iduals lost or withdrawn during the follow-up years. Instead of using 0.5 years of risk, the exact total time contributed by individuals lost or withdrawn can be directly calculated and produces the exact number persons at risk.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 In other study settings, however, individuals lost or withdrawn from follow-up may have different outcome experiences, necessitating careful selection of an adjustment method when exact values are not available.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 First, all lost and withdrawn subjects are assumed to contribute, on the average, half the survival information of an individual followed for a complete year (or complete time interval).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 The third assumption is that the lost and withdrawn individuals have the same probability of death as the individuals remaining in the follow-up data set.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,

D_x = total number of recorded deaths in the age interval x to $x + n_x$,
$D_x^{(i)}$ = number of recorded deaths from ith cause in the age interval x to $x + n_x$, and
P_x = total number of individuals at risk ages x to $x + n_x$ at midyear.

These quantities for male residents of California (1980) are given in Table 9.11.

Average age-specific mortality rates calculated from Table 9.11 are $R_x = D_x/P_x$ for the age interval x to $x + n_x$ and, similar to the single-cause, complete life table,

$$q_x = \frac{n_x R_x}{1 + 0.5 n_x R_x} \qquad (9.32)$$

is again the conditional probability of death, where n_x is the length of interval starting at age x. These probabilities are an extension of those calculated in the single-cause life table [expression (9.4)] applied to age intervals with widths of n_x years. The value q_x is, as before, the conditional probability of death between ages x and $x + n_x$ for

Table 9–11. Deaths from four causes: California, males, 1980

	P_x	$D_x^{(1)}$	$D_x^{(2)}$	$D_x^{(3)}$	$D_x^{(4)}$
Age	Population	Lung cancer	IHD	Motor	All other
0–1	193,310	1	2	3	2,507
1–4	515,150	1	3	58	375
5–9	843,750	0	2	90	195
10–14	915,240	0	1	80	248
15–19	1,091,684	3	1	523	1,162
20–24	1,213,068	4	6	965	1,507
25–29	1,132,811	3	13	627	1,665
30–34	1,008,606	12	63	437	1,547
35–39	776,545	36	136	277	1,371
40–44	629,452	85	306	201	1,510
45–49	578,420	225	567	197	2,115
50–54	578,795	445	1,050	150	3,163
55–59	573,119	786	1,807	147	4,663
60–64	467,607	1,059	2,528	129	5,603
65–69	378,259	1,297	3,328	97	7,014
70–74	269,849	1,266	3,815	89	7,423
75–79	175,580	941	3,793	99	7,508
80–84	95,767	557	3,452	44	6,202
85+	78,832	430	5,249	61	8,222
Total	11,515,844	7,151	26,122	4,274	64,000

individuals alive at age x. For example, the probability of death for individuals age 60 before age 65 is

$$q_{60} = \frac{5(0.0199)}{1 + 0.5(5)0.0199} = 0.0949, \text{ where } R_{60} = \frac{9319}{467607} = 0.0199. \qquad (9.33)$$

To "fine tune" these calculations, the 0.5 in the denominator is sometimes replaced by better estimates of the average time lived by those who died. The use of values other than 0.5, however, has little impact on the final calculations for data covering the entire life span.

To compute the cause-specific conditional probabilities of death, the q_x values are distributed proportionally (prorated) by the observed numbers of death. Since

$$q_x^{(i)} = \frac{n_x D_x^{(i)}}{P_x + 0.5 n_x D_x} \quad \text{and} \quad q_x = \frac{n_x D_x}{P_x + 0.5 n_x D_x}, \qquad (9.34)$$

then

$$q_x^{(i)} = \frac{D_x^{(i)}}{D_x} q_x. \qquad (9.35)$$

Continuing the illustration for the age interval 60 to 65, the probability of dying from lung cancer between ages 60 and 65 for individuals age 60 is

$$q_{60}^{(\text{lung})} = \frac{1059}{9319} 0.0949 = 0.0108. \qquad (9.36)$$

The value $q_x^{(i)}$ is the age-, cause-specific conditional probability of death before age $x + n_x$ for those alive at age x. These conditional probabilities for the illustrative data are given in Table 9.12.

Since all causes of death are included, $q_x = \Sigma\, q_x^{(i)}$. The $q_x^{(i)}$ values calculated from the California mortality data indicate that the cause-specific conditional probabilities for lung cancer ($q_x^{(1)}$) increase rapidly after age 40 until about age 70 and then increase less rapidly in the older ages. The same probabilities for ischemic heart disease (IHD) ($q_x^{(2)}$) also increase sharply at about age 70 but are generally associated with older individuals (shifted to the right). The conditional probabilities describing deaths from motor vehicle accidents ($q_x^{(3)}$), however, increase until ages 20 to 25, decrease and remain fairly constant until age 70 and then sharply increase again. The cause-specific probabilities for three causes of death are shown in Figure 9.4 (smoothed).

