WORLD HEALTH ORGANIZATION

LIFE TABLE AND MORTALITY ANALYSIS

CHIN LONG CHIANG

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 5 (continued)

CHAPTER 7 (continued)

CHAPTER 8 - THE LIFE TABLE WHEN A PARTICULAR CAUSE IS ELIMINATED

CHAPTER 8 (continued)

CHAPTER 9 - MEDICAL FOLLOW-UP STUDIES

CHAPTER 9 (continued)

APPENDIX V (continued)

LIFE TABLE AND MORTALITY ANALYSIS

FOREWORD

This publication on advanced methods of analysis of the mortality of populations is the second in a series of teaching aids addressed to a wide spectrum of professionals in the health field, statistics and demography. The first, a Manual on methods of analysis of national mortality statistics for public health purposes, was published in 1977 and focussed on basic methods of analysis which are commonly used in National Statistical Departments.

Public Health and Demography owe so much to the quantitative study of mortality. For centuries, the primary determinant of population trends has been mortality and it still remains so in many less developed countries; it was mortality that formed the primary challenge to the medical professions; it was the prevention of early death that was the primary objective of public health workers and of social legislation. Nowadays, this central role of the study of mortality has gradually yielded way to concern for other phenomena such as fertility and morbidity and the definition of positive health and the study of the provision and use of health services. Nonetheless, the analysis of mortality data is still an indispensable part of informed decision-making and of the evaluation of policies on health services. New problems have arisen even in the area of mortality analysis; the growing importance of chronic diseases have raised new issues and problems; demands for statistical analysis have become ever more sophisticated; the improved quality of certification of the causes of death has created a demand for a detailed study of the difficulties encountered in their interpretation; the use of computers has changed the problems of data processing and facilitated more complex methods of analysis. It was with these considerations in mind that the work on an up-to-date publication on mortality analysis was initiated.

This volume emphasizes the more advanced methods in the study of survival and mortality. The life table method of analysis, historically rooted in the actuarial and demographic sciences has by now become an indispensable tool for investigators in other disciplines such as epidemiology, zoology, manufacturing etc. The classical concept of counting risks is introduced and integrated into a coherent probabilistic approach to the study of a broad range of processes with a stochastic distribution of exit from one or more competing causes with the life table as central theme. Follow-up studies with due attention to truncated information are of great practical importance not only for medical research but may prove particularly useful for health statisticians in less developed countries who - in the absence of complete nation-wide vital statistics - concentrate on the study of the survival experience of relatively small population groups.

It is hoped that this volume will be of use for post-graduate courses in biometry, demography and epidemiology, and together with the manual will also serve as a background for training activities and refresher courses in health statistics organized or sponsored by the World Health Organization. In fact, part of the manuscript has been tested in courses organized by the World Health Organization with the financial support of UNFPA; the experience gained in this practical application is reflected in the text.

This volume has been prepared by Professor Chin Long Chiang, University of California, Berkeley (U.S.A.) an outstanding authority and pioneer in the application of the stochastic approach to the study of death processes. The manuscript has also profited from the comments of the United Nations' Population Division, Professors H. Campbell (U.K.) and S. Koller (Federal Republic of Germany) and various staff members of the World Health Organization such as the statistical officers in the Regional Offices. Dr H. Hansluwka, World Health Organization was most actively involved in the design of this volume and coordinated the various activities which led to the production of this volume.

ACKNOWLEDGEMENTS

The Support of the United Nations Fund for Population Activities is gratefully acknowledged.

This manual was made possible with the help of the following persons: Flora Fung, Bonnie Hutchings, Linda Kwok, Carol Langhauser, Patrick Wong, and Rodney Wong. Their assistance is very much appreciated.

CHAPTER 1

ELEMENTS OF PROBABILITY

1. Introduction

A good understanding of the basic concept of probability is essential for proper analysis of mortality data. Because of its potential as an analytic tool, probability has been increasingly used in vital statistics and life table analyses. As a result, studies of vital data are no longer limited to a mere description or interpretation of numerical values; statistical inference can be made regarding mortality and survival patterns of an entire population. While it is a mathematical concept, probability has an interesting intuitive appeal. Many natural phenomena can be described by means of probability laws; occurrence of daily events also seems to follow a definite pattern. Even such spontaneous events as accidents can be predicted in advance with a certain degree of accuracy. Mortality laws proposed by Benjamin Gompertz in 1825 and by W. M. Makeham in 1860 have been used in studies of human survival and death both in the field of health and in the actuarial sciences. It is appropriate then to begin this manual by introducing the fundamental probability concept, related formulas and illustrative examples.

2. Elements of Probability

2.1. Components. The concept of probability involves three components: (a) a random experiment, (b) possible outcomes, and (c) an event of interest. A random experiment is an experiment that has a number of possible outcomes, but it is not certain which of the outcomes will occur before the experiment is performed. Thus, in speaking of probability, one must have in mind a random experiment under consideration and an event of interest.

2.2. Definition of probability. The probability of the occurrence of event A is defined as the ratio of the number of outcomes where event A occurs to the total number of outcomes. For simplicity, we shall use the term "the probability of event A" for "the probability of the occurrence of event A."

Suppose that a random experiment may result in a number n of possible (and equally likely) outcomes, and in n(A) of these outcomes event A occurs. Then the probability of event A is defined as follows:

$$\Pr\{A\} = \frac{n(A)}{n} \quad . \tag{2.1}$$

Thus, the probability of event A in a random experiment is a measure of the likelihood of occurrence of the event.

2.3. Examples. The following examples may elucidate the concept of probability.

Example 1. In tossing a fair coin once, what is the probability of a head turning up? Here, tossing a fair coin once is the random experiment, and the possible outcomes are a head and a tail. Let event A be "a head." The number of possible outcomes, n, is 2, and the number of outcomes where a head occurs, n(A), is 1. Therefore, the probability is

$$\Pr\{A\} = \frac{n(A)}{n} = \frac{1}{2} \quad .$$

Example 2. In rolling a fair die once, there are 6 possible outcomes. Let event A be 3 dots. Here n=6 and n(A) = 1; therefore:

$$\Pr\{A\} = \frac{n(A)}{n} = \frac{1}{6} \quad .$$

Let event B be an even number of dots, with n(B) = 3. The probability of B is:

$$\Pr\{B\} = \frac{n(B)}{n} = \frac{3}{6} = \frac{1}{2}$$

Example 3. A name is drawn at random from a group of 120 people consisting of 39 females and 81 males. Let event A be the drawing of a female name. The probability of event A is:

$$\Pr\{A\} = \frac{n(A)}{n} = \frac{39}{120} = \frac{13}{40} \quad .$$

Example 4. A list of n=100 names consists of n(s) = 98 names of survivors and n(d) = 2 of those who have died. A name is drawn at random from the list. The probability that the name drawn will be that of a survivor is

$$P(s) = \frac{n(s)}{n} = \frac{98}{100} = .98$$

and that of one who has died is

$$P(d) = \frac{n(d)}{n} = \frac{2}{100} = .02 \quad .$$

Clearly, the sum of the two probabilities is unity:

$$P(s) + P(d) = .98 + .02 = 1 \quad .$$

2.4. Values of a probability. From the definition we see that the probability of an event A is an (idealized) proportion or relative frequency. Thus, a probability can only take on values between zero and one, i.e.,

$$0 \leq \Pr\{A\} \leq 1 \quad . \tag{2.2}$$

2.5. Sure event and impossible event. A sure event is an event that always occurs. If I is a sure event, then

$$\Pr\{I\} = 1 \quad . \tag{2.3}$$

An impossible event is an event that never occurs. If $\emptyset$ is an impossible event, then

$$\Pr\{\emptyset\} = 0 \quad . \tag{2.4}$$

2.6. <u>Complement of an event</u> (or negation of an event) can be best illustrated with examples. Let $\bar{A}$ be the complement of event A.

Example	A	$\bar{A}$
Sex of a baby	male	female
Toss of a coin	head	tail
Toss of a die	3 dots	anything but 3 dots
Toss of a die	even no. of dots	odd number of dots
Survival analysis	survival	death

Thus, the complement $\bar{A}$ occurs when and only when event A does not occur. In a random experiment the total number of outcomes can be divided into two groups according to the occurrence of A or of $\bar{A}$,

$$n = n(A) + n(\bar{A}) \quad .$$

The probability of $\bar{A}$ in a random experiment is, by definition,

$$\Pr\{\bar{A}\} = \frac{n(\bar{A})}{n} \quad .$$

It is clear then that, whatever event A may be,

$$\Pr\{A\} + \Pr\{\bar{A}\} = 1 \tag{2.5}$$

or

$$\Pr\{\bar{A}\} = 1 - \Pr\{A\} \quad . \tag{2.5a}$$

In words, the probability of the complement of A is equal to the complement of the probability of A.

2.7. Composite event (A and B). Given two events A and B, we define a composite event A and B (or AB for simplicity) by saying that the event AB occurs if both event A and event B occur.

Example 5. Consider a group of 200 newborn babies divided according to sex and prematurity as shown in the following 2x2 table:

	Male A	Female $\bar{A}$	Marginal row total
Premature B	11 $n(AB)$	9 $n(\bar{A}B)$	20 $n(B)$
Full term $\bar{B}$	93 $n(A\bar{B})$	87 $n(\bar{A}\bar{B})$	180 $n(\bar{B})$
Marginal Column Total	104 $n(A)$	96 $n(\bar{A})$	200 n

Let A = male, $\bar{A}$ = female, B = premature, $\bar{B}$ = full term.

A baby is picked at random from the group; the composite event AB is a premature boy. The corresponding probability is

$$\Pr\{AB\} = \frac{n(AB)}{n} = \frac{11}{200} \quad . \tag{2.6}$$

Other possible composite events are

$A\bar{B}$ = a full term boy

$\bar{A}B$ = a premature girl

$\bar{A}\bar{B}$ = a full term girl

The probabilities $\Pr\{A\bar{B}\}$, $\Pr\{\bar{A}B\}$, and $\Pr\{\bar{A}\bar{B}\}$ can be computed from the above table.

If I is a sure event, then

$$\Pr\{AI\} = \Pr\{A\} \quad . \tag{2.7}$$

If $\emptyset$ is an impossible event, then

$$\Pr\{A\emptyset\} = 0 \quad . \tag{2.8}$$

2.8. Conditional probability. The conditional probability of B given that A has occurred is defined by:

$$\Pr\{B|A\} = \frac{\Pr\{AB\}}{\Pr\{A\}} \quad . \tag{2.9}$$

Since

$$\Pr\{AB\} = \frac{n(AB)}{n} \qquad \text{and} \qquad \Pr\{A\} = \frac{n(A)}{n} \quad ,$$

we have

$$\Pr\{B|A\} = \frac{n(AB)/n}{n(A)/n} = \frac{n(AB)}{n(A)} \quad . \tag{2.10}$$

In terms of the previous example, $\Pr\{B|A\}$ is the probability that a baby chosen at random from the boys will be premature. Since there are $n(A) = 104$ boys, and among them $n(AB) = 11$ are premature, we have

$$\Pr\{B|A\} = \frac{n(AB)}{n(A)} = \frac{11}{104} \; ,$$

or, using

$$\Pr\{AB\} = \frac{n(AB)}{n} = \frac{11}{200} \quad \text{and} \quad \Pr\{A\} = \frac{n(A)}{n} = \frac{104}{200} \quad ,$$

and

$$\Pr\{B|A\} = \frac{\Pr\{AB\}}{\Pr\{A\}} = \frac{11/200}{104/200} = \frac{11}{104},$$

we obtain the same value.

It is clear that the conditional probability $\Pr\{B|A\}$ is different from the conditional probability $\Pr\{A|B\}$. In the above example the probability that a premature baby will be a boy is computed from

$$\Pr\{A|B\} = \frac{n(AB)}{n(B)} = \frac{11}{20} \quad .$$

The reader is advised to use the above example to compute and interpret the following conditional probabilities: $\Pr\{B|\bar{A}\}$, $\Pr\{\bar{B}|A\}$, $\Pr\{\bar{B}|\bar{A}\}$, $\Pr\{A|\bar{B}\}$, $\Pr\{\bar{A}|B\}$, and $\Pr\{\bar{A}|\bar{B}\}$.

In applying conditional probability to a practical problem, one should beware of a sequence that may exist in the occurrence of events. If event A occurs before event B, then the conditional probability $\Pr\{A|B\}$ may not be meaningful, whereas the conditional probability $\Pr\{B|A\}$ is meaningful. For example, in a study of sex differential infant mortality, sex of infant, male (A) or female ($\bar{A}$), is determined before mortality in the first year of life (denoted by B) occurs. Comparison of infant mortality of males with that of females requires the conditional probabilities $\Pr\{B|A\}$ and $\Pr\{B|\bar{A}\}$. But it may be difficult to comprehend the conditional probability $\Pr\{A|B\}$ that an infant who dies will be male.

2.9. Independence. Event B is said to be independent of event A if the conditional probability of B given A is equal to the (absolute) probability of B. In formula

$$\Pr\{B|A\} = \Pr\{B\} \quad . \tag{2.11}$$

This means that the likelihood of the occurrence of B is not influenced be the occurrence of A. Clearly, if B is independent of A, B is also independent of $\bar{A}$, or

$$\Pr\{B|A\} = \Pr\{B\} = \Pr\{B|\bar{A}\} \quad . \tag{2.12}$$

Let A = male, B = prematurity. If

$$\Pr\{\text{premature baby}|\text{male}\} = \Pr\{\text{premature baby}\},$$

then

$$\Pr\{\text{premature baby}|\text{female}\} = \Pr\{\text{premature baby}\},$$

and we say that prematurity is independent of sex of the baby.

To verify whether an event B is independent of an event A in a particular problem, we compute separately

$$\Pr\{B|A\} \qquad \text{and} \qquad \Pr\{B\} \quad .$$

If the two numerical values are equal, we say that B is independent of A.

In the example in section 2.7

$$\Pr\{B|A\} = \frac{11}{104} \qquad \text{and} \qquad \Pr\{B\} = \frac{20}{200} \quad .$$

Since 11/104 is not equal to 20/200, according to the information given in this example, prematurity is dependent on the sex of a baby.

2.10. <u>Multiplication theorem</u>. The probability of AB is equal to the product of probability of A and the conditional probability of B given A, or

$$\Pr\{AB\} = \Pr\{A\} \times \Pr\{B|A\} \quad . \tag{2.13}$$

<u>Proof</u>:

$$\Pr\{AB\} = \frac{n(AB)}{n} = \frac{n(A)}{n} \times \frac{n(AB)}{n(A)} = \Pr\{A\} \times \Pr\{B|A\} \quad .$$

With reference to the 2x2 table in example 5, we see that

$$\Pr\{AB\} = \frac{11}{200}$$

and

$$\Pr\{A\} \times \Pr\{B|A\} = \frac{104}{200} \times \frac{11}{104} = \frac{11}{200} \quad ;$$

therefore

$$\Pr\{AB\} = \Pr\{A\} \times \Pr\{B|A\} \quad .$$

Since event AB is the same as event BA, the multiplication theorem has an alternative formula:

$$\Pr\{AB\} = \Pr\{B\} \times \Pr\{A|B\} \quad . \tag{2.14}$$

The formulas of the multiplication theorem for three and four events are

$$\Pr\{ABC\} = \Pr\{A\} \times \Pr\{B|A\} \times \Pr\{C|AB\} \tag{2.15}$$

and

$$\Pr\{ABCD\} = \Pr\{A\} \times \Pr\{B|A\} \times \Pr\{C|AB\} \times \Pr\{D|ABC\} \quad . \tag{2.16}$$

2.11. <u>Multiplication theorem (continuation)</u>. If events are independent, then the formulas of the multiplication theorem become

$$\Pr\{AB\} = \Pr\{A\} \times \Pr\{B\} \quad , \tag{2.17}$$

$$\Pr\{ABC\} = \Pr\{A\} \times \Pr\{B\} \times \Pr\{C\} \quad , \tag{2.18}$$

$$\Pr\{ABCD\} = \Pr\{A\} \times \Pr\{B\} \times \Pr\{C\} \times \Pr\{D\} \quad . \tag{2.19}$$

2.12. <u>A theorem of (pairwise) independence</u>. If B is independent of A, then A is independent of B, and A and B are said to be independent events. Symbolically, the theorem may be stated as follows:

If $Pr\{B|A\} = Pr\{B\}$

then

$$Pr\{A|B\} = Pr\{A\} \quad .$$

Proof: According to the multiplication theorem,

$$Pr\{AB\} = Pr\{A\} \times Pr\{B|A\} \text{ and } Pr\{AB\} = Pr\{B\} \times Pr\{A|B\} \quad .$$

It follows that

$$Pr\{A\} \times Pr\{B|A\} = Pr\{B\} \times Pr\{A|B\} \quad . \tag{2.20}$$

If B is independent of A so that $Pr\{B|A\} = Pr\{B\}$, then (2.20) becomes

$$Pr\{A\} \times Pr\{B\} = Pr\{B\} \times Pr\{A|B\} \quad ,$$

and consequently

$$Pr\{A\} = Pr\{A|B\} \quad .$$

Conversely, if B is dependent of A, then A is dependent of B.

In the example in part 7,

$$Pr\{B|A\} = \frac{11}{104} \quad \text{and} \quad Pr\{B\} = \frac{20}{200}$$

so that B is dependent of A, while

$$Pr\{A|B\} = \frac{11}{20} \quad \text{and} \quad Pr\{A\} = \frac{104}{200} \quad ,$$

so that A is dependent of B.

2.13. Composite event (A or B). By a composite event A or B we mean either A or B or both. Thus the event A or B occurs if either A occurs, or B occurs, or AB occurs.

2.14. Mutual exclusiveness. Two events are said to be mutually exclusive if the occurrence of one implies the non-occurrence of the other; in other words, they cannot occur simultaneously in a single experiment. If A and B are mutually exclusive events, then $n(AB) = 0$ and $\Pr\{AB\} = 0$.

2.15. Addition theorem.

$$\Pr\{A \text{ or } B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{AB\} \quad .$$

Proof: Using the example in part 2.7 again and by direct enumeration, we see that

$$\Pr\{A \text{ or } B\} = \frac{n(A) + n(B) - n(AB)}{n} \quad .$$

Dividing every term in the numerator by the denominator, we have

$$\Pr\{A \text{ or } B\} = \frac{n(A)}{n} + \frac{n(B)}{n} - \frac{n(AB)}{n}$$

$$= \Pr\{A\} + \Pr\{B\} - \Pr\{AB\} \quad . \tag{2.21}$$

Example: Let A = male, B = prematurity. From example 5 in section 2.7, we compute

$$\Pr\{A \text{ or } B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{B\}$$

$$= \frac{104}{200} + \frac{20}{200} - \frac{11}{200} = \frac{113}{200}$$

The formulas of the addition theorem for three and four events are

$$\Pr\{A \text{ or } B \text{ or } C\} = \Pr\{A\} + \Pr\{B\} + \Pr\{C\}$$

$$- \Pr\{AB\} - \Pr\{BC\} - \Pr\{CA\} + \Pr\{ABC\} \tag{2.22}$$

$$\Pr\{A \text{ or } B \text{ or } C \text{ or } D\} = \Pr\{A\} + \Pr\{B\} + \Pr\{C\} + \Pr\{D\}$$

$$- \Pr\{AB\} - \Pr\{AC\} - \Pr\{AD\} - \Pr\{BC\} \quad \Pr\{BD\} - \Pr\{CD\}$$

$$+ \Pr\{ABC\} + \Pr\{ABD\} + \Pr\{ACD\} + \Pr\{BCD\} - \Pr\{ABCD\}$$

(2.23)

2.16. Addition theorem (continuation). When events are mutually exclusive so that $Pr\{AB=0\}$, etc., then the formulas of the addition theorem become

$$Pr\{A \text{ or } B\} = Pr\{A\} + Pr\{B\} \quad , \tag{2.24}$$

$$Pr\{A \text{ or } B \text{ or } C\} = Pr\{A\} + Pr\{B\} + Pr\{C\} \tag{2.25}$$

$$Pr\{A \text{ or } B \text{ or } C \text{ or } D\} = Pr\{A\} + Pr\{B\} + Pr\{C\} + Pr\{D\} \tag{2.26}$$

and so on.

2.17. Summary of the addition and multiplication theorems. Simple as they may appear to be, the addition and multiplication theorems are indispensible in computing probabilities. The following table is prepared to facilitate the applications of these two theorems.

Which theorem	Multiplication theorem	Addition theorem
When to use	A and B	A or B
Theorem	$Pr\{AB\} = Pr\{A\} \times Pr\{B\|A\}$	$Pr\{A \text{ or } B\} = Pr\{A\} + Pr\{B\} - Pr\{AB\}$
Are the events	independent?	mutually exclusive?
Particular form of theorem	If independent, then $Pr\{AB\} = Pr\{A\} \times Pr\{B\}$	If mutually exclusive, then $Pr\{A \text{ or } B\} = Pr\{A\} + Pr\{B\}$

2.18. The distributive law. When the computation of a probability requires both the addition and multiplication theorems, the rule of application of the two theorems is similar to that in an arithmetic problem. The most useful rule of operation is the distributive law:

$$2(3 + 4) = 2 \times 3 + 2 \times 4$$

in an arithmetic problem, and

$$Pr\{A(B \text{ or } C)\} = Pr\{AB \text{ or } AC\} \tag{2.27}$$

in probability; or

$$(2 + 3)(4 + 5) = 2 \times 4 + 2 \times 5 + 3 \times 4 + 3 \times 5$$

and

$$\Pr\{(A \text{ or } B)(C \text{ or } D)\} = \Pr\{AC \text{ or } AD \text{ or } BC \text{ or } BD\} \quad . \tag{2.28}$$

Using example 5 once again, we have

$$\Pr\{A(B \text{ or } \overline{B})\} = \Pr\{AB \text{ or } A\overline{B}\}$$

$$= \Pr\{AB\} + \Pr\{A\overline{B}\}$$

$$= \frac{11}{200} + \frac{93}{200} = \frac{104}{200} \quad .$$

In this case, $(B \text{ or } \overline{B}) = I$ is a sure event,

$$\Pr\{A(B \text{ or } \overline{B})\} = \Pr\{AI\} = \Pr\{A\} = \frac{104}{200} \quad .$$

2.19. <u>An example from the life table</u>

Table 1. The number of survivors and the number died out of 100,000 live births

Age Interval (in years)	Number living at age x_i	Number dying in interval (x_i, x_{i+1})
x_i to x_{i+1}	ℓ_i	d_i
(1)	(2)	(3)
0– 1	100000	1801
1– 5	98199	316
5–10	97883	184
10–15	97699	183
15–20	97516	550
20–25	96966	750
25–30	96216	681
30–35	95535	766
35–40	94769	1060
40–45	93709	1[illegible]3
45–50	92126	24[illegible]
50–55	89672	3631
55–60	86041	5341
60–65	80700	7171
65–70	73529	9480
60–75	64049	11562
75–80	52487	14192
80–85	38295	14752
85+	23543	23543

Example 6. Table 1 is a part of a life table for the 1970 California, USA, population. Column (1) shows the age intervals in years. Column (2) is the number of (life table) people living at the beginning of each age interval. Thus, the column shows that there are 100,000 (life table) people alive at the exact age 0 (that is, the population size at birth); of these 98,199 survive to the exact age of 1 year (the first birthday), 97,883 survived to the exact age of 5 years, etc., and finally 23,543 survived to the exact age of 85 years. Each figure in column (3) is the number of people dying within the corresponding age interval. Among the 100,000 living at age 0, 1801 died during the age interval (0,1), 316 died between ages 1 and 5, etc., and 23543 died beyond age 85 years.

For the purpose of illustration, we consider 100,000 newborns who are subject to the mortality experience of the 1970 California population. What is the probability that a newborn will survive to his first birthday? In this example, the "random experiment" is the baby's first year of life; possible outcomes are survival or death of the 100,000 infants; the event A of interest is a newborn's survival to his first birthday. Since 98,199 of the 100,000 newborns (the possible number of survivors) actually survived (event A occurred), the probability that a newborn will survive to his first birthday is

$$\frac{n(A)}{n} = \frac{98,199}{100,000} = .98199 \quad \text{or } 981.99 \text{ per } 1,000 \quad .$$

Similarly, the probability that a newborn will survive to the fifth birthday is 97883/100,000 = .97883, to the 10th birthday is 97699/100,000 = .97699.

For the probability of death, we use the corresponding number of deaths in the numerator of the formula. Thus we have

$$\Pr\{\text{a newborn will die in the first year of life}\} = \frac{1801}{100{,}000}$$

$$= .01801 \text{ or } 18.01 \text{ per } 1{,}000$$

and

$$\Pr\{\text{a newborn will die in interval } (1,5)\} = \frac{316}{100{,}000}$$

$$= .00316 \text{ or } 3.16 \text{ per } 1{,}000 \quad .$$

2.19.1. Conditional Probability. The probabilities computed above are absolute probabilities based on the 100,000 live births. When the base population is changed, we have conditional probabilities:

$$\Pr\{\text{a child alive at age 1 will die in interval } (1,5)\}$$

$$= \Pr\{\text{a child will die in interval } (1,5) \mid \text{he is alive at age } 1\}$$

$$= \frac{\text{number dying in } (1,5)}{\text{number living at age } 1} = \frac{316}{98199} = .00322 \quad \text{or } 3.22 \text{ per } 1{,}000 \ ,$$

and

$$\Pr\{\text{a child alive at age 5 will die in interval } (5,10)\}$$

$$= \frac{\text{number dying in } (5,10)}{\text{number living at age } 5} = \frac{184}{97883} = .00188 \quad \text{or } 18.8 \text{ per } 1{,}000 \ .$$

These conditional probabilities, which are based on the number of individuals living at the beginning of the corresponding age interval, are known as the age-specific probabilities of dying. Other conditional probabilities are possible, depending upon the given condition and the event of interest. The following are a few examples:

$$\Pr\{\text{an individual of age 25 will survive to age 50}\}$$

$$= \frac{\text{number living at age } 50}{\text{number living at age } 25} = \frac{89672}{96216} = .93199$$

and

Pr{an individual of age 25 will die before age 50}

$$= \frac{\text{number dying between ages 25 and 50}}{\text{number living at age 25}} = \frac{96216-89672}{96216}$$

$$= \frac{6544}{96216} = .06801$$

where the number 6544 can be determined also from the number of deaths in all the intervals from 25 to 50:

$$6544 = 681 + 766 + 1060 + 1583 + 2454 \ .$$

Since an individual alive at age 25 will either survive to age 50 or die before age 50, the corresponding probabilities must add to unity:

$$.93199 + .06801 = 1.00000$$

For an individual alive at age 20, the corresponding probabilities are:

Pr{an individual of age 20 will survive to age 45}

$$= \frac{92126}{96966} = .95009$$

and

Pr{an individual of age 20 will die before age 45} = 1 - .95009 = .04991.

2.19.2. Probabilities of Composite Events. Let A be an event that a male of age 25 survives to age 50 and $\overline{A}$ he dies before age 50; let B be an event that a female of age 20 survives to age 45 and $\overline{B}$ she dies before age 45. If they are subject to the probability of dying shown in the above table and if their survival is independent of one another, then we can use the multiplication theorem to compute the following probabilities:

$\Pr\{$both male and female live for 25 years$\}$

$$= \Pr\{A \text{ and } B\} = \Pr\{A\} \times \Pr\{B\} = .93199 \times .95008 = .88547,$$

$\Pr\{$both die within 25 years$\}$

$$= \Pr\{\bar{A} \text{ and } \bar{B}\} = \Pr\{\bar{A}\} \times \Pr\{\bar{B}\} = .06801 \times .04991 = .00339$$

$\Pr\{$male lives and female dies in 25 years$\}$

$$= \Pr\{A \text{ and } \bar{B}\} = \Pr\{A\} \times \Pr\{\bar{B}\} = .93199 \times .04991 = .04652$$

and

$\Pr\{$male dies and female lives for 25 years$\}$

$$= \Pr\{\bar{A} \text{ and } B\} = \Pr\{\bar{A}\} \times \Pr\{B\} = .06801 \times .95009 = .06462 .$$

Since either both male and female will survive a period of 25 years, or one of them dies, or both die, the sum of the above probabilities is equal to one:

$$.88547 + .00339 + .04652 + .06462 = 1 .$$

The reader may wish to compute similar probabilities for other ages or for a period different from 25 years.

