

CHAPTER
NINE

THE LIFE TABLE

An Introduction

If you are not familiar with the life table, you are about to be introduced to one of demography's most powerful and most interesting tools. This apparently simple aid to presenting mortality information has been known and used for generations, both within and outside of academic circles. We say *apparently* simple because, although it is not difficult to understand or even to construct a table (once you know the procedure), the more you contemplate and use it, the more it reveals about the dynamics of populations and the principles of population science. As we noted in Chapter 1, the origins of the life table extend back to the very foundations of demography in the seventeenth century, and to the work of Francis Place, John Graunt, and Edmund Halley, in particular. From that era to ours, most of the leading demographers have contributed to its refinement and application (see Namboodiri 1987).

This chapter is intended to provide an overview of the life table and its applications; principally so that you can understand its basic structure and functions, but also to point the way toward more advanced work on the subject. Anyone familiar with the life table is well aware of its practical value, in the fields of public health, actuarial science (Underwood 1950; Bell and Wade 1998), and in other areas as well. Toward the end of the chapter we provide some illustrations taken from realms that are somewhat remote from medicine and actuary, and that may at first seem unusual contexts in which this tool may be applied. In this same applied spirit, Chapter 10, which follows, demonstrates the use of life-table concepts and functions in developing projections of populations to future dates.

Before turning to these practical matters, however, we begin with some definitions and illustrations of the principles of the life table. Next, we define and derive the main functions: the surviving population, the **stationary population**, and so on. These explorations of the foundations of life table analysis are then used as the basis for considering several refinements: **abridged** and **unabridged tables**, survival rates, period versus cohort accounts, and the **Lexis diagram**. With these matters touched upon, we then move to the chapter's conclusion and to the topic of life table applications.

MEASURING MORTALITY

In our discussion of measures of mortality in Chapter 6, we began with the crude death rate (CDR), but soon observed that its value is limited. You will recall that this rate is derived by first establishing the number of deaths, which we will here label “D,” that occur in a particular population during a time period: one year, five years, 10 years, and so on. This total is then divided by the size of the population at midinterval (for example, midyear), symbolized by “P.” The resulting ratio is labeled “M” (for the force of mortality), and it is multiplied by 1,000 to produce the CDR: $D/P = M$; $M \times 1,000 = \text{CDR}$. For example, in the United States during 2010, there were approximately 2.47 million deaths throughout the country. The midyear population size, as established by the decennial enumeration, was 308.75 million. Solving for M and CDR, we find:

$$M = D/P = 2.47/308.75 = .008; \text{CDR} = M \times 1,000 = 0.008 \times 1,000 = 8$$

The symbol M expresses the probability that an individual alive at midyear (or other interval to which the data apply) will die at some point during the interval. Another way to think about what M measures is that it characterizes the average risk of death to which members of the population are exposed during the year. It is certainly an important and widely used variable, and one for which data are readily available.

However, M and CDR in these expressions are by no means the most accurate measures of the force of mortality. As noted in Chapter 6, they “average out” relevant characteristics of groups whose exposure rates are known to differ, principally age cohorts and the two sexes. Thus, if population A has a higher CDR than population B, it might be because A is a higher-risk aggregate, more prone to disease or other kinds of factors that cause death. Or it might be because A has more members in high-risk categories such as very young or very old age groups, even though overall it is less prone to risk factors than B. Moreover, because M and CDR do not account for age, they are insensitive to the differences and interactions between age in years, on one hand, and calendar date, on the other. For risk is related not only to whether or not a person is, say, 30 years old, it also depends on *when* the person is 30, in 1905, in 1955, or in 2015.

This chapter continues along these lines, analyzing and dissecting M and CDR into increasingly refined components. As we discuss the life table, its construction, and its derivations, it will help to keep this introductory discussion in view. If every individual in a given population had the same chances of dying as every other, regardless of age, or gender, or social characteristics, there would be no need for life table analysis. But, of course, risk factors vary significantly from person to person and group to group. Granting the many purposes to which the life table may be put, it is essentially designed to account for such variations, and to correct for the deficiencies of crude measures of mortality.

Definitions and Illustrations

A life table is a rectangular array of demographic information, consisting of rows and columns. In mathematics, such an array is called a **matrix** and each row and column is referred to as a **row vector** and a **column vector**, respectively. In addition to the single rows and columns, there are other ways to subdivide a life table into

parts that contain several rows, several columns, or both. The process of subdividing is called *partitioning*. Another important distinction made in life table analysis is that between a period and a cohort (or generation) table. As in the case of fertility measures (Chapter 7), the former contains information for one date, or period, and refers to all cohorts at that date; the latter contains information about one cohort as it ages through several dates. In this section, only period tables are considered.

Observed Data

The numbers that make up a (period) life table refer to age structure, mortality, and usually sex structure. These are of two types: (1) data based on observations of an actual population at a given date, such as the U.S. population in the year 2015; and (2) sums, proportions, and probabilities derived from the actual population data. The first category includes the following items (see Figure 9.1)

- the ages that at least one person is observed to have attained by midyear of the relevant date, recorded as 1-year, 5-year, 10-year, and so on, cohorts and symbolized by “ x ”;
- the number of persons alive at midyear for each age, symbolized as “ ${}_nK_x$ ”; and
- the number of deaths that occurred among persons at each age during the course of the relevant year, symbolized by “ ${}_nD_x$.”

In general, these four data vectors occupy the first four columns of the array. Otherwise, the second and third are listed separately and do not even appear in the life table as such. In this chapter, we will keep them in the first three columns so that their relationship to the remaining vectors can readily be seen. The little “ n ” notation that precedes the K_x s and D_x s is included in some of the other **life table functions** and variables as well. It indicates the size of the cohort, such that ${}_5K_x$ means that we have a five-year cohort and ${}_5K_{10}$ includes all individuals between 5 and 10 years of age. Similarly, ${}_5K_{45}$ refers to everyone who has reached their 45th birthday but not their 50th at midyear, and so on. When cohorts are one year in length, the “ n ” is ordinarily omitted, and we simply use K_x , D_x , and the like. Thus, we know that D_0 refers to the number of deaths to individuals below age one (infant deaths), without having to write “ ${}_1D_0$.” Thus in most tables, a number is included under the symbol “ n ” to indicate the sizes of the cohorts: 1, 5, 10, and so on.