Again parallel to the single-cause life table, an arbitrary number of individuals (l_0) can be distributed according to the conditional probabilities of death to produce the distribution of the number of life-

Table 9-12. Conditional probabilities: California, males, 1980

	q_x	$q_x^{(1)}$	$q_x^{(2)}$	$q_x^{(3)}$	$q_x^{(4)}$
Age	Total	Lung cancer	IHD	Motor	All others
0–1	0.01292	0.00001	0.00001	0.00002	0.01289
1–4	0.00339	0.00001	0.00002	0.00045	0.00291
5–9	0.00170	0.00000	0.00001	0.00053	0.00115
10–14	0.00180	0.00000	0.00001	0.00044	0.00135
15–19	0.00771	0.00001	0.00000	0.00239	0.00530
20–24	0.01018	0.00002	0.00002	0.00396	0.00618
25–29	0.01014	0.00001	0.00006	0.00275	0.00731
30–34	0.01016	0.00006	0.00031	0.00216	0.00763
35–39	0.01165	0.00023	0.00087	0.00177	0.00878
40–44	0.01656	0.00067	0.00241	0.00158	0.01190
45–49	0.02648	0.00192	0.00484	0.00168	0.01804
50–54	0.04069	0.00377	0.00889	0.00127	0.02677
55–59	0.06256	0.00664	0.01527	0.00124	0.03941
60–64	0.09492	0.01079	0.02575	0.00131	0.05707
65–69	0.14397	0.01591	0.04082	0.00119	0.08604
70–74	0.20896	0.02101	0.06330	0.00148	0.12317
75–79	0.29891	0.02279	0.09187	0.00240	0.18185
80–84	0.42235	0.02294	0.14217	0.00181	0.25543
85+	1.00000	0.03080	0.37595	0.00437	0.58888

Table 9-13. Deaths from four causes: California, males, 1980

	l_x	$d_x^{(1)}$	$d_x^{(2)}$	$d_x^{(3)}$	$d_x^{(4)}$
Age	Total	Lung cancer	IHD	Motor	All other
0–1	1,000,000	5	10	15	12,885
1–4	987,084	8	23	444	2,869
5–9	983,740	0	12	524	1,136
10–14	982,069	0	5	429	1,329
15–19	980,305	13	4	2,339	5,197
20–24	972,751	16	24	3,849	6,012
25–29	962,850	13	55	2,651	7,040
30–34	953,091	56	296	2,054	7,272
35–39	943,412	217	821	1,673	8,280
40–44	932,421	624	2,248	1,476	11,091
45–49	916,982	1,760	4,435	1,541	16,543
50–54	892,703	3,362	7,933	1,133	23,896
55–59	856,379	5,689	13,078	1,064	33,748
60–64	802,800	8,659	20,671	1,055	45,814
65–70	726,601	11,560	29,663	865	62,517
70–74	621,996	13,066	39,374	919	76,611
75–79	492,026	11,214	45,203	1,180	89,476
80–84	344,954	7,913	49,042	625	88,111
85+	199,263	6,137	74,913	871	117,343

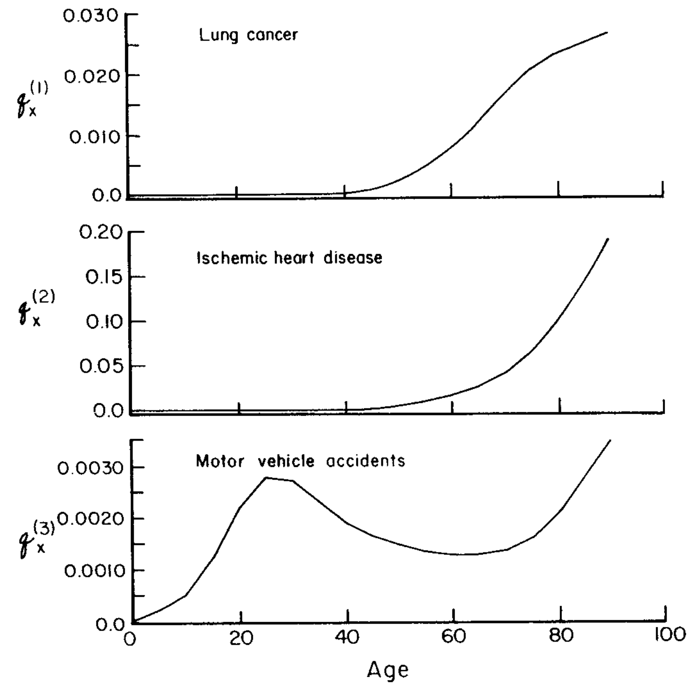

Figure 9-4. Cause-specific probabilities of death for three specific causes (lung cancer, ischemic heart disease, and motor vehicle accidents) for California males, 1980.

table "deaths" for a population with a pattern of age-specific mortality described by the estimated $q_x^{(i)}$ values. The cohort constructed from the California data is shown in Table 9.13.