2.19.3. Probability of Dissolution of Marriage. The above probabilities can be used to compute joint life insurance premiums or dissolution of marriages. For example, if a husband is of age 25 and his wife of age 20, the probability that their marriage will be dissolved in 25 years due to death may be computed as follows:

$\Pr\{$dissolution of marriage in 25 years due to death$\}$

$$= \Pr\{\text{one or both of them die in 25 years}\}$$

$$= \Pr\{(A \text{ and } \bar{B}) \text{ or } (\bar{A} \text{ and } B) \text{ or } (\bar{A} \text{ and } \bar{B})\} .$$

Here the three events (A and $\overline{B}$), ($\overline{A}$ and B), and ($\overline{A}$ and $\overline{B}$) are mutually exclusive; we use the addition theorem and the above numerical values to obtain the probability

$$\Pr\{(A \text{ and } \overline{B})\} + \Pr\{(\overline{A} \text{ and } B)\} + \Pr\{(\overline{A} \text{ and } \overline{B})\}$$

$$= .04652 + .06462 + .00339 = .11453 \quad .$$

Thus the probability of dissolution of their marriage is better than 10 percent. On the other hand,

$$\Pr\{\text{their marriage will not be dissolved in 25 years}\}$$

$$= \Pr\{\text{both live for 25 years}\} = \Pr\{A \text{ and } B\} = .88547$$

Obviously, the two probabilities are complementary to each other, and

$$.11453 + .88547 = 1.00000 \ .$$

CHAPTER 2

DEATH RATES AND ADJUSTMENT OF RATES

1. Age Specific Death Rates

For a specific age interval (x_i,x_{i+1}), the death rate, M_i, is defined as follows:

$$M_i = \frac{\text{Number dying in } (x_i,x_{i+1})}{\text{Number of years lived in } (x_i,x_{i+1}) \text{ by those alive at } x_i} \qquad (1.1)$$

Suppose that of ℓ_i people living at exact age x_i, d_i die between age x_i and x_{i+1}, and each of d_i people lives on the average a fraction, a_i, of the interval (x_i,x_{i+1}). Then the death rate M_i defined in (1.1) may be expressed in the formula

$$M_i = \frac{d_i}{n_i(\ell_i-d_i) + a_i n_i d_i} \quad , \qquad (1.2)$$

where $n_i = x_{i+1}-x_i$ is the length of the interval (x_i,x_{i+1}), $n_i(\ell_i-d_i)$ is the number of years lived in (x_i,x_{i+1}) by the (ℓ_i-d_i) survivors, and $a_i n_i d_i$ is the number of years lived by the d_i people who die in the interval. The unit of a death rate is the number of deaths per person-years. The corresponding estimate of probability of dying, given by

$$\hat{q}_i = \frac{d_i}{\ell_i} \quad , \qquad (1.3)$$

is a pure number. From (1.2) and (1.3), we find a relationship between q_i and M_i

$$\hat{q}_i = \frac{n_i M_i}{1+(1-a_i)n_i M_i} \quad . \qquad (1.4)$$

We see then that the age-specific death rate and the probability of dying are two different concepts and they are related by formula (1.4). Consider as an example the age interval (1,5) in the 1970 California life table population. Here $x_1=1$, $x_5 = 5$, and $n_1=5-1=4$. From Section 2, Table 1, we find $\ell_1 = 98199$,

$d_1 = 316$, and from Appendix [V], $a_1 = .41$. The death rate is

$$M_1 = \frac{316}{4(98199-316) + .41 \times 4 \times 316}$$

$$= .000806$$

and the estimate of the probability is

$$\hat{q}_1 = \frac{316}{98199} = .00322 \quad .$$

Formula (1.2) of the age-specific death rate is expressed in terms of a life table framework where ℓ_i people are followed for n_i years to determine the number of deaths (d_i) and the number of survivors ($\ell_i - d_i$) at the end of n_i years. In a current population, such as the 1970 California population, an age specific death rate is computed from the mortality and population data during a calendar year (1970). Instead of d_i defined in a life table, we have D_i, the observed number of deaths occurring to people in the age group (x_i, x_{i+1}) during a calendar year. To derive a formula for the death rate as in (1.2), we let N_i be the (hypothetical) number of people alive at exact age x_i; among them D_i deaths occur. Then we have the death rate

$$M_i = \frac{D_i}{n_i(N_i - D_i) + a_i n_i D_i} \tag{1.2a}$$

and an estimate of the probability q_i,

$$\hat{q}_i = \frac{D_i}{N_i} \quad . \tag{1.3a}$$

They also have the relationship in (1.4).

Since N_i is a hypothetical number, the denominator of (1.2a) and the death rate for a current population cannot be computed from (1.2a). Customarily, the denominator of (1.2a) is estimated by the midyear (calendar year) population P_i for age group (x_i, x_{i+1}), and hence the age-specific death rate is given by

$$M_i = \frac{D_i}{P_i} \quad . \tag{1.5}$$

Although it is a well known and accepted definition of age-specific death rates, formula (1.5) is much more meaningful when P_i is interpreted as an estimate of the denominator in (1.1).

In California 1970, there were $D_1 = 1049$ deaths occurring in age interval (1,5), and $P_1 = 1{,}302{,}198$ people of ages 1 to 5 at midyear. Therefore, the corresponding death rate is

$$M_1 = \frac{D_1}{P_1} = \frac{1049}{1{,}302{,}198} = .000806 \quad .$$

A death rate usually is a small number; its significance is not easily appreciated. To remedy this, the numerical value of a death rate is multiplied by a number, such as 1000, which is called the base. The formula of a death rate often appears as

$$M_i = \frac{D_i}{P_i} \times 1000 \quad . \tag{1.5a}$$

Thus, instead of $M_i = .000806$, we have $M_1 = .806$ per 1000 person-years.[1/] It should be clear that in formula (1.5) and (1.5a) the number of deaths D_i in the numerator and the midyear population P_i in the denominator refer to the same population, such as the 1970 California population between ages 1 and 5. The population and the base must be clearly stated in a death rate. For example, the death rate for the age group 1 to 5 years in the 1970 California population is .806 per 1000.

[1/] The words "person-years" are often deleted.

When a death rate is for an entire life, it is called the crude death rate. In formula:

$$M = \frac{D}{P} \times 1000 \tag{1.6}$$

where

$$D = \sum_i D_i \tag{1.7}$$

is the total number of deaths occurring during a calendar year, and

$$P = \sum_i P_i \tag{1.8}$$

is the total midyear population of a community, or a country, in question.

Death rates may be computed for any specific category of people in a population. Sex-specific death rates, occupation-specific death rates, age-sex-specific death rates, are examples. In each case, the specific rate is defined as the number of deaths occurring to people in the stated category during a calendar year divided by the midyear population of the same category.

Death rates may also be computed for specific causes such as death rates from cancer, tuberculosis, or heart diseases. These are known as cause-specific death rates. Here it is deaths, rather than population, that is divided into categories. A cause-specific death rate is defined as the number of deaths from the specific cause divided by the midyear population. In formula, the death rate from cause R_δ is given by:

$$M_\delta = \frac{D_\delta}{P} \times 100{,}000 \; . \tag{1.9}$$

Here D_δ is the number of deaths from cause R_δ during a calendar year in question, the base is 100,000 because of the small magnitude of the rate.

Prevalence of diseases varies with age. Cardiovascular disease, for example, is more prevalent among the aged than among young people; the converse is true for infectious diseases. Therefore, age-cause-specific death rates are in common use. For the age interval (x_i, x_{i+1}) and cause R_δ, the specific death rate, $M_{i\delta}$, is computed from

$$M_{i\delta} = \frac{D_{i\delta}}{P_i} \times 100{,}000 \quad , \tag{1.10}$$

where $D_{i\delta}$ is the number of deaths from cause R_δ occurring to people in age group (x_i, x_{i+1}) during a calendar year, and P_i is the midyear population of the same age group. Here a base, 100,000, is used.

2. Infant Mortality

In the human population, mortality is the highest among newborns and among the elderly. Infant mortality also has a great impact on the population distribution in later years of life. Various efforts have been made in different countries to reduce infant deaths, and many of these efforts have resulted in a considerable amount of success. Mortality in the first year of life has been decreasing, especially in the developed countries prior to 1950. Since many different causes affect mortality from conception to the end of the first year of life, this period of human life has been divided into subintervals and designated by special names, as shown in the following table.

Table 1. Fetal death and infant mortality

Designation	Interval
Early fetal death	Under 20 weeks of gestation
Intermediate fetal death	20–27 weeks of gestation
Late fetal death	28 or more weeks of gestation
Neonatal death	Under 28 days of age
Post neonatal death	28 days to end of first year of life
Infant death	Under one year of age

The corresponding definitions of death rates differ somewhat from the definition of the age-specific death rate discussed in the preceding section. The following rates are measures of mortality for a defined population during a given calendar year:

2.1. Fetal death rate (alias "stillbirth rate"). Two definitions are available:

$$\frac{\text{Number of fetal deaths of 28 or more weeks of gestation}}{\text{Number of live births + fetal deaths of 28 or more weeks of gestation}} \times 1000 \qquad (2.1)$$

$$\frac{\text{Number of fetal deaths of 20 or more weeks of gestation}}{\text{Number of live births + fetal deaths of 20 or more weeks of gestation}} \times 1000 \qquad (2.2)$$

2.2. Neonatal mortality rate.

$$\frac{\text{Number of deaths under 28 days of age}}{\text{Number of live births}} \times 1000 \qquad (2.3)$$

2.3. Perinatal mortality rate. There are two definitions in common use:

$$\frac{\text{Number of deaths under 7 days + fetal deaths of 28 or more weeks of gestation}}{\text{Number of live births + fetal deaths of 28 or more weeks of gestation}} \qquad (2.4)$$

$$\frac{\text{Number of deaths under 28 days of life + fetal deaths of 20 or more weeks of gestation}}{\text{Number of live births + fetal deaths of 20 or more weeks of gestation}} \times 1000 \qquad (2.5)$$

The second definition covers a longer period both in gestation and after birth.

2.4. Post neonatal mortality rate.

$$\frac{\text{Number of deaths at age 28 days through one year}}{\text{Number of live births - neonatal deaths}} \times 1000 \qquad (2,6)$$

It is incorrect not to subtract neonatal deaths from live births in the denominator. Difference in numerical value due to this error depends on neonatal mortality; the difference may be considerable when neonatal mortality is high.

2.5. Infant mortality rate.

$$\frac{\text{Number of deaths under one year of age}}{\text{Number of live births}} \times 1000 \qquad (2.7)$$

Mortality rates defined above are closer to probability than to age-specific death rates, since in each instance the numerator is a part of the denominator. There are measures of mortality which resemble neither probability nor age specific death rates. Nevertheless, they are quite useful in mortality analysis. Some examples follow.

2.6. Fetal death ratio.

$$\frac{\text{Number of fetal deaths of 20 or more weeks of gestation}}{\text{Number of live births}} \times 1000 \qquad (2.8)$$

2.7. Maternal mortality rate.

$$\frac{\text{Number of maternal deaths}}{\text{Number of live births}} \times 1000 \qquad (2.9)$$

A maternal death is a death occurring to women due to complications of pregnancy, childbirth and the peurperium (period after delivery). While not strictly a measure of risk, the maternal mortality rate indicates a "price" in terms of mother's life that a human population pays for every infant brought into the world.

It was indicated at the beginning of this section that fetal death and infant mortality have experienced a constant decline. We shall now substantiate this statement by citing a report prepared by Helen C. Chase in 1967. She states:

> "One of the notable health accomplishments in the 20th century has been the decline in infant mortality. Over the first half of the century the rapid decline in mortality among infants became an accepted component of the Nation's health. In the past decade, it has become difficult to adjust to the idea that infant mortality in the United States is no longer declining at its former rate."

The deceleration of the rate of decline in infant mortality, however, was not peculiar to the United States. Similar changes in trend have appeared in several European countries. Tables 2 and 3 summarize these findings. It may be noted that even during the period from 1950 to 1962, the reduction in fetal death and infant mortality was still substantial. Table 4 shows the fetal and infant mortality in the United States from 1960 to 1970. The reductions in all categories are still quite considerable.

Table 2. Infant mortality rates and percent reduction: Selected countries, 1935, 1950, and 1962

	Infant Mortality Rate			Percent Reduction		
Country	1935	1950	1962	1935-62	1935-50	1950-62
Denmark	71.0	30.7	20.0	72	57	35
England & Wales	56.9	29.9	21.7	62	47	27
Netherlands	40.0	26.7	17.0	57	33	36
Norway	44.4	28.8	17.7	60	35	39
Scotland	76.8	37.6	26.5	65	51	30
Sweden	45.9	21.0	15.3	67	54	27
United States	55.7	29.2	25.3	55	48	13

Rates per 1,000 live births.

SOURCE: Helen C. Chase, "International Comparision of Perinatal and Infant Mortality: The United States and Six West European Countries," Vital and Health Statistics, Series 3, No. 6, pp. 1-97, U.S. Government Printing Office.

Table 3. Fetal mortality rates* and percent reduction, selected countries, 1955 and 1963

Country	Fetal Mortality Rates 1955	Fetal Mortality Rates 1963	Difference	Percent Reduction
Denmark	17.9	11.4	6.5	36.3
England & Wales	23.2	17.2	6.0	25.9
Netherlands	17.0	14.3	2.7	15.9
Norway	14.9	12.6	2.3	15.4
Scotland	24.6	19.1	5.5	22.4
Sweden	16.7	12.0	4.7	28.1
United States	12.6	11.3	1.3	10.3

*Fetal deaths of 28 or more weeks of gestation. Rates per 1,000

SOURCE: Helen C. Chase, ibid

Table 4. Fetal and infant mortality and percent reduction, United States, 1960 and 1970

	1960	1970	Difference	Percent Reduction
Fetal death rate (20 weeks + gestation)	15.8	14.0	1.8	11.4
Neonatal mortality rate	18.7	15.1	3.6	19.3
Postneonatal mortality rate	7.5	4.9	2.6	34.7
Infant mortality rate	26.0	20.0	6.0	23.1
Fetal death ratio	16.1	14.2	1.9	11.8
Maternal mortality rate (per 100,000)	37.1	21.5	15.6	42.0

Rates per 1,000

3. Adjustment of Rates

Specific death rates presented in Section 3 are essential in mortality analysis. Individually, these rates describe mortality experience within respective categories of people. Collectively, they represent a mortality pattern of the population in question. When a collective measure of mortality of an entire population is required, specific rates provide the fundamental components. One of the central tasks in statistical analysis of mortality data is making comparisons of experiences of various communities or countries; summarization of specific rates in a single number is extremely important. Since age-sex distribution varies from one community to another, and from one country to another, adjustment for such variation will have to be made in summarizing specific rates. The resulting single figure is called the adjusted rate. Adjustment can be made with respect to age, sex, occupation and possibly others. For simplicity, we shall consider only age-adjusted rates. Adjusted rates for other variables, such as sex-adjusted rates, age-sex-adjusted rates, etc., can be computed similarly. Various methods of adjustment have been proposed; some of these are listed in Table 5. It is the purpose of this section to review them. But first, let us introduce some notations.

In the adjustment of rates, two populations are usually involved: A community, u, during a calendar year (the population of interest) and a standard population, s. For each age interval (x_i, x_{i+1}) in the community, u, let D_{ui} be the number of deaths; P_{ui}, the midyear population; M_{ui}, its specific death rate; and let $n_i = x_{i+1} - x_i$ be the length of the interval. The sum

$$\sum_i D_{ui} = D_u \tag{3.1}$$

is the total number of deaths occuring in the community during the calendar year. The sum

$$\sum_i P_{ui} = P_u \qquad (3.2)$$

is the total midyear population. For the standard population, the symbols D_{si}, P_{si}, M_{si}, D_s and P_s are defined similarly. These symbols are similar to those used in Section 1 except for the addition of the subscripts u and s.

Table 5. Age-adjusted death rates and mortality indices

Title	Formula	Reference
Crude death rate (C.D.R.)	$\frac{\Sigma P_{ui}M_{ui}}{P_u}$	Linder, F. E. and Grove, R. D. (1943)
Direct method of adjustment (D.M.D.R.)	$\frac{\Sigma P_{si}M_{ui}}{P_s}$	"The Registrar General's Statistical Revies of England & Wales for the Year 1934"
Comparative mortality rate (C.M.R.)	$\frac{1}{2}\Sigma\left(\frac{P_{ui}}{P_u}+\frac{P_{si}}{P_s}\right)M_{ui}$	Ibid
Indirect method of adjustment (I.M.D.R.)	$\frac{(D_s/P_s)(D_u/P_u)}{\Sigma P_{ui}M_{si}/P_u}$	"The Registrar General's Decennial Supplement, England and Wales, 1921, Part III."
Life table death rate (L.T.D.R.)	$\frac{\Sigma L_i M_{ui}}{\Sigma L_i}$	Brownlee, J. (1913) (1922)
Equivalent average death rate (E.A.D.R.)	$\frac{\Sigma n_i M_{ui}}{\Sigma n_i}$	Yule, G. U. (1934)
Relative mortality index (R.M.I.)	$\frac{\Sigma P_{ui}\frac{M_{ui}}{M_{si}}}{P_u}$	Linder, F.E. and Grove, R. D. (1943)
Mortality index (M.I.)	$\frac{\Sigma n_i\frac{M_{ui}}{M_{si}}}{\Sigma n_i}$	Yerushalmy, J. (1951)
Standardized mortality ratio (S.M.R.)	$\frac{\Sigma P_{ui}M_{ui}}{\Sigma P_{ui}M_{si}}$	"The Registrar General's Statistical Review of England and Wales , 1958

3.1. Crude death rate. As was mentioned in Section 1, the crude death rate is the ratio of the total number of deaths occurring in a community during a calendar year to the community's total midyear population:

$$\text{C.D.R.} = D_u/P_u \quad . \tag{3.3}$$

The crude death rate, which is the most commonly used and conveniently computed single value, bears a close relationship to age-specific death rates. The numerator in (3.3) is the sum of the number of deaths occurring in all age categories:

$$D_u = \sum_i D_{ui} \quad . \tag{3.4}$$

By definition, the age-specific death rate for age interval (x_i, x_{i+1}) is given by

$$M_{ui} = D_{ui}/P_{ui} \quad , \tag{3.5}$$

so that the number of deaths (D_{ui}) is the product of the age-specific death rate (M_{ui}) and the corresponding midyear population (P_{ui}):

$$D_{ui} = P_{ui}M_{ui} \quad . \tag{3.6}$$

Therefore, the total number of deaths in (3.4) may be rewritten as

$$D_u = \sum_i P_{ui}M_{ui} \quad . \tag{3.7}$$

Substituting (3.7) in (3.3) yields

$$\text{C.D.R.} = \sum_i \frac{P_{ui}}{P_u} M_{ui} \quad , \tag{3.8}$$

where the summation is taken over the entire life span. Thus the C.D.R. is a weighted mean of age-specific death rates with the actual population proportions P_{ui}/P_u experiencing the mortality used as weights. From this viewpoint, the C.D.R. is the most meaningful single figure summarizing the mortality experience of a given population.

The C.D.R., however, is not without deficiencies. The quantity on the right-hand side of (3.8) is a function of both the age-specific death rates and the age-specific population proportions. As a weighted mean of age-specific death rates, the C.D.R. is affected by the population composition of the community in question. This disadvantage becomes apparent when the C.D.R. is used as a common measure to compare the mortality experience of several communities. The example in Table 6 illustrates this point.

Table 6. Age-specific death rates and crude death rates for communities A and B.

	Community A			Community B		
	Population	Deaths	Rate per 1000	Population	Deaths	Rate per 1000
Children	10,000	80	8.0	25,000	250	10.0
Adults	15,000	165	11.0	15,000	180	12.0
Senior citizens	25,000	375	15.0	10,000	160	16.0
Total	50,000	620	12.4	50,000	590	11.8

Although the age-specific death rate for each age group in Community A is lower than that for the corresponding age group in Community B, the crude death rate for Community A, (12.4), is higher than that for Community B

(11.8). This inconsistency is explained by differences in the population composition of the two communities. Community A consists of a larger percentage of older people, who are subject to a high mortality and contribute more deaths. As a result, Community A's overall crude death rate is higher than that of the more youthful Community B.

3.2. Direct Method Death Rate (D.M.D.R.). One way of adjusting for peculiarities of population composition is to introduce a standard population common to all the communities. When the age-specific death rates of a community are applied to such a standard population, we obtain a death rate adjusted by the direct method:

$$\text{D.M.D.R.} = \sum_i \frac{P_{si}}{P_s} M_{ui} \; . \tag{3.9}$$

The D.M.D.R. is thus a weighted mean of the age-specific death rates M_{ui} of a community with standard population proportions, P_{si}/P_s, applied as weights. If formula (3.9) is rewritten as

$$\text{D.M.D.R.} = \frac{\sum_i P_{si} M_{ui}}{P_s} \; , \tag{3.10}$$

the numerator becomes the number of deaths that would occur in the standard population if it were subject to the age-specific rates of the community. The ratio of the total "expected deaths" to the entire standard population yields the D.M.D.R. However, the D.M.D.R., as well as other age adjusted rates which follow, is not designed to measure the mortality experience of a community. It is simply a means for evaluating mortality experience of one community relative to another. An age-adjusted rate should be considered with this understanding.

Computation of the D.M.D.R. based on the example in Table 6 is given in Table 7. In this illustration, the combined population of the two communities is used as the standard population shown in column (1) in Table 7. The age-specific rates in the two communities are recorded in columns (2) and (3), respectively. Each of the specific rates is then applied to the standard population in the same age group to obtain the number of deaths expected in the standard population shown in columns (4) and (5). Summing these expected numbers of deaths over all age groups yields the total number of deaths, 1,135 and 1,270, respectively. When the total number of deaths is divided by the total standard population, we obtain the D.M.D.R.

Table 7. Direct method age-adjusted rates for Communities A and B

Standard Population	Age Specific Rates		Expected No. of Deaths	
	Community A	Community B	Community A	Community B
(1)	(2)	(3)	(4)	(5)
35,000	8.0	10.0	280	350
30,000	11.0	12.0	330	360
35,000	15.0	16.0	525	560
100,000			1,135	1,270

Adjusted Rate: Community A = 11.35/1,000
Community B = 12.70/1,000

Using a single standard population, the direct method of adjustment eliminates the effect of differences in age-composition of the communities under study; the result nevertheless depends upon the composition of the population selected as a standard. When communities with very different mortality patterns are compared, different standard populations may even produce contradictory results. In computing the age-adjusted rate for the 1940 white male population of Louisiana and New Mexico, Yerushalmy (1951) found that the age-adjusted rate for Louisiana (13.06 per 1,000) was slightly higher than the rate for New Mexico (13.05 per 1,000) when the 1940 U.S. population was used as the standard; but the rate for Louisiana (10.14 per 1,000) was lower than the rate for New Mexico (11.68 per 1,000) when the 1901 population of England and Wales was used as the standard. This kind of dilemma has led to the development of other methods of adjustment.

3.3. Comparative Mortality Rate (C.M.R.). In this method of adjustment, both the age composition of the community and that of the standard population are taken into account. The formula is

$$\text{C.M.R.} = \tfrac{1}{2} \sum_i \left(\frac{P_{ui}}{P_u} + \frac{P_{si}}{P_s} \right) M_{ui} \tag{3.11}$$

Easy computations show that the first sum is the crude death rate of the community,

$$\sum_i \frac{P_{ui}}{P_u} M_{ui} = \sum_i \frac{D_{ui}}{P_u} = \frac{D_u}{P_u} \quad ,$$

while the second sum is the direct method death rate. Thus the C.M.R. is simply the mean of the C.D.R. and D.M.D.R. Using the previous example once again, we find

C.M.R.(community A) = $\tfrac{1}{2}$(12.4 + 11.35) = 11.87

C.M.R.(community B) = $\tfrac{1}{2}$(11.8 + 12.70) = 12.25

3.4. Indirect Method Death Rate (I.M.D.R.). In the age-adjusted rate by the indirect method, the crude death rate of the community is multiplied by the ratio of the crude death rate of a standard population to the death rate that would be expected in the standard population if it had the same composition as the given community.[1/] The formula for the I.M.D.R. is

$$\text{I.M.D.R.} = \frac{D_s/P_s}{\sum_i P_{ui}M_{si}/P_u} \; \frac{D_u}{P_u} \quad . \tag{3.12}$$

The denominator of the first factor in (3.12)

$$\sum_i \frac{P_{ui}M_{si}}{P_u}$$

is in effect a D.M.D.R. when the position of a community and a standard population is interchanged: the age-specific death rates of a standard population (M_{si}) are applied to a community population (P_{ui}).

When the population composition of a community and a standard population are the same, so that

$$\frac{P_{ui}}{P_u} = \frac{P_{si}}{P_s}$$

for every interval (x_i, x_{i+1}), then the first factor in (3.12) becomes unity,

$$\frac{D_s/P_s}{\sum_i P_{ui}M_{si}/P_u} = \frac{\sum_i P_{si}M_{si}/P_s}{\sum_i P_{ui}M_{si}/P_u} = 1 \; ,$$

and the I.M.D.R. is equal to the C.D.R. of the community. If a community should have a higher proportion of old people than the standard population, then for the old age group

---

1/ A method suggested by Herald Westergaard is also used in the study of death rates. Westergaard's formula, however, can be derived from the indirect method and vice versa.

$$\frac{P_{ui}M_{si}}{P_u} > \frac{P_{si}M_{si}}{P_s}$$

and the crude death rate of the community will be greater than the I.M.D.R.

Formula (3.12) can be written also as

$$\text{I.M.D.R.} = \frac{D_s}{(\sum_i P_{ui}M_{si})P_s} D_u$$

$$= \frac{D_s}{(\sum_i P_{ui}M_{si})P_s} \sum_i P_{ui}M_{ui}$$

$$= \sum_i w_i M_{ui} \tag{3.13}$$

where

$$w_i = \frac{D_s}{(\sum_i P_{ui}M_{si})P_s} P_{ui} \quad . \tag{3.14}$$

Here the weights w_i do not add to unity unless the community and the standard population have the same composition. Therefore, generally the I.M.D.R. is not an average of the specific death rates, and is not directly comparable with the C.D.R. or the D.M.D.R.

One advantage of the indirect method of adjustment may be noted. Since only the total number of deaths in a community (D_u) is in the formula, this method of adjustment requires less information from a community than the direct method.

3.5. <u>Life Table Death Rate</u> (L.T.D.R.). Most of the methods of adjustment rely on a standard population or its rates. One exception is the L.T.D.R. which is defined as

$$\text{L.T.D.R.} = \sum_i \frac{L_i}{T_0} M_{ui} \quad , \tag{3.15}$$

where L_i is the number of years spent in (x_i, x_{i+1}) by a life table population and

$$T_0 = L_0 + L_1 + \ldots \tag{3.16}$$

A full appreciation of this method of adjustment requires the knowledge of the life table discussed in Chapter 5; a brief discussion of formula (3.15) follows. Given ℓ_0 people alive at age 0 who are subject to the age-specific death rates of the community, L_i/T_0 is the proportion of their life time spent in the age interval (x_i, x_{i+1}). In other words, the L.T.D.R. shown in formula (3.15) is a weighted mean of the age specific death rates (M_{ui}) with the proportion of life time spent in (x_i, x_{i+1}) being used as weights. Since the weights L_i/T_0 depend solely on the age-specific death rates, the L.T.D.R. is independent of the population composition either of a community or a standard population.

As we will see in Chapter 5, the age specific death rate M_{ui} is equal to the ratio d_i/L_i,

$$M_{ui} = d_i/L_i \quad ,$$

hence

$$L_i M_{ui} = d_i \tag{3.17}$$

where d_i is the life table deaths in age interval (x_i, x_{i+1}). The sum,

$$d_0 + d_1 + \ldots = \ell_0 \quad , \tag{3.18}$$

is equal to the total number of individuals ℓ_0 at age 0. Substituting (3.17) in (3.15) and recognizing (3.18), we have

$$\text{L.T.D.R.} = \sum_i d_i/T_0 = \ell_0/T_0 \quad . \tag{3.19}$$

The inverse

$$T_0/\ell_0 = \hat{e}_0 \tag{3.20}$$

is known as the (observed) expectation of life at age 0; therefore

$$\text{L.T.D.R.} = \frac{1}{\hat{e}_0} \quad . \tag{3.21}$$

3.6. Equivalent Average Death Rate (E.A.D.R.). In this method of adjustment each age-specific rate is weighted with the corresponding interval length rather than the number of people for which the rate is computed. In formula, it is:

$$\text{E.A.D.R.} = \sum_i \frac{n_i}{\sum_i n_i} M_{ui} \quad , \tag{3.22}$$

where $n_i = x_{i+1} - x_i$. The last age interval is an open interval, such as 60 and over, and the corresponding death rate is usually high. An upper limit must be set for the last interval in order to prevent the high death rate of the elderly from asserting an undue effect on the resulting adjusted rate. G. U. Yule, the original author of the index, suggested that the limit of the last age interval be set at 65 years. It may be observed that since there are fewer people in the old age group, the E.A.D.R. places more emphasis on old ages than the C.R.D. or the D.M.D.R.

