and leaving the number of survivors shown in the next line of column B. Thus, in their first year of life, the 100,000 live newborns experience a risk of death of 0.00835 (835/100,000), so that 835 die (B x C) and 99,165 survive (B - D) to enter age interval 1-5 years.

Between ages one and five, the 99,165 babies who attained age one year are subjected to a five-year risk of death of 0.00177 (177/100,000), so that 176 die ( $0.0017 \times 99,165$ ) and 98,989 ( $99,165 - 176$ ) attain age six.

Notice that the age-specific risks of death (proportion dying, column B) increase from their lowest value at age 5-10 years, at first gradually, then increasingly steeply until during the age interval 80-85 nearly one-third of cohort survivors are expected to die. Correspondingly, the numbers in column D (deaths) also increase gradually, then more steeply—but not quite as steeply as do the risks in column B. The reason is that the actual number of deaths depends also on the number of people at risk of death (survivors, column C) which drops gradually at first, then more and more rapidly as the risks increase. Notice also the very high risk of death for infants: the 0.0085 means that 835 of 100,000 infants—nearly 1% --die during just one year. In contrast, only 177 of the surviving infants die during the following four years.

### ***Death risks versus death rates***

An important technical issue to consider at this point is that the risks in column B are not the same as the age-specific death rates discussed above, though the latter are the basis for deriving the risks in column B. There are two reasons. First, all but the first two of the values in column B show the risk for a five-year interval. Second, an (annual) death rate is an average value over an interval, based on the average population at risk for the interval, typically estimated by the mid-year population (which is why such death rates are called “central death rates”). In contrast, the risks in column B apply not to the average population or mid-year population but to the population at the start of the interval, which in a life-table is always greater than the average population size during the interval.

Assume that the death rate during an age interval remains fixed, so that the cohort experiences deaths during each month of the interval. Cohort members who die in the first months of the interval are obviously no longer at risk of dying later during the interval. A decreasing population with fixed death rates means that the number of deaths in each month of the interval also decreases. The calculation of the risk for the interval takes into account the fact that the cohort shrinks during the interval. At young ages, when age-specific death rates are small, the shrinkage is slight so the one-year risk is very close to the annual death rate and the five-year risk is very close to five times the average annual death rate. But at older ages, substantial shrinkage occurs and the risk is therefore less than the number of years times the average annual death rate.

To illustrate:

During infancy, the cohort loses 835 members, so that it shrinks from 100,000 to 99,165. The average or mid-year population, then, is approximately  $(.5)(100,000 + 99,165)$  or, equivalently,  $100,000 - .5(835) = 99,582.5$ . This number is very close to 100,000, so it is easy to see why the death rate during the first year (835 deaths divided by  $99,582.5 = 0.00839$ ) is almost identical to the first-year risk (0.00835). Similarly, during the next four years (ages 1-5), the average annual death rate during the interval is approximately 0.000444 (176 deaths/4 years, divided by 99,077, the average population during the interval). Multiplying this rate by four years gives 0.00178, nearly identical to the four year risk (0.00177).



At the other end of the life table, cohort size loses 15,472 members, declining from 47,900 at age 80 to 32,428 at age 85. The average annual death rate is 0.07704 (15,472 / 5 years divided by the average size of the cohort, 40,164). Multiplying this rate by five years gives 0.38522, which is considerably greater than the five-year risk in column B (0.32300). We can come much closer to this five-year risk if we treat the five-year interval like a miniature life-table by dividing up the five-year interval into single years and applying the average annual death rate (0.07704) to each year of the cohort:

Age	Annual Death Rate	Proportion surviving that year	Cumulative Proportion surviving	Cumulative Proportion dying
80-81	0.07704	0.92296	0.92296	0.07704
81-82	0.07704	0.92296	0.85186	0.14814
82-83	0.07704	0.92296	0.78623	0.21377
83-84	0.07704	0.92296	0.72566	0.27434
84-85	0.07704	0.92296	0.66975	0.33025

The cumulative 5-year risk calculated from the cumulative proportion dying comes very close to the value figure in column B of the table (0.32300). If we divide each year into 12 months, or 52 weeks, or 365.25 days, the life-table-type calculation comes even closer. (Using calculus, it can be shown that in the limit, as the number of units becomes infinite and their size approaches zero, the life-table computation of the 5-year =  $1 - \exp(-5 \times 0.07704) = 0.3197$ .)

### ***Deriving life expectancies***

Now we present the rest of the NCHS (abridged) U.S. 1993 life table, by including its three right-most columns.