If we sum all of the entries in the second column to produce $\sum {}_nK_x$, we have the total population size—or the total size of the aggregate of concern, such as all females. Solving for $\sum {}_nD_x$ gives us the total number of deaths. So, the quotient, referred to as the “central death rate.”

$$(\sum {}_nD_x / \sum {}_nK_x) = {}_nM_x; \text{ and } \sum {}_nD_x / \sum {}_nK_x \times 1,000 \text{ is the CDR}$$

The fact that we need to build *up* to find the CDR should already suggest how the life table dissects gross statements of risk to create smaller and more realistic indicators.

The column following the ${}_nK_x$ vector does not contain directly observed data. Rather, each entry, ${}_nq_x$, is a probability that is based on the direct observations of each ${}_nD_x$, ${}_nK_x$, and therefore the ${}_nM_x$ vector. This variable plays a pivotal role. We have just noted that often the second and third columns are not displayed with the rest of the life table. This is because they do contain empirical information whereas the other columns do not, and it is held that the two types of entries should not

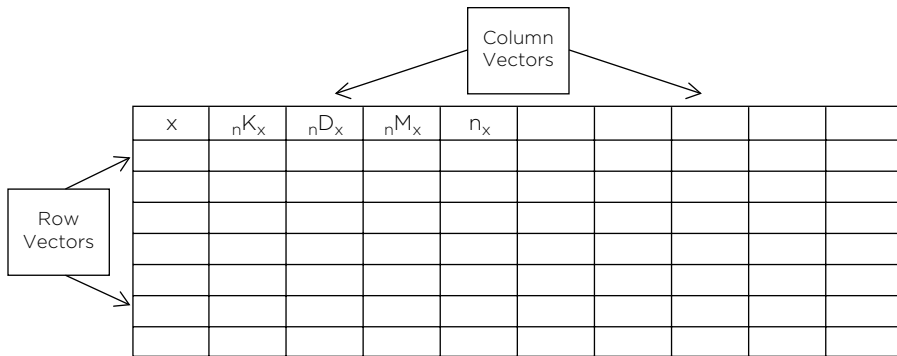


Figure 9.1. Basic Outline of Life Table

be confused with one another. However, ${}_nq_x$ is almost always included in the table, despite the fact that it is more closely related to the observed data than the other functions. The reason for this special treatment will become obvious in a moment.

The ${}_nM_x$ s give the probability of dying at age x , and they are derived in the following two steps: (1) We divide an element in the third column, an ${}_nD_x$, by its corresponding entry in the second column, ${}_nK_x$. This gives us ${}_nM_x$, the observed age-specific death rate—not multiplied by 1,000. (2) We solve for ${}_nq_x$ by assuming that the deaths to persons age x occur on the average at age $x^{1/2}$ years.

That is, we spread the deaths evenly between those who have just reached the given age and those who are about to reach the next age. Spreading the deaths evenly is done for the sake of convenience, despite the fact that the younger the infant—or an individual in any age category—the greater the probability of dying. A more precise application would use a “separation factor,” a set of numbers that add to 1.0. These would include .7 and .3, or .6 and .4, rather than the .5 and .5 indicated here; see Table 9.1. The formula for this is:

$${}_nq_x = (2 \times {}_nM_x) / (2 + {}_nM_x)$$

Table 9.1. Separation Factors for Ages 0 and 1-4, Selected European Countries

		Separation Factor for Age 0			Separation Factor for Age 1-4		
	Zones	Men	Women	Both Sexes	Men	Women	Both Sexes
Infant Mortality Rate > 0.100	I	0.33	0.35	.3500	1.558	1.570	1.5700
	II	0.29	0.31	.3100	1.313	1.324	1.3240
	III	0.33	0.35	.3500	1.240	1.239	1.2390
Infant Mortality Rate < 0.100	I	.043	0.05	.0500	1.859	1.733	1.7330
	II	.0025	0.01	.0100	1.614	1.487	1.4870
	III	.0425	0.05	.0500	1.541	1.402	1.4020

(I) Iceland, Norway, and Switzerland; (II) Austria, Czechoslovakia, North-Central Italy, Poland, and Hungary; (III) South Italy, Portugal, and Spain

Sources: (1) PAHO. Technical Information System: Regional Mortality Database, AIS, Washington, D.C., 2003. (2) United Nations Population Division, World Population Prospects: The 2002 Revision, New York, 2003.

For example, the U.S. Census estimated that a total of 3,570,923 persons (male and female) ages 20–21 were living in the United States at midyear 1996; and vital statistics indicate that a total of 3,716 deaths occurred in that cohort during the year. This produces a death rate of:

$$(1) M_{20} = D_{20}/K_{20} = 3,716/3,570,923 = 0.001040$$

Next, we find the probability of dying at that age,

$$(2) q_{20} = (2 \times M_{20})/(2 + M_{20}) = (2 \times 0.001040)/(2 + 0.001040) = 0.00208/2.001040 = .001039$$

To review, life table analysis begins with observed data arranged into columns, as follows:

Column	1	2	3	4	5	6
Variable	Age	Number of persons	Number of deaths	Size of cohort	Central death rate	Probability of dying
Symbol	x	${}_nK_x$	${}_nD_x$	n	${}_nM_x$	${}_nq_x$
Data entries for each variable						

Occasionally, the ${}_nM_x$ vector is shown in the table, following the ${}_nD_x$ s or at another point. This is ordinarily not done with unabridged tables, because q_x and M_x are typically very close in value. However, with abridged tables, in which mortality over several successive years is treated as a single entry, the ${}_nM_x$ and ${}_nq_x$ vectors can differ considerably and the former is included more often.