3.7. Relative Mortality Index (R.M.I.). The basic quantities used in the relative mortality index are the ratios of specific rates of a community to the corresponding rates of a standard population. The index is a weighted mean of these ratios, obtained by using the community age-specific population proportions as weights. The formula for the R.M.I. is

$$\text{R.M.I.} = \sum_i \frac{P_{ui}}{P_u} \frac{M_{ui}}{M_{si}} . \tag{3.23}$$

The R.M.I. strongly reflects the mortality pattern of young age groups where small changes in the specific rates may produce large differences in the value of the index.

When (3.23) is rewritten as

$$\text{R.M.I.} = \frac{1}{P_u} \sum_i \frac{D_{ui}}{M_{si}} ,$$

we see that the R.M.I. may be computed without knowledge of the community's population by age.

3.8. Mortality Index (M.I.). This index is also a weighted average of the ratios of community age-specific death rates to the corresponding rates of a standard population. It differs from the relative mortality index in that the weights used here are the lengths of age intervals. The formula for the index is

$$\text{M.I.} = \frac{\Sigma n_i \frac{M_{ui}}{M_{si}}}{\Sigma n_i} \tag{3.24}$$

Generally, the M.I. is affected more by the death rates in old age groups than is the R.M.I. A main feature of this method is that, for intervals of the same length, a constant change of the ratio M_{ui}/M_{si} has an equal effect on the value of the index.

3.9. Standardized Mortality Ratio (S.M.R.). The General Register Office of Great Britain has used the S.M.R. in the Statistical Review of England and Wales since 1958. It is a ratio of the number of deaths occurring in a community to the expected number of deaths in the community if it were subject to the age-specific rates of a standard population. In formula,

$$\text{S.M.R.} = \frac{\sum_i D_{ui}}{\sum_i P_{ui} M_{si}} = \frac{\Sigma P_{ui} M_{ui}}{\Sigma P_{ui} M_{si}} \tag{3.25}$$

Since the numerator is the total deaths in the community, (3.25) can be rewritten as

$$\text{S.M.R.} = \frac{D_u}{\sum_i P_{ui} M_{si}} \quad , \tag{3.26}$$

or

$$\text{S.M.R.} = \frac{D_u / P_u}{\sum_i P_{ui} M_{si} / P_u} \tag{3.27}$$

Thus, the S.M.R. is the crude death rate of a community divided by the direct method death rate, when standard population age-specific death rates are applied to a community population.

CHAPTER 3

STANDARD ERROR OF MORTALITY RATES

1. Introduction

An age-specific death rate is a measure of the mortality experience of a defined population group over a given period of time. An age-adjusted death rate, as a function of age-specific rates, is designed to summarize the mortality experience of an entire population for the purpose of comparing it with that of other populations. As with any observable statistical quantity, both the specific rate and the adjusted rate are subject to random variation (random error) and any expression of the rates must take this variation into account. A measure of the variation is the standard deviation, or the standard error, of a rate. We need the standard deviation in order to use the rates in estimation, for testing hypotheses, or for making other statistical inferences concerning the mortality of a population. With the standard deviation one can assess the degree of confidence that may be placed in the findings and conclusions reached on the basis of these rates. With the standard deviation one can also measure the quality of the vital statistics and, in fact, evaluate the reliability of the rates themselves.

Since a death rate is often determined from the mortality experience of an entire population rather than from a sample, it is sometimes argued that there is no sampling error; and therefore the standard deviation, if it exists, can be disregarded. This point of view, however, is static. Statistically speaking, human life is a random experiment and its outcome, survival or death, is subject to chance. If two people were subjected to the same risk of dying (force of mortality) during a calendar year, one might die during the year and the other survive. If a person was allowed to relive the year he survived the first time, he might not survive the second time. Similarly, if a population were allowed to live the same year over again, the total number of deaths occuring during the second time would assume a different value and so, of course, would the corresponding death rate. It is in this sense that a death rate is subject to random variation even though it is based on the total number of deaths and the entire population.

From a theoretical viewpoint, a death rate is an estimate of certain functions of the force of mortality acting upon each individual, and may assume different values with correspondingly different probabilities, even if the force of mortality remains constant. Therefore, it is natural and meaningful to study the standard deviation of a rate.

Age-specific death rates, when they are determined from a sample, are subject to sampling variation in addition to random variation. The standard deviation of a death rate assumes different forms, depending upon the sampling unit and sampling procedure used. But generally it consists of two components: one due to sampling, and the other due to experimentation (the chance of surviving the year). The standard deviation of a death rate based on a sample will be discussed in Section 4. At present, we will discuss the standard deviation of death rates subject to random variation only.

REMARK. The terms "standard deviation" (of a death rate) and "standard error" (of a death rate) have the same meaning. They are the square root of the variance (of a death rate), and both are commonly used in statistics and in mortality analysis. To acquaint the reader with both terms, we shall use "standard deviation" and "standard error" alternately in this manual.

2. The Binomial Distribution

The basic concept used in application of statistical inference to death rates is the binomial distribution and the central limit theorem. Consider a sequence of independent trials, each trial having either of two possible outcomes, i.e., "success" or "failure," with the corresponding probabilities remaining the same for all trials. Such trials are called Bernoulli trials. Tossing a coin is a familiar example: each toss of a coin constitutes a trial (a random experiment) with either of two possible outcomes, heads or tails. A person's life over a year is another example with the corresponding outcomes of survival or death during the year. The binomial random variable is the number of "successes" in a number of independent and identical trials, each trial can result either in a "success" or a "failure" and the probability of a "success" is the same for all trials. Thus a binomial random variable is the number of "successes" in a number of Bernoulli trials. The number of heads shown in a number of tosses of a coin is a binomial random variable. If N_i people alive at exact age x_i are subject to the same probability q_i of dying in the age interval (x_i, x_{i+1}), the number of people D_i dying in the interval is also a binomial random variable. The expected number of deaths, denoted by $E(D_i)$ is

$$E(D_i) = N_i q_i \tag{2.1}$$

and the variance of D_i is

$$\sigma^2_{D_i} = N_i q_i (1-q_i). \tag{2.2}$$

The proportion of deaths, or the binomial proportion,

$$\frac{D_i}{N_i} = \hat{q}_i \tag{2.3}$$

is an unbiased estimate of the probability q_i in the sense that its expected value is equal to q_i

$$E(\hat{q}_i) = E\left(\frac{D_i}{N_i}\right) = \frac{1}{N_i} E(D_i) = \frac{1}{N_i} N_i q_i = q_i . \tag{2.4}$$

The variance of $\hat{q}_i$, which may be derived from (2.2), is given by

$$\sigma^2_{\hat{q}_i} = \frac{1}{N_i} q_i(1-q_i). \tag{2.5}$$

When the probability q_i is unknown, its estimate $\hat{q}_i$ is substituted in (2.5) to give the "sample" variance of $\hat{q}_i$,

$$S^2_{\hat{q}_i} = \frac{1}{N_i} \hat{q}_i(1-\hat{q}_i) . \tag{2.6}$$

Both the variance in (2.5) and the sample variance in (2.6) are measures of variation associated with the proportion $\hat{q}_i$ and play an important role in making inferences concerning the unknown probability q_i. The fundamental theorem needed in this situation is the central limit theorem. According to the theorem, when N_i is sufficiently large, the standardized form of the random variable $\hat{q}_i$,

$$Z = \frac{\hat{q}_i - q_i}{\sqrt{q_i(1-q_i)/N_i}} \tag{2.7}$$

has the standard normal distribution with a mean of zero and a variance of one.

Formula (2.7) expresses the deviation of the random variable $\hat{q}_i$ from its expected value q_i in units of the standard deviation $\sigma_{\hat{q}_i}$. Using formula (2.7), one can test a hypothesis concerning the probability q_i or estimate q_i by means of a confidence interval.

Suppose a study of infant mortality in a community suggests a decline in infant deaths. A hypothesis concerning the probability of death in the first year of life, $q_0 = .028$ (or 28 per 1,000), is to be tested against an alternative hypothesis $q_0 < .208$. The statistic used to test the hypothesis is the quantity in (2.7) with the substitution of $q_0 = .028$, or

$$Z = \frac{\hat{q}_0 - .028}{\sqrt{(.028)(1-.028)/N_0}} \tag{2.8}$$

where N_0, the number of newborns in the study and $\hat{q}_0 = D_0/N_0$, the proportion of infant deaths, can be determined from the data observed, and the quantity in (2.8) can be computed. Rejection or acceptance of the hypothesis $q_0 = .028$ is based on the computed value of (2.8). At the 5% level of significance, for example, the hypothesis is rejected if the computed value of Z is less than -1.645, the fifth percentile in the standard normal distribution.

One may also use (2.7) and the normal distribution percentiles to determine confidence intervals for the probability q_i. For a .95 confidence coefficient, for example, we use the 2.5 percentile of -1.96 and the 97.5 percentile of +1.96. This means that

$$\Pr\left\{-1.96 < \frac{\hat{q}_i - q_i}{\sqrt{q_i(1-q_i)/N_i}} < 1.96\right\} = .95 \quad . \tag{2.9}$$

The inequalities inside the braces are approximately equivalent to

$$\hat{q}_i - 1.96\, S_{\hat{q}_i} < q_i < \hat{q}_i + 1.96\, S_{\hat{q}_i} \quad , \tag{2.10}$$

where the sample standard deviation, $S_{\hat{q}_i}$, is the square root of the variance in (2.6). The inequalities in (2.10) provide the fundamental formula for the 95% confidence interval for the probability q_i.

3. Probability of Death and the Age-specific Death Rate

The probability of death and the age-specific death rate are two measures of the risk of mortality acting on individuals in the population. While the probability of death is an established concept in the field of statistics, analytic meaning of the age-specific death rate is not fully appreciated. The age-specific death rate either is regarded as an ill-defined statistical quantity, or else it is treated as if it were another name for the probability of death. These misconceptions need be corrected. The age-specific death rate is just as meaningful analytically as the probability. The exact meaning of the age-specific death rate and its relationship with the probability of death have been given in Chapter 2 and will be discussed in more detail in Chapter 5. For easy reference, we state again the estimate of the probability and the age-specific death rate below.

Let N_i be the number of individuals alive at the exact age x_i, among them a number D_i dying during the interval (x_i, x_{i+1}). Then the estimate of the probability of dying in (x_i, x_{i+1}) is given by (cf. equation (2.3)),

$$\hat{q}_i = \frac{D_i}{N_i} \quad . \tag{3.1}$$

On the other hand, the age-specific death rate, M_i, is the ratio of the number of deaths, D_i, to the total number of years lived in the interval (x_i, x_{i+1}) by the N_i people. In formula

$$M_i = \frac{D_i}{n_i(N_i-D_i) + a_i n_i D_i} \quad . \tag{3.2}$$

Solving equations (3.1) and (3.2) yields the basic relationship between $\hat{q}_i$ and M_i

$$\hat{q}_i = \frac{n_i M_i}{1 + (1-a_i)n_i M_i} \quad . \tag{3.3}$$

Here $n_i = x_{i+1} - x_i$, and a_i is the average fraction of the age interval (x_i, x_{i+1}) lived by individuals dying at any age included in the interval. The fraction a_i has been computed for a number of countries whose population and mortality data are available; the values of a_i are given in Appendix V.

For a current population, the age-specific death rates are determined from the vital and population statistics,

$$M_i = \frac{D_i}{P_i} \tag{3.4}$$

where D_i is the number of deaths occurring in age group (x_i, x_{i+1}) during a calendar year and P_i is the corresponding mid-year population. The probability of death $\hat{q}_i$ is computed from formula (3.3).

To determine the variance of $\hat{q}_i$, we start with formula (2.6)

$$S^2_{\hat{q}_i} = \frac{1}{N_i} \hat{q}_i(1-\hat{q}_i) \ . \tag{2.6}$$

Since equation (3.1) implies that

$$\frac{1}{N_i} = \frac{1}{D_i} \hat{q}_i \ ,$$

we have the desired formula for the sample variance of $\hat{q}_i$:

$$S^2_{\hat{q}_i} = \frac{1}{D_i} \hat{q}_i^2 \ (1-\hat{q}_i) \quad . \tag{3.5}$$

The exact formula for the variance of the age-specific death rate is difficult to derive. However, since the population size P_i in (3.4) usually is large, we use Taylor's expansion to establish the following relationship between the variance of M_i and the variance of D_i:

$$S_{M_i}^2 = \frac{1}{P_i^2} S_{D_i}^2 , \tag{3.6}$$

where the sample variance of D_i is

$$S_{D_i}^2 = N_i \hat{q}_i (1-\hat{q}_i) = D_i (1-\hat{q}_i) \quad . \tag{3.7}$$

Substituting formula (3.7) in (3.6) yields the required formula for the sample variance of the age-specific death rate

$$S_{M_i}^2 = \frac{1}{P_i} M_i (1-\hat{q}_i) \ . \tag{3.8}$$

When $\hat{q}_i$ is very small so that $1-\hat{q}_i$ is close to one, formulas (3.5) and (3.8) may be approximated by

$$S_{q_i}^2 = \frac{1}{D_i} \hat{q}_i^2 \tag{3.9}$$

and

$$S_{M_i}^2 = \frac{M_i}{P_i} , \tag{3.10}$$

respectively.

4. The Death Rate Determined from a Sample

It should be emphasized that although N_i and P_i in the above discussion both refer to the numbers of people in a population, the formulas of sample variances of $\hat{q}_i$ and M_i in (3.5) and (3.8) hold also when N_i and P_i are the numbers of people in a sample. To verify this, suppose a random sample of N people is taken from an entire population. In the sample there are N_i people of age x_i, D_i of whom die during the year, and */

$$\frac{D_i}{N_i} = \hat{q}_i \tag{4.1}$$

is an estimate of the probability q_i. We are interested in the sample variance of $\hat{q}_i$. In formula (4.1) both the numerator and the denominator are random variables; N_i is subject to sampling variation in the sense that the number of people of age x_i included in the sample varies from one sample to another, while D_i is subject to sampling variation and random variation (survival or death during the year). The formula for the variance of the ratio in (4.1) thus can be expressed in terms of the variance of N_i and of D_i. However, the variance of D_i consists of two components: the random component and the sampling component. The derivation of the variance of $\hat{q}_i$ through the variance of D_i is lengthy. To save space, we use the following simpler approach to derive the variance of $\hat{q}_i$ directly.

It is easy to verify that given N_i the conditional expectation and conditional variance of $\hat{q}_i$ are, respectively,

$$E(\hat{q}_i|N_i) = q_i \tag{4.2}$$

and

$$\sigma^2_{\hat{q}_i|N_i} = \frac{1}{N_i} q_i(1-q_i). \tag{4.3}$$

*/ For simplicity in demonstrating our reasoning, but at the expense of a certain degree of reality, we assume N_i people of exact age x_i.

On the other hand, because of (4.2), the variance of $\hat{q}_i$ is equal to the expected value of the conditional variance of $\hat{q}_i$ given N_i,

$$\sigma^2_{\hat{q}_i} = E(\sigma^2_{\hat{q}_i} | N_i). \tag{4.4}$$

Substituting (4.3) in (4.4) gives

$$\sigma^2_{\hat{q}_i} = E(\frac{1}{N_i}) q_i (1-q_i). \tag{4.5}$$

Using the sample information, we obtain the sample variance of $\hat{q}_i$

$$S^2_{\hat{q}_i} = \frac{1}{N_i}\hat{q}_i (1-\hat{q}_i) = \frac{1}{D_i} \hat{q}_i^2 (1-\hat{q}_i) , \tag{4.6}$$

since N_i is given in (4.1). This shows that although $\hat{q}_i$ in (4.1) is computed from a sample, its sample variance has the same expression as the variance of $\hat{q}_i$ based on a total population. It is easy to justify now that the variance of the age-specific death rate in (3.8) holds true also when the death rate is computed on the basis of a sample.

5. Age-Adjusted Death Rates and Mortality Indices

In Chapter 2 several methods of adjustment of age-specific death rates were presented. Although each method was developed on the basis of a specific philosophic argument and designed to serve a definite purpose, they all assume a general form of a weighted mean of the age-specific death rates. These methods of adjustment are reproduced in Table 1 for easy reference.

With the exception of the indirect method of adjustment, the weights add to unity. The sum of the weights in the indirect method can be greater or less than unity, depending upon the difference between community and standard populations in age composition. For this reason, the indirect method is not strictly comparable with any other adjusted rate, and neither is its standard error.

The inclusion of the crude death rate in the list of adjusted rates is of significance. Since it is usually expressed as the ratio of all deaths to the total midyear population, the crude death rate is occasionally treated as a binomial proportion, which leads to an incorrect formula for the standard deviation. Individuals differing in age and sex obviously do not have the same probability of dying, and the notion of an average probability is incomprehensible; therefore, a direct application of the binomial theory is inappropriate. If, however, it is visualized as the weighted mean of specific death rates, with the actual population size employed as weights, then the crude death rate is perhaps the most meaningful measure of mortality for a single community. This way of viewing the crude rate is also essential in the derivation of its standard deviation.

In all the adjusted rates, the choice of weights applied to specific rates is based on: (1) the proportion of those in a specific age group to the total population, i.e., population proportion, and (2) the relative interval length of a specific age group. For the crude rate, the weights used are the community population proportions in specific age groups (P_{ui}/P_u); for the direct method of adjustment, the standard population proportions in specific age groups (P_{si}/P_s); for the comparative mortality rate, the average of the two population proportions; for the life table death rate, the life table population proportions for specific age groups (L_i/T_0); and for the equivalent average death rate, the relative interval lengths of the age groups $(n_i/\Sigma n_i)$. The weights used in the indirect method of adjustment are functions of the age-specific rate for the standard population, community population proportions, and standard population proportions.

The methods of adjustment listed in Table 1 also include two indices, the relative mortality index, the mortality index, and the standardized mortality ratio. As seen from the second panel of Table 1, the two indices are weighted means of the ratios of a community's specific death rates to the corresponding specific rates for the standard population. The difference is that the relative mortality index uses the population proportions of the community for specific age groups as weights, while the mortality index uses the relative lengths of the age intervals. In the derivation of their standard deviations, however, we shall consider them as linear functions of age-specific death rates of a community with coefficients as listed in the weight column.

Table 1. Formulas and weights used to compute the crude death rate, age-adjusted rates and mortality indices

Title	Formula	Weight (w_i)
Crude death rate (C.D.R.)	$\dfrac{\Sigma\, P_{ui}M_{ui}}{P_u}$	$\dfrac{P_{ui}}{P_u}$
Direct method of adjustment (D.M.D.R.)	$\dfrac{\Sigma P_{si}M_{ui}}{P_s}$	$\dfrac{P_{si}}{P_s}$
Comparative mortality rate (C.M.R.)	$\frac{1}{2}\Sigma\left(\dfrac{P_{ui}}{P_u}+\dfrac{P_{si}}{P_s}\right)M_{ui}$	$\frac{1}{2}\left(\dfrac{P_{ui}}{P_u}+\dfrac{P_{si}}{P_s}\right)$
Indirect method of adjustment (I.M.D.R.)	$\dfrac{(D_s/P_s)(D_u/P_u)}{\Sigma P_{ui}M_{si}\ /P_u}$	$\dfrac{D_s/P_s}{\Sigma P_{ui}M_{si}}\ P_{ui}$
Life table death rate (L.T.D.R.)	$\dfrac{\Sigma L_i M_{ui}}{\Sigma L_i}$	$\dfrac{L_i}{\Sigma L_i}$
Equivalent average death rate (E.A.D.R.)	$\dfrac{\Sigma n_i M_{ui}}{\Sigma n_i}$	$\dfrac{n_i}{\Sigma n_i}$
Relative mortality index (R.M.I.)	$\dfrac{\Sigma P_{ui}\ \dfrac{M_{ui}}{M_{si}}}{P_u}$	$\dfrac{P_{ui}}{P_u M_{si}}$
Mortality index (.M.I.)	$\dfrac{\Sigma n_i\ \dfrac{M_{ui}}{M_{si}}}{\Sigma n_i}$	$\dfrac{n_i}{(\Sigma n_i)M_{si}}$
Standardized mortality ratio (S.M.R.)	$\Sigma\ \dfrac{P_{ui}}{\Sigma P_{ui}M_{si}}\ M_{ui}$	$\dfrac{P_{ui}}{\Sigma P_{ui}M_{si}}$

6. Sample Variance of the Age-Adjusted Death Rate

To derive the formula for the sample variance of an adjusted death rate, it is first essential to identify the random variables involved. Clearly, M_{ui}, the age-specific death rates for a community, are random variables while n_i, the interval length for an age group, is a constant. Community and standard population proportions for specific age groups will not be treated as random variables for the reason that the random event under study is death, not population. The age-specific death rates of the standard population are random variables, just as are the community age-specific death rates. However, since adjusted death rates are derived for the purpose of testing hypotheses concerning the mortality experience of communities, only that part of the random variation associated with the communities in question should be taken into consideration. In other words, random variations attributable to the age-specific rates for the standard population should not be included in the variance of the adjusted rates. Life table population proportions, on the other hand, are derived from the age-specific death rates for a community; therefore, they should be treated as random variables. To summarize, we shall consider only the community age-specific death rates and the life table population proportions for specific age groups as random variables in the derivation of the sample variance.

With this understanding, and making an exception of the life table death rate, we shall write adjusted rates and mortality indices as linear functions of the basic random variables, the age-specific death rates of a community. The general formula for an adjusted death rate or mortality index R takes the form

$$R = \sum_i w_i M_{ui} \tag{6.1}$$

with the coefficient w_i as given in Table 1. The general rules for the variance of a linear function of random variables may now be applied and the variance of the adjusted rate, R, may be expressed as follows:

$$S_R^2 = \sum_i w_i^2 S_{M_{ui}}^2 + \sum_{i \neq j}\sum w_i w_j S_{M_{ui}, M_{uj}} \quad , \tag{6.2}$$

where $S_{M_{ui}}^2$ is the sample variance of the age-specific death rate for age group (x_i, x_{i+1}) in the community u, and $S_{M_{ui}, M_{uj}}$ is the sample covariance between the age-specific death rates, M_{ui} and M_{uj}.

The age-specific death rate, M_{ui}, is a function of the corresponding estimated probability of death, $\hat{q}_i$; and the covariance between death rates is also a function of the covariance between the two corresponding estimated probabilities. It has been proved [cf. Section 5, Appendix II] that the estimated probabilities for two nonoverlapping age intervals have a zero covariance. Thus, two death rates will also have a zero covariance. It follows that all the covariances in formula (6.2) will vanish, and the formula for the sample variance of R becomes

$$S_R^2 = \sum_i w_i^2 S_{M_{ui}}^2 \quad . \tag{6.3}$$

Using (3.8) for the variance of M_{ui}, we have

$$S_R^2 = \sum_i w_i^2 \frac{M_{ui}}{P_{ui}} (1-\hat{q}_{ui}) \quad , \tag{6.4}$$

or using the approximate formula, (3.10), we have

$$S_R^2 = \sum_i w_i^2 \frac{M_{ui}}{P_{ui}} \quad . \tag{6.5}$$

7. Computation of the Sample Variance of the Direct Method Age-Adjusted Death Rate

The computation of the sample variance of the age-specific death rate is the common essential part to all the methods of adjustment except for the life table death rate. Therefore, it is sufficient to use only the direct method of adjustment (D.M.D.R.) as an example. The formula for the sample variance of D.M.D.R. is obtained from (6.4) with $w_i = P_{si}/P_s$:

$$S_R^2 = \sum_i \left(\frac{P_{si}}{P_s}\right)^2 \frac{M_{ui}}{P_{ui}} [1-\hat{q}_{ui}] \tag{7.1}$$

For this illustration, we use the death rates of the total California population of 1970, and the United States 1970 population as the standard population. The steps involved in the computation are shown in Table 2.

The age group 85 and over presents a problem which needs special treatment. Because it is an open-ended group, the interval length is not determinable. The average number of years, $a_{85}n_{85}$, lived by individuals may be estimated by the reciprocal of the central death rate,

$$a_{85}n_{85} = \frac{1}{M_{85}} \quad . \tag{7.2}$$

Justification of (7.2) is given in Appendix II (cf., equation (8)) on life table construction. Equation (7.2) implies

$$1-a_{85}n_{85}M_{85} = 0 \ ,$$

which means that the sample variance of M_{85}, as given in (3.8), is equal to zero. Intuitively, the zero variance can be justified as follows: Each individual alive at age 85 has a future life time of, say, y years. The sample variance of y is the mean-square deviation of each y from the sample mean. If in a group of individuals alive at age 85, the only information available is that each y assumes an average of $a_{85}n_{85}$ years, then the deviation of each y from the sample mean is zero. The sample variance is also zero. This implies that the sample variance of M_{85} is zero. It may be noted also that $\hat{q}_{85} = 1$ so that the variance in (3.8) is equal to zero.

For each of the remaining age groups (x_i, x_{i+1}) we compute the sample variance of the death rate M_{ui} [cf., equation (3.8)]

$$S^2_{M_{ui}} = \frac{M_{ui}}{P_{ui}} (1-\hat{q}_{ui}) \tag{7.3}$$

as shown in column (7) in Table 2 and the corresponding weight squared $(P_{si}/P_s)^2$ in column (8), and find the product

$$\left(\frac{P_{si}}{P_s}\right)^2 S^2_{M_{ui}} = \left(\frac{P_{si}}{P_s}\right)^2 \frac{M_{ui}}{P_{ui}} (1-\hat{q}_{ui}) \; . \tag{7.4}$$

Adding the products in (7.4) over all age groups, we obtain the sample variance of R in formula (7.1).

For the California 1970 population the age-adjusted death rate is

$$R = \Sigma \frac{P_{si}}{P_s} M_{ui}$$

$$= \frac{1,787,768.98}{203,211,926} = .0087976 \; .$$

The computation in Table 2 shows that the sample variance is

$$S^2_R = 340.631 \times 10^{-12} \; .$$

and the standard deviation is

$$S_R = \sqrt{340.631 \times 10^{-2}} = 18.456 \times 10^{-6}$$

In comparison, the standard deviation S_R is much smaller than the age-adjusted death rate. Formally, the magnitude of a standard deviation is measured by the coefficient of variation, which is defined as the ratio

$$\text{Coeff. of variation of } R = \frac{S_R}{R} \quad . \tag{7.5}$$

In this case

$$\text{Coeff. of variation of } R = \frac{.018456}{8.7976}$$

$$= .0020979 = .21 \text{ percent}$$

The small magnitude of the coefficient of variance is mainly due to the large population sizes P_{ui}.

The age-adjusted death rate of 8.7976 per 1000 for the California 1970 population may be compared with the total United States population, 1970, death rate, 9.453 per 1000, since both are based on the same population distribution. Because of the small standard deviation (S_R = .018456 per 1000), we conclude that in 1970 the California population had a significantly lower mortality than the United States as a whole.

Table 2. Computation of sample standard error of the age-adjusted death rate for total population, California, 1970.