Column E shows the sum of the number of years lived by all members of the cohort during each age interval. During a five-year interval, most cohort members will live for five years, but those who die during the interval will live fewer years. During the lowest risk five years (ages 5-10), nearly all of the 98,989 cohort members who enter the interval (column C) will live 5 years, for a total number of years of life of 494,945, which is just slightly above the value in column E. Between ages 80 and 85, however, only about two-thirds of the entering cohort live all five years, so the number in column E (201,029) is much lower than five times column C (239,500). However, if we use the average population size (40,164) to estimate years of life lived during ages 80-85, we obtain  $5 \times 40,164 = 200,820$ , which is very close to the number in Column E. (The numbers in column E also can be explained in terms of the concept of a “stationary population”.)

The next column (F) gives the sum of the number of years of life during the specific age interval and the remaining intervals. For example, the 395,851 total years of life remaining for the cohort members who attain age 80 are the sum of the 201,029 years to be lived during 80-85 plus the 194,822 years left for those who survive to age 85. The 669,345 years for cohort members reaching age 75 are the sum of the 273,494 years to be lived during the age 75-80 interval plus the 395,851 years remaining for members who reach age 80.

**U.S. 1993 abridged life table (total population)**  
**(Source: National Center for Health Statistics)**

Age Interval (years)	Risk of death	Number still alive	Deaths	Years lived	Years remaining	Life expectancy
x-x+n	${}_nQ_x$	$l_x$	${}_nD_x$	${}_nL_x$	$T_x$	
(A)	(B)	(C)	(D)	(E)	(F)	(G)
<= 1 yr	.00835	100,000	835	99,290	7,553,897	75.5
1-5	.00177	99,165	176	396,248	7,454,607	75.2
5-10	.00106	98,989	105	494,659	7,058,359	71.3
10-15	.00126	98,884	125	494,177	6,563,700	66.4
15-20	.00431	98,759	426	492,829	6,069,523	61.5
20-25	.00545	98,333	536	490,352	5,576,694	56.7
25-30	.00612	97,797	599	487,486	5,086,342	52.0
30-35	.00797	97,198	775	484,098	4,598,856	47.3
35-40	.01031	96,423	994	479,771	4,114,758	42.7
40-45	.01343	95,429	1,282	474,168	3,634,987	38.1
45-50	.01842	94,147	1,734	466,717	3,160,819	33.6
50-55	.02808	92,413	2,595	455,985	2,694,102	29.2
55-60	.04421	89,818	3,971	439,733	2,238,117	24.9
60-65	.06875	85,847	5,902	415,279	1,798,384	20.9
65-70	.10148	79,945	8,113	380,318	1,383,105	17.3
70-75	.14838	71,832	10,658	333,442	1,002,787	14.0
75-80	.21698	61,174	13,274	273,494	669,345	10.9
80-85	.32300	47,900	15,472	201,029	395,851	8.3
>= 85	1.00000	32,428	32,428	194,822	194,822	6.0

Life expectancy, then, the average number of years of life remaining after a given age, is the total years of life left (column F) divided by the number of cohort members who have attained that age (column C). Since the cohort numbers 100,000 at birth, life expectancy at birth is simply  $7,553,897 / 100,000 = 75.5$ . The 89,818 cohort members who attain age 55 years have a total of 2,238,117 total years of life remaining, or an average of 24.9 years.

An advantage of surviving is that the average age the cohort will expect to attain keeps rising also. Fifty-year-olds have an average life expectancy of 29.2, for an expected age at death of 79.2; 70-year-olds have an average life expectancy of 14.0, for an expected age at death of 84 years. The reason, of course, is that cohort members who live shorter lives bring down the average; when they drop out the average is reduced by less than the number of years of the interval.

Because of the method of computation, the mathematical structure of life expectancies is not readily apparent. Life expectancies are simply the average number of years lived across a range of ages. For example, if in the imaginary cohort of 100,000 people, the number who remain alive to the end of each year of age is  $n_i$ , then (ignoring for the moment the fact that people generally die during a year rather than at the very end) the total number of years lived from age x to age y is simply the sum of

the  $n_i$  from age  $x$  to  $y$ , or  $\sum (n_i)$  where the summation runs from  $x$  to  $y$ . The life expectancy to age  $y$  for someone age  $x$  is  $\sum (n_i)/n_x$ .

