ıphs in Epidemiology and Biostatistics
ı Jennifer L. Kelsey, Michael G. Marmot,
Stolley, Martin P. Vessey

EPIDEMIOLOGY OF DEMENTIA
Mortimer and Leonard M. Schuman 1981

.-CONTROL STUDIES
nduct, Analysis
Schlesselman 1982

EMIOLOGY OF MUSCULOSKELETAL DISORDERS
.. Kelsey 1982

ANIZATION AND CANCER MORTALITY
d States Experience, 1950–1975
R. Greenberg 1983

NTRODUCTION TO EPIDEMIOLOGIC METHODS
.. Kahn 1983

LEUKEMIAS
gic Aspects
. Linet 1984

EENING IN CHRONIC DISEASE
Morrison 1985

ICAL TRIALS
nduct, and Analysis
Meinert 1986

CINATING AGAINST BRAIN DYSFUNCTION SYNDROME
aign Against Rubella and Measles
. Gruenberg 1986

THODS IN OBSERVATIONAL EPIDEMIOLOGY
L. Kelsey, W. Douglas Thompson, Alfred S. Evans 1986

NICAL EPIDEMIOLOGY:
UDY OF THE OUTCOME OF ILLNESS
Veiss 1986

.TISTICAL METHODS IN EPIDEMIOLOGY
. Kahn and Christopher T. Sempos 1989

EARCH METHODS IN OCCUPATIONAL EPIDEMIOLOGY
Checkoway, Neil E. Pearce, Douglas J. Crawford-Brown 1989

NCEPTION TO BIRTH
gy of Prenatal Development
line, Zena Stein, Mervyn Susser 1989

TRITIONAL EPIDEMIOLOGY
Villett 1990

.TISTICAL MODELS FOR LONGITUDINAL STUDIES OF HEALTH
. Dwyer, Manning Feinleib, Peter Lippert, Hans Hoffmeister 1991

Monographs in Epidemiology and Biostatistics

Volume 17

Statistical Analysis of Epidemiologic Data

Steve Selvin, Ph.D.
University of California, Berkeley

New York Oxford
OXFORD UNIVERSITY PRESS
1991

9 Life Tables: An Introduction

Life tables (of sorts) date back to third-century Roman records of the age at death. The development of the formal life table is usually attributed to Edmund Halley (1693) and John Graunt (1662). By the end of the nineteenth century, life tables were routinely computed as part of a generally emerging awareness of the importance of mortality statistics. The first official U.S. (death registration states) life table published in 1900 showed the expected length of life for white males as 46.6 years and for white females as 48.7 years.

A life table is a systematic way to keep track of the mortality experience of a group. A cohort life table is constructed from the mortality records of individuals followed from the birth of the first to the death of the last member of a group. Such life tables are constructed from animal and insect data. For human populations, it is obviously not practical to construct a life table by following a cohort of individuals from birth until all have died. Instead a life table is constructed from current mortality rates. These rates do not apply to past populations and undoubtedly will not apply to future populations. Nevertheless, patterns of mortality can be seen from a current life table, and the comparison of life tables calculated for different groups is a basic strategy for analyzing certain types of epidemiologic data.

Complete, Current Life Table: Construction

The word *complete* when applied to a life table means that ages are not grouped but recorded in 1-year intervals. The actual construction of a complete life table is rather mechanical and embraces seven basic elements:

Age interval (x to $x + 1$): Each age interval consists of 1 year (age denoted by x) except the last age interval, which is left open ended (e.g., 90+ years).

Number alive (l_x): The number of individuals alive at exactly age x. The number l_x is the life table population at risk for the interval x

to $x + 1$. The number alive at age 0 (l_0) is set at some arbitrary value, such as 100,000, and called the radix.

Deaths(d_x): The number of individuals who died between the ages of x and $x + 1$.

Probability of death (q_x): The conditional probability that an individual who is alive at age x dies before age $x + 1$. That is, $q_x = d_x/l_x$. The probability of death within a specific age interval is related to a hazard rate and is distinct from the probability of dying before a specific age x, which is expressed in terms of a survival curve.

Years lived (L_x): The cumulative time lived by the entire cohort between the ages of x and $x + 1$. Each individual alive at age x contributes to the total time lived by all individuals either 1 year or the proportion of the year lived if the person died in the interval. The value L_x is the life table person-years of risk for the interval x to $x + 1$.

Total time lived (T_x): The total time lived beyond age x by all individuals alive at age x is $T_x = L_x + L_{x+1} + L_{x+2} + \cdots$. The value T_x is primarily a calculational step in the life-table construction.

Expectation of life (e_x): The average number of additional years expected to be lived by those individuals alive at age x and $e_x = T_x/l_x$.

The following relationships are direct consequences of these definitions:

1. Number dying in the interval x to $x + 1 = d_x = q_x l_x = l_x - l_{x+1}$;
2. Number surviving at age $x + 1 = l_{x+1} = p_x l_x = l_x - d_x$;
3. Probability of dying in the interval x to $x+1 = q_x = (l_x - l_{x+1})/l_x = d_x/l_x$;
4. Probability of surviving from x to $x+1 = p_x = 1 - q_x = (l_x - d_x)/l_x = l_{x+1}/l_x$.

These definitions apply to a complete life table, using age intervals of 1 year.

The total person-years at risk for the interval x to $x + 1$ includes 1 year of survival for each person who did not die during the interval. Individuals who die contribute the proportion of the year they were alive to the total time lived. The average time contributed by those who died in the interval x to $x + 1$ is represented by $\bar{a}_x$. The value $\bar{a}_x$ is close to 0.5 for all ages except the first few years of life. For years 0 to 4 the

values of $\bar{a}_x$ are: $\bar{a}_0 = 0.09$, $\bar{a}_1 = 0.43$, $\bar{a}_2 = 0.45$, $\bar{a}_3 = 0.47$, and $\bar{a}_4 = 0.49$ (determined empirically by Chiang [Ref. 1]). These values make logical sense—the distribution of survival times in the first year of life is skewed towards the beginning of the interval since most deaths in the interval 0 to 1 year are within the first month. Therefore, the average contribution of time lived by those who died to the total years lived is low for the interval 0 to 1. For ages 2 to 4, the mean $\bar{a}_x$ shows slightly earlier deaths within the interval, but these $\bar{a}_x$ values are much closer to 0.50. For all other age intervals the average value of $\bar{a}_x$ is essentially 0.5 years.

The value $\bar{a}_x$ takes on importance in calculating the person-years of life for a life table since

$$L_x = (l_x - d_x) + \bar{a}_x d_x \tag{9.1}$$

estimates the life table person-years of risk for the age interval x to $x + 1$. Using L_x, the life-table age-specific mortality rate becomes d_x/L_x, providing a link to observed age-specific mortality rates. This life-table person-years calculation does not differ from the person-years calculation in Chapter 1 [expression (1.5)].

The starting point for construction of a life table is a set of age-specific probabilities of death (q_x). These probabilities can be derived by equating the life-table age-specific mortality rates to the age-specific mortality rates from the population of interest, or

$$\text{life-table mortality rate} = \frac{d_x}{L_x} = R_x = \text{observed mortality rate}, \tag{9.2}$$

where R_x is the age-specific rate for age x calculated from observed mortality data. A value for q_x follows from R_x since

$$\text{life-table mortality rate} = \frac{d_x}{(l_x - d_x) + \bar{a}_x d_x} = \frac{q_x}{1 - (1 - \bar{a}_x)q_x} = R_x \tag{9.3}$$

and solving for q_x gives

$$q_x = \frac{R_x}{1 + (1 - \bar{a}_x)R_x}. \tag{9.4}$$

A set of observed mortality rates (R_x) produce a set of life-table probabilities (q_x). The probabilities q_x generate the rest of the life-table functions (l_x, d_x, L_x, T_x, and e_x) with one exception.