The Life Table-Proper

Beginning with the fifth column, we find a set of symbols known as the *life table functions*, which are derived from the ${}_nq_x$ s. These functions create a model population that behaves as an actual population would if certain conditions were to prevail.¹ We emphasize that it is a model to distinguish it from the aggregate from which the ${}_nK_x$ s, the ${}_nD_x$ s, and thus the ${}_nq_x$ s were taken, although it is closely tied to the actual population. Therefore, the fourth column divides the life table into two parts, the observed data and the life table-proper.

When we move from column 4 to column 5 and beyond, we say that we are “entering the life table,” even though, loosely speaking, the first four columns are also part of the array. The major life table functions are: ${}_n l_x$, the surviving population;² ${}_n d_x$, the number of deaths in the surviving population; ${}_n L_x$, the number of person years lived, or the life table population; ${}_n T_x$, the number of life-years remaining, and, **average life expectancy**. Following are the life table functions as such:

Column	7	8	9	10	11
Variable	Number of survivors to age x	Number of deaths at age x	Number of years lived between l_{x-n} and l_x	Total years lived between 0 and x	Average number of additional years expected at x and beyond
Symbol	${}_n l_x$	${}_n d_x$	${}_n L_x$	${}_n T_x$	${}_n \hat{e}_x$
Data entries for each variable					

As noted, life tables that use one-year age intervals, with $n = 1$, are referred to as unabridged or, sometimes, as *complete*; and those that use larger intervals, with $n > 1$, are called *abridged*³ (and the functions in unabridged tables are usually not preceded by a subscript “ n ”). The main techniques for constructing and interpreting life tables apply equally to unabridged and abridged versions, but the latter present practical and technical challenges that have occupied researchers throughout the years. For this reason, we begin with unabridged data, usually focusing on only a portion of the complete table to avoid the unwieldiness of 85 or more rows. Then we move on to discuss entire tables with abridged data. For the sake of comparison, Table 9.2 shows a portion of an unabridged table; Table 9.3 shows a complete abridged table from a different source. In the next section, we discuss the life table functions in detail and demonstrate how they are derived.

DATA SOURCES AND FUNCTIONS

The data upon which a life table is based can come from any of several sources, depending on the purpose to which the table is to be put (Stolnitz 1956). In later sections, we focus on health and actuarial applications, and adaptations to educational administration. Some of these applications deal with special populations, for which information may best be drawn from surveys, official records, and/or local-area data sources. For most general purposes, however, the definitive source is national government census/vital statistics offices.

As discussed in Chapter 2, in the United States, the Census Bureau collects and publishes information on cohort sizes, ${}_n K_x$, and the CDC publishes mortality information, ${}_n D_x$, along with complete tables (see <http://wonder.cdc.gov/>). States and other local areas keep such records for their jurisdictions, and they also have access to the records of Federal agencies. Mortality data are collected continuously and are published in annual editions of *Vital Statistics of the United States*. These are usually presented in abridged format, so that the research often must go directly to NCHS or the Census Bureau to obtain one-year breakdowns. Census data are, of course, officially collected every 10 years. Thus, unless one happens to be creating a life table for 1990, 2000, or 2010, population estimates are required. These, too, are available

Table 9.2. Portion of an Unabridged Life Table, United States, 2010

Age	Probability of Dying between Ages x to $x+1$: q_x	Number Surviving to Age x : l_x	Number Dying between Ages x to $x+1$: d_x	Person Years Lived between Ages x to $x+1$: L_x	Total Number of Person Years Lived above Age x : T_x	Average Life Expectancy at Age x : e_x
0-1	0.006123	100,000	612	99,465	7,866,027	78.7
1-2	0.000428	99,388	43	99,366	7,766,561	78.1
2-3	0.000275	99,345	27	99,331	7,667,195	77.2
3-4	0.000211	99,318	21	99,307	7,567,864	76.2
4-5	0.000158	99,297	16	99,289	7,468,556	75.2
5-6	0.000145	99,281	14	99,274	7,369,267	74.2
6-7	0.000128	99,267	13	99,260	7,269,993	73.2
7-8	0.000114	99,254	11	99,249	7,170,733	72.2
8-9	0.000100	99,243	10	99,238	7,071,484	71.3
9-10	0.000087	99,233	9	99,229	6,972,246	70.3
10-11	0.000079	99,224	8	99,220	6,873,017	69.3
11-12	0.000086	99,216	9	99,212	6,773,797	68.3
12-13	0.000116	99,208	12	99,202	6,674,585	67.3
13-14	0.000175	99,196	17	99,188	6,575,383	66.3
14-15	0.000252	99,179	25	99,167	6,476,195	65.3
15-16	0.000333	99,154	33	99,138	6,377,028	64.3
16-17	0.000412	99,121	41	99,101	6,277,891	63.3
17-18	0.000492	99,080	49	99,056	6,178,790	62.4
18-19	0.000573	99,032	57	99,003	6,079,734	61.4
87-88	0.105525	34,351	3,625	32,539	195,275	5.7
88-89	0.117007	30,726	3,595	28,929	162,736	5.3
89-90	0.129450	27,131	3,512	25,375	133,807	4.9
90-91	0.142873	23,619	3,375	21,932	108,432	4.6
91-92	0.157280	20,245	3,184	18,653	86,500	4.3
92-93	0.172661	17,061	2,946	15,588	67,847	4.0
93-94	0.188988	14,115	2,668	12,781	52,259	3.7
94-95	0.206214	11,447	2,361	10,267	39,478	3.4
95-96	0.224274	9,087	2,038	8,068	29,211	3.2
96-97	0.243080	7,049	1,713	6,192	21,144	3.0
97-98	0.262527	5,335	1,401	4,635	14,951	2.8
98-99	0.282492	3,935	1,112	3,379	10,316	2.6
99-100	0.302838	2,823	855	2,396	6,937	2.5
100+	1.000000	1,968	1,968	4,542	4,542	2.3

Source: *National Vital Statistics Report* 63, no. 7 (November 6, 2014).

at *census.gov*. Alternatively, researchers can create their own estimates, following the techniques to be presented in Chapter 10.