From this we can see several relations among life expectancies:

1. The life expectancy for a population can be expressed as a weighted average of the life expectancy for men and the life expectancy for women;
2. The life expectancy from birth can be expressed as a weighted average of the life expectancy at birth for people who die by a given age and the life expectancy at birth for people who die after that age.

For example, consider the “paradox” that the life expectancy at birth in 2000 in the U.S. was 76.9 years and at age 65 was 17.9 years. So 65 year-olds in 2000 could “expect” to live to age 82.9 (65+17.9) years of age if death rates remained constant. But that figure is much greater than their life expectancy at birth (76.9 years). Based on the second relation above, if life expectancy from birth is a weighted average, then unless the life expectancies into which it can be composed are equal, some will be smaller and some will be greater. Life expectancy at birth for those who die by 65 will certainly be lower than the overall life expectancy at birth, so life expectancy at birth for those who die after age 65 must be greater. Life expectancy at birth for those who live beyond age 65 will be 65 years plus life expectancy at age 65. For an algebraic demonstration of this relation, see the appendix.

## **Cohort life tables**

Because the current life table uses risks derived from current (or recent) death rates at each age, the life expectancies are simply a technique for summarizing them more meaningfully than if we took a simple average of age-specific death rates. Of course, in actual fact, age-specific death rates are likely to change, hopefully to decline. If they do, then by the time a cohort of newborns reach age 20, they will experience not the 1993 death rates for 20-year-olds but those in effect in 2013. Similarly, they will experience the death rates for 30-year-olds in effect in 2023, for 40-year-olds in 2033, and so forth.

The cohort life table is constructed to take account of changing death rates. Of course, if such a life table is to be based on observed death rates, it can apply only to a cohort born sufficiently in the past. If, for example, we create a cohort life table for persons born in 1880, then we can use the observed death rates for the appropriate age for each year or interval beginning in 1880. Average years of life remaining at each age of a life table constructed from historical death rates summarizes the actual mortality experience of past birth cohorts. In epidemiology, cohort life tables are used much more often than current life tables, because the life table technique is often useful for analyzing data collected during the follow-up of a cohort (some authors call these follow-up life tables).

The cohort in a current or cohort life table loses members only to death, so that everyone who survives an interval is included in the next one. The cohorts studied by epidemiologists, on the other hand, can lose members who become lost to follow-up so that their vital status cannot be

determined. Moreover, epidemiologists usually study outcomes other than all-cause mortality, so epidemiologic cohorts lose members who migrate or withdraw from the study or who become ineligible to have the outcome of interest (e.g., due to such reasons as death from another cause, surgical removal of an organ prior to the development of the disease of interest, or discontinuance of a medication being studied). In addition, the members of an epidemiologic cohort may not enter the cohort at the same calendar time or age.

A follow-up life table provides a way of representing and analyzing the experience of an epidemiologic cohort. In one common type of follow-up life table, people being studied are entered into the cohort on the basis of an event, such as employment, illness onset, surgery, attaining age 18, or sexual debut, and are then followed forward in time. Their time in the cohort (and in the life table) is computed with respect to their enrollment event. At each time interval following initiation of follow-up, the number of outcomes observed is analyzed in relation to the cohort members whose status is observed for all or part of the interval. Where the precise time of follow-up for each cohort member is unknown, then some intermediate number is used, in analogy to the use of the mid-year population for a central death rate.

## **Cohort effects**

The life table and the TFR are both based on the concept of a cohort proceeding through time, and both employ the assumption that age-specific rates remain constant. In actuality, of course, age-specific rates do change over secular time, and populations are composed of many cohorts, not only one. Since age, secular time, and cohort are fundamentally tied to one another - as time advances, cohorts age - it can be difficult to ascertain whether an association with one of these aspects of time reflects the influence of that aspect or of another.

When we look at a single age-specific rate for a given year, we have no indication of the extent to which that rate reflects the influences of chronological age, calendar time-associated changes in the social and physical environment, or characteristics of the cohort that happens to be passing through that age during that year. Even if we look at a given age interval across a span of calendar years or at multiple ages in a given year, there is no way for us to know whether what appear to be changes associated with aging or the passage of time are really reflections of the characteristics of different cohorts (i.e., characteristics acquired due to environmental experiences at a formative period of life, such as exposure to lead in infancy or to radiation in adolescence).