The person-years of life (L_x) for the last interval cannot be calculated directly since a value for $\bar{a}_x$ is not generally available. The individuals who are present at the start of the last interval all die ($q_{x'} = 1.0$) so that $l_{x'} = d_{x'}$, where x' symbolizes the final age interval (e.g., if the last interval is 90+, then $x' = 90$). Therefore, again equating the observed mortality rate with the life-table mortality rate for this last age interval gives

$$\frac{d_{x'}}{L_{x'}} = \frac{l_{x'}}{L_{x'}} = R_{x'} \tag{9.5}$$

and then solving for $L_{x'}$ yields

$$L_{x'} = \frac{l_{x'}}{R_{x'}}. \tag{9.6}$$

Therefore, an observed set of age-specific mortality rates is all that is needed to calculate a complete life table.

Specifically, consider the age interval 65 to 66 for white males, 1980, California:

1. $$q_{65} = \frac{R_{65}}{1 + 0.5R_{65}} = \frac{0.0284}{1 + 0.5(0.0284)} = 0.0280,$$

   since $R_{65} = \dfrac{2097}{73832} = 0.0284$ (note: $\bar{a}_{65} = 0.5$),

2. $d_{65} = l_{65}q_{65} = 69728(0.0280) = 1953,$
3. $L_{65} = l_{65} - d_{65} + 0.5d_{65} = 69728 - 1953 + 0.5(1953) = 68752,$
4. $T_{65} = L_{65} + L_{66} + \cdots + L_{90+}$
   $= 68752 + 66757 + \cdots + 9126 + 41616 = 1011356,$

and

5. $$e_{65} = \frac{T_{65}}{l_{65}} = \frac{1011356}{69728} = 14.504.$$

These five steps are repeated for each age interval, starting at age 0, resulting in the entire current life table from a set of mortality rates (R_x) and an arbitrary starting value (l_0).

Two example life tables are given in Tables 9.1 and 9.2 for male and female residents of California for the year 1980. The expected number of years of life remaining after the age x is an effective summary of the entire mortality pattern described by a life table (e_x; last column in Tables 9.1 and 9.2). The expectation of life is not more than a special mean value and is calculated in the same way as most mean values, where

$$\text{mean years remaining} = e_x = \frac{\text{total years lived beyond age } x}{\text{number of individuals age } x} = \frac{T_x}{l_x}. \tag{9.7}$$

Perhaps the most common single summary value calculated from a life table is the expectation of life at birth (e_0). For the California data, $e_0 = 69.61$ years for males and $e_0 = 76.93$ years for females, based on 1980 mortality patterns.

Table 9–1. California 1980 population of white males

$x - x+1$	Population	Deaths	R_x^*	q_x	d_x	l_x	L_x	T_x	e_x
0 1	129,602	2,166	1,671.3	0.01647	1,647	100,000	98,518	6,960,692	69.61
1 2	117,753	123	104.5	0.00104	103	98,353	98,295	6,862,175	69.77
2–3	115,003	73	63.5	0.00063	62	98,251	98,217	6,763,880	68.84
3–4	113,314	60	53.0	0.00053	52	98,188	98,161	6,665,663	67.89
4–5	110,822	41	37.0	0.00037	36	98,137	98,118	6,567,502	66.92
5–6	110,548	55	49.8	0.00050	49	98,076	98,100	6,469,384	65.95
6–7	106,857	42	39.3	0.00039	39	98,051	98,032	6,371,308	64.98
7–8	112,184	58	51.7	0.00052	51	98,013	97,988	6,273,276	64.00
8–9	116,423	44	37.8	0.00038	37	97,962	97,944	6,175,288	63.04
9–10	132,952	52	39.1	0.00039	38	97,925	97,906	6,077,344	62.06
10–11	134,266	48	35.7	0.00036	35	97,887	97,869	5,979,438	61.09
11–12	128,938	60	46.5	0.00047	46	97,852	97,829	5,881,569	60.11
12–13	125,502	52	41.4	0.00041	41	97,806	97,786	5,783,740	59.13
13–14	128,212	82	64.0	0.00064	63	97,766	97,735	5,685,954	58.16
14–15	132,775	129	97.2	0.00097	95	97,703	97,656	5,588,219	57.20
15–16	143,600	233	162.3	0.00162	158	97,608	97,529	5,490,563	56.25
16–17	151,840	290	191.0	0.00191	186	97,450	97,357	5,393,034	55.34
17–18	157,365	400	254.2	0.00254	247	97,264	97,141	5,295,677	54.45
18–19	159,476	415	260.2	0.00260	252	97,017	96,891	5,198,535	53.58
19–20	171,235	416	242.9	0.00243	235	96,765	96,648	5,101,644	52.72
20–21	173,682	418	240.7	0.00240	232	96,530	96,414	5,004,996	51.85
21–22	172,656	436	252.5	0.00252	243	96,298	96,177	4,908,582	50.97
22–23	176,544	400	226.6	0.00226	217	96,056	95,947	4,812,405	50.10
23–24	175,732	410	233.3	0.00233	223	95,838	95,726	4,716,458	49.21
24–25	174,780	409	234.0	0.00234	223	95,615	95,503	4,620,731	48.33
25–26	173,214	393	226.9	0.00227	216	95,391	95,283	4,525,228	47.44
26–27	169,980	400	235.3	0.00235	224	95,175	95,063	4,429,944	46.55
27–28	168,369	366	217.4	0.00217	206	94,951	94,848	4,334,881	45.65
28–29	157,189	330	209.9	0.00210	199	94,745	94,646	4,240,033	44.75
29–30	162,394	346	213.1	0.00213	201	94,547	94,446	4,145,387	43.84
30–31	161,191	329	204.1	0.00204	192	94,345	94,249	4,050,941	42.94
31–32	154,874	355	229.2	0.00229	216	94,153	94,045	3,956,692	42.02
32–33	162,136	338	208.5	0.00208	196	93,937	93,840	3,862,647	41.12
33–34	163,065	305	187.0	0.00187	175	93,742	93,654	3,768,807	40.20
34–35	127,624	267	209.2	0.00209	196	93,567	93,469	3,675,153	39.28
35–36	128,890	296	229.7	0.00229	214	93,371	93,264	3,581,684	38.36
36–37	127,933	302	236.1	0.00236	220	93,157	93,047	3,488,420	37.45
37–38	127,923	334	261.1	0.00261	242	92,937	92,816	3,395,373	36.53
38–39	109,718	281	256.1	0.00256	237	92,695	92,576	3,302,557	35.63
39–40	108,168	325	300.5	0.00300	277	92,458	92,319	3,209,981	34.72
40–41	104,314	338	324.0	0.00324	298	92,180	92,031	3,117,662	33.82
41–42	100,059	342	341.8	0.00341	314	91,882	91,725	3,025,630	32.93
42–43	97,330	344	353.4	0.00353	323	91,569	91,407	2,933,905	32.04
43–44	92,394	356	385.3	0.00385	351	91,246	91,070	2,842,497	31.15
44–45	91,741	431	469.8	0.00469	426	90,895	90,682	2,751,427	30.27
45–46	92,331	438	474.4	0.00473	428	90,469	90,255	2,660,745	29.41
46–47	88,150	522	592.2	0.00590	532	90,041	89,775	2,570,491	28.55
47–48	90,475	559	617.9	0.00616	551	89,509	89,233	2,480,716	27.71
48 49	90,095	650	721.5	0.00719	639	88,958	88,638	2,391,483	26.88
49–50	97,275	696	715.5	0.00713	630	88,318	88,003	2,302,845	26.07
50–51	98,008	734	748.9	0.00746	654	87,688	87,361	2,214,841	25.26