The life-table deaths given in Table 9.13 come from applying the relationship

$$d_x^{(i)} = l_x q_x^{(i)} \tag{9.37}$$

where, as before, l_x represents the number of persons alive at the beginning of age interval x. For example, the number of persons age 60 who die from lung cancer between age 60 to 65 is

$$d_{60}^{(\text{lung})} = 802800(0.0108) = 8659. \tag{9.38}$$

An additional table calculated by accumulating the deaths in each cause-specific category is also a useful description of the life-table population. These sums represent the number of individuals who reach age x and will ultimately die of a specific cause. In symbols,

$$W_x^{(i)} = d_x^{(i)} + d_{x+n_x}^{(i)} + \cdots + d_{x'}^{(i)} \quad (9.39)$$

and to illustrate

$$W_{60}^{(\text{lung})} = 8659 + 11560 + \cdots + 7913 + 6137 = 58550 \quad (9.40)$$

is the number of individuals who reach age 60 who will eventually die of lung cancer. Again for the California data, see the values in Table 9.14.

The cumulative numbers of deaths provide the values necessary to estimate the probability of death before age x for each cause. That is, for the ith cause

$$F_x^{(i)} = 1 - \frac{W_x^{(i)}}{W_0^{(i)}} \quad (9.41)$$

is the probability of dying before age x. Among individuals dying of lung cancer, the probability of dying before age 60 is

$$F_{60}^{(\text{lung})} = 1 - \frac{58550}{70313} = 0.1673, \quad (9.42)$$

Table 9-14. Expected number of deaths after age x: California, males, 1980

	$W_x^{(1)}$	$W_x^{(2)}$	$W_x^{(3)}$	$W_x^{(4)}$
Age	Lung cancer	IHD	Motor	All other
0–1	70,313	287,809	24,707	617,171
1–4	70,308	287,799	24,691	604,285
5–9	70,301	287,776	24,248	601,416
10–14	70,301	287,765	23,723	600,280
15–19	70,301	287,759	23,295	598,951
20–24	70,287	287,755	20,955	593,754
25–29	70,271	287,731	17,106	587,742
30–34	70,259	287,676	14,455	580,702
35–39	70,202	287,380	12,401	573,430
40–44	69,985	286,558	10,728	565,151
45–49	69,360	284,311	9,251	554,059
50–54	67,601	279,876	7,711	537,516
55–59	64,239	271,943	6,577	513,620
60–64	58,550	258,865	5,513	479,872
65–69	49,891	238,194	4,459	434,058
70–74	38,330	208,531	3,594	371,540
75–79	25,264	169,157	2,676	294,929
80–84	14,050	123,955	1,496	205,454
85+	6,137	74,913	871	117,343

or about 17% of the lung cancer deaths occur before age 60. Table 9.15 shows cumulative probabilities of death (F_x values) for the California 1980 data.

The age structure for each cause of death throughout the life span is apparent from the F_x values and the patterns for separate causes of death can be contrasted. For example, 78% of all motor vehicle accident deaths occur by age 60, while 17% of lung cancer deaths occur before age 60. These cumulative distributions are shown in Figure 9.15, and a few representative summary values are given in Table 9.16.

The cumulative distributions reveal the distinct pattern of mortality associated with three specific causes. Motor vehicle accidents, expectedly, have the greatest impact at the younger ages, while, perhaps less expectedly, the ischemic heart disease is associated with the older ages, producing a median age at death of 78.8 years.

Lifetime Probability of Death

A multiple-cause life table allows a direct calculation of the lifetime probability of death from a specific cause, which is occasionally a useful summary of risk. The probability of dying from a specific cause is

Table 9-15. Cumulative distributions for four causes of death: California, males, 1980