(Adjustment made by the direct method. The standard population used is the total population of the United States, enumerated as of April 1, 1970)

Age Interval (in years)	Length of Interval	Mid-year Population in Interval (x_i,x_{i+1})	Death Rate	Fraction of Last Age Interval of Life	Probability of Dying in Interval	Sample Variance of Age Specific Death Rate $\frac{M_{ui}}{P_{ui}}(1-\hat{q}_{ui})$	Square of Standard Population Proportion (U.S., 1970)
x_i to x_{i+1}	n_i	P_{ui}	M_{ui}	a_{ui}	$\hat{q}_{ui}$	$10^{12}\ S^2_{M_{ui}}$	$10^8(P_{si}/P_s)^2$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0-1	1	340483	.018309	.09	.01801	52805.15	29415.50
1-5	4	1302198	.000806	.41	.00322	616.96	452458.67
5-10	5	1918117	.000377	.44	.00188	196.18	964404.79
10-15	5	1963681	.000374	.54	.00187	190.10	1046618.41
15-20	5	1817379	.001130	.59	.00564	618.27	880681.46
20-25	5	1740966	.001552	.49	.00773	884.57	649012.63
25-30	5	1457614	.001421	.51	.00708	967.98	439832.81
30-35	5	1219389	.001611	.52	.00802	1310.56	316393.25
35-40	5	1149999	.002250	.53	.01119	1934.63	298733.21
40-45	5	1208550	.003404	.54	.01689	2769.03	347603.72
45-50	5	1245903	.005395	.53	.02664	4214.84	355480.49
50-55	5	1083852	.008256	.53	.04049	7308.85	298580.84
55-60	5	933244	.012796	.52	.06207	12860.25	240855.01
60-65	5	770770	.018565	.52	.08886	21945.99	179800.97
65-70	5	620805	.027526	.51	.12893	38622.55	118374.42
70-75	5	484431	.039529	.52	.18052	66868.60	71764.70
75-80	5	342097	.062336	.51	.27039	132947.58	35611.87
80-85	5	210953	.095419	.50	.38521	278083.97	12636.07
85+	-	142691	.157564	-	1.00000	0.00	5528.07

8. Sample Variance of the Life Table Death Rate

The life-table death rate is a special case in that the weights $L_x / \sum_x L_x$, as functions of the age-specific death rates, are themselves random variables and are correlated not only with each other, but with the specific death rates as well. Obviously, a derivation of the sample variance of the life-table death rate based on the approach presented in the previous sections will involve a series of complicated and difficult computations.

The derivation can be simplified by making use of the inverse relationship between R, the life-table death rate, and $\hat{e}_0$, the observed expectation of life at birth:

$$R = \frac{\Sigma L_x M_{ux}}{\Sigma L_x} = \frac{\Sigma d_x}{\Sigma L_x} = \frac{\ell_0}{T_0} = \frac{1}{\hat{e}_0} \quad . \tag{8.1}$$

Employing the general rule on the variance of the inverse of a random variable, we have

$$S_R^2 = \frac{1}{\hat{e}_0^4} S_{\hat{e}_0}^2 \quad . \tag{8.2}$$

Here the sample variance of $\hat{e}_0$, which may be found in Chapter 4, is

$$S_{\hat{e}_0}^2 = \sum_{x \geq 0} \hat{p}_{0x}^2 \, [(1-a_x)n_x + \hat{e}_{x+n_x}]^2 \, S_{\hat{q}_x}^2 \, . \tag{8.3}$$

Substituting (8.3) in (8.2) gives the required formula

$$S_{\hat{e}_0}^2 = \frac{1}{\hat{e}_0^4} \sum_{x \geq 0} \hat{p}_{0x}^2 \, [(1-a_x)n_x + \hat{e}_{x+n_x}]^2 \, S_{\hat{q}_x}^2 \, , \tag{8.4}$$

where

$$S_{\hat{q}_x}^2 = \frac{\hat{q}_x^2 \, [1-\hat{q}_x]}{D_x}$$

as given in Section 3 .

CHAPTER 4

THE LIFE TABLE AND ITS CONSTRUCTION

An Historical Note

Long before the development of modern probability and statistics, men were concerned with the length of life and constructed tables to measure longevity. Particular interest has been expressed to the longevity of famous persons or to individuals who were reported to have died at an extreme old age. A crude table, credited to the Roman Praetorian Praefect Ulpianus, was constructed in the middle of the third century A.D., and indicates an expectation of life of thirty years. But since its purpose was to serve as a basis for determining annuity grants, it is unlikely that it reflects mortality in the general population. Nevertheless, it continued in official use in northern Italy until the end of the eighteenth century. John Graunt's Bills of Mortality, published in 1662, and Edmund Halley's famous table for the city of Breslau, published in 1693, mark the beginning of modern life tables. In Bills of Mortality, Graunt introduced the proportion surviving to various ages, while Halley's table already contained most of the columns in use today. Rough calculation of the average length of life from Graunt's data for seventeenth century London gives a figure of 18.2 years, whereas Halley's estimate for Breslau near the end of the century was 33.5 years. During the next hundred years several life tables were constructed, including the French tables of Deparcieux (1746), of Buffon (1749), of Mourgue and Duvillard (both published in the 1790's), the Northampton table of Richard Price (1783), and in the United States Wigglesworth's table for Massachusetts and New Hampshire (1793). The first official English life table was published in 1843 during William Farr's term as Compiler of Abstracts in the General

Records Office. Several countries in Continental Europe have established series of life tables dating back almost two centuries. Sweden, for example, began a series of life tables in 1755, Netherlands in 1816, France in 1817, Norway in 1821, Germany in 1871 and Switzerland in 1876. Reliable mortality statistics for the construction of United States life tables did not become available until 1900; from there J. W. Glover, of the Bureau of the Census, determined that the expectation of life at birth was 46.07 years for males and 49.42 for females.

1. Introduction

The life table is largely a product of actuarial science, but its application is not limited to the computation of insurance premiums. Recent advances in theoretical statistics and stochastic processes have made it possible to study the length of life from a purely statistical point of view, making the life table a valuable analytical tool for demographers, epidemiologists, physicians, and research workers in other areas of public health.

There are two forms of the life table in general: the cohort (or generation) life table and the current life table. In its strictest form, a cohort life table records the actual mortality experience of a particular group of individuals (the cohort) from birth to the death of the last member of the group. The difficulties involved in constructing a cohort life table for a human population are apparent. Statistics covering a period of 100 years are available for only a few populations and even those are likely to be less reliable than current statistics. Individuals in a given cohort may have emigrated or died unrecorded, and the life expectancy of a group of people already dead is of little more than historical interest. However, cohort life tables do have practical applications in studying animal populations and have even been extended to access the durability of inanimate objects such as engines, electric light bulbs, etc. Modified or adapted cohort tables have been useful in epidemiological, sociological, and medical and para-medical studies with human subjects. Extensive use of life table methods has been made in the analysis of chance and duration of patient-survival in studies of treatment effectiveness. These will be discussed in more detail with some examples in Chapter 9.

The current life table, as the name implies, gives a cross-section view of the mortality and survival experience of all ages in a population during one short period of time, for example, the California population of 1970. It is dependent entirely on the age-specific death rates prevailing in the year for which it is constructed. Such tables project the life span of each individual in a hypothetical cohort on the basis of the actual death rates in a given population. When we speak of the life expectancy of an infant born in a current year, for example, we mean the life expectancy that would be obtained if he were subjected throughout his life to the same age-specific mortalities prevailing in the current year. The current life table is then a fictitious pattern reflecting the mortality experience of a real population during a calendar year. However, it is the most effective means of summarizing mortality and survival experience of a population, and is a sound basis for making statistical inference about the population under study. The reader can no doubt confirm from his own experience that the current life table is a standard and useful tool for comparing international mortality data, and for assessing mortality trends on the national level.

A current life table may be based on the deaths occurring over three, instead of one, calendar years; e.g., years 1969, 1970, 1971. For each age group the average number of deaths per year is then divided by the corresponding population size of the middle of the three years (1970, in this example) to obtain the age-specific death rate. Usually, the middle year is a census year, so that population figures are available and more accurate.

The advantage of such a table is to reduce the possible abnormalities in mortality pattern which may exist in a single calendar year.

Data for constructing life tables are sometimes refined by graduation or other methods for smoothing or reducing the effect of extreme values. Techniques for refinement of life table data were developed by acuarial scientists. While refinement of data has its merit in smoothing data, it is difficult to make proper statistical inference of life table functions which is based on such information.

This chapter will describe a general form of the life table with interpretations of its various functions and present a method of constructing a current life table. Theoretical aspects of life table functions will be discussed in detail in Appendix II.

Cohort and current life tables may be either complete or abridged. In a complete life table the functions are computed for each year of life; an abridged life table differs only in that it deals with age intervals greater than one year, except possibly the first year of the first five years of life. A typical set of intervals is 0-1, 1-5, 5-10, 10-15, etc.

2. Description of the Life Table

Cohort and current life tables are identical in appearance but different in construction. The following discussion refers to the complete current life table. The function of each column is defined and its relation to the other columns explained; conventional symbols have been modified for the sake of simplicity. The complete current life table for the California total population in 1970, presented in Table 2, will serve as an example.

Column 1. Age interval, (x,x+1) -- As with the cohort table, each interval in this column is defined by the two exact ages stated except for the final age interval, which is open-ended such as 85 and over. The starting point for the final age interval is denoted by w.

Column 2. Proportion (of those alive at age x) dying in interval (x,x+1), $\hat{q}_x$ -- Each $\hat{q}_x$ is an estimate of the probability that an individual alive at the exact age x will die during the interval. These proportions are the basic quantities from which figures in other columns of the table are computed. They are derived from the corresponding age-specific death rates of the current population, using formulas that will be explained in the next section. To avoid decimals, the proportions are sometimes expressed as the number of deaths per 1,000 population, and the column is headed, "1000 $\hat{q}_x$."

Column 3. Number alive at age x, ℓ_x -- The first number in this column, ℓ_0, is an arbitrary figure called the "radix," while each successive figure represents the number of survivors at the exact age x. Thus the figures in this column have meaning only in conjunction with the radix ℓ_0, and do not describe any observed population. The radix is usually assigned a convenient number, such as 100,000. Table 2 shows that ℓ_2 or 98,088 of every 100,000 persons born alive will survive to the second birthday, provided they are subject to the same mortality experience as that of the 1970 California population.

Column 4. Number dying in interval $(x,x+1)$, d_x -- The figures in this column are the product of ℓ_x and $\hat{q}_x$ and thus also depend upon the radix ℓ_0. Again using the 1970 California experience, we see that out of $\ell_0 = 100,000$ born alive, $d_0 = 1801$ will die in the first year of life. But the number 1801 is meaningless by itself, and is certainly not the number of infant deaths occurring in California in 1970. For each age interval $(x,x+1)$, d_x is merely the number of life table deaths.

The figures in the columns ℓ_x and d_x are computed from the values of $\hat{q}_0, \hat{q}_1, \ldots, \hat{q}_w$ and the radix ℓ_0 by using the relations

$$d_x = \ell_x \hat{q}_x, \qquad x=0,1,\ldots,w, \tag{2.1}$$

and

$$\ell_{x+1} = \ell_x - d_x, \qquad x=0,1,\ldots,w-1 \quad . \tag{2.2}$$

Starting with the first age interval, we use equation (2.1) for $x=0$ to obtain the number d_0 dying in the interval $(0,1)$ and equation (2.2) for $x=0$ to obtain the number ℓ_1 who survive to the end of the interval. With ℓ_1 persons alive at the exact age 1, we again use the relations (2.1) and (2.2) for $x=1$ to obtain the corresponding figures for the second interval. By repeated applications of (2.1) and (2.2) we compute all the figures in columns 3 and 4.

Column 5. Fraction of last year of life for age x, a_x' -- Each of the d_x people who die during the interval $(x,x+1)$ has lived x complete years plus some fraction of the year $(x,x+1)$. The average of these fractions, denoted by a_x', plays an important role in the construction of life tables, and in the theoretical studies of life table functions as presented in Appendix II. This will be explained more fully in the next section.

Column 6. Number of years lived by the total cohort in interval (x,x+1), L_x -- Each member of the cohort who survives the year (x,x+1) contributes one year to L_x, while each member who dies during the year (x,x+1) contributes, on the average, a fraction a_x' of a year, so that

$$L_x = (\ell_x - d_x) + a_x' d_x \qquad x=0,1,\ldots,w-1, \tag{2.3}$$

where the first term on the right side is the number of years lived in the interval (x,x+1) by the $(\ell_x - d_x)$ survivors, and the last term is the number of years lived in (x,x+1) by the d_x persons who died during the interval. When a_x' is assumed to be 1/2 (which is usually the case for ages greater than 5), then

$$L_x = \ell_x - \frac{1}{2} d_x \ . \tag{2.4}$$

The similarity of L_x to the concept of "person years" may be recognized by the reader.

Column 7. Total number of years lived beyond age x, T_x -- This total is essential for computation of the life expectancy. It is equal to the sum of the number of years lived in each age interval beginning with age x, or

$$T_x = L_x + L_{x+1} + \ldots + L_w \ , \qquad x=0,1,\ldots,w \ , \tag{2.5}$$

with an obvious relationship

$$T_x = L_x + T_{x+1} \ . \tag{2.6}$$

Column 8. Expectation of life at age x, $\hat{e}_x$ -- This is number of years, on the average, yet to be lived by a person now aged x. Since the total number of years of life remaining to the ℓ_x individuals is T_x,

$$\hat{e}_x = \frac{T_x}{\ell_x} \ , \qquad x=0,1,\ldots,w. \tag{2.7}$$

Each $\hat{e}_x$ summarizes the mortality experience of persons beyond age x in the current population under consideration, making this column the most important in the life table. Further, this is the only column in the table other than $\hat{q}_x$ and a_x' that is meaningful without reference to the radix ℓ_0. As a rule, the expection of life $\hat{e}_x$ decreases as the age x increases, with the single exception of the first year of life where the reverse is true due to the high mortality during the first year. In the 1970 California population, for example, the expectation of life at birth is $\hat{e}_0 = 71.90$ years whereas at age one $\hat{e}_1 = 72.22$ The symbol $\hat{e}_x$, denoting the observed expectation of life, is computed from the actual mortality data and is an estimate of e_x, the true unknown expectation of life at age x.[1/]

Remark 1: Useful quantities which are not listed in the conventional life table are

$$\hat{p}_x = 1 - \hat{q}_x , \tag{2.8}$$

the proportion of survivors over the age interval (x,x+1), and

$$\hat{p}_{xy} = \hat{p}_x \hat{p}_{x+1} \cdots \hat{p}_{y-1} = \frac{\ell_y}{\ell_x} , \tag{2.9}$$

the proportion of those living at age x who will survive to age y. When x=0, $\hat{p}_{0y}$ becomes the proportion of the total born alive who survive to age y; clearly

$$\hat{p}_{0y} = \ell_y / \ell_0 .$$

3. Construction of the Complete Current Life Table

In the construction of current life tables, we are mainly concerned with the computations of $\hat{q}_x$, the proportion dying in the age interval $(x,x+1)$, and L_x, the number of years lived by the radix ℓ_0 in the interval $(x,x+1)$.

An important element in complete life table construction as described in this section is the fraction of the last year of life lived by those who die at each age; for example, a man who dies at age 30 has lived 30 complete years plus a fraction of the 31st year. The average value of this fraction is denoted by a'_x where x refers to the age at the last birthday. It might reasonably be expected that the average value of this fraction is equal to one half on the assumption that there are as many deaths at 30 years plus one month as at 30 years plus two months, and at each month thereafter through the 11th; or, in other words, on the assumption that deaths occur uniformly throughout each year of age. Extensive studies of the fraction have been made using the 1960 California mortality data (Chiang, et al [1961]) collected by the State of California Department of Public Health and the 1963 U.S. data collected by the National Vital Statistics Division of the National Center for Health Statistics, respectively. The results obtained so far show that from age 5 on the fractions a'_x are invariant with respect to race, sex, and age, and that the assumed value of .5 is then valid. But a much smaller value has been observed for the first year because of the large proportion of infant deaths occurring in the first weeks of life. A brief description of the analysis regarding a'_x using the California data is given in Section 5.

To return to the computation of $\hat{q}_x$, readers familiar with vital statistics terminology will recognize the resemblance between $\hat{q}_x$ and M_x, the age specific death rate. The essential step in constructing a complete life table for a current population is to establish a relationship between $\hat{q}_x$ and M_x so that the probability of dying q_x can be computed from the death rate M_x for each age x. These two quantities can both be expressed in terms of the number of observed deaths of age x (D_x) that occur during the calendar year and the corresponding midyear (calendar year) population (P_x). Let N_x be the number of people alive at the exact age x, among whom D_x deaths occur in (x,x+1). Then, by definition, the proportion died is given by

$$\hat{q}_x = \frac{D_x}{N_x} \quad . \tag{3.1}$$

The age specific death rate, M_x, is the ratio of the number of deaths (D_x) to the total number of years lived by the N_x people during the interval (x,x+1). This total number is composed of (N_x-D_x) years lived by the survivors and the number of years by those dying during the year. Let a_x' be the fraction of the year (x,x+1) lived by a person who dies during the year; then D_x people as a group will live $a_x'D_x$ years. Hence the total number of years lived in (x,x+1) is $(N_x-D_x) + a_x'D_x$ and the formula

$$M_x = \frac{D_x}{(N_x-D_x) + a_x'D_x} \quad . \tag{3.2}$$

When the denominator is estimated with the corresponding mid-year population P_x,

$$(N_x-D_x) + a_x'D_x = P_x \ , \tag{3.3}$$

we have the familiar formula

$$M_x = \frac{D_x}{P_x} \quad . \tag{3.4}$$

Now, N_x, which was introduced to establish a relationship between $\hat{q}_x$ and M_x, is nevertheless an unknown quantity. By eliminating N_x from (3.1) and (3.2) and using (3.4), we obtain the desired relationship. Formally, we derive N_x from (3.3):

$$N_x = P_x + D_x - a'_x D_x \quad , \tag{3.5}$$

or

$$N_x = P_x + (1-a'_x)D_x \quad ,$$

and substituting (3.5) in (3.1) to obtain

$$\hat{q}_x = \frac{D_x}{P_x + (1-a'_x)D_x} \quad . \tag{3.6}$$

Since the age-specific death rate is usually available, we may divide both the numerator and denominator of (3.6) by P_x to obtain the basic formula

$$\hat{q}_x = \frac{M_x}{1 + (1-a'_x)M_x} \quad . \tag{3.7}$$

As it was noted earlier, the fraction a'_x is subject to little variation. The California data suggest values: $a'_0 = .09$, $a'_1 = .43$, $a'_2 = .45$, $a'_3 = .47$, $a'_4 = .49$, $a'_x = .50$ for $x \geq 5$. Formula (3.7) is fundamental in the construction of complete life tables by the present method and was suggested in Chiang [1960b], [1961].

To illustrate, let us consider the 1970 California population as shown in Table 1. For the first year of life we have $P_0 = 340{,}483$ in Column 2 and $D_0 = 6{,}234$ in Column 3. Thus, the age-specific rate for x=0 is

$$M_0 = \frac{D_0}{P_0} = \frac{6,234}{340,483} = .018309 \qquad \text{or } 18.309 \text{ per } 1000.$$

The average fraction of the year lived by an infant who dies in his first year of life is $a_0' = .09$. Therefore, the estimate of the probability of dying is computed from (3.1):

$$\hat{q}_0 = \frac{.018309}{1 + (1 - .09)\ .018309} = .01801\ .$$

When all the values of $\hat{q}_x$ have been computed and ℓ_0 has been selected, d_x and ℓ_x for successive values of x are determined from equations (2.1) and (2.2) as shown in Table 2. For the 1970 California population we determine first the number of life table infant deaths with $\ell_0 = 100,000$,

$$d_0 = \ell_0 \hat{q}_0 = 100,000 \times .01801 = 1801\ ,$$

and the life table survivors at age one,

$$\ell_1 = \ell_0 - d_0 = 100,000 - 1801 = 98199\ .$$

The formula for the number L_x of years lived in the age interval (x,x+1) is derived also with the aid of a_x', the fraction of the last year of life as given in Section 2:

$$L_x = (\ell_x - d_x) + a_x' d_x\ , \qquad x=0,1,\ldots\ . \tag{2.3}$$

To take again the example of the first year of life, $a_0' = .09$ and

$$L_0 = 98199 + .09 \times 1801 = 98361\ .$$

Remark 3: The ratio d_x/L_x is shown as the life-table death rate for age x. Since a life table is entirely based on the age-specific death rates of a current population, the death rates computed from the life table should be identical to the corresponding rates of the current population; symbolically

$$\frac{d_x}{L_x} = M_x = \frac{D_x}{P_x} \qquad x=0,1,\ldots \quad . \tag{3.8}$$

To prove equation (3.8), we substitute (2.3) in the left side of (3.8) and divide the resulting expression by ℓ_x to obtain

$$\frac{d_x}{L_x} = \frac{\hat{q}_x}{1-(1-a_x{}')\hat{q}_x} \tag{3.9}$$

Substituting (3.6) for $\hat{q}_x$ in (3.9) and simplifying the resulting expression give D_x/P_x, proving the assertion (3.8).

The final age interval in a life table is a half-open interval, such as age 85 and over. The values of D_w, P_w, M_w, ℓ_w, d_w, and T_w all refer to the open interval age w and over, and $\hat{q}_w = 1$ (since there can be no survivors). The length of the interval is infinite and the necessary information for determining the average number of years lived by an individual beyond age w is unavailable. We must therefore use an approach other than equation (2.3) to determine L_w. Writing the first equation in (3.8) for x=w, we have

$$L_w = \frac{d_w}{M_w} \quad . \tag{3.10}$$

Since each one of the ℓ_w people alive at w will eventually die, $\ell_w = d_w$, and from (3.10) we have the required formula

$$L_w = \frac{\ell_w}{M_w} \quad , \tag{3.11}$$

where ℓ_w, survivors to age w, is computed from the preceding interval and M_w is the mortality rate for age interval w and over. The quantities T_w and $\hat{e}_w$ can be computed as follows:

$$T_w = L_w \qquad \text{and} \qquad \hat{e}_w = \frac{T_w}{\ell_w} = \frac{L_w}{d_w} = \frac{1}{M_w} \quad . \tag{3.12}$$

In the 1970 California life table w = 85, and $\ell_{85} = 23274$. The death rate for age 85 and over is $M_{85} = .157564$; therefore

$$L_{85} = \frac{\ell_{85}}{M_{85}} = \frac{23274}{.157564} = 147711$$

and

$$T_{85} = 147711 \quad \text{and} \quad \hat{e}_{85} = 6.35 \quad .$$

Figures 1 to 4 show graphically the probability of dying ($\hat{q}_x$), the number of survivors (ℓ_x), the number of deaths (d_x), and the expectation of life ($\hat{e}_x$), for each age x for the total California population, 1970. Figures 5 to 8 show the corresponding four sets of quantities for the total United States population, 1970. As we see from Figure 1 that the probability of dying is extremely high for the first year of life. It decreases sharply after the first year and reaches a minimum at the age of 10 years. From there, the probability rises gradually and reaches the same magnitude of q_0 around age 65 and it continues to increase monotonically and later drastically. The pattern of $\hat{q}_x$ is also reflected in the ℓ_x, d_x and $\hat{e}_x$. Since the life tables for the California population and for the United States population end at age 85, the graphics also stop at that age.

In this manual we have introduced two sets of terms, one for the theoretical quantities and the other for the estimates of the theoretical quantities. The theoretical quantities include the probability of dying q_x, the expectation of life e_x for each age x, and many others; the corresponding estimates are $\hat{q}_x$ and $\hat{e}_x$. For simplicity of reading and when there would be no confusion, we shall drop the words "estimate of" and refer to $\hat{q}_x$ as the probability of dying and $\hat{e}_x$ as the expectation of life, etc.

Table 1.

Construction of Complete Life Table for Total California Population, USA, 1970

Age interval (in years)	Midyear population in interval (x,x+1)	Number of deaths in interval (x,x+1)	Death rate in interval (x,x+1)	Fraction of last year of life	Probability of dying in interval (x,x+1)
x to x+1	P_x	D_x	M_x	a'_x	$\hat{q}_x$
(1)	(2)	(3)	(4)	(5)	(6)
0- 1	340483	6234	.018309	.09	.01801
1- 2	326154	368	.001128	.43	.00113
2- 3	313699	269	.000858	.45	.00086
3- 4	323441	237	.000733	.47	.00073
4- 5	338904	175	.000516	.49	.00052
5- 6	362161	179	.000494	.50	.00049
6- 7	379642	171	.000450	.50	.00045
7- 8	386980	131	.000339	.50	.00034
8- 9	391610	121	.000309	.50	.00031
9-10	397724	121	.000304	.50	.00030
10-11	406118	126	.000310	.50	.00031
11-12	388927	127	.000327	.50	.00033
12-13	395025	138	.000349	.50	.00035
13-14	388526	158	.000407	.50	.00041
14-15	385085	186	.000483	.50	.00048
15-16	377127	235	.000623	.50	.00062
16-17	368156	344	.000934	.50	.00093
17-18	366198	385	.001051	.50	.00105
18-19	354932	506	.001426	.50	.00142
19-20	350966	584	.001664	.50	.00166
20-21	359833	583	.001620	.50	.00162
21-22	349557	562	.001608	.50	.00161
22-23	365839	572	.001564	.50	.00156
23-24	370548	564	.001522	.50	.00152
24-25	295189	421	.001426	.50	.00143
25-26	304013	416	.001368	.50	.00137
26-27	305558	391	.001280	.50	.00128
27-28	310554	461	.001484	.50	.00148
28-29	275897	411	.001490	.50	.00149
29-30	261592	392	.001499	.50	.00150

Table 1. (continued)

Construction of Complete Life Table for Total California Population, USA, 1970

x to x+1	P_x	D_x	M_x	a'_x	q_x
(1)	(2)	(3)	(4)	(5)	(6)
30-31	264083	399	.001511	.50	.00151
31-32	247777	378	.001526	.50	.00152
32-33	241726	388	.001605	.50	.00160
33-34	232025	365	.001573	.50	.00157
34-35	233778	434	.001856	.50	.00185
35-36	234338	439	.001873	.50	.00187
36-37	224302	475	.002118	.50	.00212
37-38	228652	519	.002270	.50	.00227
38-39	226727	549	.002421	.50	.00242
39-40	235980	606	.002568	.50	.00256
40-41	249027	665	.002670	.50	.00267
41-42	232893	719	.003087	.50	.00308
42-43	239747	863	.003600	.50	.00359
43-44	238783	874	.003660	.50	.00365
44-45	248100	993	.004002	.50	.00399
45-46	253828	1140	.004491	.50	.00448
46-47	249857	1268	.005075	.50	.00506
47-48	247955	1362	.005493	.50	.00548
48-49	252137	1422	.005640	.50	.00562
49-50	242126	1530	.006319	.50	.00630
50-51	243799	1594	.006538	.50	.00652
51-52	220599	1710	.007752	.50	.00772
52-53	213448	1793	.008400	.50	.00837
53-54	203618	1870	.009184	.50	.00914
54-55	202388	1981	.009788	.50	.00974
55-56	201750	2217	.010989	.50	.01093
56-57	193828	2333	.012036	.50	.01196
57-58	187257	2483	.013260	.50	.01317
58-59	178602	2392	.013393	.50	.01330
59-60	171807	2517	.014650	.50	.01454
60-61	174613	2733	.015652	.50	.01553
61-62	157734	2743	.017390	.50	.01724
62-63	154174	2911	.018881	.50	.01870
63-64	144149	2968	.020590	.50	.02038
64-65	140100	2954	.021085	.50	.02086
65-66	135857	3391	.024960	.50	.02465
66-67	129386	3278	.025335	.50	.02502
67-68	123925	3352	.027049	.50	.02669
68-69	112574	3331	.029589	.50	.02916
69-70	119063	3736	.031378	.50	.03089

Table 1. (continued)

Construction of Complete Life Table for Total California Population, USA, 1970

x to x+1	P_x	D_x	M_x	a'_x	$\hat{q}_x$
(1)	(2)	(3)	(4)	(5)	(6)
70-71	114066	3846	.033717	.50	.03316
71-72	100781	3704	.036753	.50	.03609
72-73	93031	3706	.039836	.50	.03906
73-74	89992	3830	.042559	.50	.04167
74-75	86561	4063	.046938	.50	.04586
75-76	81003	4275	.052776	.50	.05142
76-77	73552	4383	.059590	.50	.05787
77-78	70516	4259	.060398	.50	.05863
78-79	60616	4181	.068975	.50	.06668
79-80	56410	4227	.074934	.50	.07223
80-81	57646	4424	.076744	.50	.07391
81-82	48299	4288	.088780	.50	.08501
82-83	39560	3995	.100986	.50	.09613
83-84	34439	3753	.108975	.50	.10334
84-85	31009	3669	.118320	.50	.11171
85+	142691	22483	.157564		1.00000

Table 2.