Attempts to disentangle the interwoven effects of age, secular time, and cohort are referred to as “age-period-cohort” analyses. The most straightforward approach involves assembling data from more than one period and from a broad range of ages, and then examining the data in relation to age, period, and cohort. For example:

**Age-period-cohort analysis of mean serum cholesterol (mg/dL, hypothetical data)**

60-69	200 <sup>A</sup>	210 <sup>B</sup>	235 <sup>C</sup>	240 <sup>D</sup>	<u>230</u> <sup>E</sup>
50-59	205 <sup>B</sup>	230 <sup>C</sup>	235 <sup>D</sup>	<u>225</u> <sup>E</sup>	215 <sup>F</sup>
40-49	240 <sup>C</sup>	230 <sup>D</sup>	<u>220</u> <sup>E</sup>	210 <sup>F</sup>	200 <sup>G</sup>
30-39	225 <sup>D</sup>	<u>215</u> <sup>E</sup>	205 <sup>F</sup>	195 <sup>G</sup>	185 <sup>H</sup>
20-29	<u>210</u> <sup>E</sup>	200 <sup>F</sup>	190 <sup>G</sup>	180 <sup>H</sup>	170 <sup>I</sup>
	1950-59	1960-69	1970-79	1980-89	1990-96

Birth cohorts:

A - 1890-1899	D - 1920-1929	G - 1950-1959
B - 1900-1909	E - 1930-1939 (underlined)	H - 1960-1969
C - 1910-1919	F - 1940-1949	I - 1970-1979

From the columns (calendar decades), it appears that serum cholesterol increases by 15 mg/dL per decade of age. If we had only one calendar decade of data, this observation is all that we can make, leading us to overstate the relationship between age and cholesterol. With the full data, we can follow the birth cohorts longitudinally, which reveals that for a given cohort cholesterol rises by 5 mg/dL per decade of age, but that also each new cohort has 10 mg/dL lower average cholesterol than the previous one.

This observation can be labeled a “cohort effect” and has the capability to confuse interpretation of cross-sectional (one point in time) data. (The reason that the 15 mg/dL increase does not continue at the older ages in the earlier decades is that I decided to precede the secular decline in cholesterol with a secular rise, so that the earliest cohorts had lower cholesterol levels than the ones that came afterwards.)

Thought question: Professors typically comment that with each entering class (i.e., cohort), students seem to be younger. Is this an effect of age, secular time, or cohort? (See bottom of page for the answer.)

## Appendix — Weighted averages of life expectancies

DRAFT

### **Es**

Life expectancy equals total years lived divided by the population starting the period . . .

## Bibliography

Colton, Theodore. in *Statistics in Medicine*, pp. 237-250.

De Vita, Carol J. The United States at mid-decade. *Population Bulletin*, Vol. 50, No. 4 (Washington, D.C.: Population Reference Bureau, March 1996).

Falkenmark, Malin and Carl Widstrand. Population and water resources: a delicate balance. *Population Bulletin*, Vol. 47, No. 3 (Washington, D.C.: Population Reference Bureau, November 1992).

Gil, Piédrola. *Medicina preventiva y salud pública*. 10<sup>a</sup> edición. Barcelona, España, Masson., 2001. (thanks to Maria Soledad Velázquez for this information)

Lutz, Wolfgang. The future of world population. *Population Bulletin*, Vol. 49, No. 1, June 1994.

McFalls, Joseph A., Jr. Population: a lively introduction. *Population Bulletin*, Vol. 46, No. 2 (Washington, D.C.: Population Reference Bureau, October 1991)

McMichael, Anthony J. Planetary overload. Global environmental change and the health of the human species. NY, Cambridge, 1993.

Mosley, W. Henry and Peter Cowley. The challenge of world health. *Population Bulletin*, Vol. 46, No. 4 (Washington, D.C.: Population Reference Bureau, December 1991).

Petersen, William. *Population* 2<sup>nd</sup> ed. Macmillan, London, 1969

Remington, Richard D. and M. Anthony Schork. Statistics with applications to the biological and health sciences. Englewood Cliffs, NJ, Prentice-Hall, 1970.

Roudi, Farzaneh. Iran's revolutionary approach to family planning. *Population Today*, July/August 1999; 27(7): 4-5. (Washington, D.C.: Population Reference Bureau)

Web sites: Data and publications are now widely available over the World Wide Web. A list of useful sites (e.g., [www.ameristat.org](http://www.ameristat.org), [www.cdc.gov](http://www.cdc.gov) and [www.cdc.gov/nchswww](http://www.cdc.gov/nchswww), [www.who.int](http://www.who.int)) is available at [www.sph.unc.edu/courses/epid168/](http://www.sph.unc.edu/courses/epid168/).

**Answer:** Age - the aging of the professors!