	$F_x^{(1)}$	$F_x^{(2)}$	$F_x^{(3)}$	$F_x^{(4)}$
Age	Lung cancer	IHD	Motor	All other
0–1	0.00000	0.00000	0.00000	0.00000
1–4	0.00007	0.00004	0.00062	0.02088
5–9	0.00018	0.00012	0.01859	0.02553
10–14	0.00018	0.00016	0.03980	0.02737
15–19	0.00018	0.00017	0.05716	0.02952
20–24	0.00037	0.00019	0.15184	0.03794
25–29	0.00060	0.00027	0.30764	0.04768
30–34	0.00078	0.00046	0.41494	0.05909
35–39	0.00158	0.00149	0.49809	0.07087
40–44	0.00467	0.00435	0.56580	0.08429
45–49	0.01355	0.01216	0.62555	0.10226
50–54	0.03858	0.02757	0.68792	0.12906
55–59	0.08640	0.05513	0.73379	0.16778
60–64	0.16730	0.10057	0.77685	0.22246
65–69	0.29045	0.17239	0.81954	0.29670
70–74	0.45486	0.27545	0.85453	0.39799
75–79	0.64069	0.41226	0.89171	0.52213
80–84	0.80018	0.56932	0.93946	0.66710
85+	0.91272	0.73971	0.96476	0.80987

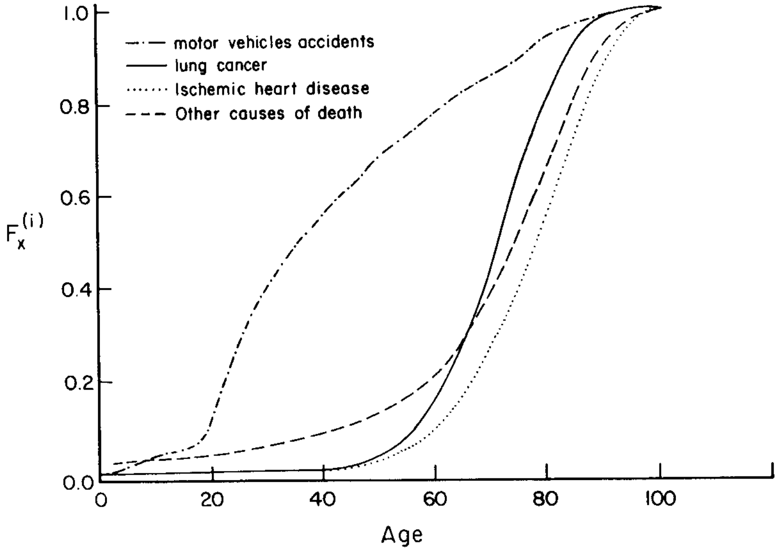

Figure 9-5. Cumulative distributions of age at death for three specific causes (lung cancer, ischemic heart disease, and motor vehicle accidents) for California males, 1980.

estimated by the number of people who died of that cause divided by the number of persons who could have died (those at risk). The table of the expected numbers of deaths after a specific age contains this information (Table 9.14). The first row in the table contains the total number of individuals ultimately dying from each cause over the entire life span. Since 1,000,000 males make up the 1980 California life-table "population at risk" (sum of the first row of Table 9.14), then

P(dying from lung cancer) = 70,313/1,000,000 = 0.070
P(dying from ischemic heart disease) = 287,809/1,000,000 = 0.288
P(dying from motor vehicle accident) = 24,707/1,000,000 = 0.025
P(dying from other causes) = 617,170/1,000,000 = 0.617

Table 9-16. Median age (as well as 25th and 75th percentiles) at death

	Median	25th percentile	75th percentile
Lung cancer	71.98	64.45	78.77
Ischemic heart disease	78.82	69.34	86.40
Motor vehicle accidents	36.40	24.20	58.09
Other causes	75.03	63.16	83.79
All causes	74.64	63.37	83.86

are the lifetime probabilities of dying from any one of the three specific causes.

Each row in the table allows the estimation of the lifetime probability associated with individuals of a specific age. For example, for males age 60, the lifetime probability of dying of lung cancer is 58,550/802,800 = 0.073, where 802,800 is the number of individuals alive at the beginning of the age interval 60–65 (the sum of the row age 60–65) and 58,550 is the number who died of lung cancer after age 60. Three cause-specific conditional probabilities for the 1980 California data are:

P(dying from lung cancer after age 60) = 58,550/802,800 = 0.073
P(dying from ischemic heart disease after age 60) = 258,865/802,800 = 0.322
P(dying from motor vehicle accident after age 60) = 5,513/802,800 = 0.007
P(dying from other causes after age 60) = 479,872/802,800 = 0.598.

The cumulative probability of death from a multiple-cause life table is related to the lifetime probability of death from a specific cause. The probability $1 - F_x^{(i)}$ is the conditional probability of death after age x among those who ultimately die of cause i. The lifetime probability of death from a specific cause i is the conditional probability of death from cause i for all individuals who reach age x. That is, the first probability is P(death after age x|death from cause i) and the second is P(death from cause i|death after age x). Specifically, $1 - F_{60}^{(\text{lung})} = P$(death after 60|death from lung cancer) = 58,550/70,313 = 0.833 and P(death from lung cancer|death after 60) = 58,550/802,800 = 0.073.