Complete Life Table for Total California Population, USA, 1970

Age interval (in years)	Probability of dying in intervals (x,x+1)	Number living at age x	Number dying in interval (x,x+1)	Fraction of last year of life	Number of years lived in interval (x,x+1)	Total number of years lived beyond age x	Observed Expectation of life at age x
x to x+1	$\hat{q}_x$	ℓ_x	d_x	a'_x	L_x	T_x	$\hat{e}_x$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0- 1	.01801	100000	1801	.09	98361	7190390	71.90
1- 2	.00113	98199	111	.43	98136	7092029	72.22
2- 3	.00086	98088	84	.45	98042	6993893	71.30
3- 4	.00073	98004	72	.47	97966	6895851	70.36
4- 5	.00052	97932	51	.49	97906	6797885	69.41
5- 6	.00049	97881	48	.50	97857	6699979	68.45
6- 7	.00045	97833	44	.50	97811	6602122	67.48
7- 8	.00034	97789	33	.50	97772	6504311	66.51
8- 9	.00031	97756	30	.50	97741	6406539	65.54
9-10	.00030	97726	29	.50	97711	6308798	64.56
10-11	.00031	97697	30	.50	97682	6211087	63.58
11-12	.00033	97667	32	.50	97651	6113405	62.59
12-13	.00035	97635	34	.50	97618	6015754	61.61
13-14	.00041	97601	40	.50	97581	5918136	60.64
14-15	.00048	97561	47	.50	97538	5820555	59.66
15-16	.00062	97514	60	.50	97484	5723017	58.69
16-17	.00093	97454	91	.50	97408	5625533	57.73
17-18	.00105	97363	102	.50	97312	5528125	56.78
18-19	.00142	97261	138	.50	97192	5430813	55.84
19-20	.00166	97123	161	.50	97043	5333621	54.92
20-21	.00162	96962	157	.50	96884	5236578	54.01
21-22	.00161	96805	156	.50	96727	5139694	53.09
22-23	.00156	96649	151	.50	96574	5042967	52.18
23-24	.00152	96498	147	.50	96424	4946393	51.26
24-25	.00143	96351	138	.50	96282	4849969	50.34
25-26	.00137	96213	132	.50	96147	4753687	49.41
26-27	.00128	96081	123	.50	96020	4657540	48.48
27-28	.00148	95958	142	.50	95887	4561520	47.54
28-29	.00149	95816	143	.50	95745	4465633	46.61
29-30	.00150	95673	144	.50	95601	4369888	45.68

Table 2. (continued)

Complete Life Table for Total California Population, USA, 1970

Age interval (in years)	Probability of dying in interval (x,x+1)	Number living at age x	Number dying in interval (x,x+1)	Fraction of last year of life	Number of years lived in interval (x,x+1)	Total number of years lived beyond age x	Observed Expectation of life at age x
x to x+1	q_x	x	d_x	a'_x	L_x	T_x	$\acute{e}_x$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
30-31	.00151	95529	144	.50	95457	4274287	44.74
31-32	.00152	95385	145	.50	95312	4178830	43.81
32-33	.00160	95240	152	.50	95164	408318	42.88
33-34	.00157	95088	149	.50	95014	3988354	41.94
34-35	.00185	94939	176	.50	94851	3893340	41.01
35-36	.00187	94763	177	.50	94674	3798489	40.08
36-37	.00212	94586	201	.50	94486	3703815	39.16
37-38	.00227	94385	214	.50	94278	3609329	38.24
38-39	.00242	94171	228	.50	94057	3515051	37.33
39-40	.00256	93943	240	.50	93823	3420994	36.42
40-41	.00267	93703	250	.50	93578	3327171	35.51
41-42	.00308	93453	288	.50	93309	3233593	34.60
42-43	.00359	93165	334	.50	92998	3140284	33.71
43-44	.00365	92831	339	.50	92661	3047286	32.83
44-45	.00399	92492	369	.50	92307	2954625	31.94
45-46	.00448	92123	413	.50	91916	2862318	31.07
46-47	.00506	91710	464	.50	91478	2770402	30.21
47-48	.00548	91246	500	.50	90996	2678924	29.36
48-49	.00562	90746	510	.50	90491	2587928	28.52
49-50	.00630	90236	568	.50	89952	2497437	27.68
50-51	.00652	89668	585	.50	89376	2407485	26.85
51-52	.00772	89083	688	.50	88739	2318109	26.02
52-53	.00837	88395	740	.50	88025	2229370	25.22
53-54	.00914	87655	801	.50	87255	2141345	24.43
54-55	.00974	86854	846	.50	86431	2054090	23.65
55-56	.01093	86008	940	.50	85538	1967659	22.88
56-57	.01196	85068	1017	.50	84559	1882121	22.12
57-58	.01317	84051	1107	.50	83497	1797562	21.39
58-59	.01330	82944	1103	.50	82393	1714065	20.67
59-60	.01454	81841	1190	.50	81246	1631672	19.94

Table 2. (continued)

Complete Life Table for Total California Population, USA, 1970

Age interval (in years)	Probability of dying in interval (x,x+1)	Number living at age x	Number dying in interval (x,x+1)	Fraction of last year of life	Number of years lived in interval (x,x+1)	Total number of years lived beyond age x	Observed Expectation of life at age x
x to x+1	q_x	x	d_x	a'_x	L_x	T_x	$\hat{e}_x$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
60-61	.01553	80651	1253	.50	80025	1550426	19.22
61-62	.01724	79398	1369	.50	78713	1470401	18.52
62-63	.01870	78029	1459	.50	77299	1391688	17.84
63-64	.02038	76570	1560	.50	75790	1314389	17.17
64-65	.02086	75010	1565	.50	74228	1238599	16.51
65-66	.02465	73445	1810	.50	72540	1164371	15.85
66-67	.02502	71635	1792	.50	70739	1091831	15.24
67-68	.02669	69843	1864	.50	68911	1021092	14.62
68-69	.02916	67979	1982	.50	66988	952181	14.01
69-70	.03089	65997	2039	.50	64978	885193	13.41
70-71	.03316	63958	2121	.50	62897	820215	12.82
71-72	.03609	61837	2232	.50	60721	757318	12.25
72-73	.03906	59605	2328	.50	58441	696597	11.69
73-74	.04167	57277	2387	.50	56083	638156	11.14
74-75	.04586	54890	2517	.50	53632	582073	10.60
75-76	.05142	52373	2693	.50	51026	528441	10.09
76-77	.05787	49680	2875	.50	48243	477415	9.61
77-78	.05863	46805	2744	.50	45433	429172	9.17
78-79	.06668	44061	2938	.50	42592	383739	8.71
79-80	.07223	41123	2970	.50	39638	341147	8.30
80-81	.07391	38153	2820	.50	36743	301509	7.90
81-82	.08501	35333	3004	.50	33831	264766	7.49
82-83	.09613	32329	3108	.50	30775	230935	7.14
83-84	.10334	29221	3020	.50	27711	200160	6.85
84-85	.11171	26201	2927	.50	24738	172449	6.58
85+	1.00000	23274	23274		147711	147711	6.35