Table 9-1. *(Continued)*

$x-x+1$	Population	Deaths	R_x^*	q_x	d_x	l_x	L_x	T_x	e_x
51–52	93,134	825	885.8	0.00882	768	87,034	86,650	2,127,480	24.44
52–53	94,496	875	926.0	0.00922	795	86,267	85,869	2,040,830	23.66
53–54	93,239	1,010	1,083.2	0.01077	921	85,472	85,011	1,954,960	22.87
54–55	96,443	1,126	1,167.5	0.01161	981	84,551	84,060	1,869,949	22.12
55–56	97,763	1,197	1,224.4	0.01217	1,017	83,569	83,061	1,785,889	21.37
56–57	96,823	1,272	1,313.7	0.01305	1,077	82,552	82,014	1,702,829	20.63
57–58	96,189	1,334	1,386.9	0.01377	1,122	81,475	80,914	1,620,815	19.89
58–59	98,518	1,553	1,576.4	0.01564	1,257	80,353	79,724	1,539,901	19.16
59–60	96,154	1,564	1,626.6	0.01613	1,276	79.096	78,458	1,460,177	18.46
60–61	88,552	1,472	1,662.3	0.01649	1,283	77,820	77,820	1,381,719	17.76
61–62	83,814	1,684	2,009.2	0.01989	1,522	76,537	75,776	1,304,541	17.04
62–63	81,464	1,763	2,164.1	0.02141	1,606	75,014	74,211	1,228,766	16.38
63–64	76,317	1,871	2,451.6	0.02422	1,778	73,408	72,519	1,154,554	15.73
64–65	75,505	2,032	2,691.2	0.02656	1,902	71,630	70,679	1,082,035	15.11
65–66	73,832	2,097	2,840.2	0.02801	1,953	69,728	68,752	1,011,356	14.50
66–67	69,480	2,121	3,052.7	0.03007	2,038	67,776	66,757	942,604	13.91
67–68	65,690	2,130	3,242.5	0.03191	2,098	65,738	64,689	875,847	13.32
68–69	62,557	2,256	3,606.3	0.03542	2,254	63,640	62,513	811,159	12.75
69–70	57,412	2,327	4,053.2	0.03973	2,439	61,386	60,166	748,646	12.20
70–71	53,926	2,205	4,088.9	0.04007	2,362	58,947	57,766	688,479	11.68
71–72	50,402	2,376	4,714.1	0.04606	2,606	56,585	55,282	630,713	11.15
72–73	47,213	2,342	4,960.5	0.04840	2,613	53,979	52,673	575,431	10.66
73–74	42,931	2,233	5,201.4	0.05070	2,604	51,366	50,064	522,759	10.18
74–75	39,611	2,300	5,806.5	0.05643	2,751	48,762	47,386	472,694	9.69
75–76	36,306	2,408	6,632.5	0.06420	2,954	46,011	44,534	425,308	9.24
76–77	33,386	2,251	6,742.3	0.06523	2,808	43,057	41,653	380,774	8.84
77–78	30,141	2,102	6,973.9	0.06739	2,712	40,249	38,892	339,121	8.43
78–79	26,432	2,272	8,595.6	0.08241	3,094	37,536	35,990	300,229	8.00
79–80	26,264	2,093	7,969.1	0.07664	2,640	34,443	33,123	264,239	7.67
80–81	21,846	1,958	8,962.7	0.08578	2,728	31,803	30,439	231,117	7.27
81–82	18,868	1,947	10,319.1	0.09813	2,853	29,075	27,648	200,677	6.90
82–83	16,653	1,802	10,820.9	0.10265	2,692	26,222	24,876	173,029	6.60
83–84	14,825	1,751	11,811.1	0.11153	2,624	23,530	22,218	148,153	6.30
84–85	13,137	1,689	12,856.8	0.12080	2,525	20,906	19,643	125,935	6.02
85–86	11,350	1,622	14,290.7	0.13338	2,452	18,380	17,155	106,292	5.78
86–87	9,442	1,426	15,102.7	0.14042	2,237	15,929	14,811	89,137	5.60
87–88	8,047	1,198	14,887.5	0.13856	1,897	13,692	12,744	74,327	5.43
88–89	6,091	1,072	17,599.7	0.16176	1,908	11,795	10,841	61,583	5.22
89–90	5,382	897	16,666.7	0.15385	1,521	9,887	9,126	50,742	5.13
90+	17,346	3,487	20,102.6	1.00000	8,366	8,366	41,616	41,616	4.97

*Rate per 100,000 person years of risk.

Expectations of life from birth are compared among countries and among groups within a country. The U.S. life expectancy e_0 has steadily increased over the last 80 years, and the difference between males and females has also increased, as Table 9.3 shows.

The expectation of life has a geometric interpretation related to the

Table 9-2. California 1980 population of white females

$x-x+1$	Population	Deaths	R_x^*	q_x	d_x	l_x	L_x	T_x	e_x
0–1	123,342	1,635	1325.6	0.01310	1,310	100,000	98,821	7,693,461	76.93
1–2	111,520	64	57.4	0.00057	57	98,690	98,658	7,594,641	76.95
2–3	109,200	41	37.5	0.00038	37	98,633	98,613	7,495,983	76.00
3–4	108,749	22	20.2	0.00020	20	98,596	98,586	7,397,370	75.03
4–5	105,698	41	38.8	0.00039	38	98,576	98,557	7,298,784	74.04
5–6	105,801	37	35.0	0.00035	34	98,538	98,521	7,200,227	73.07
6–7	101,630	37	36.4	0.00036	36	98,504	98,486	7,101,706	72.10
7–8	106,850	32	29.9	0.00030	29	98,468	98,453	7,003,220	71.12
8–9	110,410	32	29.0	0.00029	29	98,438	98,424	6,904,767	70.14
9–10	127,237	33	25.9	0.00026	26	98,410	98,397	6,806,342	69.16
10–11	128,916	33	25.6	0.00026	25	98,384	98,372	6,707,945	68.18
11–12	124,123	32	25.8	0.00026	25	98,359	98,347	6,609,573	67.20
12–13	119,672	28	23.4	0.00023	23	98,334	98,322	6,511,227	66.22
13–14	123,652	48	38.8	0.00039	38	98,311	98,292	6,412,905	65.23
14–15	127,869	68	53.2	0.00053	52	98,273	98,247	6,314,613	64.26
15–16	139,122	98	70.4	0.00070	69	98,220	98,186	6,216,366	63.29
16–17	146,318	93	63.6	0.00064	62	98,151	98,120	6,118,180	62.33
17–18	150,163	132	87.9	0.00088	86	98,089	98,046	6,020,059	61.37
18–19	152,382	121	79.4	0.00079	78	98,003	97,964	5,922,014	60.43
19–20	162,203	138	85.1	0.00085	83	97,925	97,883	5,824,050	59.47
20–21	162,313	118	72.7	0.00073	71	97,842	97,806	5,726,167	58.52
21–22	162,709	104	63.9	0.00064	62	97,771	97,739	5,628,360	57.57
22–23	167,087	96	57.5	0.00057	56	97,708	97,680	5,530,621	56.60
23–24	168,874	121	71.7	0.00072	70	97,652	97,617	5,432,940	55.64
24–25	168,959	119	70.4	0.00070	69	97,582	97,548	5,335,324	54.68
25–26	168,414	110	65.3	0.00065	64	97,513	97,481	5,237,776	53.71
26–27	165,167	141	85.4	0.00085	83	97,450	97,408	5,140,295	52.75
27–28	164,403	123	74.8	0.00075	73	97,366	97,330	5,042,887	51.79
28–29	154,062	137	88.9	0.00089	86	97,294	97,250	4,945,557	50.83
29–30	158,102	135	85.4	0.00085	83	97,207	97,166	4,848,307	49.88
30–31	157,975	134	84.8	0.00085	82	97,124	97,083	4,751,141	48.92
31–32	153,534	134	87.3	0.00087	85	97,042	97,000	4,654,058	47.96
32–33	160.016	157	98.1	0.00098	95	96,957	96,910	4,557,058	47.00
33–34	160,299	127	79.2	0.00079	77	96,862	96,824	4,460,149	46.05
34–35	125,826	144	114.4	0.00114	111	96,785	96,730	4,363,324	45.08
35–36	126,747	158	124.7	0.00125	120	96,675	96,614	4,266,594	44.13
36–37	125,960	155	123.1	0.00123	119	96,554	96,495	4,169,980	43.19
37–38	127,942	161	125.8	0.00126	121	96,436	96,375	4,073,485	42.24
38–39	109,358	169	154.5	0.00154	149	96,314	96,240	3,977,110	41.29
39–40	106,481	196	184.1	0.00184	177	96,166	96,077	3,880,870	40.36
40–41	103,828	171	164.7	0.00165	158	95,989	95,910	3,784,793	39.43
41–42	99,325	205	206.4	0.00206	198	95,831	95,732	3,688,883	38.49
42–43	96,380	228	236.6	0.00236	226	95,633	95,520	3,593,151	37.57
43–44	93,276	256	274.5	0.00274	261	95,407	95,276	3,497,631	36.66
44–45	92,873	258	277.8	0.00277	264	95,146	95,014	3,402,355	35.76
45–46	92,183	246	266.9	0.00267	253	94,882	94,755	3,307,341	34.86
46–47	88,595	274	309.3	0.00309	292	94,629	94,483	3,212,586	33.95
47–48	91,046	323	354.8	0.00354	334	94,337	94,170	3,118,103	33.05
48–49	89,588	384	428.6	0.00428	402	94,003	93,802	3,023,934	32.17
49–50	97,274	398	409.2	0.00408	382	93,601	93,409	2,930,132	31.30