Competing Risks

British statistician William Farr (1875) was among the first to discuss the problem of estimating the risk of one disease while other risks are operating in the studied population. This problem was also explored by the early French mathematicians Bernoulli and D'Alembert and later by a British actuary Makeham. The issues are neatly summarized by the following simple example given by J. Berkson and L. Elveback [Ref. 5]:

> Two marksmen shoot at a range of targets under conditions in which, if a target is struck, it instantly drops from view so that it cannot be struck again. Represent the striking rate of marksman 1, that is the probability of a hit when he is firing alone, as Q_1 and similarly the rate of marksman 2 when he is firing alone as Q_2. The probability when one risk operates alone is called the net risk or rate and is represented by upper case Q; when it operates together with another risk it is called the crude risk or rate and is represented by lower case q.

Suppose N targets are exposed and marksman 1 shoots first, followed by marksman 2:
Rate for 1 is $q_1 = Q_1$;
Rate for 2 is $q_2 = (1 - Q_1)Q_2$;
Total rate is $q = q_1 + q_2 = Q_1 + Q_2 - Q_1Q_2$.
Suppose marksman 2 shoots first, followed by marksman 1, then:
Rate for 2 is $q_2 = Q_2$;
Rate for 1 is $q_1 = (1 - Q_2)Q_1$;
Total rate is $q = q_1 + q_2 = Q_1 + Q_2 - Q_1Q_2$.

It is seen that the total crude rate with both marksmen shooting is the same, whichever marksmen shoots and assuming independence of the net probabilities Q_1 and Q_2, this will be true in general. Regardless of the ordering of the shooting or whether the two marksmen shoot together, the total crude rate is given by the "total rate," which, of course, can be derived as the complement of the product of the probabilities, $P_1 = 1 - Q_1$ and $P_2 = 1 - Q_2$, of not being struck (survival rate).

If, from independent trials, we know Q_1, the net rate of marksman 1, and have a record of q, the crude rate when both shot together, we can derive the net rate Q_2 from "total rate":

$$Q_2 = \frac{q - Q_1}{1 - Q_1}. \qquad (9.43)$$

Rarely are the net probabilities Q_1 or Q_2 known, but, rather, the crude probabilities q_1, q_2, and q can be estimated from collected data. Manipulation of these crude probabilities, under specific conditions, allows estimation of the net probabilities from observed data.

For the following discussion of competing risks, it is assumed that only two causes of death are of interest and only a single age interval is considered (simply 0 to 1). These two assumptions do not affect the principles underlying the competing risk argument (mathematicians say, "there is no loss of generality") and simplify the notation.

The formal definitions of the two central probabilities are:

Crude probability: q_i = the probability an individual who is alive at the start of the interval dies from cause i in the presence of cause j, sometimes called the mixed probability of death.
Net probability: Q_i = the probability an individual who is alive at the start of the interval dies from cause i when cause j is not present, sometimes called the pure probability of death.

The marksman example shows a relationship between the net and crude probabilities [expression (9.43)], but is not much use unless one of the net probabilities is known. To estimate the net probabilities further statistical structure is needed. First, assume that the net probabilities are described by exponential functions, where λ_1 and λ_2 are hazard rates associated with causes 1 and 2, respectively, and where

$$Q_1 = 1 - e^{-\lambda_1} \quad \text{and} \quad Q_2 = 1 - e^{-\lambda_2} \qquad (9.44)$$

and, second, that the probability of surviving the interval is

$$P(\text{surviving}) = P_1P_2 = (1 - Q_1)(1 - Q_2) = (e^{-\lambda_1})(e^{-\lambda_2}) = e^{-\lambda_1 + -\lambda_2} = e^{-\lambda}, \qquad (9.45)$$

where $\lambda = \lambda_1 + \lambda_2$. That is, cause 1 and cause 2 are statistically independent. Cause 2 can be thought of as a specific cause of death and cause 1 as all the other causes combined. Then, the net probability Q_1 describes the likelihood of death as if death from cause 2 was not possible (cause 2 "removed"). The exponential survival model will be explored in more detail in the next chapter.

Expression (9.45) for the probability of surviving the interval is valid only when cause 1 and 2 are statistically independent. Although death from cause 1 is mutually exclusive of death from cause 2, it is still important that the mechanisms underlying these two events act independently. In terms of the marksman example, independence means that the hits and misses of one marksman do not influence the accuracy of the other marksman and conversely. Equivalently, cause of death 1 is assumed not to be related in any way to cause of death 2. Independence of causes of death is certainly not a realistic assumption for some diseases, particularly chronic diseases. The influence of non-independence of diseases on the estimate of the net probabilities has not been extensively studied.