FIGURE I. PROBABILITY OF DYING TOTAL CALIFORNIA POPULATION, 1970

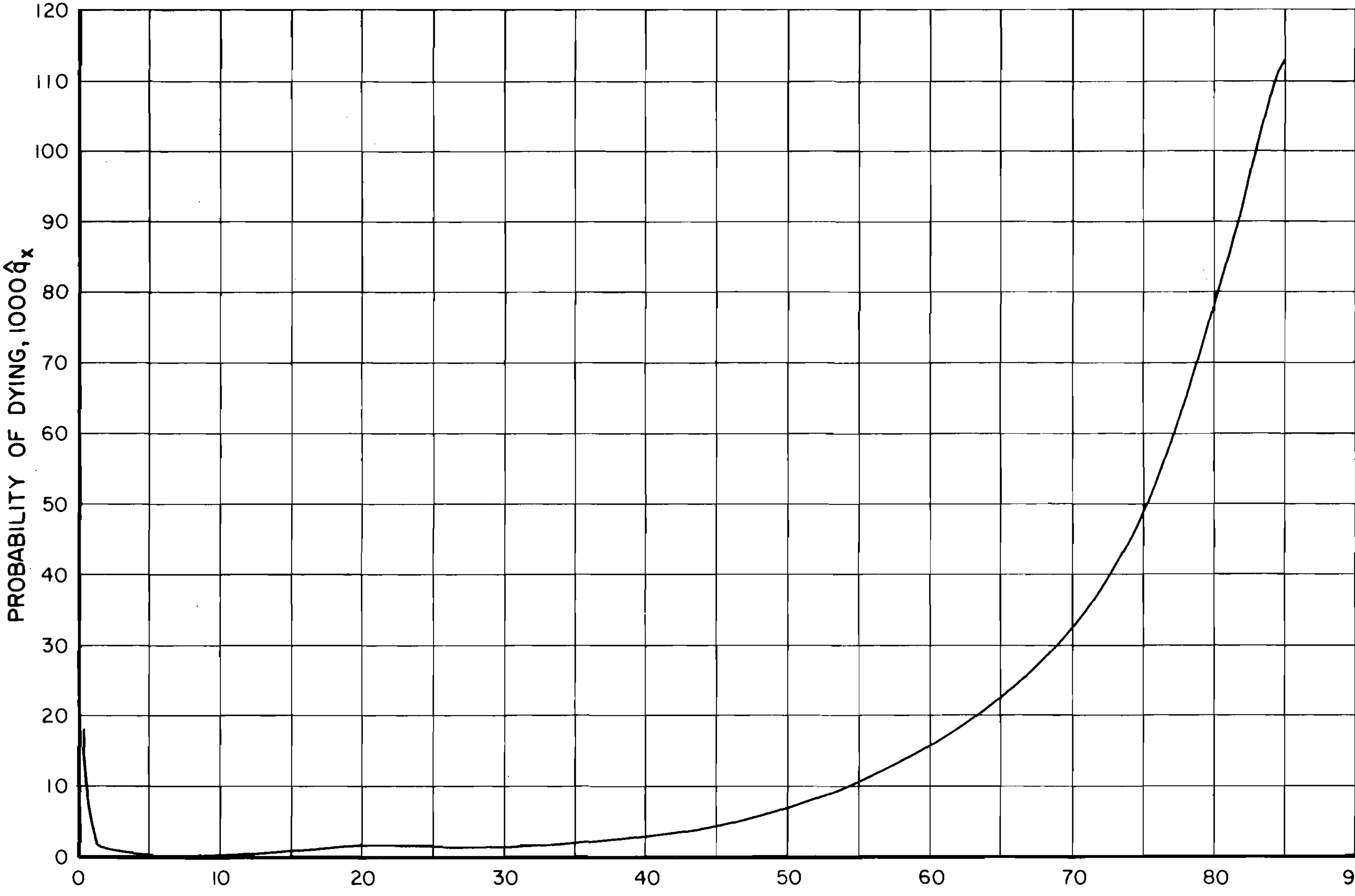


FIGURE 2. NUMBER OF SURVIVORS OUT OF 100,000 LIVE BIRTHS
TOTAL CALIFORNIA POPULATION, 1970

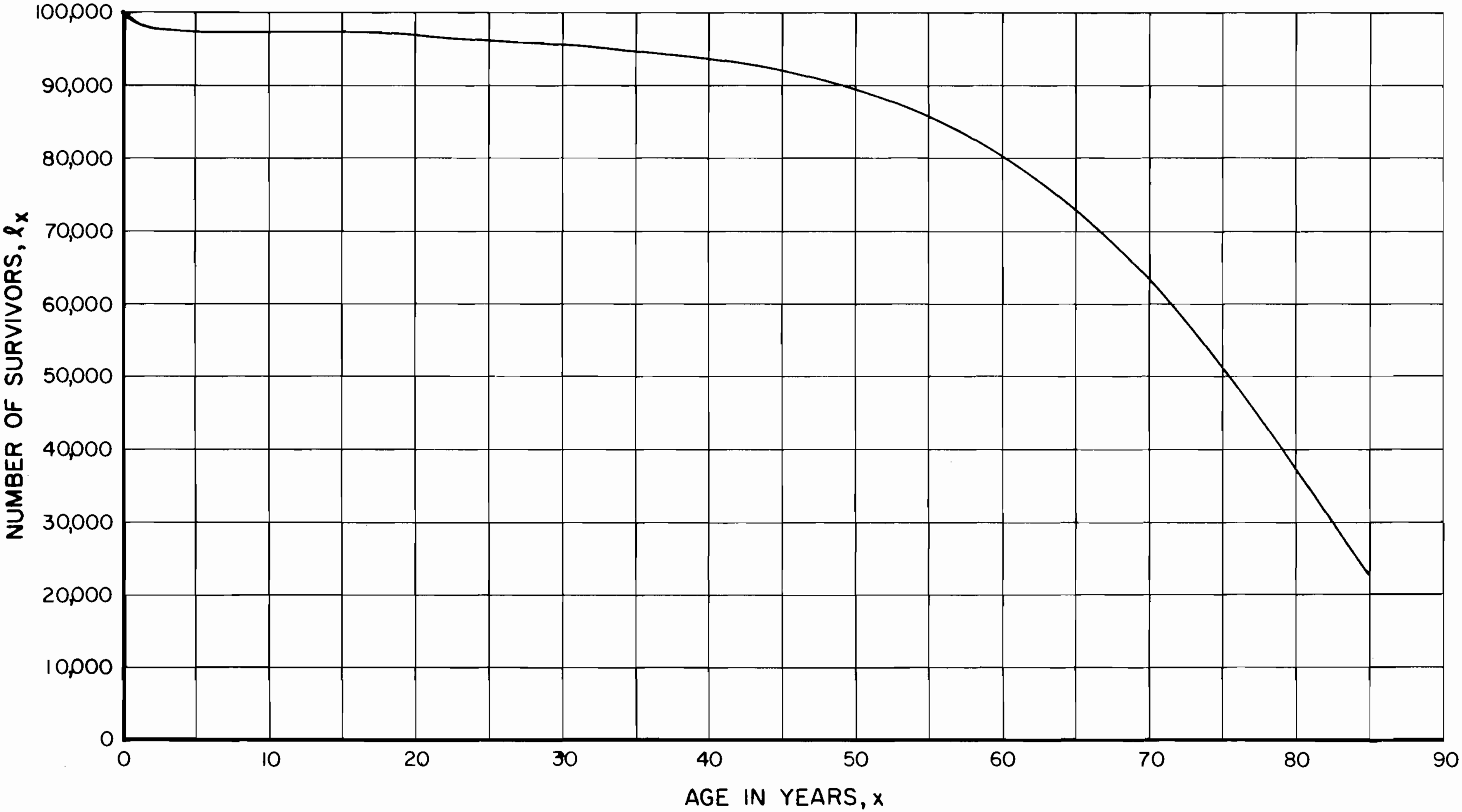


FIGURE 3. NUMBER OF DEATHS OUT OF 100,000 LIVE BIRTHS
TOTAL CALIFORNIA POPULATION, 1970

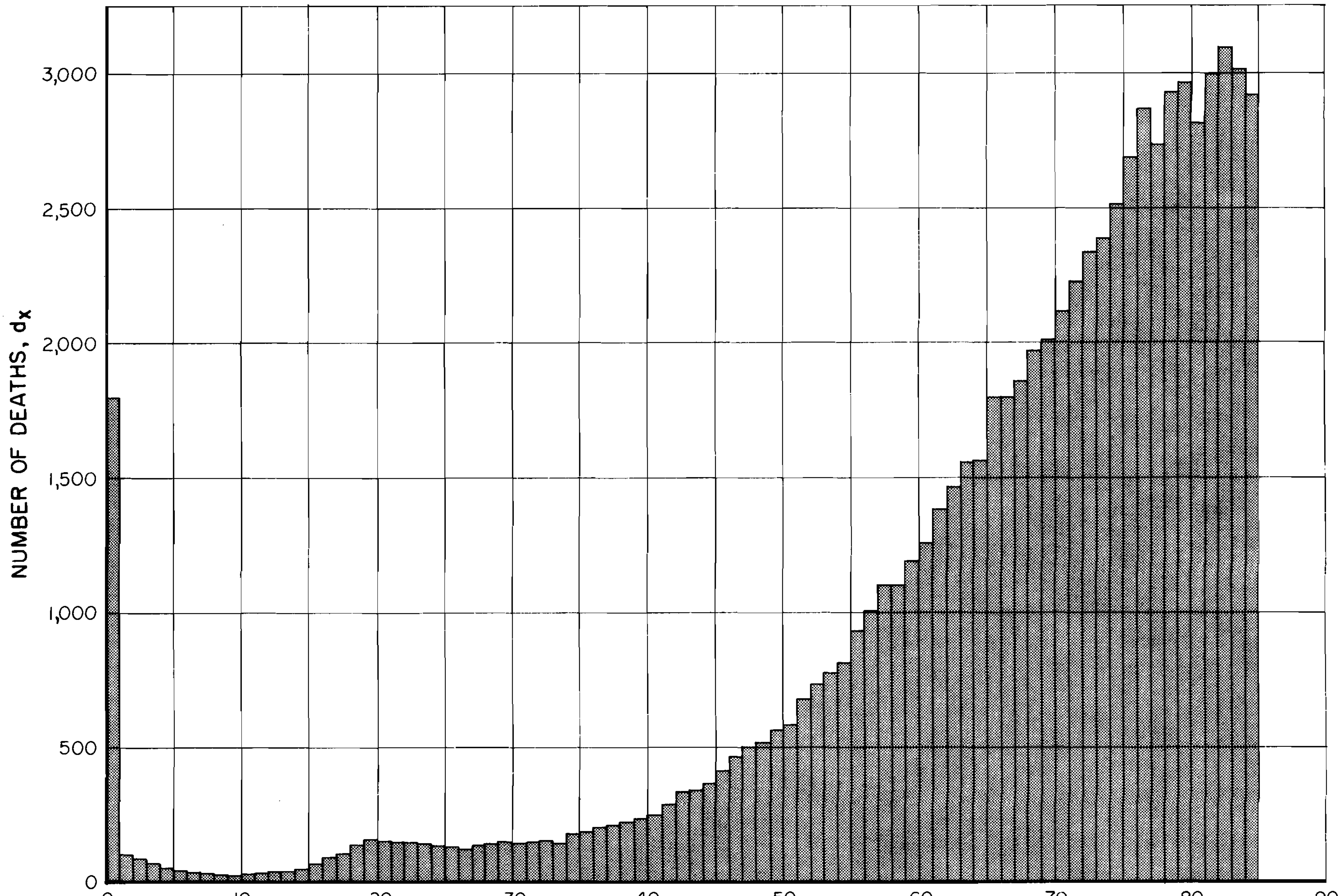


FIGURE 4. EXPECTATION OF LIFE
TOTAL CALIFORNIA POPULATION, 1970

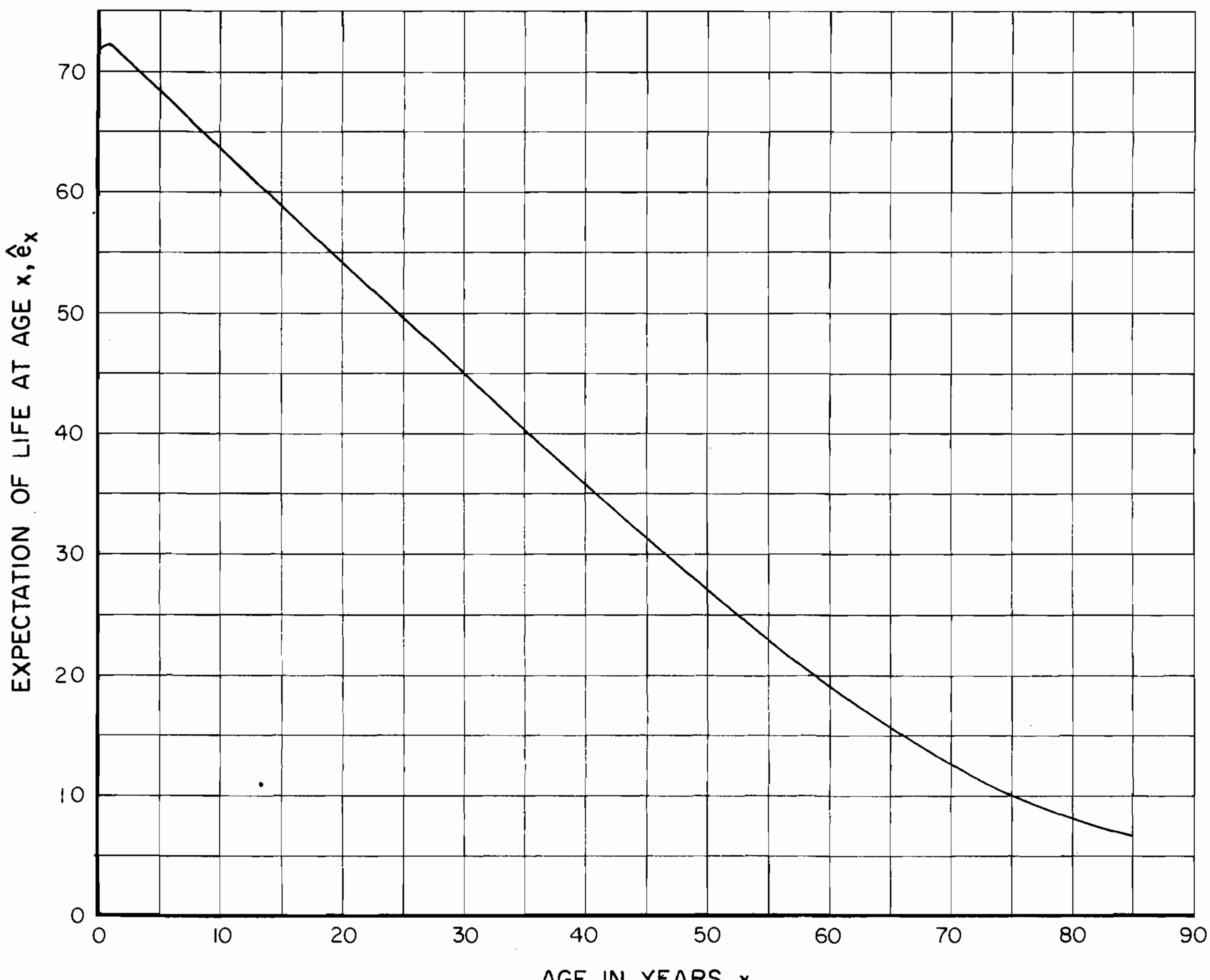


FIGURE 5. PROBABILITY OF DYING
TOTAL UNITED STATES POPULATION, 1970

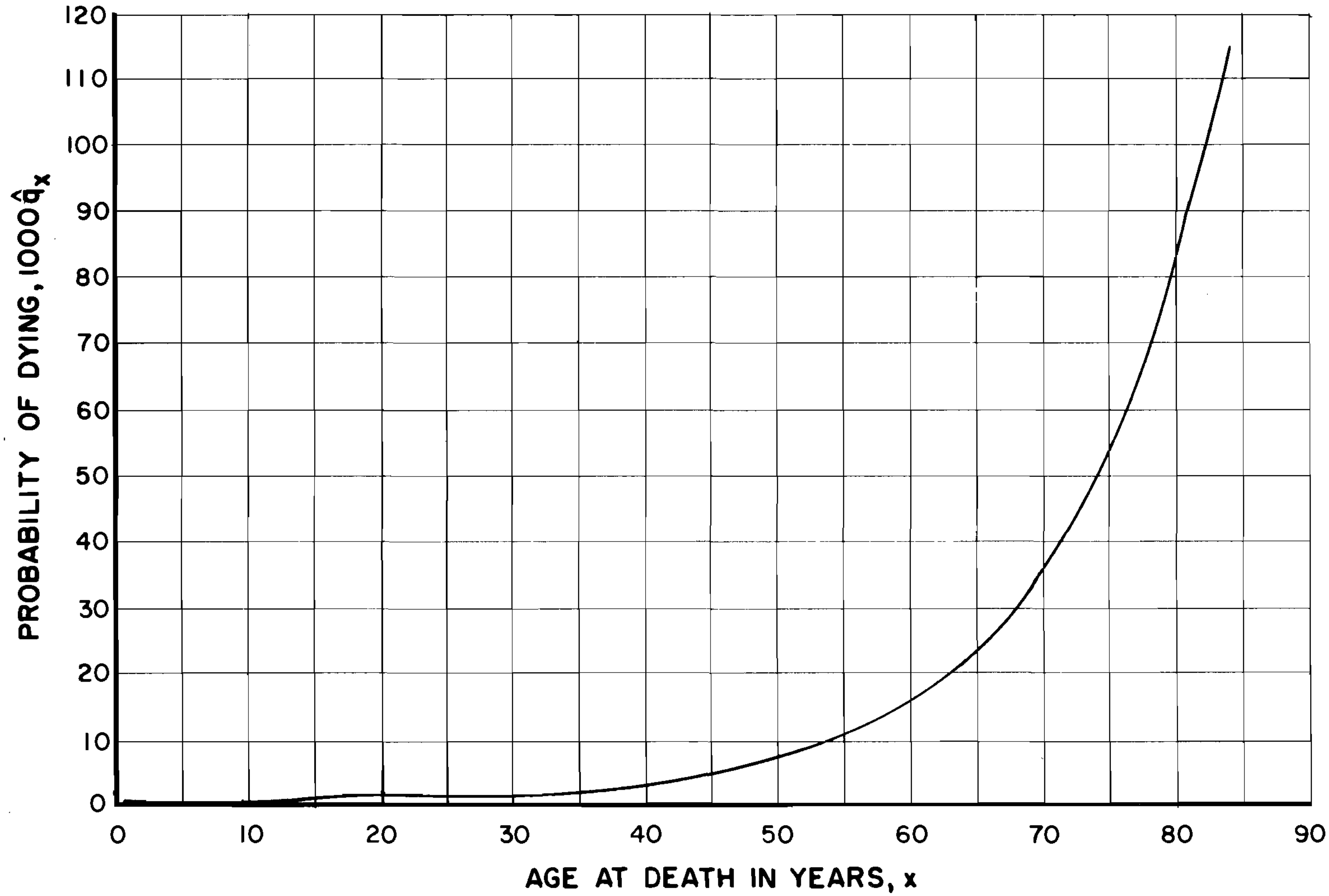


FIGURE 6. NUMBER OF SURVIVORS OUT OF 100,000 LIVE BIRTHS
TOTAL UNITED STATES POPULATION, 1970

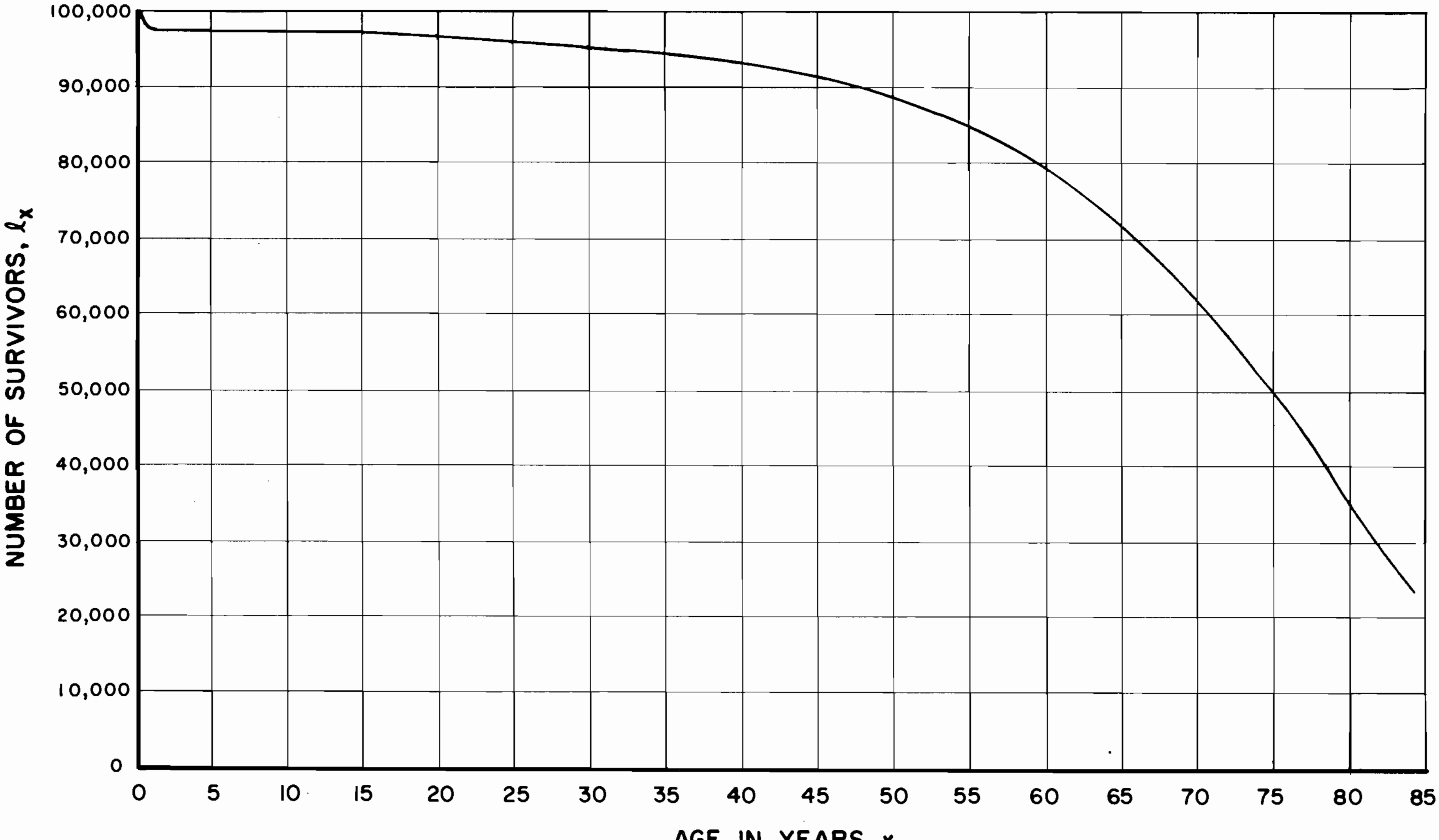


FIGURE 7. NUMBER OF DEATHS OUT OF 100,000 LIVE BIRTHS

TOTAL UNITED STATES POPULATION, 1970

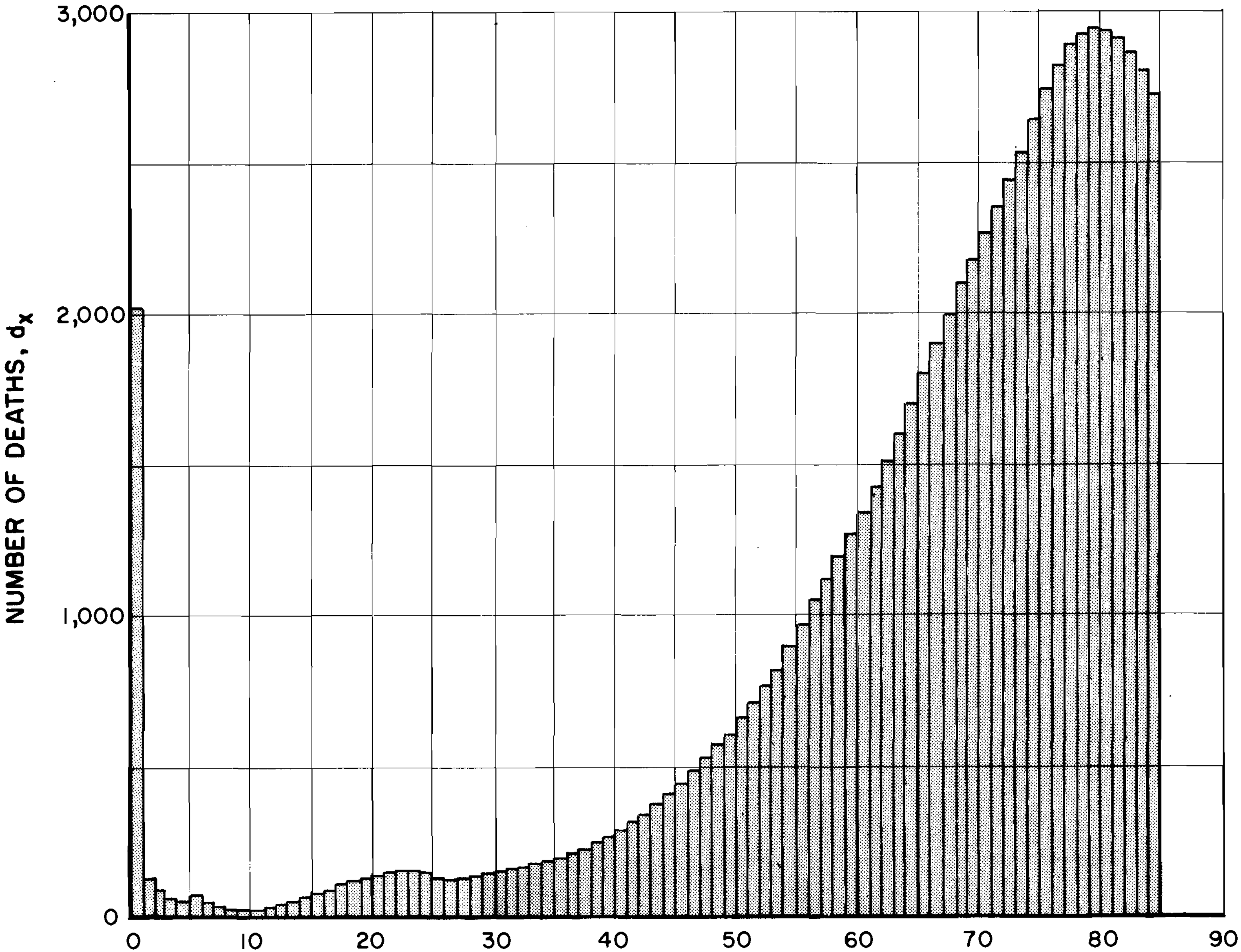
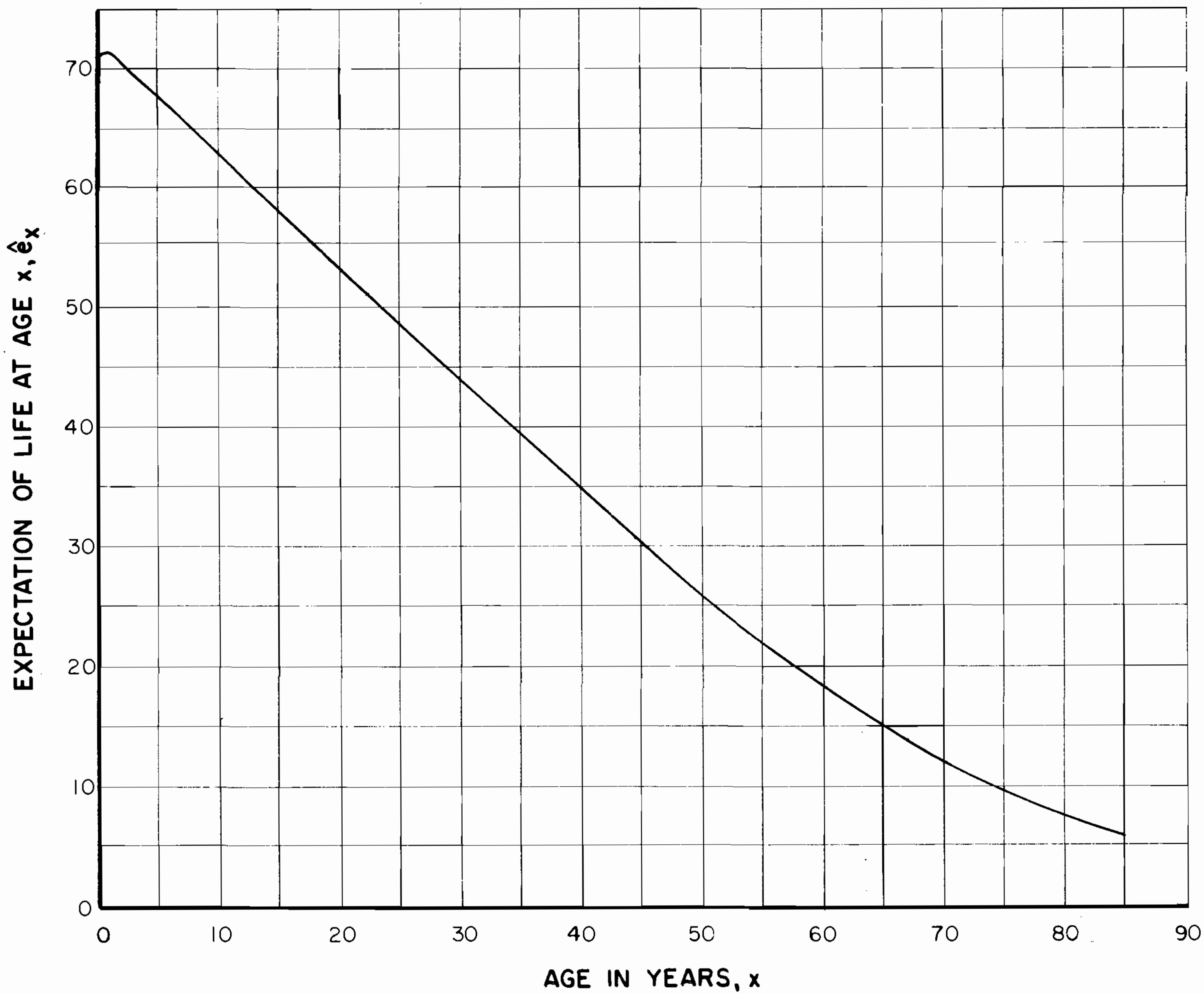


FIGURE 8. EXPECTATION OF LIFE

TOTAL UNITED STATES POPULATION, 1970



CHAPTER 5

THE LIFE TABLE AND ITS CONSTRUCTION - ABRIDGED LIFE TABLES

1. Introduction

Clearly, the current life table furnishes information not obtainable from other sources. It provides the public health worker, demographers and other research workers with tools for making international comparisons as well as for comparing contemporary groups within a country or for assessing trends within a given population. The life table death rate has the advantage over other mortality indices of being independent of age and sex distributions. This, of course, is also true for $\hat{e}_0$, the average length of life, or for $\hat{e}_x$, the average remaining lifetime at any age x. The ratio ℓ_k/ℓ_j gives a convenient measurement for comparing the survival of selected age segments of two populations; for example, one might want to know if Swedish women who survive to age 20 have as good a chance of surviving to age 45 as do their Italian counterparts by comparing the ratios ℓ_{45}/ℓ_{20}.

Life table estimates have the disadvantage of any statistics based on the population census and vital records. Individuals or entire households may be missed by the census taker or overlooked by the informant. Cross-checking with birth and death certificates show that young children, even when they survive infancy, are sometimes forgotten; migratory segments of the population (particularly young males) are subject to marked under-enumeration. Misstatements of age are clearly discernible in bar graphs of the age distribution particularly the overstatement of the ages of young children (followed by an understatement in the middle years), of persons approaching retirement age, and of the very old; in addition, a heaping is found for ages in multiples of five and at even ages.

Completeness of birth registration varies from country to country and must occasionally be checked. Death registration can be improved by the requirement that it be filed before a burial permit is issued. These defects in mortality data and population census have a marked effect on the complete life table.

There are three other disadvantages of complete life tables that are more closely related to the tables themselves. (1) The data necessary for intervals of one year of age is frequently not available; (2) Computations are tedious and time-consuming when computer services are not available; (3) A table consisting of 85 or 95 age groups does not present a concise picture of the mortality experience of a population.

These objections can be obviated by constructing an abridged rather than the complete life table. The computations are discussed in the following paragraphs.

2. A Method of Life Table Construction

An abridged life table contains columns similar to those described for the complete life table. The limits of age intervals are denoted by x_i, $i=0,1,\ldots,w$, and the length of the interval by n_i so that $x_{i+1}-x_i = n_i$. Thus we have

Column 1. Age interval (x_i,x_{i+1})

Column 2. Proportion dying in interval (x_i,x_{i+1}), $\hat{q}_i$.

Column 3. Number alive at age x_i, ℓ_i.

Column 4. Number dying in interval (x_i,x_{i+1}), d_i.

Column 5. The average fraction of interval (x_i,x_{i+1}) lived by an individual dying at an age included in the interval a_i.

Column 6. Total number of years lived in interval (x_i,x_{i+1}), L_i.

Column 7. Total number of years lived beyond age x_i, T_i.

Column 8. Observed expectation of life age at x_i, $\hat{e}_i$.

The present method of construcing the abridged life table was proposed by Chiang [1960b], [1961] and was used for the 1959-61 California abridged Life Tables [Norris]. The idea and procedure involved are the same as those used in the construction of the complete life table described in Chapter 4 with differences due only to the length of intervals. The length of the typical interval (x_i, x_{i+1}) in the abridged table is $n_i = x_{i+1} - x_i$, which is greater than one year (commonly, $n_i = 5$ years, see Table 2). The essential element here is the average fraction of the interval lived by each person who dies at an age included in the interval. This fraction, called the fraction of last age interval of life, denoted by a_i, is conceptually a logical extension of the fraction of the last year of life, a'_x. Determination and discussion of a_i will be presented in Section 4. We use a_i as the point of departure.

Starting with the values of a_i we can construct the abridged life table by following the steps in Chapter 4. Because of its importance, however, we repeat the previous argument to derive the formula for $\hat{q}_i$, the estimate of the probability that an individual alive at age x_i will die in the interval (x_i, x_{i+1}). Let D_i be the number of deaths occurring in the age interval (x_i, x_{i+1}) during the calendar year under consideration, M_i the corresponding age-specific death rate. To derive a relationship between $\hat{q}_i$ and M_i, we introduce N_i, the number of individuals alive at exact age x_i, such that among the N_i persons D_i will die in the interval. Then by definition the proportion dying in (x_i, x_{i+1}) is given by

$$\hat{q}_i = \frac{D_i}{N_i} \quad . \tag{2.1}$$

The age specific death rate M_i is the ratio of D_i to the total number of years lived by the N_i individuals during the interval (x_i, x_{i+1}), or

$$M_i = \frac{D_i}{(N_i - D_i)n_i + a_i n_i D_i} \quad . \tag{2.2}$$

The first term in the denominator of (2.2) is the number of years lived by the $(N_i - D_i)$ survivors, while the second term is the number of years lived by those who die in (x_i, x_{i+1}). Eliminating N_i from (2.1) and (2.2) yields the basic formula in the construction of an abridged life table

$$\hat{q}_i = \frac{n_i M_i}{1 + (1-a_i) n_i M_i} \quad . \tag{2.3}$$

The age-specific death rate M_i may be estimated from

$$M_i = \frac{D_i}{P_i} \tag{2.4}$$

with P_i being the mid-year population.

All other quantities in the table are functions of $\hat{q}_i$, a_i and the radix ℓ_0. The number d_i of deaths in (x_i, x_{i+1}) and the number ℓ_{i+1} of survivors at age x_{i+1} are computed from

$$d_i = \ell_i \hat{q}_i, \qquad i=0,1,\ldots,w-1, \tag{2.5}$$

and

$$\ell_{i+1} = \ell_i - d_i, \qquad i=0,1,\ldots,w-1, \tag{2.6}$$

respectively. The number of years lived in the interval (x_i, x_{i+1}) by the ℓ_i survivors at age x_i is

$$L_i = n_i(\ell_i - d_i) + a_i n_i d_i, \qquad i=0,1,\ldots,w-1 \; . \tag{2.7}$$

The final age interval is again an open interval, and L_w is computed exactly as in the complete life table [cf., equation (3.6) in Chapter 4]:

$$L_w = \frac{\ell_w}{M_w} \quad , \tag{2.8}$$

where M_w is again the specific death rate for people of age x_w and over.

The total number T_i of years remaining to all the people attaining age x_i is the sum of L_j for $j=i,i+1,\ldots,w$. The observed expectation of life $\hat{e}_i$ at age x_i is the ratio T_i/ℓ_i, or

$$\hat{e}_i = \frac{L_i + L_{i+1} + \ldots + L_w}{\ell_i} \quad , \qquad i=0,\ldots w \ . \tag{2.9}$$

As an example, the abridged life table for the California 1970 total population is given in Tables 1 and 2. The required data for constructing an abridged life table is the death rate (M_i) and the fraction of last age interval of life (a_i) for each age group. The death rate may be computed from the mid-year population (P_i column (2) in Table 1) and the number of deaths (D_i, column (3)) of the population in question using formula (2.4). For the

California 1970 total population, for example, the death rate M_0 is computed from

$$M_0 = \frac{D_0}{P_0} = \frac{6234}{340483} = .018309 \ . \tag{2.4a}$$

The fraction of the last age interval of life, a_i, remains relatively constant over time for a given age interval (x_i,x_{i+1}). The only exception is a_0, which may be computed from the readily available published data on infant deaths. The value of a_i, for $i=0,1,\ldots$, have been computed for several countries and are given in Appendix V. They may be revised every 10 years.

When death rate (M_i) for each age group of a population is determined, one uses the corresponding a_i and formula (2.3) to compute $\hat{q}_i$. The figures in other columns in the life table can be obtained by using formulas (2.5) through (2.9). The reader should use these formulas to verify the numerical values in Tables 1 and 2.

Table 1.

Construction of Abridged Life Table for Total California Population, USA, 1970

Age interval (in years)	Mid-year population in interval (x_i, x_{i+1})	Number of deaths in interval (x_i, x_{i+1})	Death rate	Fraction of last age interval of life	Probability of dying in interval $(x_i, x_{i+1}$
x_i to x_{i+1}	P_i	D_i	M_i	a_i	$\hat{q}_i$
(1)	(2)	(3)	(4)	(5)	(6)
0- 1	340483	6234	.018309	.09	.0181
1- 5	1302198	1049	.000806	.41	.00322
5-10	1918117	723	.000377	.44	.00188
10-15	1963681	735	.000374	.54	.00187
15-20	1817379	2054	.001130	.59	.00564
20-25	1740966	2702	.001552	.49	.00773
25-30	1457614	2071	.001421	.51	.00708
30-35	1219389	1964	.001611	.52	.00802
35-40	1149999	2588	.002250	.53	.01119
40-45	1208550	4114	.003404	.54	.01689
45-50	1245903	6722	.005395	.53	.02664
50-55	1083852	8948	.008256	.53	.04049
55-60	933244	11942	.012796	.52	.06207
60-65	770770	14309	.018565	.52	.08886
65-70	620805	17088	.027526	.51	.12893
70-75	484431	19149	.039529	.52	.18052
75-80	342097	21325	.062336	.51	.27039
80-85	210953	20129	.095419	.50	.38521
85+	142691	22483	.157564		1.00000

Table 2.

Abridged Life Table for Total California Population, USA, 1970

Age interval (in years)	Probability of dying in interval (x_i, x_{i+1})	Number living at age x_i	Number dying in interval (x_i, x_{i+1})	Fraction of last year of life	Number of years lived in interval (x_i, x_{i+1})	Total number of years lived beyond age x_i	Observed Expectation of life at age x_i
x_i to x_{i+1}	q_i	ℓ_i	d_i	a_i	L_i	T_i	$\mathring{e}_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0- 1	.01801	100000	1801	.09	98361	7195221	71.95
1- 5	.00322	98199	316	.41	392050	7096860	72.27
5-10	.00188	97883	184	.44	488900	6704810	68.50
10-15	.00187	97699	183	.54	488074	6215910	63.62
15-20	.00564	97516	550	.59	486452	5727836	58.74
20-25	.00773	96966	750	.49	482917	5241384	54.05
25-30	.00708	96216	681	.51	479412	4758467	49.46
30-35	.00802	95535	766	.52	475837	4279055	44.79
35-40	.01119	94769	1060	.53	471354	3803218	40.13
40-45	.01689	93709	1583	.54	464904	3331864	35.56
45-50	.02664	92126	2454	.53	454863	2866960	31.12
50-55	.04049	89672	3631	.53	439827	2412097	26.90
55-60	.06207	86041	5341	.52	417387	1972270	22.92
60-65	.08886	80700	7171	.52	386290	1554883	19.27
65-70	.12893	73529	9480	.51	344419	1168593	15.89
70-75	.18052	64049	11562	.52	292496	824174	12.87
75-80	.27039	52487	14192	.51	227665	531678	10.13
80-85	.38521	38295	14752	.50	154595	304013	7.94
85+	1.00000	23543	23543		149418	149418	6.35

3. The Fraction of the Last Year of Life, a'_x, and the Fraction of the Last Age Interval of Life, a_i

3.1 The fraction of the last year of life, a'_x. The description of life table construction in the preceding section clearly indicates that the main ingredient in the construction of complete life tables is the fraction of the last year of life. Computation of the fraction is quite easy when the necessary data are available. Since we need to know only the exact number of days lived past the final birthday, which may be obtained from the date of birth and the date of death. In the State of California, both dates are key punched into tabulation cards, from which a computer can make the required subtraction to get the number of days lived during the last year of life for each person who died, sum the number of days lived, divide by the number of deaths to obtain the average number of days lived, and divide by 365 (or 366) days to give the desired fraction of year lived, a'_x. Various statistical tests have been performed regarding the fraction a'_x using 132,205 California resident death records in 1960. Some of the results are briefly described below.

First the hypothesis of uniform distribution of deaths was tested for each year of age. The number of deaths by days lived during the last year of life was tabulated for each race and sex group and for each year of age. The year was divided into 26 intervals of 14 days each, except for the first interval which was 15 days. A frequency distribution of the number of deaths by intervals of days lived for each of the selected ages is shown in Table 5. The Chi-square method was used to test the uniform distribution of deaths. For the 10 ages shown, Chi-square values were significant for ages 0, 1, and 59.

The distribution of the deaths in the first year of life is highly skewed with the first interval of 15 days accounting for almost 70 percent of the total 8624 deaths. The second interval of 14 days contains only about 3.5 percent of the deaths, and this percentage decreases with the increase in age. The distribution of deaths in the second year also shows a decrease in the percentage of deaths with increasing age, although in a much smaller degree. No definite pattern can be ascertained for the distribution of deaths for age 59.

The t-test and the F-test were also performed for the difference between the observed fraction (a_x') and the hypothesized value of .5, and for the difference between sex and race groups for each year of life. The results show that from age 5 on, the fraction a_x' is invariant with respect to sex and race and the assumed value of .5 is accepted. For the first 5 years of life, the data suggest the values of $a_0' = .09$, $a_1' = .43$, $a_2' = .45$, $a_3' = .47$, and $a_4' = .49$ for both sexes. These values, except for a_0', may be assumed for other countries. The value of a_0', however, needs to be computed for each country. The data required for the computation of a_0' are usually available in vital statistics publications.

Computation of a_0' is shown in Table 6, where the 1970 United States infant death data are used for illustration. The number of deaths in column (3) by age at death are usually available in vital statistics publications. The average point for each interval [Column (2)], takes into account the distribution of deaths in each interval. The product of the figures in columns (2) and (3), recorded in column (4), is the number of days lived by individuals who died in each interval. The sum of the products, appearing in the lower right hand corner (2,464,403.7 in this case) is the total number of days lived by the (74,667) infants who died during the first year of life. This total, when divided by 74,667 x 365, gives the fraction of a'_0, the fraction of the year lived by an infant who dies during the first year of life. Since both the complete life table and the abridged life table begin with the age interval (0,1), $a_0 = a'_0$.

3.2 The fraction of the last age interval of life, a_i. This fraction is as essential in the construction of the abridged life table as the fraction of the last year of life in the complete table. Conceptually, it is an extension of the latter. When a person dies at age 23, for example, he has lived a certain fraction of the age interval (20,25). The average fraction lived in each interval (x_i, x_{i+1}), which is called the fraction of the last age interval of life, depends on the probability of dying and the corresponding fraction of last year of life a'_x for each year of age within the interval. The relation between a_i and the probabilities q_x and $p_x(= 1-q_x)$ and a'_x is derived as follows.

Consider the age interval (1, 5) and the fraction of the interval a_1 that a person will live if he dies between ages one and five. For a person alive at the exact age of one year [i.e., the beginning of the interval (1, 5)], there is a probability q_1 that he will die during the year (1, 2), a probability $(1-q_1)q_2 = p_1q_2$ that he will die in (2, 3), a probability $p_1p_2q_3$ of dying in (3, 4), and a probability $p_1p_2p_3q_4$ of dying in (4, 5). The corresponding periods of time that he lives are a'_1, $(1+a'_2)$, $(2+a'_3)$, and $(3+a'_4)$, respectively. For example, if he dies during the year (2, 3), he will have lived one complete year (1, 2) and a fraction a'_2 of the year (2, 3), therefore he will have lived a total of $1+a'_2$ years. The probability of dying at any time during the interval (1, 5) is $1 - p_1p_2p_3p_4$, and the length of the interval is 5-1 = 4 years, therefore the formula for the fraction a_1 is

$$a_1 = \frac{q_1a'_1 + p_1q_1(1+a'_2) + p_1p_2q_3(2+a'_3) + p_1p_2p_3q_4(3+a'_4)}{4(1 - p_1p_2p_3p_4)} . \quad (3.1)$$

Table 3. Frequency distribution of deaths by interval of days lived in the last year of life for selected ages, total population, California, 1960

Interval (in completed days)	Age at Death (in completed years)									
	0	1	2	3	9	19	29	39	49	59
0-14	6,091	29	20	19	5	11	14	21	50	83
15-28	305	27	15	10	5	9	8	22	48	63
29-42	228	33	22	11	2	9	7	23	43	63
43-56	247	25	11	15	7	12	13	23	46	64
57-70	233	21	16	12	3	8	9	20	44	71
71-84	190	20	10	8	5	7	16	22	41	68
85-98	152	31	13	13	3	4	9	32	49	87
99-112	153	24	17	13	4	18	2	27	43	69
113-126	152	22	13	14	3	8	12	25	42	56
127-140	114	28	16	13	1	9	5	29	49	77
141-154	89	25	16	9	4	8	10	27	46	93
155-168	91	28	17	8	5	11	10	15	49	80
169-182	61	28	16	11	7	12	10	20	50	71
183-196	55	18	9	12	3	14	7	20	42	72
197-210	55	11	15	9	5	9	14	33	50	84
211-224	54	13	11	8	4	9	6	21	57	84
225-238	47	14	12	9	6	11	13	21	43	63
239-252	50	8	6	13	4	7	12	25	43	68
253-266	35	12	10	9	2	7	10	24	53	74
267-280	41	14	16	8	5	9	12	28	53	87
281-294	36	12	14	8	4	4	14	20	55	87
295-308	31	13	11	6	3	13	9	24	46	85
309-322	24	28	11	9	1	5	6	26	45	95
323-336	41	13	6	7	3	6	11	28	43	92
337-350	28	11	10	15	3	8	14	20	54	91
351-364	21	18	8	3	3	11	13	16	59	90
Total Deaths	8,624	526	341	272	100	239	266	612	1,243	2,017
x^2	97,299.5**	68.4**	29.4	22.9	15.8	26.6	27.4	21.0	13.2	40.9*

* Significant at the 5 percent level
** Significant at the 1 percent level

Table 4. Computation of the fraction a_0 based on infant deaths, United States total population, 1970

Age interval at death (1)	Average point in interval (in days) (2)	Number of deaths in interval* (3)	Number of days lived (2) x (3) (4)
<1 hour	.02	6,485	129.7
1-24 hours	.5	26,425	13,212.5
1-2 days	1.5	7,944	11,916.0
2-3 days	2.5	4,761	11,902.5
3-4	3.5	2,163	7,570.5
4-5	4.5	1,346	6,057.0
5-6	5.5	984	5,412.0
6-7	6.5	713	4,634.5
7-14	10.0	2,722	27,220.0
14-21	17.0	1,461	24,837.0
21-28	24.0	1,275	30,600.0
28-60	42.0	4,662	195,804.0
2-3 mos.	73.0	3,561	259,953.0
3-4	103.0	2,586	266,358.0
4-5	134.0	1,866	250,044.0
5-6	164.0	1,379	226,156.0
6-7	195.0	1,065	207,675.0
7-8	225.0	874	196,650.0
8-9	256.0	678	173,568.0
9-10	287.0	597	171,339.0
10-11	318.0	565	179,670.0
11-12	349.0	555	193,695.0
Total		74,667	2,464,403.7

*Source: U.S. Department of Health, Education and Welfare, Public Health Service, National Center for Health Statistics, Vital Statistics, of the U.S., 1970, Vol. II, Part A, pp. 2-10, 11.

$$a_0 = \frac{2,464,403.7}{365 \text{x} 74,667} = .09$$

Using the established values of a_1', a_2', a_3' and a_4', we have

$$a_1 = \frac{.43q_1 + 1.45p_1q_2 + 2.47p_1p_2q_3 + 3.49p_1p_2p_3q_4}{4(1 - p_1p_2p_3p_4)} \quad . \qquad (3.2)$$

For a given country, the given probabilities, q_1, q_2. q_3 and q_4 can be determined. Therefore, the fraction a_1 for the interval (1, 5) may be computed from formula (3.2). The computation of a_1 for California population, 1970, is demonstrated in Table 5.