Table 9-2. (*Continued*)

$x-x+l$	Population	Deaths	R_x^*	q_x	d_x	l_x	L_x	T_x	e_x
50–51	98,371	449	456.4	0.00455	425	93,218	93,006	2,836,722	30.43
51–52	95,717	474	495.2	0.00494	458	92,794	92,565	2,743,716	29.57
52–53	99,570	557	559.4	0.00558	515	92,335	92,078	2,651,152	28.71
53–54	101,653	687	675.8	0.00674	618	91,820	91,511	2,559,074	27.87
54–55	105,815	675	637.9	0.00636	580	91,202	90,912	2,467,563	27.06
55–56	108,657	737	678.3	0.00676	613	90,622	90,316	2,376,651	26.23
56–57	106,689	784	734.8	0.00732	659	90,009	89,680	2,286,336	25.40
57–58	106,142	842	793.3	0.00790	706	89,350	88,997	2,196,656	24.58
58–59	107,384	929	865.1	0.00861	764	88,644	88,263	2,107,659	23.78
59–60	103,981	1,007	968.4	0.00964	847	87,881	87,457	2,019,396	22.98
60–61	97,063	964	993.2	0.00988	860	87,034	86,604	1,931,939	22.20
61–62	93,115	1,033	1,109.4	0.01103	951	86,174	85,698	1,845,335	21.41
62–63	90,046	1,070	1,188.3	0.01181	1,007	85,223	84,720	1,759,637	20.65
63–64	86,916	1,141	1,312.8	0.01304	1,098	84,216	83,667	1,674,917	19.89
64–65	85,726	1,282	1,495.5	0.01484	1,234	83,118	82,501	1,591,250	19.14
65–66	86,996	1,387	1,594.3	0.01582	1,295	81,884	81,237	1,508,749	18.43
66–67	83,258	1,400	1,681.5	0.01668	1,344	80,589	79,917	1,427,513	17.71
67–68	79,961	1,428	1,785.9	0.01770	1,403	79,245	78,544	1,347,595	17.01
68–69	78,039	1,485	1,902.9	0.01885	1,467	77,842	77,109	1,269,052	16.30
69–70	74,389	1,617	2,173.7	0.02150	1,642	76,375	75,554	1,191,943	15.61
70–71	70,163	1,614	2,300.4	0.02274	1,700	74,733	73,883	1,116,389	14.94
71–72	67,599	1,816	2,686.4	0.02651	1,936	73,033	72,065	1,042,506	14.27
72–73	65,045	1,813	2,787.3	0.02749	1,954	71,097	70,120	970,441	13.65
73–74	60,676	1,905	3,139.6	0.03091	2,137	69,143	68,074	900,320	13.02
74–75	57,975	1,889	3,258.3	0.03206	2,148	67,006	65,931	832,246	12.42
75–76	54,912	1,995	3,633.1	0.03568	2,314	64,857	63,700	766,315	11.82
76–77	51,217	2,089	4,078.7	0.03997	2,500	62,543	61,293	702,615	11.23
77–78	48,251	1,993	4,130.5	p.04047	2,430	60,043	58,828	641,322	10.68
78–79	43,234	2,344	5,421.7	0.05279	3,041	57,613	56,093	582,494	10.11
79–80	47,158	2,399	5,087.2	0.04961	2,707	54,572	53,218	526,401	9.65
80–81	39,462	2,318	5,874.0	0.05706	2,960	51,865	50,385	473,183	9.12
81–82	36,295	2,416	6,656.6	0.06442	3,151	48,905	47,330	422,798	8.65
82–83	31,875	2,360	7,403.9	0.07140	3,267	45,755	44,121	375,468	8.21
83–84	30,470	2,535	8,319.7	0.07987	3,394	42,488	40,791	331,347	7.80
84–85	27,904	2,540	9,102.6	0.08706	3,404	39,094	37,392	290,556	7.43
85–86	24,712	2,458	9,946.6	0.09475	3,382	35,690	34,000	253,163	7.09
86–87	21,302	2,383	11,186.7	0.10594	3,423	32,309	30,597	219,164	6.78
87–88	19,402	2,120	10,926.7	0.10361	2,993	28,886	27,389	188,567	6.53
88–89	14,905	1,993	13,371.4	0.12533	3,245	25,893	24,270	161,177	6.22
89–90	13,873	1,900	13,695.7	0.12818	2,903	22,648	21,196	136,907	6.05
90+	47,650	8,131	17,064.0	1.00000	19,745	19,745	115,710	115,710	5.86

*Rate per 100,000 person years of risk.

survival curve (see the next section). The expectation of life (e_0) is approximately equal to the area under the survival curve. In Chapter 10 this interpretation is discussed further [expression (10.36)].

The crude mortality rate associated with a life table is the total

Table 9-3. United States life expectancy for white males and females (1900–80)

Year	1900	1910	1920	1930	1940	1950	1960	1970	1980
Male	46.6	48.6	54.5	59.7	62.1	66.5	67.4	68.0	70.7
Female	48.7	52.0	55.6	63.5	66.6	72.2	74.1	75.6	78.1

Source: Vital Statistics of the United States, 1983, U.S. Department of Health and Human Services.

number of persons who died divided by the total number of person-years lived by the entire life-table population or

$$\text{crude mortality rate} = \frac{\Sigma d_x}{T_0} = \frac{l_0}{T_0}. \tag{9.8}$$

The crude mortality rate is the reciprocal of the expectation of life at birth or

$$\text{crude mortality rate} = \frac{l_0}{T_0} = \frac{1}{e_0} \quad \text{or} \quad e_0 = \frac{1}{l_0/T_0}. \tag{9.9}$$

Referring to the life table for males (Table 9.1), the crude mortality rate is $100{,}000/6{,}960{,}692 = 0.0144$, or 1,437 deaths per 100,000 person-years of life and $1/0.0144 = 69.607$ years of life are expected to be lived by a newborn male infant who experiences the exact 1980 age-specific mortality rates. A life table formally shows the expected relationship that survival time (average expected lifetime) is inversely related to risk (average rate of death).