Regardless of the source of information, an annual period life table is *dated*, for example, “for the year 2000,” and is built upon data on the size of each cohort at midyear (July 1). These begin with age 0–1 and extend to the oldest cohort in which at least one person was alive at midyear. The oldest cohort is usually symbolized with a “ ω ,” the lower-case omega (Ω), the last letter in the Greek alphabet. In practice, ${}_nK_{\omega}$, ${}_nD_{\omega}$, and so on, usually refer to an open interval that includes all individuals ages 65+ or 85+. Also required, from NCHS or another agency, is an accurate count of the number of deaths that occurred from the beginning to the end of the year of interest (from January 1 to December 31).

Table 9.3. A Complete Abridged Life Table, Brazil, 2002

x	n	D _x	P _x	nM_x	nq_x	l_x	nd_x	nL_x	nT_x	$n\hat{e}_x$
0-1	0	65,532	3,205,108	0.02045	0.02006	100,000	2,006	98,095	7,196,592	71.97
1-4	1	11,271	13,084,650	0.00086	0.00344	97,994	337	391,143	7,098,498	72.44
5-9	4	5,366	16,533,114	0.00032	0.00162	97,657	158	487,891	6,707,355	68.68
10-14	5	6,294	17,406,984	0.00036	0.00181	97,499	176	487,055	6,219,463	63.79
15-19	5	19,255	17,847,032	0.00108	0.00538	97,323	524	485,306	5,732,408	58.90
20-24	5	26,620	16,500,057	0.00161	0.00803	96,799	778	482,053	5,247,103	54.21
25-29	5	25,404	14,534,868	0.00175	0.00870	96,022	835	478,020	4,765,050	49.62
30-34	5	28,162	13,533,472	0.00208	0.01035	95,186	985	473,468	4,287,030	45.04
35-39	5	33,578	12,953,294	0.00259	0.01288	94,201	1,213	467,972	3,813,563	40.48
40-44	5	39,855	10,942,252	0.00364	0.01805	92,988	1,678	460,744	3,345,591	35.98
45-49	5	45,880	9,106,099	0.00504	0.02488	91,310	2,272	450,869	2,884,847	31.59
50-54	5	52,276	7,139,958	0.00732	0.03595	89,038	3,201	437,188	2,433,978	27.34
55-59	5	58,078	5,425,966	0.01070	0.05212	85,837	4,474	418,000	1,996,790	23.26
60-64	5	72,044	4,553,017	0.01582	0.07611	81,363	6,192	391,334	1,578,790	19.40
65-69	5	81,641	3,365,780	0.02426	0.11435	75,171	8,596	354,365	1,187,456	15.80
70-74	5	93,339	2,588,020	0.03607	0.16541	66,575	11,012	305,345	833,091	12.51
75-79	5	90,927	1,602,984	0.05672	0.24839	55,563	13,801	243,310	527,746	9.50
80-84	5	80,847	857,170	0.09432	0.38161	41,761	15,937	168,965	284,436	6.91
85+	+	103,085	460,928	0.22365	1.00000	25,825	25,825	115,471	115,471	4.47

Source: World Health Organization, *Epidemiological Bulletin* 24, no. 4 (December 2003).

The use of midyear cohort sizes incorporates the assumption that on July 1 the number of persons alive was the average between the number alive on January 1 and the number alive on December 31. Thus, we assume that one-half of the deaths in each cohort occurred before July 1 and one-half occurred after that date. This is a fairly reasonable assumption for most cohorts, but when translated into life table functions, it is likely to be inaccurate for those ages 0 to 1 at the beginning of the year. The reason, discussed in Chapter 6, is that IMRs are much higher during the early months, especially during the first four weeks of life, for which special neonatal rates are derived. Thus, depending on how we treat the distribution of the ages of infants (month-by-month), as of July 1, the 50/50 assumption may need to be adjusted.

Age: x

In the life table and in demographic applications based on it, age is always symbolized by a lower case “x.” This symbol refers to the exact age at midyear, for example, 20 years old on July 1, even for those who celebrate their 20th birthday on June 30 or their 21st on July 2. In unabridged tables, x stands for a single year: 1, 26, 38, and so forth; in abridged tables it stands for a range: 1-5, 5-10, and so on. As seen in Tables 9.1-9.3, the ages create the rows of the table, with each row containing a complete set of information for one cohort, from its size to its average life expectancy. This row-wise structure is divided into three segments that are treated somewhat differently: the top segment containing the youngest cohort or cohorts; the middle, which makes up the bulk of the table; and the bottom, consisting of the oldest cohort (for which $x = \omega$). We discuss the reasons for this partition below.

The Observed Cohort Sizes, Deaths, and Probabilities of Dying:

${}_nK_x$, ${}_nD_x$, and ${}_nq_x$

As noted, the ${}_nK_x$ column lists the midyear sizes of each cohort in the observed population: ${}_nK_{20}$ is the number of 20-year-olds, ${}_nK_{50}$, the number of 50-year-olds, and ${}_nK_0$ is the number of infants. In each case, we refer to those who survived to July 1 but did not yet reach their next (21st, 51st, 1st) birthday by that date. The ${}_nD_x$ column contains the number of deaths at each age, and ${}_nq_x$ is the set of age-specific probabilities of dying derived from ${}_nM_x$ (${}_nD_x/{}_nK_x$) as discussed above. Even without going further, it is evident that this technique produces results that are more highly refined and more informative about the manner in which the force of mortality operates than do M_x or the CDR.

Using the observed data on which Table 9.2 is based (not shown), the portion of an unabridged table for the U.S. population that covers ages 0 through 14, we did some calculations and found that $\sum {}_nK_x = 57,708,194$ and that $\sum {}_nD_x = 45,373$. Dividing $\sum {}_nD_x/\sum {}_nK_x$, we see that the force of mortality, $M = .00079$ and that the CDR (for this segment of the whole population) = 0.79 deaths per 1,000 persons. However, the ${}_nq_x$ column indicates that *none* of the 15 cohorts shown was actually exposed to that risk. The 1–2 group came close, with $q_1 = .000428$, but this and nearly every other probability listed is lower than the general M . Two cohorts, the 10- and 11-year-olds, are much lower—with less than one-fifth the risk of dying than the average of the 15 cohorts. The exception, and it is a significant one, is q_0 , infant mortality. Here, the IMR in the observed population = 8.01 per 1,000, which is more than 10 times higher than the CDR. Thus we see that the general measures of mortality for a set of cohorts often conceal more than they reveal. In this case, what is concealed is the well-established fact that the probability of dying before age 15 is very low once individuals reach their first birthday.