Table 5. Computation of the fraction a_1 for age interval (1, 5) based on California mortality data, 1970

Year of Age	Conditional Probability of Dying in year (x, x+1) given alive at age 1	Expected length of time lived in interval: Length of time lived (1, 5)	Expected length of time lived in interval: (2 x 3)
1	2	3	4
1-2	q_1 = .00113	.43	.000486
2-3	p_1q_2 = (.99887)(.00086) = .000859	1.45	.001246
3-4	$p_1p_2q_3$ = (.99887)(.99914)(.00073) = .000729	2.47	.001800
4-5	$p_1p_2p_3q_4$ = (.99887)(.99914)(.99927)(.00052) = .000519	3.49	.001810
Total	$1 - p_1p_2p_3p_4$ = .003236		.00534

$$a_1 = \frac{.00534}{4\text{x}.003236}$$

$$= .41$$

From age 5 to the last interval in the life table, the length of each age interval is 5 years and the fraction of last year of life for each year is $a_x' = 1/2$. The formula for the fraction a_i for interval (x_i, x_i+5) can be simplified somewhat. For age interval (5, 10) for example, we have

$$a_2 = \frac{.5q_5 + (1+.5)p_5q_6 + (2+.5)p_5p_6q_7 + (3+.5)p_5p_6p_7q_8 + (4+.5)p_5p_6p_7p_8q_9}{5(1 - p_5p_6p_7p_8p_9)}$$

$$= \frac{p_5q_6 + 2p_5p_6q_7 + 3p_5p_6p_7q_8 + 4p_5p_6p_7p_8q_9}{5(1 - p_5p_6p_7p_8p_9)} + .1 \quad , \tag{3.3}$$

since

$$q_5 + p_5q_6 + p_5p_6q_7 + p_6p_6p_7q_8 + p_5p_6p_7p_8q_9 = 1 - p_5p_6p_7p_8p_9 \; . \tag{3.4}$$

The values of the fraction a_i for the abridged life table have been computed from formulas (3.2) and (3.3) for selected countries for which the required information is available, and are listed in Appendix V. These values of a_i can be used directly in constructing life tables for the respective countries.

Remark 4. Formulas (3.2) and (3.3) show that a_i does not depend on the absolute values of q_x or p_x but rather on the trend of mortality within the interval. For example, if $q_5 > q_6 > q_7 > q_6 > q_9$, then $a_5 < 1/2$, regardless of the absolute values of these q_x's.

Remark 5. The probabilities q_x and p_x are computed from the mortality data of a population in question, the value of a_i represents the mortality trend in each interval prevailing in the population. Since the mortality trend does not vary much over time (although death rates do), the a_i values may be regarded as constant and may be used for the construction of abridged life tables of the subsequent years of the population.

The invariant property of a_i not only holds over time, but is also true for countries with similar mortality patterns. Table 8 shows a remarkable agreement of the five sets of a_i values. For countries with a similar mortality pattern, the same set of a_i values may be used.

Remark 6: The assumption that $a'_x = 1/2$ for each year of age within an interval (x_i, x_{i+1}) does not necessarily imply that $a_i = 1/2$ for the entire interval. As formula (3.3) shows, the value of the fraction a_i depends on the mortality pattern over an entire interval and not on the mortality rate for any single year. When the mortality rate increases with age in an interval, the fraction $a_i > 1/2$; then the reverse pattern prevails, $a_i < 1/2$. Consider, for example, the age intervals (5, 10) and (10, 15) in 1970 California population. Although $a'_x = 1/2$ for each age in the two intervals, $a_2 = .44$ for interval (5, 10) and $a_3 = .54$ for interval (10, 15) due to the changing mortality pattern, as shown in Table 7.

Table 6

Fraction of last age interval of life, a_i, for selected populations

Age	Austria 1969	California 1970	France 1969	Finland 1968	U.S.A. 1970
0-1	.12	.09	.16*	.09	.09
1-5	.37	.41	.38	.38	.40
5-10	.47	.44	.46	.49	.46
10-15	.51	.54	.54	.52	.55
15-20	.58	.59	.56	.53	.54
20-25	.48	.49	.51	.51	.51
25-30	.51	.51	.51	.51	.51
30-35	.53	.52	.53	.52	.52
35-40	.53	.53	.53	.54	.53
40-45	.52	.54	.53	.55	.54
45-50	.54	.53	.54	.53	.54
50-55	.52	.53	.52	.54	.53
55-60	.53	.52	.53	.53	.53
60-65	.54	.52	.53	.53	.52
65-70	.53	.51	.53	.52	.52
70-75	.52	.52	.52	.52	.51
75-80	.51	.51	.51	.51	.51
80-85	.48	.50	.49	.47	.49
85-90	.45		.46		
90-95	.40		.41		

*A large a_0 value for the France 1969 population is due to the fact that infants who die before 3 days old are not recorded. Age at death of these infants are not included in the calculation of a_0.

Table 7

Computation of a_i for age intervals (5, 10) and (10, 15) based on California population, 1970

Age interval	Fraction of the last year of life	Proportion dying in age interval	Fraction of last age interval
x to x+1	a'_x	q_x	a_i
(1)	(2)	(3)	(4)
5-6	.50	.00049	
6-7	.50	.00045	
7-8	.50	.00034	.44
8-9	.50	.00031	
9-10	.50	.00030	
10-11	.50	.00031	
11-12	.50	.00033	
12-13	.50	.00035	.54
13-14	.50	.00041	
14-15	.50	.00048	

4. Significant Historical Contributions to the Construction of Abridged Life Tables

The history of life table construction reflects increasing refinement of the method. For instance, although the earliest tables (see Introduction to this chapter) were based solely on recorded deaths, Milne's table of 1815 took into account population figures as well. In 1839 the English Life Tables were constructed using only registered births and deaths since, due to the influence of William Farr, census figures were found to be unreliable. Other significant contributions and refinements followed, in particular those of Moore, Day, Wickens, Pell, King, Derksen, Greville, Reed-Merrell, Wiesler, Keyfitz, and Sirken. We shall briefly discuss some of these methods below.

4.1. King's Method. This method was introduced by George King in the construction of the Seventh English Life Table at the turn of the century and has been used by many English-speaking countries for fifty years or so. Using this method, data are arranged in five-year age groups. Population figures and the number of deaths are calculated for the central year of age (pivotal age) of each age group by a graduation process, yielding the values of q_x for the pivotal age. Using the complement of q_x at pivotal ages and finite difference formulas, the number of survivors (ℓ_x) are obtained. T. N. E. Greville adapted this method for the 1939-41 United States Life Tables.

4.2. Reed-Merrell Method. In the search for a relation between the probability q_i and the mortality rate, M_i, Lowell J. Reed and Margaret Merrell studied extensively some thirty-three tables in J. W. Glover's 1910 series of United States Life Tables. Their findings were published in 1939 showing that the following equation describes satisfactorily the entire range of observations in Glover's tables:

$$q_i = 1-e^{-n_i M_i - .008 n_i^3 M_i^2} .$$

Many formulas are also given to determine the L_i column from the number of survivors ℓ_i in the life table.

4.3. Greville's Method. T.N.E. Greville used a mathematical approach to derive a relation between q_i and M_i. He started with the equation

$$M_i = - \frac{d}{dx_i} \log L_i$$

After integrating both sides of the equation, thus yielding L_i, and applying the Euler-Maclaurin summation formula, Greville was able to express T_i in terms of a series of exponential functions of M_i. He then used quite skilful mathematical manipulations, and arrived at the formula:

$$q_i = \frac{M_i}{1/n_i + M_i[1/2 + n_i/12(M_i - \log c)]}$$

where the constant c is the constant in the Gompertz' law of mortality:

$$\mu_x = Bc^x$$

Greville also suggested a number of formulas to compute the life table population L_i.

4.4. Wiesler's Method. This method, proposed in "Une méthods simple pour la construction de tables de mortalité abrégées," World Population Conference, 1954, Volume IV, United Nations, in essence uses age specific death rates M_i as the probability of dying $\hat{q}_i$. For an age interval (x_i, x_i+n_i), let D_i be the number of deaths during a calendar year and P_i be the total of living people in the age group (x_i, x_i+n_i). Then Weisler suggests that

$$\hat{p}_i = 1 - \frac{D_i}{P_i} \quad \text{or} \quad \hat{q}_i = \frac{D_i}{P_i} \quad ,$$

and ℓ_1, ℓ_5, ℓ_{10}, etc., are computed successively from

$$\ell_1 = \ell_0 \hat{p}_0$$

$$\ell_5 = \ell_1 (p_{1-4})^{t_{1-4}}$$

$$\ell_{10} = \ell_5 (p_{5-9})^{t_{5-9}} \quad , \ldots \quad ,$$

where t_{1-4}, t_{5-9}, etc., are assumed to be the same for all mortality tables. The expectation of life at x_α is computed from

$$\hat{e}_\alpha = \frac{1}{2} + \frac{\ell_{\alpha+1} + \ell_{\alpha+2} + \ldots}{\ell_\alpha} \quad .$$

4.5. Sirken's Method. Monroe Sirken distinguishes the age specific death rate M_i from a current population:

$$M_i = \frac{D_i}{P_i}$$

and the rate m_i defined in the life table quantities:

$$m_i = \frac{d_i}{L_i} \quad .$$

Using the observed death rate M_i, one derives q_i from the equation

$$q_i = \frac{n_i M_i}{1+\alpha_i M_i} \tag{A}$$

where the constant α_i is assumed to be the same as in a standard table. Using q_i, one completes the columns ℓ_i and d_i. To compute L_i, Sirken considers another equation

$$q_i = \frac{n_i m_i}{1+a_i m_i} \quad . \tag{B}$$

Substituting $q_i = d_i/\ell_i$ and $m_i = d_i/L_i$ in (B) yields

$$\frac{d_i}{\ell_i} = \frac{n_i \; d_i/L_i}{1+a_i \; d_i/L_i} \quad .$$

Solving the last equation for L_i one gets

$$L_i = n_i \ell_i - a_i d_i$$

where the constant a_i is assumed to be the same as in a standard table but is different from α_i.

4.6. Keyfitz's Method. This is an iterative procedure using the basic relationship between the probability q_i and the age-specific mortality rate m_i or M_i

$$q_i = \frac{n_i M_i}{1+(n_i - {}_n a_i) M_i} \tag{A}$$

where ${}_n a_i$ is the number of years lived in the age interval $(x_i, x_i + n_i)$ by an individual who dies in the interval. In addition to ${}_n a_i$, Keyfitz introduces a quantity ${}_n A_i$, the average number of years already lived within the interval $(x_i, x_i + n_i)$ by a stationary population aged $(x_i, x_i + n_i)$.

Taking ${}_n a_i = n_i/2$ on the first cycle to obtain first approximation of q_i using formula (A), then using

$$ {}_n a_i = \frac{n_i}{2} + \frac{n_i}{24} \left(\frac{d_{i+1} - d_{i-1}}{d_i} \right) \quad , $$

$$ L_i = \frac{n_i}{2} (\ell_i + \ell_{i+1}) + \frac{n}{24} (d_{i+1} - d_{i-1}) \quad , $$

and

$$ {}_n A_i = \frac{n_i}{2} + \frac{n_i}{24} \left(\frac{L_{i+1} - L_{i-1}}{L_i} \right) $$

and other formulas to arrive at a second approximation of q_i. After each iteration, a life table is constructed and the age-specific mortality rate is compared with those observed, and an adjustment made for the next iteration. The iterative process continues until the life table age specific rates agree with the corresponding observed rates.

5. Cohort (Generation) Life Table

It has been pointed out in Section 1 of this chapter that a cohort life table describes the actual mortality experience of a particular group of individuals (the cohort) from birth to the death of the last member of the group. The subject involved need not be human beings. Cohort life tables for various animal populations have appeared frequently in the literature. In fact, the cohort life table has been widely used for years in studies of animal populations, in biological control, in ecology, and in population dynamics. Cohort life tables have been constructed for inanimate objects as well.

For simplicity of formulas, the length of age interval is assumed to be constant and denoted by n. The basic variables involved in a cohort life table are the number of survivors (ℓ_x) at each age x, the number of deaths (d_x) within each age interval (x,x+n). The unit of x depends on the problem in question. In any case, the numbers ℓ_x and d_x satisfy the obvious relationship

$$\ell_x - \ell_{x+n} = d_x \tag{5.1}$$

or

$$\ell_{x+n} = \ell_x - d_x \quad . \tag{5.2}$$

The number of survivors at age x+n is equal to the number alive at the beginning of the interval (x, x+n) minus those who died during the interval. The probability q_x for each interval is estimated by dividing d_x by ℓ_x,

$$\hat{q}_x = \frac{d_x}{\ell_x} \quad . \tag{5.3}$$

When ℓ_x, d_x, and $\hat{q}_x$ are determined, the remaining part of the life table can be completed in exactly the same way as in the current life table in Section 1 and 2. Assuming $q_x = 1/2$, we have

$$L_x = n\ell_{x+n} + (1-\frac{1}{2})nd_x = \frac{n}{2}(\ell_x + \ell_{x+n}) \tag{5.4}$$

In a cohort life table, observations are usually made throughout the life span of subjects under study. Therefore, the values ℓ_x, d_x, L_x and formulas from (5.1) to (5.5) are all applicable to the last age interval (ℓ_x are over). The quantity T_x as before is equal to the sum of L_x, i.e.,

$$T_x = L_x + \ \ldots + L_w \tag{5.5}$$

where the symbol w indicates the beginning of the last age interval. Finally, the expectation of life ℓ_x is given by

$$\hat{e}_x = \frac{T_x}{\ell_x}, \quad x = 0,1,\ldots,w \ . \tag{5.6}$$

An example of a life table for adult Drosophila Melanogaster is presented in Table 10 for illustration [Miller and Thomas]. A group of $\ell_0 = 270$ male fruit flies were followed from the time they became adults. The number of survivors at each five day period and the number of deaths occurring within each age interval of five days are recorded in columns (2) and (3) respectively. Dividing d_x by ℓ_x for each age interval gives the probability of dying $\hat{q}_x$ in column (4). Using relations (5.4), (5.5), and (5.6), we computed the quantities L_x, T_x, and $\hat{e}_x$ for each age, and recorded the numerical values in columns (5), (6), and (7), respectively assuming $a_x = .05$ over all intervals. A similar table has been constructed for female adult Drosophila. Comparison between the two sexes with respect to the expectation of life, the survival probability, or the probability of death, can easily be made with the aid of the corresponding standard deviations (c.f., Chapter 6).

Table 8

Life Table of Adult Male Drosophila Melanogaster

Age Interval (Days)	Number living at age x	Number dying in (x, x+n)	Probability of dying in (x, x+n)	Days lived in (x, x+n)	Days lived beyond age x	Observed Expectation of life at age x
(x, x+n)	ℓ_x	d_x	$\hat{q}_x$	L_x	T_x	$\hat{e}_x$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0-5	270	2	.00741	1345	11660	43.2
5-10	268	4	.01493	1330	10315	38.5
10-15	264	3	.01136	1312	8985	34.0
15-20	261	7	.02682	1288	7673	29.4
20-25	254	3	.01181	1262	6385	25.1
25-30	251	3	.01195	1248	5123	20.4
30-35	248	16	.06452	1200	3875	15.6
35-40	232	66	.28448	995	2675	11.5
40-45	166	36	.21687	740	1680	10.1
45-50	130	54	.41538	515	940	7.2
50-55	76	42	.55263	275	425	5.6
55-60	34	21	.61765	118	150	4.4
60 +	13	13	1.00000	32	32	2.5

Table 9

Life Table of Adult Female Drosophila Melanogaster

Age Interval (Days)	Number living at age x	Number dying in (x, x+n)	Probability of dying in (x, x+n)	Days lived in (x, x+n)	Days lived beyond x	Observed Expectation of life at age x
(x, x+n)	ℓ_x	d_x	$\hat{q}_x$	L_x	T_x	$\hat{e}_x$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0-5	275	4	.01455	1365	10303	37.5
5-10	271	7	.02583	1338	8938	33.0
10-15	264	3	.01136	1312	7600	28.8
15-20	261	7	.02682	1288	6288	24.1
20-25	254	13	.05118	1238	5000	19.7
25-30	241	22	.09129	1150	3762	15.6
30-35	219	31	.14155	1018	2612	11.9
35-40	188	68	.36170	770	1594	8.5
40-45	120	51	.42500	472	824	6.9
45-50	69	38	.55072	250	352	5.1
50-55	31	26	.83871	90	102	3.3
55 +	5	5	1.00000	12	12	2.5

Footnotes

1/ During the early development of the concept of expectation of life, a curtate expectation of life, defined as

$$e_x = \frac{\ell_{x+1} + \ell_{x+2} + \cdots}{\ell_x}$$

was first introduced. This expectation considers only the completed years of life lived by survivors, whereas the complete expectation of life takes into account also the fractional years of life lived by those who die in any year. Under the assumption that each person who dies during any year of age lives half of the year on the average, the complete expectation of life is given by

$$\mathring{e}_x = e_x + \frac{1}{2} \quad .$$

Since the curtate expectation is no longer in use, in this book the symbol e_x is used to denote the true expectation of life at age x.

CHAPTER 6

STATISTICAL INFERENCE REGARDING LIFE TABLE FUNCTIONS

1. Introduction

Each figure in a life table as described in the preceding chapters is an estimate of the corresponding unknown true value. Statistical inference regarding these unknown values may be made on the basis of the observed quantities. An essential element required in making statistical inference, as indicated in Chapter 3, is the standard error of the estimate. The purpose of this chapter is (1) to derive formulas for the sample variances (or their square roots, standard error) of the life table functions, and (2) to demonstrate with numerical values how to construct confidence intervals and how to test statistical hypotheses. Specifically, inference will be made about three categories of parameters: (i) q_i, the probability of dying in an age interval (x_1, x_{i+1}); (ii) p_{ij}, the survival probability from age x_i to x_j; and (iii) e_α, the expectation of life at age x_α, for $\alpha = 0,1,\ldots,w$.

2. The Probability of Dying q_i and the Probability of Surviving p_i

The probability of dying and the probability of surviving an age interval are complementary to one another; therefore their estimates have the same sample variance. Denoting the sample variances by $S^2_{\hat{q}_i}$ and $S^2_{\hat{p}_i}$, respectively, we have

$$S^2_{\hat{q}_i} = S^2_{\hat{p}_i} \quad . \tag{2.1}$$

In a current life table, the estimate $\hat{q}_i$ is derived from the corresponding mortality information, in terms of which the sample variance of $\hat{q}_i$ should be expressed. We have found in Chapter 3, equation (3.5) that

$$S^2_{\hat{q}_i} = \frac{1}{D_i} \hat{q}_i^2 (1-\hat{q}_i) , \tag{2.2}$$

and the 95% confidence interval for the probability q_i:

$$\Pr\{\hat{q}_i - 1.96\, S_{\hat{q}_i} < q_i < \hat{q}_i + 1.96\, S_{\hat{q}_i}\} = .95 \quad . \tag{2.3}$$

For a given problem, $\hat{q}_i$ and $S_{\hat{q}_i}$ can be determined, and the two limits, $\hat{q}_i - 1.96\, S_{\hat{q}_i}$ and $\hat{q}_i + 1.96\, S_{\hat{q}_i}$, can be found. These limits are called the confidence limits, and the interval extending from the lower limit $\hat{q}_i - 1.96\, S_{\hat{q}_i}$ to the upper limit $\hat{q}_i + 1.96\, S_{\hat{q}_i}$ is the 95% confidence interval.

As an example, consider the probability of dying in the first year of life, q_0. In the 1970 California experience, the estimate $\hat{q}_0 = .01801$, the number of deaths, $D_0 = 6234$, and hence the standard error of $\hat{q}_0$ is:

$$S_{\hat{q}_0} = \sqrt{\frac{\hat{q}_0^2(1-\hat{q}_0)}{D_0}}$$

$$= \sqrt{\frac{(.01801)^2(1-.01801)}{6234}}$$

$$= .000226 \ .$$

Substituting these values in (2.3) yields the 95% confidence limits for the probability q_0:

$$\hat{q}_0 - 1.96\, S_{\hat{q}_0} = .01801 - 1.96(.000226) = .01757$$

$$\hat{q}_0 + 1.96\, S_{\hat{q}_0} = .01801 + 1.96(.000226) = .01845 \ .$$

Thus we conclude with a 95% confidence that, if the California 1970 mortality experience prevails in a population, the probability that a newborn will not survive to the first birthday is between .01757 and .01845.

The logic of the preceding statement needs some explanation. Formula (2.3) indicates that, before information is gathered, the chances are 95 out of 100 that the interval ($\hat{q}_i$ - 1.96 $S_{\hat{q}_i}$, $\hat{q}_i$ + 1.96 $S_{\hat{q}_i}$) to be determined will contain the unknown quantity q_i. After the information is gathered, and the numerical values of the limits (.01757 and .01845) are obtained, we certainly have confidence in the statement that the quantity q_0 is between .01757 and .01845; a measure of this confidence is the value of the probability .95. This measure of confidence (.95, in this case) is called the confidence coefficient. The essential point to be recognized is that a probability is a measure of likelihood of occurrence of an event (death, for example) before the event takes place, whereas a confidence coefficient is a measure of confidence one has in a statement about an unknown quantity after the corresponding event has occurred.

A second use of the sample variance of the estimate $\hat{q}_i$ is testing a hypothesis concerning either the probability of dying in one age interval or the comparison of two or more probabilities. Suppose we want to know if the force of mortality has decreased over the past decade so that a new born in 1970 has a better chance of surviving the first year of life than that in 1960. Here we are testing the hypothesis that $q_0(1970)$ is the same as $q_0(1960)$ against the alternative hypothesis that $q_0(1970)$ is smaller than $q_0(1960)$. The statistics for the test is

$$Z = \frac{\hat{q}_0(1960) - \hat{q}_0(1970)}{\text{S.E.}[\hat{q}_0(1960) - \hat{q}_0(1970)]} \tag{2.4}$$

where the standard error of the difference is given by

$$\text{S.E.}[\hat{q}_0(1960) - \hat{q}_0(1970)] = \sqrt{\frac{\hat{q}_0^2(1960)[1-\hat{q}_0(1960)]}{D_0(1960)} + \frac{\hat{q}_0^2(1970)[1-\hat{q}_0(1970)]}{D_0(1970)}} \tag{2.5}$$

Using California experience again, we have the required information given in the following table

Table 1

Estimate of probability of dying in the first year of life and the standard error, California, 1960 and 1970

	1960	1970	$\hat{q}_0(1960) - \hat{q}_0(1970)$
$\hat{q}_0$	.02378	.01801	.00577
D_0	8663	6234	
s^2	$6.3724\text{x}10^{-8}$	$5.109\text{x}10^{-8}$	$11.481\text{x}10^{-8}$
S.E.	$2.524\text{x}10^{-4}$	$2.260\text{x}10^{-4}$	$3.388\text{x}10^{-4}$

From Table 1 we compute the statistic

$$Z = \frac{.00577}{3.388\text{x}10^{-4}} = 17.03 \ ,$$

which is significantly greater than the 99th percentile in the standard normal distribution. We conclude that a newborn in California in 1970 has a smaller probability of dying in the first year of life than that in 1960.

Remark 1. In a cohort (generation) life table both the number living (ℓ_i) at age x_i and the number of deaths (d_i) occurring in the interval (x_i, x_{i+1}) are directly observed. The probabilities are estimated by

$$\hat{q}_i = \frac{d_i}{\ell_i} \quad \text{and} \quad \hat{p}_i = 1 - \frac{d_i}{\ell_i} \ . \tag{2.6}$$

Here we have a binomial situation, so that the variance of $\hat{q}_i$ ($\hat{p}_i$) is given by

$$S^2_{\hat{q}_i} = \frac{1}{\ell_i}\hat{q}_i(1-\hat{q}_i) = S^2_{\hat{p}_i} \ . \tag{2.7}$$

3. The Survival Probability, p_{ij}

The probability that a person of age x_i will survive to age x_j is an important quantity in the survival analysis. It provides an investigator with critical information that he seeks in his study. This probability can be obtained directly from the life table. Since the survival of a person from age x_i to x_j means the survival of every single intermediate age interval, the probability p_{ij} is given by the equation

$$p_{ij} = p_i \, p_{i+1} \cdots p_{j-1} \tag{3.1}$$

or

$$p_{ij} = (1-q_i)(1-q_{i+1}) \ \ldots \ (1-q_{j-1}) \ . \tag{3.2}$$

A case of particular interest is when $x_i = 0$. Here we have the probability of surviving from age 0 to a specified age x_j

$$\begin{aligned} p_{0j} &= p_0 \, p_1 \cdots p_{j-1} \\ &= (1-q_0)(1-q_1) \ \ldots \ (1-q_{j-1}) \quad . \end{aligned} \tag{3.3}$$

To obtain the estimate of the survival probability, we only need to substitute the estimates of $\hat{q}_i$ in the formulas (3.2) and (3.3). When the information is taken from a life table, computations can be simplified. For example,

$$\begin{aligned} \hat{p}_{0j} &= \hat{p}_0 \, \hat{p}_1 \cdots \hat{p}_{j-1} \\ &= \frac{\ell_1}{\ell_0} \ \frac{\ell_2}{\ell_1} \cdots \frac{\ell_j}{\ell_{j-1}} = \frac{\ell_j}{\ell_0} \ , \end{aligned} \tag{3.4}$$

similarly

$$\hat{p}_{ij} = \frac{\ell_j}{\ell_i} \quad . \tag{3.5}$$

In the current life table the individual estimates,

$$\hat{p}_h = 1-\hat{q}_h\ , \quad h = i,\ldots,j-1\ , \tag{3.6}$$

are based on the corresponding age-specific death rates, the sample variance of $\hat{p}_{ij}$ should be expressed in terms of the sample variance of each $\hat{q}_h$. Since the individual estimates $\hat{q}_h$ are based on mortality information of separate age groups, they are statistically independent of one another. Using a theorem on the variance of a product of independent random variables, the sample variance of $\hat{p}_{ij}$ may be determined from the formula:

$$S^2_{\hat{p}_{ij}} = \hat{p}^2_{ij} \sum_{h=i}^{j-1} \hat{p}_h^{-2} S^2_{\hat{p}_h} \tag{3.7}$$

with the sample variance of $\hat{p}_h$ given in (2.2).

For the 1960 United States data and for the 1970 California data, the probability p_{0i} and the corresponding sample variances and standard errors have been computed. The numerical results are given in Table 3 and Table 4, respectively. The mean steps in the computation are as follows:

(1) Record the number of deaths (D_i) occurring in each age interval in the population in Column 2, and the probability of dying in Column 3.

(2) Use formula (2.2) to compute the sample variance of $\hat{q}_i$ and enter it in Column 4.

(3) Use formula (3.3) to compute the probability of surviving age interval $(0,x_i)$ $\hat{p}_{0i}$, and record it in Column 5. $\hat{p}_{00}$ is 1 by definition.

(4) Use formula (3.7) or

$$S^2_{\hat{p}_{0i}} = \hat{p}^2_{0i}[\hat{p}_0^{-2} S^2_{\hat{p}_0} + \hat{p}_1^{-2} S^2_{\hat{p}_1} + \ldots + \hat{p}_{i-1}^{-2} S^2_{\hat{p}_{i-1}}]$$

to compute the variance of $\hat{p}_{0i}$ and record it in Column 6.

Table 2

Abridged life table for total United States population, 1960

Age Interval (in years)	Proportion Dying in Interval (x_i, x_i+1)	Number Living at Age x_i	Number Dying in Interval (x_i, x_i+1)	Fraction of Last Age Interval of Life	No. Years Lived in Interval (x_i, x_{i+1})	Total No. Years Lived Beyond Age x_i	Observed Expectation of Life at Age x_i
x_i to x_{i+1}	$\hat{q}_i$	ℓ_i	d_i	a_i	L_i	T_i	$\hat{e}_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0-1	.02623	100,000	2,623	.10	97,639	6,965,395	69.65
1-5	.00436	97,377	425	.39	388,471	6,867,756	70.53
5-10	.00245	96,952	238	.46	484,117	6,479,285	66.83
10-15	.00219	96,714	212	.54	483,082	5,995,168	61.99
15-20	.00458	96,502	442	.57	481,560	5,512,086	57.12
20-25	.00616	96,060	592	.49	478,790	4,030,526	52.37
25-30	.00652	95,468	622	.50	475,785	4,551,736	47.68
30-35	.00800	94,846	759	.52	472,408	4,075,951	42.97
35-40	.01159	94,087	1,090	.54	467,928	3,603,543	38.30
40-45	.01840	92,997	1,711	.54	461,050	3,135,615	33.72
45-50	.02902	91,286	2,649	.54	450,337	2,674,565	29.30
50-55	.04571	88,637	4,052	.53	433,663	2,224,228	25.09
55-60	.06577	84,585	5,563	.52	409,574	1,790,565	21.17
60-65	.10257	79,022	8,105	.52	375,658	1,380,991	17.48
65-70	.14763	70,917	10,469	.52	329,459	1,005,333	14.18
70-75	.21472	60,448	12,979	.51	270,441	675,874	11.18
75-80	.31280	47,469	14,848	.51	200,967	405,433	8.54
80-85	.46312	32,621	15,107	.48	123,827	204,466	6.27
85-90	.61437	17,514	10,760	.45	57,980	80,639	4.60
90-95	.78812	6,754	5,323	.41	18,067	22,659	3.35
95+	1.00000	1,431	1,431		4,592	4,592	3.21

Table 3

Computation of the standard error of survival probability.
Total United States population, 1960.