Three assumptions are implicit in constructing and interpreting a life table. The life-table structure requires that the same number of births occur each year (l_0 constant). The deaths are assumed to be uniformly distributed within each interval for ages greater than four (thus resulting in $\bar{a}_x = 0.5$), and no population growth occurs (the number of births is equal to the number of deaths each year, and no immigration or emigration occurs). When a population conforms to these three properties, it is called a stationary population. Although stationary human populations do not exist, in most cases changes are sufficiently slow so that postulating that a group of individuals has an approximately stationary structure is not unreasonable, making a life table a useful tool to describe human mortality experience.

Life Table Survival Curve

A fundamental summary statistic derived from a life table is an estimate of a survival curve (introduced in Chapter 1), that is, the probability of surviving beyond a specific point in time. In symbols, $S(x)$ represents the probability of surviving beyond age x. Two identical ways of computing $S(x)$ from a life table are:

$$S(x) = \frac{l_x}{l_0} \tag{9.10}$$

or

$$S(x) = \prod_{i=0}^{x-1} (1 - q_i) = \prod_{i=0}^{x-1} p_i. \tag{9.11}$$

The equivalence of these two calculations comes from the fact that

$$S(x) = \prod_{i=0}^{x-1} p_i = \frac{l_1}{l_0}\frac{l_2}{l_1}\frac{l_3}{l_2}\frac{l_4}{l_3}\cdots\frac{l_{x-1}}{l_{x-2}}\frac{l_x}{l_{x-1}} = \frac{l_x}{l_0}, \tag{9.12}$$

since $p_i = l_{i+1}/l_i$ is the probability of surviving from age i to age $i + 1$ given that the individual is alive at the beginning of the interval. Also note that $S(0) = 1$, which is a property of survival curves in general.

The survival curves for the male (solid line) and female (dotted line) 1980 California populations are displayed in Figure 9.1 (top). A small

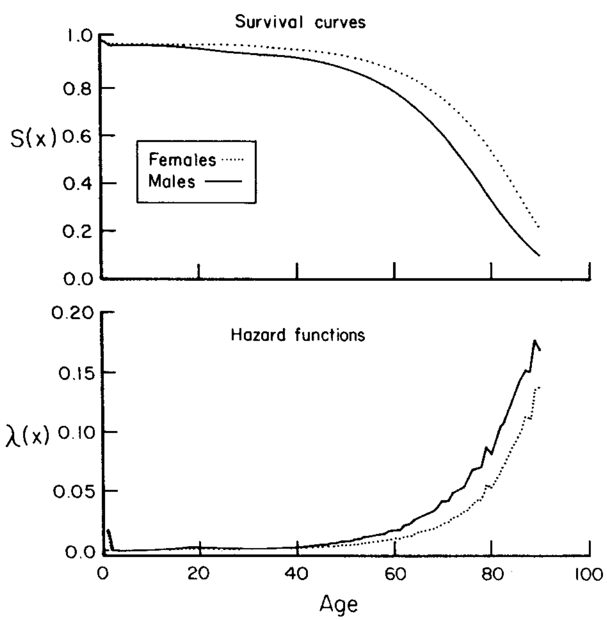

Figure 9-1. Survival curve and hazard function from the life table for white males and females, California, 1980.

decrease in $S(x)$ caused by high rates of infant mortality in the first year of life is followed by a slight and gradual decrease in the probability of survival until about ages 60 or 70, where the $S(x)$ curve begins to fall rapidly. This pattern is often observed in modern human populations. The probability of living more than 90 years is given by the values $S(90) = 0.084$ for males and $S(90) = 0.197$ for females (females are 2.4 times more likely than males to live beyond the age of 90).

Life Table Hazard Function

The slope of the survival curve or the derivative of $S(x)$ at the point x $[dS(x)/dx]$ measures the impact of mortality on a population at a specific age x. The slope indicates the rate of change (intensity of mortality) of the curve representing the probability of surviving beyond a particular point. Analogous to the definition of a mortality rate [expression (1.2)], if the instantaneous slope of the survival curve is measured relative to the proportion surviving up to age x, then the previous definition of a hazard rate emerges, given as

$$\lambda(x) = -\frac{dS(x)/dx}{S(x)}, \tag{9.13}$$

where $\lambda(x)$ represents the hazard rate and the negative sign makes it a positive quantity. A hazard rate applied to mortality data is the instantaneous rate of death, relative to being alive at age x.

To estimate the hazard rate from a life table, it is necessary to make a series of approximations to calculate this theoretical quantity. The slope of the survival curve at the midpoint of the interval x to $x + 1$ is approximately $S(x + 1) - S(x)$, and the value of the survival curve at $x + \frac{1}{2}$ is approximately $[S(x + 1) + S(x)]/2$. These two approximations are exact if the survival curve is a straight line. Combining these two quantities gives an approximate expression for the hazard rate at age $x + \frac{1}{2}$ of

$$\lambda(x + \tfrac{1}{2}) = -\frac{dS(x + \frac{1}{2})/dx}{S(x + \frac{1}{2})} \approx -\frac{S(x + 1) - S(x)}{[S(x + 1) + S(x)]/2}. \tag{9.14}$$

This expression in terms of the number of persons alive at age x (l_x) is

$$\lambda(x + \tfrac{1}{2}) \approx -\frac{l_{x+1} - l_x}{(l_{x+1} + l_x)/2} = -2\frac{p_x - 1}{p_x + 1} = \frac{2q_x}{p_x + 1}, \tag{9.15}$$

where $p_x = l_{x+1}/l_x$.

Since $\log(p) \approx 2(p - 1)/(p + 1)$ for $p > 0.7$, then

$$\lambda(x + \tfrac{1}{2}) \approx -\log(p_x), \tag{9.16}$$

which provides a useful approximation of the hazard rate for most life tables based on human mortality. An expression for the hazard rate at age x is the average of the hazard rates at age $x - \frac{1}{2}$ and $x + \frac{1}{2}$ or

$$\lambda(x) \approx \frac{-[\log(p_{x-1}) + \log(p_x)]}{2}. \tag{9.17}$$

A further simplification is achieved by using yet another approximation, that of $\log(p) \approx p - 1$ for $p > 0.9$, giving

$$\lambda(x + \tfrac{1}{2}) \approx -\log(p_x) \approx q_x \tag{9.18}$$

and, as before,

$$\lambda(x) \approx \frac{q_{x-1} + q_x}{2} \tag{9.19}$$

for age intervals with low probabilities of death. Similar to a mortality rate, a hazard rate is conceptually an instantaneous quantity and must be approximated when the survival curve $S(x)$ is not specified.

Another estimate for the hazard rate $\lambda(x + \frac{1}{2})$ can be derived by noting that a hazard rate is an instantaneous age-specific rate. An average age-specific rate from a life table is estimated by

$$\text{rate} = \frac{d_x}{l_x - 0.5d_x}. \tag{9.20}$$

For a small interval (say, 1 year), the age-specific life-table mortality rate is approximately equal to the hazard rate at the middle of an age interval or

$$\lambda(x + \tfrac{1}{2}) \approx \text{rate} = \frac{d_x}{l_x - 0.5d_x}. \tag{9.21}$$

Two other versions of this expression are used. They are

$$\lambda(x + \tfrac{1}{2}) \approx \frac{q_x}{1 - 0.5q_x} = \frac{2q_x}{p_x + 1}. \tag{9.22}$$

The last expression is the same as the previous expression for the hazard rate [expression (9.15)] derived from different considerations. Again, if d_x is small relative to l_x ($p_x \approx 1$), then $\lambda(x + \frac{1}{2}) \approx q_x$. In general, an approximate life-table hazard rate is

$$\lambda(x + n_x/2) \approx \frac{d_x}{n_x(l_x - 0.5d_x)}, \tag{9.23}$$

where n_x represents the interval length. The accuracy of this expression as an approximation for a hazard rate decreases as the interval length n_x increases for most situations.