These two assumptions (exponential survival and independence) make it possible to estimate the risk from one cause while the other cause is "removed" from consideration (net probability). To estimate the net probability of death, a bit of algebra relates the crude and net probabilities. Consider q = crude probability of death in the interval, death from either from cause 1 or 2, then

$$P(\text{death}) = q = 1 - P_1P_2 = 1 - e^{-\lambda}. \qquad (9.46)$$

Note that the crude probability has the same form as both net probabilities. Furthermore,

$$(1 - q)^{\lambda_i/\lambda} = e^{-\lambda_i} = P_i, \quad \text{giving } Q_i = 1 - P_i = 1 - (1 - q)^{\lambda_i/\lambda}. \qquad (9.47)$$

This basic relationship [expression (9.47)] allows the estimation of the net probabilities since the ratio of the two hazard rates λ_i/λ is estimated by d_i/d, where d_i represents the number of deaths from cause i and $d = d_1 + d_2$ represents the total number of deaths from both causes in the time interval being considered. The estimated net probability of death from cause i is, then,

$$\hat{Q}_i = 1 - \left(1 - \frac{d}{l}\right)^{d_i/d}, \tag{9.48}$$

where l individuals are at risk from both causes of death at the beginning of the interval.

The assumption that the net probabilities are a simple exponential function may not be appealing in some situations [expression (9.44)]. An alternative estimate of the net probability can be derived from intuitive considerations that do not involve an exponential risk model. Individuals at risk can be classified into three categories: (1) died of cause 1, (2) died of cause 2, or (3) lived through the interval. A death from cause 2 can be considered as a person "lost to follow-up" with respect to calculations for cause 1. When cause 2 is "removed," deaths from cause 1 are undercounted since the former "lost to follow-up" are then at risk. That is, the direct estimate of the net probability is too small since a proportion of the individuals who would have died of cause 2 and are "lost" can now die of cause 1. Those who would have died of cause 2 are exposed to risk, on the average, for half the interval so that $0.5d_2$ represents the additional number of individuals at risk when cause 2 is "removed." The value $0.5d_2Q_1$ estimates the number of deaths from cause 1 among the individuals who would have died from cause 2 if it were present. Therefore, "correcting" the number of deaths d_1 gives

$$\hat{Q}_1' = \frac{d_1 + 0.5d_2\hat{Q}_1'}{l} \tag{9.49}$$

and solving for the net probability Q_1' yields

$$\hat{Q}_1' = \frac{d_1}{l - 0.5d_2}. \tag{9.50}$$

The probability $\hat{Q}_1'$ is another estimate of the net probability of death from cause 1 among l individuals at risk. The net probability $\hat{Q}_1'$ is greater than crude probability q_1 since additional individuals are at risk and die of cause 1 when cause 2 is "removed." In general,

$$\text{net probability} = \hat{Q}_i' = \frac{d_i}{l - 0.5d_j} \geqslant \frac{d_i}{l} = \hat{q}_i = \text{crude probability}. \tag{9.51}$$

For most applications of competing risk calculations the crude probability and the net probability differ by very little. Expression (9.51) indicates why. For $\hat{Q}_i'$ and $\hat{q}_i$ to differ substantially, the competing cause of death must be a fairly large proportion of the individuals at risk (d_j has to be large relative to l), which is not usually the case for human mortality data.

Table 9-17. Competing risks: Exponential versus intuitive methods

	q_1	$q_2=0.05$	0.10	0.15	0.20
Exponential	0.05	0.0513	0.1027	0.1541	0.2056
Intuitive	0.05	0.0513	0.1026	0.1538	0.2051
Exponential	0.10	0.0527	0.1056	0.1585	0.2116
Intuitive	0.10	0.0526	0.1053	0.1579	0.2105
Exponential	0.15	0.0543	0.1087	0.1633	0.2182
Intuitive	0.15	0.0540	0.1081	0.1622	0.2162
Exponential	0.20	0.0559	0.1112	0.1686	0.2254
Intuitive	0.20	0.0556	0.1111	0.1667	0.2222

Although the exponential and intuitive estimates come from different considerations, they differ little in value ($\hat{Q}_i \approx \hat{Q}_i'$) for most situations. Table 9.17 illustrates the similarity of the two expressions. If $q < 0.1$, then $Q_i - Q_i' < 0.001$, showing why Q_i and Q_i' are essentially equal when applied to questions concerning competing risks among human diseases. The net probability of death from a specific cause, if other causes of death act independently, can also be estimated by considering other causes as censored survival times. The topic of censored data is developed in the next two chapters. It should simply be noted that many of the methods applicable to censored data can be applied in the context of competing risks.