The Surviving Population: ${}_nI_x$

With the set of ${}_nI_x$ s, we enter the life table proper and the column containing the surviving (stationary) population. It is defined as the hypothetical number of individuals who survive to the beginning of the age x interval during the year. The first of these, l_0 , is of special interest because it is an arbitrary figure upon which the rest of the table is based. Its function is the same as that of the “100” in percentages and the “1,000”—used in most demographic rates, in that it adjusts fractions of a whole to control for absolute size, “per 100,” “per 1,000,” and so on. Referred to as a *radix*, it could be any number; but for convenience we use 1,000, 10,000, or 100,000, depending on the size of the reference population. Unlike any other **radix** we have used to this point, however, the nlx vector is dynamic: it changes size, always decreasing, as x increases.

We understand that any ${}_nI_x$ with $x > 0$ is the number of individuals who reach age x per whatever ${}_nI_0$ equals, for example, per 100,000, as in Figures 9.1 through 9.3. When put this way, we see that l_0 is a model birth cohort, because reaching age 0 is the same as being born. That is why ${}_nI_x$ is ordinarily expressed as “the number of survivors to age x for every 100,000 (etc.) born.” This last form indicates how l_0 functions as the basis of the entire table, for it is the life table’s imaginary population at “birth.”

Assuming an unabridged table, once l_0 is selected, l_1 is derived as follows: (1) Multiply l_0 by q_0 . That is, apply the observed probability of death for persons ages

0 to 1 to the imaginary starting number of, say, 100,000. This gives the number of deaths that occur in the surviving life table population to persons ages 0 to 1, which is also a kind of infant death rate. The number can be recorded in the d_x column, or d_x can be calculated directly from the l_x s, as discussed in the next section. (2) Next, subtract this number from l_0 . Here we reduce the number “born” by the number of deaths, which yields the number of survivors to age 1 to 2, or l_1 . In Table 9.2, we begin with $l_0 = 100,000$, and multiply it by $q_0 = 0.006123$. The result is 612, which can be seen as the first entry in the d_0 column. This is then subtracted from 100,000 to yield $l_1 = 99,388$.

To find the number of survivors to age 2, repeat the operation, this time multiplying l_1 by q_1 to find the number of deaths to members of the surviving life table population ages 1 to 2. This product is subtracted from l_1 to produce l_2 . In Table 9.2, we multiply $l_1 = 99,388$ by $q_1 = 0.0428$ and get 43. This, subtracted from 99,388, is $l_2 = 99,345$. This process continues until l_ω is reached, at which point, $l_\omega = l_{\omega-1} - (q_\omega \times l_{\omega-1})$, and there are no survivors to a higher age: $l_{\omega+1} = 0$. Thus, the l_x column is completed, and all 100,000 of the original cohort have died. In general,

$$\text{for any } x > 0, l_x = l_{x-1} - (q_x \text{ times } l_{x-1})$$

For example, $l_{13} = l_{12} - (q_{12} \times l_{12})$; and, from Table 9.2, $l_{13} = 98,208 - (0.00016 \times 99,208) = 99,208 - 11.6 = 99,196$

Table 9.4. Creating an Abridged Table from Abridged Data: Age-Specific Death Rates for Adams County, Colorado, 2011

x	${}_nK_x$	${}_nD_x$	${}_nM_x$	n	${}_nq_x$	l_x	d_x	${}_nL_x$	T_x	\hat{e}_x
0-1	7,172*	43	.0060	1	.00570	10,000	57	9,972	793,729	79.37
1-4	30,088	3	.0001	4	.00039	9,943	4	39,764	783,757	78.83
5-9	37,588	4	.0001	5	.00050	9,939	5	49,670	743,993	74.86
10-14	34,533	4	.0001	5	.00050	9,934	5	49,658	694,323	69.89
15-19	31,114	11	.0004	5	.00192	9,929	19	49,598	644,665	64.92
20-24	29,130	23	.0009	5	.00449	9,910	44	49,440	595,067	60.05
25-29	35,300	37	.0010	5	.00499	9,866	49	49,208	545,627	55.30
30-34	36,679	45	.0012	5	.00598	9,817	51	48,958	496,419	50.57
35-39	34,681	44	.0013	5	.00647	9,766	63	47,923	447,461	45.81
40-44	32,539	60	.0018	5	.00950	9,403	89	46,793	399,538	42.49
45-49	30,194	109	.0036	5	.01589	9,314	148	46,200	352,745	37.87
50-54	28,996	143	.0049	5	.02421	9,166	222	45,150	306,545	33.44
55-59	24,315	174	.0072	5	.03344	8,894	297	43,728	261,395	29.39
60-64	20,147	215	.0107	5	.04897	8,597	421	41,932	207,566	24.14
65-69	12,296	200	.0150	5	.07247	8,176	593	39,410	164,634	20.13
70-74	9,353	255	.0273	5	.12722	7,588	965	35,528	125,224	16.50
75-79	7,138	326	.0457	5	.20510	6,623	1,358	45,063	89,696	13.54
80-84	4,999	312	.0624	5	.26953	5,265	1,417	42,709	44,633	8.48
85+	4,165	608	0.146	*	1.000	3,848	3,848	19,240	19,240	5.00
All ages	451,576	2,616	.0058							

Note: Reed-Merrell method used for calculating ${}_nq_x$.

Source: Colorado Health Statistics, county data, <http://www.chd.dphe.state.co.us/Resources/vs/2011/Adams.pdf>.

* infant mortality based on live births, other rates based on midyear population estimates. Assumes no survivors beyond age 95.