The hazard functions (a series of hazard rates) are plotted in Figure 9.1 (bottom) for the California 1980 life tables for males and females. Detail of the mortality pattern is clearly seen from these hazard functions. For example, an inconsistency in the rise of the hazard function for the older age groups is obvious and undoubtedly due to the lack of reliability in reporting of age for older individuals (about 80 years or so).

The shape of the curve observed for the 1980 California life-table populations is typical of most human populations over the entire age span. After the first year of life, the next 60 years are characterized by an essentially level hazard function followed by a sharp increase. However, hazard functions in other contexts take on a variety of shapes. A population subject to only accidental (random) deaths, for example, would have a mortality pattern with a constant hazard function (a horizontal line). A hazard function and a survival curve are related—higher rates of mortality imply lower probabilities of survival. The exact mathematical relationship is described in Chapter 11, and complete discussions are found in technical texts on survival analysis (e.g., [Ref. 2]).

Life tables can be constructed from small sets of data. The principles are the same as those described, but the issue of sampling variation should not be ignored. The values q_x, l_x, etc. are estimated quantities subject to sampling variation, which usually requires reporting their associated standard errors. Huge numbers of individuals make up the California life-table data sets so the precision of the estimates is not much of an issue. For a life table based on a small number of individuals, however, the variability of the estimated quantities should be taken into account. Expressions for the variances of life-table estimates are based on assuming that the probabilities of death can be modeled by binomial distributions (these expressions are presented in detail elsewhere [Ref. 1]). A life table based on small numbers of observations illustrates where 11 individuals failed to respond to a specific treatment ("died") [Ref. 2]. The survival times, amount of time to remission (in weeks), are 5, 5, 8, 8, 12, 23, 27, 30, 33, 43, 45. A life table, based on 10-week intervals, summarizing these data is given in Table 9.4.

The size of the sample used to construct this life table is small, making the variability of the estimates an issue, and, once again, categorizing a continuous variable (survival time) is not an ideal way

Table 9–4. Life table for a small set of data

Interval	Midpoint	"Deaths"	Population	q_x	p_x	l_x	$S(x)$	$\lambda(x+5)$
0–10	5	4	11	0.364	0.636	1.000	1.000	0.044
10–20	15	1	7	0.143	0.857	636	0.636	0.015
20–30	25	2	6	0.333	0.667	545	0.545	0.040
30–40	35	2	4	0.500	0.500	364	0.364	0.067
40–50	45	2	2	1.000	0.000	182	0.182	0.200

to proceed. Small sets of survival data are better analyzed by other approaches (presented in Chapters 10 and 11).

Proportional Hazard Rates—An Example

An instructive application of a life table involves an actuarial-like calculation showing the consequences of lowered hazard rates in a specific population. Suppose a hazard rate is reduced uniformly by a set proportion c [i.e., $\lambda(t) = c\lambda_0(t)$, where $\lambda_0(t)$ is a known or estimated hazard function]; construction of a life table based on such a hazard function describes the resulting mortality experience. Figure 9.2 (top) shows three hypothetical hazard functions based on the 1980 California, white male population mortality rates [$\lambda_0(t)$, top line], where c is set at 0.75, 0.50, and 0.25. The logarithms of the hazard rates clearly show the detail of these curves (Figure 9.2, bottom). Note that the logarithms of a set of proportional hazard rates produce parallel lines. The associated life table can be used to describe the impact of the lower hazard rates.

To summarize the life tables constructed from the three reduced hazard functions, the proportion of individuals alive at ages 65, 75, and 85 years along with the expected length of life from birth for the "proportional populations" are shown in Table 9.5.

It is unlikely that a decrease in mortality would be exactly proportional throughout the life span (i.e., proportional hazards rates); nevertheless, some idea of the impact of decreasing mortality risk is gained by life-table summary values. The percentage of older individuals increases markedly as the age-specific mortality decreases. For example, about 46% of the 1980 California males are older than 75 years, but, when the mortality is reduced by a factor of 4 ($c = 0.25$), this value increases to an estimated 84%. The expected length of life at birth is correspondingly increased from 69.6 to 87.4 years.

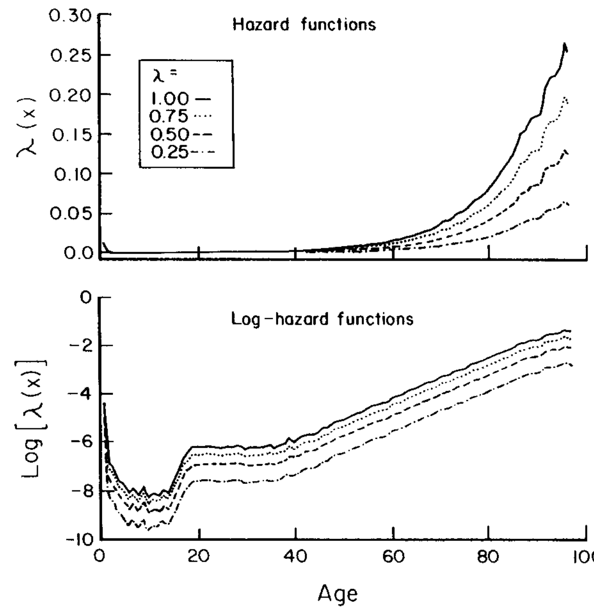

Figure 9–2. Hazard functions and the logarithm of the hazard functions for three hypothetical patterns of mortality based on the white, male mortality rate, California 1980.

The effects on a population of an increasing or decreasing hazard rate are not always clear. As the illustration shows, a hazard rate can be related to more easily interpreted measures of the impact of risk using life-table summaries. A decrease in hazard rate becomes a less abstract expression of risk when translated, for example, into an increase in the number of individuals exceeding a specific age or into an increase in the expected years of remaining life.

Table 9–5. Influence of three hypothetical hazard rates on the 1980 California male population

Hazard	%≥65 years	%≥75 years	%≥85 years	Expectation
1.00λ	69.7	46.0	18.3	69.6
0.75λ	78.7	58.7	30.0	74.7
0.50λ	85.3	70.2	45.1	79.8
0.25λ	92.3	83.8	67.4	87.4

LIFE TABLES: THREE APPLICATIONS OF LIFE TABLE TECHNIQUES

Life-Table Method for Calculating a Survival Probability

The evaluation of the treatment of chronic disease usually involves the assessment of survival (or, perhaps, remission) times. The probability of surviving 5 years after receiving a treatment is a frequent measure of efficacy. Survival data can be collected and recorded in a sequence of intervals to form a series of cohort tables (one for each year of follow-up, for example). It is this follow-up pattern of data collection that allows an efficient estimate of the 5-year survival probability or, in general, an estimate of the survival curve associated with the sampled population. The set of follow-up data in Table 9.6 concerns the survival of six cohorts of kidney cancer patients, illustrating this type of data [Ref. 3].