Applications

The estimation of the net probabilities (exponential and intuitive) are illustrated by a subset of data from a large study of the effects of smoking on coronary heart disease mortality (Hammond and Horn [Ref. 6] and reported in [Ref. 5]). A small part of these smoking and CHD data are given in Table 9.18.

As expected, the net probabilities of death from CHD for smokers and nonsmokers increase, but moderately, when competing causes of death are "removed." The increase in net risk for CHD among smokers and nonsmokers can be expressed as a difference or as a ratio (Table 9.18), providing an estimate of the "pure" impact of smoking on CHD risk. Some controversy exists over which is the "best" expression for the increased risk from smoking. The issues surrounding the choice of a ratio versus a difference as an expression of risk are basically semantic and are discussed elsewhere (see [Ref. 5 or 7]).

in the time interval being considered. The estimated net probability of death from cause i is, then,

$$\hat{Q}_i = 1 - \left(1 - \frac{d}{l}\right)^{d_i/d}, \tag{9.48}$$

where l individuals are at risk from both causes of death at the beginning of the interval.

The assumption that the net probabilities are a simple exponential function may not be appealing in some situations [expression (9.44)]. An alternative estimate of the net probability can be derived from intuitive considerations that do not involve an exponential risk model. Individuals at risk can be classified into three categories: (1) died of cause 1, (2) died of cause 2, or (3) lived through the interval. A death from cause 2 can be considered as a person "lost to follow-up" with respect to calculations for cause 1. When cause 2 is "removed," deaths from cause 1 are undercounted since the former "lost to follow-up" are then at risk. That is, the direct estimate of the net probability is too small since a proportion of the individuals who would have died of cause 2 and are "lost" can now die of cause 1. Those who would have died of cause 2 are exposed to risk, on the average, for half the interval so that $0.5d_2$ represents the additional number of individuals at risk when cause 2 is "removed." The value $0.5d_2Q_1$ estimates the number of deaths from cause 1 among the individuals who would have died from cause 2 if it were present. Therefore, "correcting" the number of deaths d_1 gives

$$\hat{Q}'_1 = \frac{d_1 + 0.5d_2\hat{Q}'_1}{l} \tag{9.49}$$

and solving for the net probability Q'_1 yields

$$\hat{Q}'_1 = \frac{d_1}{l - 0.5d_2}. \tag{9.50}$$

The probability $\hat{Q}'_1$ is another estimate of the net probability of death from cause 1 among l individuals at risk. The net probability $\hat{Q}'_1$ is greater than crude probability q_1 since additional individuals are at risk and die of cause 1 when cause 2 is "removed." In general,

$$\text{net probability} = \hat{Q}'_i = \frac{d_i}{l - 0.5d_j} \geqslant \frac{d_i}{l} = \hat{q}_i = \text{crude probability}. \tag{9.51}$$

For most applications of competing risk calculations the crude probability and the net probability differ by very little. Expression (9.51) indicates why. For $\hat{Q}'_i$ and $\hat{q}_i$ to differ substantially, the competing cause of death must be a fairly large proportion of the individuals at risk (d_j has to be large relative to l), which is not usually the case for human mortality data.

Table 9–17. Competing risks: Exponential versus intuitive methods

	q_1	$q_2=0.05$	0.10	0.15	0.20
Exponential	0.05	0.0513	0.1027	0.1541	0.2056
Intuitive	0.05	0.0513	0.1026	0.1538	0.2051
Exponential	0.10	0.0527	0.1056	0.1585	0.2116
Intuitive	0.10	0.0526	0.1053	0.1579	0.2105
Exponential	0.15	0.0543	0.1087	0.1633	0.2182
Intuitive	0.15	0.0540	0.1081	0.1622	0.2162
Exponential	0.20	0.0559	0.1112	0.1686	0.2254
Intuitive	0.20	0.0556	0.1111	0.1667	0.2222

Although the exponential and intuitive estimates come from different considerations, they differ little in value ($\hat{Q}_i \approx \hat{Q}'_i$) for most situations. Table 9.17 illustrates the similarity of the two expressions. If $q < 0.1$, then $Q_i - Q'_i < 0.001$, showing why Q_i and Q'_i are essentially equal when applied to questions concerning competing risks among human diseases. The net probability of death from a specific cause, if other causes of death act independently, can also be estimated by considering other causes as censored survival times. The topic of censored data is developed in the next two chapters. It should simply be noted that many of the methods applicable to censored data can be applied in the context of competing risks.

Applications

The estimation of the net probabilities (exponential and intuitive) are illustrated by a subset of data from a large study of the effects of smoking on coronary heart disease mortality (Hammond and Horn [Ref. 6] and reported in [Ref. 5]). A small part of these smoking and CHD data are given in Table 9.18.