This transition from empirical data (${}_nK_x$, ${}_nD_x$, ${}_nM_x$, and ${}_nq_x$) to a model population (${}_nl_x$) is the key operation in creating a life table. This is illustrated in Table 9.2. In effect, this procedure causes a model population to age and survive through time as if it were experiencing what is, in fact, a momentary force of mortality in the actual, reference population. In the reference population, q_x is the probability that persons age x will die during a specific year and q_{x+1} is the probability that persons age $x+1$ will die during that same year. However, in the case of the life table, beginning with the l_x s, we assume that l_0 , a birth cohort of 100,000 individuals, experiences the passage of one year on the calendar, and those who survive make up l_1 . Then, with the passage of another year, the survivors from l_1 become l_2 , and so forth. Thus, in the life table, q_x is the probability that whoever has survived from the original 100,000 to l_x will not survive to l_{x+1} ; and q_{x+1} is the probability that those who do survive to age $x+1$ will not reach $x+2$.

This property has led demographers to refer to l_x as the *stationary population*.⁴ Of course, “stationary,” means “unchanging,” and the analogy here is that if a real population behaved as l_x does, it would be unchanging in certain important ways. Assume, as we do in a life table, that year after year a constant number of individuals—100,000—is born, and that age-specific mortality rates remain constant as well. Then, since $\sum {}_nd_x = 100,000$, under these conditions, the population size would always remain the same. The total size would equal $\sum {}_nl_x$, and the size of each cohort would also be fixed at $l_0 = 100,000$, l_1 , l_2 , and so on.

The transition from actual to model populations is a somewhat subtle but powerful assumption that is critical to life table construction. It is quite evident when one moves from period to **cohort tables**; and it is the foundation of the components and matrix methods of projection, noted in the next chapter. In addition, the assumption effectively illustrates the paired concepts, *synchronous* and *diachronic*: “at the same time” and “through time”; and thus shows why the process is called “projection.” In the actual population, ${}_nK_x$, ${}_nD_x$, and ${}_nq_x$ are synchronous. These data all pertain to one time period, thus the term **period table**. But, in the life table, ${}_nq_x$, ${}_nl_x$, and the other functions are understood to be diachronic; they pertain to a (hypothetical) process of aging from year to year. Figure 9.2 illustrates this difference.

The rectangle on the left represents the observed data and the one on the right the life table population. Notice that the q_x vector is included in both rectangles, but that it performs a different role in each. On the left side, it summarizes the relationship between observed cohort sizes and observed deaths, but on the right it is used to reduce the size of the surviving population as it ages “year by year.” The arrow linking the two rectangles indicates that we take a vertical object and exert force, via multiplication, to create a horizontal object, as if it were a projectile being thrust ahead toward the future.

Deaths in the Life Table: ${}_nd_x$

The next column vector contains the number of deaths at each age that would occur if the observed probability of dying $2M_x/(M_x + 2)$ were experienced by the surviving population. It is derived by multiplying ${}_nq_x$ by ${}_nl_x$. For example, in Table 9.2, d_1 , the number of deaths that would occur between ages 1 and 2 is equal to:

$$q_1 \times l_1 = .00048 \times 99,388 = 48$$

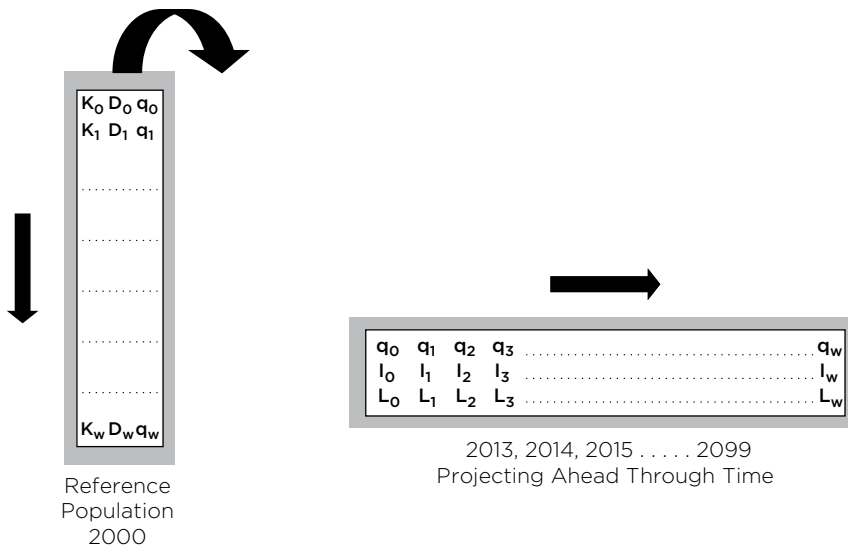


Figure 9.2. Synchronous and Diachronic Applications of Mortality Data

We can conceive of this as 48 deaths at age 1 for every 100,000 persons born or, alternatively, 48 deaths at age 1 for every 99,388 persons surviving to age 1. Because of the way the ${}_n l_x$ vector is calculated (see above), we can also find the ${}_n d_x$ s by subtracting. That is, for an unabridged table, $d_x = (l_{x+1} - l_x)$; and in general, ${}_n d_x = ({}_n l_{x+1} - {}_n l_x)$. This indicates that the number of deaths at a given age in the life table is equal to the size of the surviving cohort at the next higher age minus the size of the cohort at that age. So, in the abridged Table 9.4, we see that:

$$l_{10} - 14 - 135 - 39 = 9,934 - 9,766 = 168$$

In fact, this is what is meant by the term “surviving population”: the aggregate remaining from a younger age to an older age after the number of deaths has been subtracted.

The quantity $\sum {}_n d_x$, for all x s between two stated ages, for example, between 10–14 and 35–39, is the sum of the deaths to such individuals in the life table, in this case $4 + 11 + 23 + 37 + 45 + 44 = 154$. This can be compared to the second column of the tenth row in Table 9.2. If we divide that total by the sum of the surviving populations at those ages, we produce a probability of dying in the larger interval, a procedure that is used in constructing abridged tables. When the interval extends from 0 to w , the sum is equal to ${}_n l_0$, for example, 100,000; which is to say that the life table assumes that all members of the original birth cohort die by the time the oldest age is surpassed.