The complete display of the data set is presented to show the cohorts formed as each year new patients are added to the sample. The interval x to $x+1$ denotes the years survived since the kidney cancer was diagnosed. The column labeled l_x contains the count of the individuals alive at the beginning of the time interval x to $x+1$. The number of deaths in each interval is represented by d_x. The possibility exists that patients are "lost to follow-up" during the time period covered by the study. The count of patients lost during an interval is symbolized by u_x. The last column in the table contains the counts of patients withdrawn from study. Individuals are said to be withdrawn when they are no longer relevant to further calculations. For example, consider the 1950 cohort of 19 patients. Five patients died the first year, and one the second year; two were lost, one each year, and the remaining 11 individuals produced information about the first and second year of survival but cannot be used in calculations for the third year or beyond since they were only observed for a maximum of 2 years. The 11 ($w_2 = 11$) members of this cohort alive at the end of the second year are said to be withdrawn after 2 years and are not part of subsequent calculations. They either survived or died after 1951, but this information is not part of the collected data. The times of these four possible events (l_x, d_x, u_x, and w_x) are recorded to the nearest year in the kidney cancer follow-up data. A summary table that combines the survival experience of all kidney cancer patients for the six cohorts (Table 9.6) is given in Table 9.7. Note that

$$l_{x+1} = l_x - d_x - u_x - w_x. \tag{9.24}$$

If the entire cohort was entered into the study on the first day and followed for at least 5 years and no one was lost, then a 5-year survival probability would be the number who lived 5 years divided by the number who started the study. For most survival data, however, individuals die, are lost, or withdrawn from follow-up at different times during the study period. It is also likely that, during the course of collecting a set of follow-up data, individuals will die from causes other than the one being investigated. Somewhat pragmatically, these

Table 9-6. Calculation of a survival probability: Data

Year	x to $x+1$	l_x	d_x	u_x	w_x
1946	0–1	9	4	1	—
	1–2	4	0	0	—
	2–3	4	0	0	—
	3–4	4	0	0	—
	4–5	4	0	0	—
	5–6	4	0	0	4
1947	0–1	18	7	0	—
	1–2	11	0	0	—
	2–3	11	1	0	—
	3–4	10	2	2	—
	4–5	6	0	0	6
1948	0–1	21	11	0	—
	1–2	10	1	2	—
	2–3	7	0	0	—
	3–4	7	0	0	7
1949	0–1	34	12	0	—
	1–2	22	3	3	—
	2–3	16	1	0	15
1950	0–1	19	5	1	—
	1–2	13	1	1	11
1951	0–1	25	8	2	15

Table 9-7. Calculation of a survival probability from tabled data: Summary data

$x-x+1$	l_x	d_x	u_x	w_x
0–1	126	47	4	15
1–2	60	5	6	11
2–3	38	2	0	15
3–4	21	2	2	7
4–5	10	0	0	6
5–6	4	0	0	4

individuals are usually classified as lost (i.e., u_x is increased), which introduces no bias if these deaths are completely unrelated to the disease under study. The sequential pattern of follow-up data collection makes it necessary to piece together the followup information.

Notice that 15 individuals in the 1951 cohort were withdrawn after 1 year. If the exact time these patients were observed was known, then the total person-years of risk would be the sum of their observed individual survival times. When this information is not available, estimates of survival time must be adjusted to compensate for the incomplete nature of the data. One approach is to assume that each person withdrawn during an interval, on the average, contributes one-half an interval of time ($\bar{a}_x = 0.5$) to the total survival time. That is, it is postulated that individuals come into the study uniformly throughout the follow-up period, implying they will be withdrawn uniformly from observation. If this is the case, then attributing one-half an interval's time to each person withdrawn is "on the average" correct. A similar assumption is usually made about individuals lost from follow-up. An estimate of the probability of death (q_x) that accounts for the two types of incomplete information is made by reducing the number of persons beginning the interval (l_x) to compensate for those individuals lost (u_x) and withdrawn (w_x) during the interval. Specifically,

$$l'_x = l_x - 0.5u_x - 0.5w_x, \tag{9.25}$$

where l'_x is the "effective" persons at risk in the interval and the probability of death within an interval is then estimated by

$$q_x = \frac{d_x}{l'_x}. \tag{9.26}$$

The adjusted persons at risk (l'_x) better reflects the underlying situation.

An alternate view of this adjustment comes from noting that the observed number of deaths is understated since lost and withdrawn individuals are not followed for, on the average, half an interval and deaths occurring during that time will not be recorded. An estimate of this additional number of "deaths" is $0.5(u_x + w_x)q_x$. Adding these "deaths" to the number of observed deaths gives an estimate of the probability of death as

$$q_x = \frac{d_x + 0.5(u_x + w_x)q_x}{l_x}, \tag{9.27}$$

and solving for q_x produces the same result as before ($q_x = d_x/l'_x$).

Employing the value q_x to estimate the proportion of deaths among those who were lost or withdrawn implies that these individuals do not differ in their mortality experience from those who continued to be followed. This assumption may not be tenable in some situations. For example, it might be that lost individuals are more likely to have survived or, perhaps, more likely to have died; a suitable q_x should be used under these conditions. A more subtle implication of employing l'_x is the implicit assumption that mortality experience is unrelated to the probability that an individual is withdrawn from follow-up.

Analogous to the life-table calculation of the survival curve, the survival probabilities are

$$\hat{P}_k = \prod_{x=0}^{k-1} p_x, \tag{9.28}$$

where, as before, $p_x = 1 - q_x$. The value $\hat{P}_k$ is the probability of surviving up to the kth time interval. Applying these estimates to the kidney cancer data gives Table 9.8.

The 5-year survival probability is

$$\hat{P}_5 = (0.597)(0.903)(0.934)(0.879)(1.000) = 0.442$$

(standard error $= 0.060$). The variance of these estimates comes from the expression

$$\text{variance}(\hat{P}_k) = P_k^2 \sum_{i=0}^{k-1} \frac{q_x}{l'_x p_x}. \tag{9.29}$$

The variance estimate is often referred to as "Greenwood's formula" after Major M. Greenwood, an early biostatistician, and is used to test hypotheses or construct confidence intervals for specific estimated survival probabilities.

Another estimate of the 5-year survival probability is the number of individuals who survived 5 years divided by those who began the study at least 5 years previously. Only the 1946 cohort can be used to estimate this 5-year survival probability since the other cohorts contain

Table 9–8. Calculation of a 5-year survival rate from tabled data: Calculations

Interval	d_x	l'_x	q_x	p_x	$\hat{P}_x$	Πp_x	Std. error
0–1	47	116.5	0.403	0.597	$\hat{P}_0$	1.000	—
1–2	5	51.5	0.097	0.903	$\hat{P}_1$	0.597	0.045
2–3	2	30.5	0.066	0.934	$\hat{P}_2$	0.539	0.048
3–4	2	16.5	0.121	0.879	$\hat{P}_3$	0.503	0.051
4–5	0	7.0	0.000	1.000	$\hat{P}_4$	0.442	0.060
5–6	0	2.0	0.000	1.000	$\hat{P}_5$	0.442	0.060

individuals with less than 5 years of follow-up time. The 5-year survival probability is then $4/9 = 0.444$, with a standard error of 0.166 (assuming the lost individual survived). Using all available data rather than a single cohort produces a more precise estimate of the 5-year survival probability (ratio of standard errors $= 0.166/0.060 = 2.7$ in the kidney cancer example). However, the cost of this increased precision is possible bias from the assumption that the mortality experience over time is similar enough among cohorts that combining data for all years reflects the overall mortality experience of all observed individuals.

Another important summary of survival data is an estimate of the mean time individuals survived. This calculation is complicated by the fact that the time of death is not known for all participating individuals. For the data recorded on the 126 kidney cancer patients, the mean survival time is 3.523 years. Mean survival time calculations are discussed in Chapter 10.

Survival patterns experienced by different groups can be summarized and compared using specific survival probabilities. Two such groups from the WCGS data are those with high values of the body-mass index (greater than the 75th percentile) and those with smaller body-mass values (less than the 75th percentile). The data and the calculated "survival" probabilities are (here "survival" means free from a coronary event) given in Tables 9.9 and 9.10.