As expected, the net probabilities of death from CHD for smokers and nonsmokers increase, but moderately, when competing causes of death are "removed." The increase in net risk for CHD among smokers and nonsmokers can be expressed as a difference or as a ratio (Table 9.18), providing an estimate of the "pure" impact of smoking on CHD risk. Some controversy exists over which is the "best" expression for the increased risk from smoking. The issues surrounding the choice of a ratio versus a difference as an expression of risk are basically semantic and are discussed elsewhere (see [Ref. 5 or 7]).

Table 9–18. Competing risks: Deaths after 44 months of follow-up for ages 60–65

	Nonsmokers	Smokers
CHD = d_1	552	921
Other = d_2	714	1,095
Population	20,278	21,594
Crude	0.0272	0.0427
Exponential	0.0277	0.0438
Intuitive	0.0277	0.0438

Difference 0.0155 (crude); 0.0161 (net)
Ratio 1.567 (crude); 1.579 (net).

Occasionally the argument is put forth that cancer increases in the last three or four decades, at least in part, are due to the decrease in mortality from infectious diseases. This thought is based on the idea that deaths from infectious diseases operate early in life, thereby eliminating a proportion of individuals who would die of cancer later in life. Data for the years 1900 to 1950 that reflect on this question are given in Table 9.19.

Using competing risk estimates, the net probabilities show no reason to believe that the decreasing mortality from infectious disease plays a role in the observed increase in cancer mortality. Comparison of the crude and net probabilities (multiplied by 100,000) for cancer deaths shows essentially identical values for all six decades. That is, under the conditions for a competing risk calculation, "removing" infectious disease as a competing cause of death does not change the national mortality pattern of cancer deaths over the years 1900 to 1950.

The expression for net probabilities can be used when specific causes of death are available and the results summarized with life-table

Table 9–19. Competing risks: Total cancer and infectious disease deaths by year for the U.S.

Year	1900	1910	1920	1930	1940	1950
Infection	240,077	225,565	191,958	137,971	90,239	60,370
Cancer	48,700	70,414	88,793	119,985	158,943	208,109
Total deaths	1,308,056	1,356,535	1,382,887	1,394,611	1,422,161	1,472,842
Population	76,094	92,407	106,466	123,188	132,122	151,683
Crude	64.00	76.20	83.40	97.40	120.30	137.20
Intuitive*	64.10	76.29	83.48	97.45	120.34	137.23

Note: the crude cancer mortality rate is (d_{cancer}/population) × 100,000, and population is given in thousands.
*Net probabilities multiplied by 100,000

Table 9–20. Expectation of life for specific competing causes of death "eliminated," California, 1980

Age	No causes*	CVD*	IHD*	Lung cancer*	Motor*
0	70.92	80.63	73.79	71.80	71.81
20	52.41	62.61	55.33	53.31	53.19
40	34.49	44.71	37.49	35.41	34.68
60	18.16	28.01	20.08	18.96	18.22
80	7.07	16.56	8.07	7.18	7.07

*Cause of death eliminated (cause j).

functions. The exponential-based expression for a net probability of death from cause i at age x using life-table deaths is

$$Q_{x,i} = 1 - (1 - q_x)^{d_x^{(i)}/d_x}, \tag{9.52}$$

where $d_x^{(i)}$ represents life-table deaths from ith cause in the interval x to $x + 1$ and $d_x = d_x^{(i)} + d_x^{(j)}$ represents the total life-table deaths. The net probabilities $Q_{x,i}$ reflect the impact of mortality at age x from cause i with the cause j "removed" and can be used to calculate other life-table functions, particularly the expectation of life. For example, if all deaths from cardiovascular disease (CVD deaths = cause j) are "eliminated" and a life table based on the remaining causes of death (all non-CVD deaths = cause i) is computed, then an estimate of the years of life lost attributable to cardiovascular disease is found by comparing the "net" expectation of life with the expectation calculated when all causes of death are operating. That is, the life-table functions are based on the net probabilities $Q_{x,i}$ rather than the crude probabilities q_x. Table 9.20 gives the expectation of life for 1980 California males for five selected ages. Also included are the expectations of life when three other causes of death (ischemic heart disease, lung cancer, and motor vehicle accidents) are each "removed." The impact of cardiovascular disease on the total mortality picture is clear. The life-table competing-risk calculations indicate that the expectation of life would be increased about 10 years if cardiovascular disease was "removed" as a risk of death and a 1–4-year increase would result if ischemic heart disease was "removed." Almost no impact on the expectation of life is observed when lung cancer or motor vehicle accidents are "removed" as causes of death.