The Person Years Lived by the Life Table Population: ${}_n L_x$

Because ${}_n l_x$ is measured from the beginning of one age category to the beginning of the next, it does not take account of the deaths that occur in that cohort as of

midyear. In particular, it overestimates the size of the cohort by not subjecting it to the force of mortality during the period of interest. For this reason, it cannot be used to derive the true size of the life table population or for related purposes, such as establishing certain rates. These types of applications require a truer measure of the numbers alive in the model population and, thus, the third life-table function. This is ${}_nL_x$, which adjusts the first function (the surviving population) by the second (the number of deaths) to produce a more accurate indicator of size and structure.

In general, the value of ${}_nL_x$ is based on the assumption that the deaths that occur at a given age, x , are equally distributed among those whose ages lie in the interval between exactly x and $x + 6$ months and those whose ages are in the interval between $x + 6$ months and $x + 11$ months and 29 (or 30) days. For example, in Table 9.2, we assume that the deaths to 13-year-olds are equally divided between the group aged 13 to 13½ and that whose ages are between 13½ and 14. The adjustment mentioned is reflected, for instance, in the difference between l_{13} and L_{13} , which is 99,196 versus 99,188. The reason that L_{13} is smaller, by 8 individuals, is that it reflects the true cohort size at midyear, at which point some (one-half) of those who do not reach the next age cohort had died.

Based on these assumptions, the general formula (for an unabridged table) is:

$$L_x = (l_x + l_{x+1}) / 2; \text{ which, as we have just seen, is equal to: } l_x - (0.5 d_x)$$

This formula can also be expressed as:

$$(0.5 l_x) + (0.5 l_{x+1})$$

This indicates that the life table population at a given age, x , is equal to the average of those who reach that age and those who reach the next higher age, $x+1$. Because it is an adjusted total, it represents the true number of persons alive at age x . This is why L_x is referred to as the number of “person years lived” at that age. So, for instance, we note that 98,848 person years are lived at age 9 for every 100,000 persons born; 98,802 person years are lived at age 12; and so on.

The “0.5”s in the formula are called **separation factors**, and are symbolized as f' and f'' , respectively. These apportion the deaths as follows:

$$f' = d_t / (d_t + d_{t+1}), \text{ and } f'' = 1 - f'$$

where d_t is the number of deaths that occur to those individuals born between July 1 and December 31 of the first year, and d_{t+1} is the number of deaths that occur to those born between January 1 and June 30 of the second year. If $f'' = f'$, we are assuming that the two groups experience the same number of deaths. Thus we can say, more generally, that $L_x = f''l_x + f'l_{x+1}$, and that $f'' = f' = 0.5$. There are important exceptions, however, for which we know that the separation factors are not equal; and it is at this point that the three-part division of the life table mentioned earlier becomes effective.

The assumption of equal separation factors is valid for the middle portion of the table. But, for the upper portion, containing the youngest cohorts, we recognize that the number of deaths that occur during the first six months at a given age exceeds the number occurring in the last six months. This difference is especially

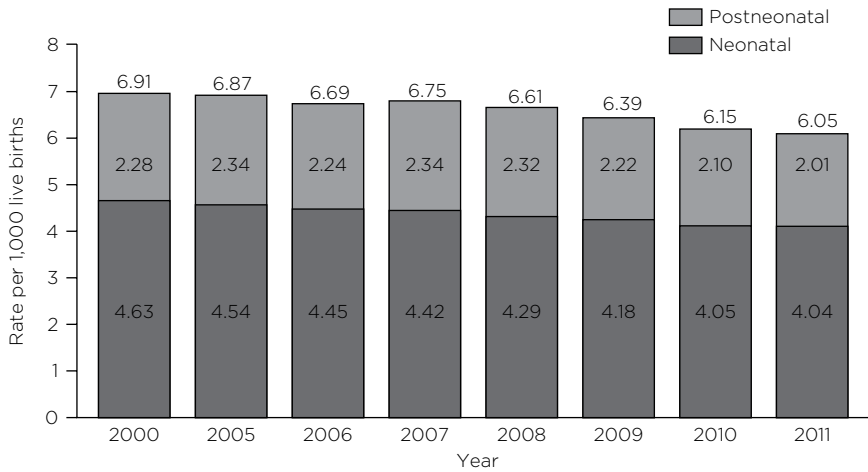


Figure 9.3. Infant, Neonatal, and Postneonatal Mortality Rates, United States, 2000 and 2005-2011

Sources: CDC/NCHS, National Vital Statistics System, mortality data set; U.S. Department of Health and Human Services, Health Resources and Services Administration, Maternal and Child Health Bureau. *Child Health USA 2011*.

pronounced in the 0–1, infant, cohort. For the lower portion of the table, containing the oldest cohorts, the opposite condition prevails: more deaths occur during the second six months at a given age than during the first. Although generally a trivial matter, the situation regarding older cohorts requires special consideration because, even in abridged tables, these tend to be stated in more-than-one year, or open, intervals.

Several methods have been devised for determining the appropriate f' and f'' for the youngest and oldest age groups. In both cases, the most reliable approach is to observe ${}_nK_x$ and ${}_nd_x$ for $n < 1$ year, for example, monthly totals. In fact, neonatal mortality rates, such as those shown in Figure 9.3, are collected by the CDC and are often used for these purposes. However, for ages 1–2, 2–3, and so on, such data are generally not available. In these cases, various approximation methods have been employed, including those based on standard estimates and on overall infant mortality.⁵ In Table 9.2, we assumed that infant mortality (ages 0 to 1) is distributed 70/30, for ages 1–2 it is 60/40, and for ages 2–3, 45/65 (with the remainder equal).

In the case of the oldest cohorts, there are two distinct approaches. In the more common situation, in which the oldest cohort is open, then the separation factors are accounted for as part of the abridging procedure, which we discuss in the next section. On the other hand, in the rare instances in which data are available for all older cohorts, including the one containing the oldest person (s) alive, then the deaths in that and possibly the next-younger group are apportioned as with the very youngest cohorts. For example, here are the last two rows of a completely unabridged table (Shryock and Siegel 1976:262, Table 15.2)

x	q_x	l_x	d_x	L_x	T_x	\hat{e}_x
108–109	.52810	1	0	1	2	1.35
109–110	.54519	1	1	1	1	1.29

You will note that although the probability of dying in the last interval is just under 0.55, the single individual to survive to age 109 from $l_0 = 100,000$ is assumed not to survive to his or her 110th birthday in that year. That is, one person year is lived at 109, but none beyond.