The comparison of these survival probabilities shows a lower probability (higher risk) of "survival" for those individuals with high body-mass indexes. For example, the 5-year survival probability of 0.940 for high values of body-mass index is less than the 0.961 observed for individuals with "normal" values of the body-mass index. The

Table 9–9. WCGS body mass > 75th percentile

$x-x+1$	l_x	d_x	w_x	q_x	$\hat{P}_x$	Std. error
0–1	871	6	0	0.0069	1.000	—
1–2	865	8	21	0.0094	0.993	0.0028
2–3	836	16	19	0.0194	0.984	0.0043
3–4	801	9	23	0.0114	0.965	0.0063
4–5	769	11	14	0.0144	0.954	0.0072
5–6	744	12	19	0.0163	0.940	0.0082
6–7	713	18	46	0.0261	0.925	0.0092
7–8	649	9	195	0.0163	0.901	0.0106
8–9	445	5	431	0.0218	0.886	0.0115
9–10	9	0	9	0.0000	0.867	0.0141

Table 9–10. WCGS body mass < 75th percentile

$x-x+1$	l_x	d_x	w_x	q_x	$\hat{P}_x$	*Std. error*
0–1	2283	9	4	0.0039	1.000	
1–2	2270	20	24	0.0089	0.996	0.0013
2–3	2226	23	50	0.0104	0.987	0.0024
3–4	2153	18	41	0.0084	0.977	0.0032
4–5	2094	18	37	0.0087	0.967	0.0037
5–6	2039	27	61	0.0134	0.961	0.0042
6–7	1951	14	99	0.0074	0.947	0.0048
7–8	1838	22	502	0.0139	0.940	0.0051
8–9	1314	12	1271	0.0177	0.927	0.0057
9–10	31	0	31	0.0000	0.911	0.0073

standard errors for these estimates indicate that this difference is not likely to have occurred by chance variation. For the >75th percentile group the approximate 95% confidence interval is (0.924, 0.956) and for the <75th percentile group it is (0.953, 0.969) based on "Greenwood's" variance [expression (9.29)]. A plot of these two sets of survival probabilities is given in Figure 9.3.

The WCGS follow-up times are recorded exactly (to the nearest day); so the probability that a coronary event does not occur ("survival") can be calculated without assumptions about the indiv-

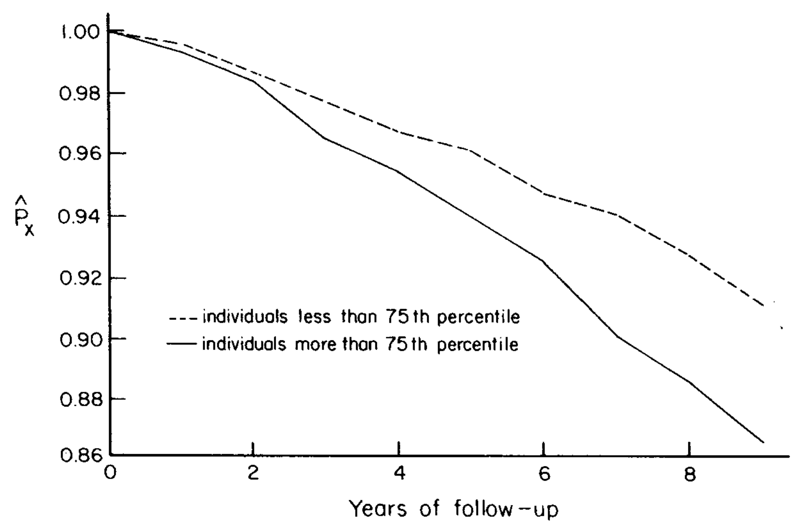

Figure 9–3. Survival probabilities for individuals with a body-mass index less than and greater than the 75th percentile (WCGS data).

iduals lost or withdrawn during the follow-up years. Instead of using 0.5 years of risk, the exact total time contributed by individuals lost or withdrawn can be directly calculated and produces the exact number persons at risk. The difference between the exact and approximate approaches is inconsequential in this example. The 9-year probability using the exact follow-up times is 0.869 for individuals with body-mass indexes in the upper quartile and 0.913 for the "normal" body-mass individuals, compared to the approximate ($\bar{a}_x = 0.5$) values 0.867 and 0.911, respectively. In other study settings, however, individuals lost or withdrawn from follow-up may have different outcome experiences, necessitating careful selection of an adjustment method when exact values are not available.

Three assumptions about the structure of the sampled population are made to calculate a survival curve using life-table techniques. First, all lost and withdrawn subjects are assumed to contribute, on the average, half the survival information of an individual followed for a complete year (or complete time interval). Second, the data collected for a number of cohorts are combined to maximize the number of observations available in each time interval to calculate the probability of death. To give an unbiased estimate of survival probabilities, all cohorts must experience the same pattern of mortality during the follow-up period (again, the absence of interaction permits the data to be combined). In terms of the kidney cancer data, the individuals who entered the study in 1947, for example, are assumed to have the same pattern of mortality as the patients who entered in 1951, which allows the data from both groups to be used in the calculation of the probability of surviving the first year after diagnosis. The third assumption is that the lost and withdrawn individuals have the same probability of death as the individuals remaining in the follow-up data set. This conjecture is probably the most tenuous when applied to individuals lost from observation. Situations certainly arise where other assumptions make sense. For example, if it is assumed that all individuals classified as lost actually survived, then

$$q'_x = \frac{d_x + 0.5w_x q'_x}{l_x} \quad \text{or} \quad q'_x = \frac{d_x}{l_x - 0.5w_x} \tag{9.30}$$

or, if all individuals lost in fact died, then

$$q''_x = \frac{d_x + 0.5(u_x + w_x q''_x)}{l_x} \quad \text{or} \quad q''_x = \frac{d_x + 0.5u_x}{l_x - 0.5w_x}. \tag{9.31}$$

The probabilities q'_x and q''_x represent the extremes in terms of the impact of the lost individuals on the calculation of the q_x. These two extremes applied to the kidney cancer data yield 5-year survival probabilities of $\hat{P}'_5 = 0.454$ if all lost patients survive and $\hat{P}''_5 = 0.387$ if all lost patients die.

Life-Table Measures of Specific Causes of Death

Hundreds of causes of death act simultaneously within human populations. Two approaches based on life-table methods provide an opportunity to isolate the individual impact of specific causes on the pattern of human mortality. These methods help resolve two questions:

1. What is the age structure throughout the life span associated with specific causes of death, taking into account other causes?
2. How does the probability of death from a specific cause change when other causes are "eliminated" from the population?

The first question is answered by applying a multiple cause life table (also called a multiple decrement life table). The second question is addressed by a competing risk analysis.

Multiple Cause Life Table

A multiple-cause life table is similar to the single-cause life table but is used to describe simultaneously the mortality patterns of a number of diseases in a population. The goal of such a table is to organize and display the age structure of individuals dying of specific causes. The mechanics of constructing these age distributions are defined and illustrated by a set of data consisting of California resident males who died during 1980. The causes of death come from death certificates, classified according to the ninth revision of the International Classification of Diseases (ICD9) [Ref. 4]. These deaths are classified into four categories—death from lung cancer (ICD9, code 162), deaths from ischemic heart disease (ICD9, codes 410 to 414), deaths from motor vehicle accidents (ICD9, codes E810 to E819), and deaths from all other causes. Also necessary is a series of age-specific population counts—the 1980 U.S. Census counts of California male residents are used. The following life-table construction is abridged, which means that the lengths of the age intervals are not consistently 1 year. Most age intervals are 5-year lengths (represented as n_x; for example, $n_{60} = 5$ years).

The basic components required to construct a multiple-cause life table are the total number of deaths, the age-, cause-specific numbers of deaths and the age-specific midyear populations. That is,