Total Person Years Remaining: ${}_n T_x$

The next life table function is the ${}_n T_x$ vector, which is interpreted as the total number of person years remaining at age x and beyond. In most applications, it is written without the preceding “ n ” subscript, although we retain it in this chapter. As the name indicates, it is a sum of ${}_n L_x$ s that cumulates according to the value of x : ${}_n L_x + {}_n L_{x+n} + \dots$. In Table 9.3, we see that $T_{30-34} = 4,278,030$. This means that just over 4.4 million person years in the life table can be counted between ages 30 and 35, 35 and 40, 45 and 50, and so on, up to and including 85+. The general formula is:

$${}_n T_a^w = \sum_{x=a}^w {}_n L_x, \text{ where } a \text{ is any given age and } w \text{ is the highest age in the table}$$

As we move down the column, we can see how this function steadily decreases as the number of survivors declines, with its rate of decrease indicating how sharply mortality levels rise with age.

Average Life Expectancy: \hat{e}_x

The final life table function is perhaps the best known. This is \hat{e}_x , the average number of person years remaining to those who survive to age x . A small “ \circ ” or “ \wedge ” is usually placed above the symbol to distinguish it from the exponential constant (“plain”) e that we encountered in Chapter 8. Commonly referred to as “average life expectancy,” this function is the basis of actuarial calculations of life insurance, and it has many other uses as well. The general formula is:

$$\hat{e}_x = {}_n T_x / l_x \text{ or, simply, } T_x / l_x$$

Table 9.4 illustrates the entire process from x to \hat{e}_x . We see that the function takes the total number of person years remaining at age x and averages them out (equally) among all of the survivors to that age. Thus, if an individual survives to a given age, say 30, then that individual is most likely to survive an additional \hat{e}_{30} years. According to Table 9.5, for example, for African American males $\hat{e}_{30} = 43.1$. If we add the age already attained, 30 years in this example, we get the total life expectancy for those who survive to that age; for example, $30 + 41.1 = 6733.1$. Those who live to age 30 will, on the average, survive to age 75 and about 2 months. Demographers stress that this is an average because we know that some individuals who reach age x will not survive the full \hat{e}_x number of years and, in contrast, some will live beyond the age indicated.

Average life expectancy at birth, that is $\hat{e}_0 = T_0 / l_0$, is of special interest for both technical and practical reasons. Because l_0 is in the denominator, this measure makes explicit use of the table’s radix of 1,000, 100,000, and so on. In this light, it can be interpreted as indicating that, on the average, a certain number of years of life remain at birth—for example, 79.6 for males and 53.56 for females in Table 9.6—to

Table 9.5. Comparative Probabilities of Dying (q_x) and Average Life Expectancies (\hat{e}_x) for Selected Ages for the United States, 2008

x	White Males		White Females		Black Males		Black Females	
	q_x	\hat{e}_x	q_x	\hat{e}_x	q_x	\hat{e}_x	q_x	\hat{e}_x
5	.000167	71.6	.000125	76.4	.000270	66.7	.000190	73.2
10	.000068	67.6	.000082	72.4	.000095	61.8	.000129	68.2
20	.001145	56.9	.000398	61.6	.001767	52.2	.000537	58.4
30	.001311	47.6	.000577	51.8	.006370	43.1	.000981	42.8
40	.002096	39.2	.001248	42.2	.003444	34.2	.00225	39.4
50	.005155	38.3	.003034	32.9	.008585	25.7	.005549	30.6
60	.010870	29.3	.006623	24.1	.018943	18.5	.010701	22.6
70	.024538	21.2	.016423	16.2	.036296	12.6	.022247	15.4
80	.063309	8.0	.055871	9.4	.077689	7.6	.05123	9.4

Source: *National Vital Statistics Reports* 61, no. 3 (September 24, 2012): Tables 1, 5, 7, 8, and 9.

the 1,000, 100,000, and so on, individuals born into the table's model population. Because T_0 is in the numerator, this measure takes account of the entire set of mortality rates in the table. That is, \hat{e}_0 is a cumulative measure of survival at each and every age attained.

It may be apparent that an individual person born into a reference population during the reference year, for example a male born in India in 1980 would probably not have the actual average life expectancy at birth calculated, for example, about 53 years. This would probably differ, unless the deaths rate at age one in 1984 were the same as the death rates at age one in 1983, the rate at age 2 in 1985 were the same as the rate at age 2 in 1983, the rate at age 3 in 1986 were the same as in 1983, and so on. If these actually decreased, which is quite likely, then the true average life span for persons born in 1983 would be greater than the calculated 53 years. The tools of the Lexis diagram and cohort life tables help to clarify and resolve this synchronous/diachronic issue, as will see in later sections (also, see Shavelle 1996).

As a practical matter, \hat{e}_0 not only provides an especially accurate measure of mortality conditions, it is also one of the most effective indicators of overall socioeconomic development. In our sample of 66 nations introduced in Chapter 3, the average life expectancy at birth ranged between 33 and 84 years, with means of 62.79 for males, 67.53 for females, and 65.14 for both sexes combined. The relationships between female life expectancy and seven crucial development variables are all very strong. In particular, where average life expectancy is high, per-capita gross national product, contraceptive use, and urbanization rates are also high, whereas death, birth, infant mortality, and TFRs are low.

Abridged and Unabridged Tables

Most life tables one is likely to encounter are abridged. Although, in principle, unabridged tables are more accurate and, of course, more complete, in human populations, the listing of information for each age, year-by-year, can be difficult to comprehend without careful study. Moreover, for most applications, the level of detail presented in unabridged tables is unnecessary; the number of deaths, probabilities of dying, and the other data contained in tables that use five- or even 10-year cohorts