Age Interval	Number of Deaths in Interval (x_i, x_{i+1})	Probability of Dying in Interval (x_i, x_{i+1})	Sample Variance of $\hat{q}_i$ $(\hat{p}_i)$	Probability of Surviving Interval $(0, x_i)$	Sample Variance of $\hat{p}_{0i}$	Standard Error of $\hat{p}_{0i}$
(x_i, x_{i+1})	D_i	$\hat{q}_i$	$10^8 x S^2_{\hat{q}_i}$	$\hat{p}_{0i}$	$10^8 x S^2_{\hat{p}_{0i}}$	$10^4 x S_{\hat{p}_{0i}}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0 - 1	110873.	0.026230	0.60426	1.000000	0.00000	0.00000
1 - 5	17682.	0.004360	0.10703	0.973770	0.60426	0.77734
5 - 10	9163.	0.002450	0.06534	0.969520	0.70049	0.83695
10 - 15	7374.	0.002190	0.06489	0.967140	0.75848	0.87091
15 - 20	12185.	0.004580	0.17136	0.965020	0.81586	0.90325
20 - 25	13348.	0.006160	0.28252	0.960600	0.96799	0.98386
25 - 30	14214.	0.006520	0.29712	0.954680	1.21680	1.10308
30 - 35	19200.	0.008000	0.33066	0.948460	1.47180	1.21317
35 - 40	29161.	0.011590	0.45530	0.940870	1.74580	1.32128
40 - 45	42942.	0.018400	0.77390	0.929970	2.10864	1.45211
45 - 50	64283.	0.029020	1.27206	0.912860	2.70107	1.64349
50 - 55	90593.	0.045710	2.20093	0.886370	3.60661	1.89910
55 - 60	116753.	0.065770	3.46131	0.845850	5.01355	2.23909
60 - 65	153444.	0.102570	6.15306	0.790220	6.85223	2.61767
65 - 70	196605.	0.147630	9.44893	0.709170	9.36099	3.05957
70 - 75	223707.	0.214720	16.18415	0.604480	11.55334	3.39902
75 - 80	219978.	0.312800	30.56591	0.474690	13.03838	3.61087
80 - 85	185231.	0.463120	62.16567	0.326210	13.04497	3.61178
85 - 90	120366.	0.614370	120.92803	0.175140	10.37583	3.22115
90 - 95	50278.	0.788120	261.75601	0.067540	5.25246	2.29182
95+	13882.	1.000000	0.00000	0.014310	1.42976	1.19572

Table 4

Computation of standard error of survival probability.
Total California population, 1970.

Age Interval	Number of Deaths in Interval (x_i, x_{i+1})	Probability of Dying in Interval (x_i, x_{i+1})	Sample Variance of $\hat{q}_i (\hat{p}_i)$	Probability of Surviving Interval $(0, x_i)$	Sample Variance of $\hat{p}_{0i}$	Standard Error of $\hat{p}_{0i}$
(x_i, x_{i+1})	D_i	$\hat{q}_i$	$10^8 x S^2_{\hat{q}_i}$	$p_{\hat{0}i}$	$10^8 x S^2_{\hat{p}_{0i}}$	$10^4 x S_{\hat{p}_{0i}}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0 - 1	6234.	0.018010	5.10937	1.000000	0.00000	0.00000
1 - 5	1049.	0.003220	0.98522	0.981990	5.10937	2.26039
5 - 10	723.	0.001880	0.48793	0.978830	6.02660	2.45491
10 - 15	735.	0.001870	0.47487	0.976990	6.47146	2.54390
15 - 20	2054.	0.005640	1.53993	0.975160	6.90051	2.62688
20 - 25	2702.	0.007730	2.19433	0.969660	8.28727	2.87876
25 - 30	2071.	0.007080	2.40325	0.962160	10.22275	3.19730
30 - 35	1964.	0.008020	3.24870	0.955350	12.30338	3.50761
35 - 40	2588.	0.011190	4.78419	0.947690	15.07196	3.88226
40 - 45	4114.	0.016890	6.81706	0.937090	19.03349	4.36273
45 - 50	6722.	0.026640	10.27645	0.921260	24.38215	4.93782
50 - 55	8948.	0.040490	17.58000	0.896720	31.82238	5.64113
55 - 60	11942.	0.062070	30.25915	0.880410	43.43359	6.59041
60 - 65	14309.	0.088860	50.27921	0.807000	60.60944	7.78520
65 - 70	17088.	0.128930	84.73635	0.735290	83.06080	9.11377
70 - 75	19149.	0.180520	139.45783	0.640490	108.83660	10.43247
75 - 80	21325.	0.270390	250.13991	0.524870	130.29900	11.41485
80 - 85	20129.	0.385210	453.21022	0.382950	138.27254	11.75893
85 +	22843.	1.000000	0.00000	0.235430	118.72215	10.89596

(5) Take the square root of the variance to obtain the standard error of $\hat{p}_{0i}$ and record it in Column 7.

Statistical inference about the unknown survival probability (p_{0i}) can now be made using the standard errors in Table 3 and Table 4. For example, the estimate of the probability of surviving from birth to age 20 is $\hat{p}_{0,20} = .96060$ for the total United States population, 1960, and $\hat{p}_{0,20} = .96811$ for the total California population, 1970. To test for the significance of difference between these two probabilities, we compute the critical ratio

$$Z = \frac{\hat{p}_{0,20}(\text{U.S.}) - \hat{p}_{0,20}(\text{Cal.})}{\text{S.E.(diff.)}} . \tag{3.8}$$

The standard error of the difference is given by

$$\text{S.E.(diff.)} = \sqrt{(.96799\text{x}10^{-8}) + (8.28727\text{x}10^{-8})}$$

$$= 3.0422\text{x}10^{-4} . \tag{3.9}$$

Substituting the numerical values of $\hat{p}_{0,20}$ and (3.8) in (3.7),

$$Z = \frac{.96060 - .96966}{3.0422\text{x}10^{-4}} = \frac{-.00906}{3.0422\text{x}10^{-4}}$$

$$= -29.78 .$$

Based on the above findings, we conclude that a newborn who is subject to California 1970 mortality experience has a greater probability of surviving to age 20 than one who is subject to United States 1960 experience.

The converse is true, however, for the probability of surviving from age 20 to age 40. Table 5 shows that $p_{20,40}(\text{U.S.}) > p_{20,40}(\text{Cal.})$ and

$$\frac{p_{20,40}(\text{U.S.}) - p_{20,40}(\text{Cal.})}{\text{S.E.(diff.)}} = \frac{.96811 - .96641}{3.6487\text{x}10^{-4}} = 4.59 \ .$$

Remark 2. The formula for the estimate $\hat{p}_{ij}$ in (3.5) applies to both the current life table and the cohort life table. However, the formula for the variance of $\hat{p}_{ij}$ assumes different forms in the two cases. In a cohort life table, ℓ_j is the number of survivors at x_j of ℓ_i individuals living at x_i, with $\hat{p}_{ij}$ being the proportion of surviving the interval (x_i, x_j). Therefore, $\hat{p}_{ij} = \ell_j/\ell_i$ is a binomial proportion with a sample variance given by

$$S^2_{\hat{p}_{ij}} = \frac{1}{\ell_i}\hat{p}_{ij}(1-\hat{p}_{ij}) \ . \tag{3.10}$$

Formula (3.10) for the sample variance of $\hat{p}_{ij}$ is equal to (3.7) in the cohort life table, where the sample variance of the proportion for each age interval $\hat{p}_h$ is computed from

$$S^2_{\hat{p}_h} = \frac{1}{\ell_h}\hat{p}_h(1-\hat{p}_h) = \frac{1}{\ell_h}\frac{\ell_{h+1}}{\ell_h}\left(1 - \frac{\ell_{h+1}}{\ell_h}\right) \tag{3.11}$$

for a cohort life table, then formula (3.7) will be reduced to formula (3.10). Substituting $\hat{p}_{ij} = \ell_j/\ell_i$, $\hat{p}_h = \ell_{h+1}/\ell_h$ and (3.11) in (3.6), we have

$$\begin{aligned} S^2_{\hat{p}_{ij}} &= \left(\frac{\ell_j}{\ell_i}\right)^2 \sum_{h=i}^{j-1} \left(\frac{\ell_h}{\ell_{h+1}}\right)^2 \frac{1}{\ell_h}\frac{\ell_{h+1}}{\ell_h}\left[1 - \frac{\ell_{h+1}}{\ell_h}\right] \\ &= \left(\frac{\ell_j}{\ell_i}\right)^2 \sum_{h=i}^{j-1} \frac{1}{\ell_{h+1}}\left[1 - \frac{\ell_{h+1}}{\ell_h}\right] \\ &= \frac{1}{\ell_i}\frac{\ell_j}{\ell_i}\left[1 - \frac{\ell_j}{\ell_i}\right] = \frac{1}{\ell_i}\hat{p}_{ij}(1-\hat{p}_{ij}) \ , \end{aligned} \tag{3.12}$$

is required to be shown.

Table 5

Statistical test for the significance of difference between survival probabilities of United States population, 1960, and California population, 1970.

Age Interval	United States 1960		California 1970		Difference*	
(x_i, x_j)	$\hat{p}_{ij}$	$10^4 S_{\hat{p}_{ij}}$	p_{ij}	$10^4 S_{\hat{p}_{ij}}$	(2)-(4)	10^4S.E.(diff.)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
(0, 20)	.96060	.98386	.96966	2.87876	-.00906	3.0422
(20, 40)	.96811	1.14104	.96641	3.4657	+.00170	3.6487

*Formula for the standard error of the difference, $\hat{p}_{ij}$(U.S.) - $\hat{p}_{ij}$(Cal.):

$$\text{S.E.(diff.)} = \sqrt{S^2_{\hat{p}_{ij}}(\text{U.S.}) + S^2_{\hat{p}_{ij}}(\text{Cal.})}$$

4. Expectation of Life at Age x_α, e_α

Expectation of life at a given age is the mean future lifetime beyond this age. In the life table, there are ℓ_α individuals living at age x_α. Let the lifetime beyond x_α of these individuals be denoted by $Y_{\alpha k}$, for $k=1,\ldots,\ell_\alpha$. Their mean value

$$\bar{Y}_\alpha = \frac{1}{\ell_\alpha} \sum_{k=1}^{\ell_\alpha} Y_{\alpha k} \tag{4.1}$$

is approximately normally distributed, with an expected value of e_α. This sample mean $\bar{Y}_\alpha$ is equal to the observed expectation of life $\hat{e}_\alpha$, or

$$\bar{Y}_\alpha = \hat{e}_\alpha \; . \tag{4.2}$$

We now show that equation (4.2) is indeed true.

As any continuous random variable, lifetime of an individual is not accurately measured. In fact, the values of the ℓ_α values are not individually recorded in the life table, but grouped in the form of a frequency distribution in which the ages x_i and x_{i+1} are the lower and upper limits for the interval i, and the deaths, d_i, are the corresponding frequencies for $i = \alpha, \alpha+1,\ldots,w$. The sum of the frequencies equals the number of survivors at age x_α, or

$$d_\alpha + \ldots + d_w = \ell_\alpha \; . \tag{4.3}$$

The total number of years remaining to the ℓ_α survivors depends on the exact age at which death occurs, that is, on the distribution of deaths within each age interval. Suppose that the distribution of death in the interval (x_i, x_{i+1}) is such that, on the average, each of the d_i individuals lives a fraction a_i of the interval, or $a_i n_i$ years in the interval (since $x_{i+1}-x_i = n_i$ is the interval length). Each thus lives

$x_i + a_i n_i$ years, or $x_i - x_\alpha + a_i n_i$ years after age x_α, and the sample mean is given by

$$\bar{Y}_\alpha = \frac{1}{\ell_\alpha} \sum_{i=\alpha}^{w} (x_i - x_\alpha + a_i n_i)\, d_i$$

$$= \frac{1}{\ell_\alpha} \left[\sum_{i=\alpha}^{w} (x_i - x_\alpha)\, d_i + \sum_{i=\alpha}^{w} a_i\, n_i\, d_i \right] \quad . \tag{4.4}$$

By definition

$$x_i - x_\alpha = n_\alpha + n_{\alpha+1} + \ldots + n_{i-1} = \sum_{j=\alpha}^{i-1} n_j \quad , \tag{4.5}$$

hence

$$\sum_{i=\alpha}^{w} (x_i - x_\alpha)\, d_i = \sum_{i=\alpha}^{w} \left(\sum_{j=\alpha}^{i-1} n_j \right) d_i$$

$$= \sum_{j=\alpha}^{w-1} n_j \sum_{i=j+1}^{w} d_i \quad . \tag{4.6}$$

Since the number of individuals living at age x_j will all eventually die,

$$\ell_j = d_j + d_{j+1} + \ldots + d_w \ , \tag{4.7}$$

or

$$\ell_j - d_j = d_{j+1} + \ldots + d_w = \sum_{i=j+1}^{w} d_i \ . \tag{4.8}$$

Therefore (4.6) may be rewritten

$$\sum_{i=\alpha}^{w} (x_i - x_\alpha)\, d_i = \sum_{j=\alpha}^{w-1} n_j (\ell_j - d_j) \ . \tag{4.9}$$

Substituting (4.9) in (4.4) gives

$$\overline{Y}_\alpha = \frac{1}{\ell_\alpha}\left[\sum_{j=\alpha}^{w-1} n_j(\ell_j - d_j) + \sum_{i=\alpha}^{w} a_i n_i\, d_i\right]$$

$$= \frac{1}{\ell_\alpha}\left[\sum_{i=\alpha}^{w-1} \{n_i(\ell_i - d_i) + a_i n_i\, d_i\} + a_w n_w\, d_w\right]. \tag{4.10}$$

The quantity inside the braces, for $i = \alpha,\ldots,w-1$,

$$n_i(\ell_i - d_i) + a_i n_i\, d_i = L_i\ , \tag{4.11}$$

is the number of years lived by the ℓ_i individuals in the interval $(x_i,\ x_{i+1})$. Also we let

$$L_w = a_w n_w\, d_w = a_w n_w\, \ell_w \tag{4.12}$$

be the number of years lived by ℓ_α beyond age x_w. Using (4.11) and (4.12) we rewrite (4.10) is

$$\overline{Y}_\alpha = \frac{L_\alpha + L_{\alpha+1} + \ldots + L_w}{\ell_\alpha} \tag{4.13}$$

which, of course, is $\hat{e}_\alpha$, the observed expectation of life at age x_α, proving (4.2).

4.1 Formula for the variance of the expectation of life. Once the equality between the observed expectation of life $\hat{e}_\alpha$ and the sample mean future lifetime $\overline{Y}_\alpha$ is established, the sample variance of $\hat{e}_\alpha$ can easily be computed. We may visualize the age and death columns in a cohort life table as a frequency distribution, with $x_i - x + a_i n_i$ being the average value and d_i the corresponding frequency so that the sample variance of Y_α is given by

$$S^2_{Y_\alpha} = \frac{1}{\ell_\alpha} \sum_{i=\alpha}^{w} [(x_i - x_\alpha + a_i n_i) - \hat{e}_\alpha]^2 \, d_i \ . \tag{4.14}$$

Consequently, we have the formula for the sample variance of $\bar{Y}_\alpha$ (or $\hat{e}_\alpha$)

$$S^2_{\hat{e}_\alpha} = S^2_{\bar{Y}_\alpha} = \frac{1}{\ell_\alpha} S^2_{Y_\alpha} \ , \tag{4.15}$$

or, by substitution of (4.14),

$$S^2_{\hat{e}_\alpha} = \frac{1}{\ell^2_\alpha} \sum_{i=\alpha}^{w} [(x_i - x_\alpha + a_i n_i) - \hat{e}_\alpha]^2 \, d_i \ . \tag{4.16}$$

In formula (4.14), a_i, n_i, $\hat{e}_\alpha$, d_i and ℓ_α are all given in a life table; the sample variance of $\hat{e}_\alpha$ can be determined.

Formula (4.14), however, is not applicable for the current life table for a number of reasons. First of all, figures d_i and ℓ_α are dependent upon the choice of the radix ℓ_0, and therefore are not meaningful quantities when they appear without reference to ℓ_0. Secondly, basic variables in a current life table are the $\hat{q}_i$. Therefore, the sample variance estimates of $\hat{e}_\alpha$ should be expressed in terms of the variance of $\hat{q}_i$.

Formula (4.10) for the observed expectation of life, with the substitution of $\ell_{j+1} = \ell_j - d_j$ and $\ell_j - \ell_{j+1} = d_j$, may be rewritten as

$$\hat{e}_\alpha = \frac{1}{\ell_\alpha} \left[\sum_{j=\alpha}^{w-1} \{ n_j \ell_{j+1} + a_j n_j (\ell_j - \ell_{j+1}) \} + a_w n_w d_w \right] \tag{4.15}$$

or

$$\hat{e}_\alpha = \frac{1}{\ell_\alpha} \left[a_\alpha n_\alpha \ell_\alpha + \sum_{j=\alpha+1}^{w} [(1-a_{j-1}) n_{j-1} + a_j n_j] \, \ell_j \right] \ . \tag{4.16}$$

Now we let

$$c_j = (1-a_{j-1}) n_{j-1} + a_j n_j \tag{4.17}$$

and write

$$\hat{e}_\alpha = a_\alpha n_\alpha + \frac{c_{\alpha+1}\,\ell_{\alpha+1} + \dots + c_w \ell_w}{\ell_\alpha} \tag{4.18}$$

or, since $\ell_j/\ell_\alpha = \hat{p}_{\alpha j}$,

$$\hat{e}_\alpha = a_\alpha n_\alpha + \sum_{j=\alpha+1}^{w} c_j\,\hat{p}_{\alpha j} \quad . \tag{4.19}$$

Thus, the observed expectation of life $\hat{e}_\alpha$ is a linear function of $\hat{p}_{\alpha j}$, which in the current life table is computed from

$$\hat{p}_{\alpha j} = \hat{p}_\alpha\,\hat{p}_{\alpha+1} \cdots \hat{p}_{j-1}\ , \qquad j = \alpha+1, \dots, w\ . \tag{4.20}$$

Clearly, the derivatives of $\hat{p}_{\alpha j}$ with respect to $\hat{p}_i$ is given by

$$\begin{aligned} \frac{\partial}{\partial \hat{p}_i}\,\hat{p}_{\alpha j} &= \hat{p}_{\alpha i}\,\hat{p}_{i+1,j} \qquad \text{for } \alpha \le i < j \\ &= 0 \qquad\qquad\quad \text{otherwise} \quad . \end{aligned} \tag{4.21}$$

Hence, from (4.19)

$$\begin{aligned} \frac{\partial}{\partial \hat{p}_i}\,\hat{e}_\alpha &= \sum_{j=\alpha+1}^{w} c_j \frac{\partial}{\partial \hat{p}_i}\,\hat{p}_{\alpha j} \\ &= \sum_{j=i+1}^{w} c_j\,\hat{p}_{\alpha i}\,\hat{p}_{i+1,j} \\ &= p_{\alpha i}\left[c_{i+1} + \sum_{j=i+2}^{w} c_j\,\hat{p}_{i+1,j}\right] \quad . \end{aligned} \tag{4.22}$$

Using relation (4.19), or

$$\hat{e}_{i+1} = a_{i+1}n_{i+1} + \sum_{j=i+2}^{w} c_j\, p_{i+1,j}\ , \tag{4.23}$$

and

$$c_{i+1} = (1-a_i)n_i + a_{i+1}\, n_{i+1} \tag{4.24}$$

we rewrite the derivative in (4.22) as follows

$$\frac{\partial}{\partial \hat{p}_i}\ \hat{e}_\alpha = \hat{p}_{\alpha i}\ [(1-a_i)n_i + \hat{e}_{i+1}]\ . \tag{4.25}$$

Because of (4.21), the derivative (4.25) vanishes when i = w. Now the estimated probabilities ($\hat{p}_i$) for two nonoverlapping age intervals are based on mortality experience of two distinct groups of people, and therefore are not correlated. Consequently, the variance of the expectation of life may be computed from the following

$$S^2_{\hat{e}_\alpha} = \sum_{i=\alpha}^{w-1} \left\{\frac{\partial}{\partial \hat{p}_i}\ \hat{e}_\alpha\right\}^2\ S^2_{\hat{p}_i}\ . \tag{4.26}$$

Substituting (4.25) in (4.26) yields the desired formula for the sample variance of $\hat{e}_\alpha$:

$$S^2_{\hat{e}_\alpha} = \sum_{i=\alpha}^{w-1} \hat{p}^2_{\alpha i}\ [(1-a_i)n_i + \hat{e}_{i+1}]^2\ S^2_{\hat{p}_i} \tag{4.27}$$

where the variance of $\hat{p}_i$ is given in (2.2).

$$S^2_{\hat{p}_i} = \frac{\hat{q}_i^2(1-\hat{q}_i)}{D_i}\ . \tag{2.2}$$

4.2 <u>Computation of the variance of the expectation of life in a current life table</u>. Formula (4.27), which holds true for any age x_α in the life table, contains terms that appear repeatedly for different values of α. Therefore, the variances of $\hat{e}_\alpha$ for all ages x_α in the life table can be calculated by a single computation program. Using formula (4.27) and referring to Table 6, the essential steps in the computation of the sample variance of $\hat{e}_\alpha$ are as follows.

1. Designate age interval in Column 1.

2. Record the length of age interval n_i in Column 2, and the fraction of last age interval of life a_i in Column 3.

3. Compute the sample variance of $\hat{p}_i$ $(\hat{q}_i)$ from formula (2.2) and record it in Column 4.

4. Compute for each age interval the quantity

$$\ell_i^2\ [(1-a_i)n_i + \hat{e}_{i+1}]^2\ S^2_{\hat{p}_i}$$

and record it in Column 5.

5. Sum the products in Column 5 from the bottom of the table up to x_α and enter the sum in Column 6.

6. Divide the sum in Column 6 by ℓ_i^2 to obtain the sample variance of the observed expectation of life in Column 7.

7. Take the square root of the sample variance to obtain the sample standard error of the observed expectation of life, as shown in Column 8.

4.3 Statistical inference about expectation of life. An observed expectation of life, as shown earlier in this section, is a sample mean of future lifetime. Therefore, statistical tests based on normal distribution may be used in making inference regarding expectation of life at a particular age, or in comparing expectation of life of two or more populations. In Table 7 the expectations of life for the United States 1960 population are compared with those for the California 1970 population. For each age the expectation of life and the standard errors are recorded in Columns 2 through 5. The difference of the expectations is given in Column 6.

The standard error of the difference computed from

$$\text{S.E.(diff.)} = \sqrt{S^2_{\hat{e}_i}(\text{Cal.}) + S^2_{\hat{e}_i}(\text{U.S.})} \qquad (4.28)$$

is recorded in Column (7). The ratio of the difference to the **corresponding** standard error is recorded in Column 8.

The critical ratio for each age **far exceeds** the critical value of 2.33 in the normal distribution corresponding to the one percent level of significance. This means that a person of any age, who is subject to the California 1970 mortality experience, has a greater expectation of life than one who is subject to the United States 1960 experience.

Table 6

Computation of the sample variance of the observed expectation of life

Total U.S. Population, 1960

Age interval (in years)	Length of interval	Fraction of last age interval of life	Sample variance of $\hat{p}_i$			Sample variance of $\hat{e}_i$	Sample Standard Error of $\hat{e}_i$
x_i to x_{i+1}	n_i	a_i	$10^8 S^2_{\hat{p}_i}$	$\ell_i^2[(1-a_i)n_i+\hat{e}_{i+1}]^2 S^2_{\hat{p}_i}$	$\sum_{j>i} \ell_j^2[(1-a_j)n_j + \hat{e}_{j+1}]^2 S^2_{\hat{p}_j}$	$10^4 S^2_{\hat{e}_i}$	$S_{\hat{e}_i}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0-1	1	.10	.6043	308,328.66	1,384,473.52	1.3845	.012
1-5	4	.39	.1070	48,684.08	1,076,144.86	1.1349	.011
5-10	5	.46	.0653	25,686.27	1,027,460.78	1.0931	.010
10-15	5	.54	.0649	21,433.28	1,001,774.51	1.0710	.010
15-20	5	.57	.1714	47,445.51	980,341.23	1.0527	.010
20-25	5	.49	.2825	65,770.32	932,895.72	1.0110	.010
25-30	5	.50	.2971	55,984.55	867,125.40	.9514	.010
30-35	5	.52	.3307	49,278.91	811,140.85	.9017	.009
35-40	5	.54	.4553	52,293.09	761,861.94	.8606	.009
40-45	5	.54	.7739	66,833.91	709,568.85	.8205	.009
45-50	5	.54	1.2721	79,526.83	642,734.94	.7713	.009
50-55	5	.53	2.2009	95,654.42	563,208.11	.7169	.008
55-60	5	.52	3.4613	97,872.06	467,553.69	.6535	.008
60-65	5	.52	6.1531	105,623.14	369,681.63	.5920	.008
65-70	5	.52	9.4489	87,635.79	264,058.49	.5250	.007
70-75	5	.51	16.1842	71,425.04	176,422.70	.4828	.007
75-80	5	.51	30.5659	52,370.93	104,997.66	.4660	.007
80-85	5	.48	62.1657	34,293.39	52,626.73	.4946	.007
85-90	5	.45	120.9280	13,802.48	18,333.34	.5977	.008
90-95	5	.41	261.7560	4,530.86	4,530.86	.9932	.010

Table 7

Expectation of life and the standard error, total United States Population, 1960, and total California population, 1970.

Age Interval (x_i, x_{i+1})	United States $\hat{e}_i$	United States $S_{\hat{e}_i}$	California $\hat{e}_i$	California $S_{\hat{e}_i}$	Difference $\hat{e}_i(\text{Cal.})-\hat{e}_i(\text{U.S.})$ (4)-(2)	Difference S.E.(diff.)	Critical Ratio * $\frac{\hat{e}_i(\text{Cal.})-\hat{e}_i(\text{U.S.})}{\text{S.E.(diff.)}}$ (6) ÷ (7)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0-1	69.65	.012	71.95	.037	2.30	.039	59.0
1-5	70.53	.011	72.27	.034	1.74	.036	48.3
5-10	66.83	.010	68.50	.033	1.67	.035	47.7
10-15	61.99	.010	63.62	.033	1.63	.035	46.6
15-20	57.12	.010	58.74	.033	1.62	.035	46.3
20-25	52.37	.010	54.05	.032	1.68	.034	49.4
25-30	47.68	.010	49.46	.032	1.78	.033	53.9
30-35	42.97	.009	44.79	.031	1.82	.033	55.2
35-40	38.30	.009	40.13	.030	1.83	.032	57.2
40-45	33.72	.009	35.56	.030	1.84	.031	59.4
45-50	29.30	.009	31.12	.030	1.82	.030	60.7
50-55	25.09	.008	26.90	.029	1.81	.029	62.4
55-60	21.17	.008	22.92	.028	1.75	.028	62.5
60-65	17.48	.008	19.27	.027	1.79	.027	66.3
65-70	14.18	.007	15.89	.026	1.71	.026	65.8
70-75	11.18	.007	12.87	.024	1.69	.024	70.4
75-80	8.54	.007	10.13	.023	1.59	.023	69.1
80-85	6.27	.007	7.94	.021	1.69	.020	84.5

*Formula for the standard error of the difference $\hat{e}_i(\text{Cal.}) - \hat{e}_i(\text{U.S.})$

$$\text{S.E.(diff.)} = \sqrt{S^2_{\hat{e}_i(\text{Cal.})} + S^2_{\hat{e}_i(\text{U.S.})}}$$

CHAPTER 7

MULTIPLE DECREMENT TABLE FOR A CURRENT POPULATION

1. Introduction

The multiple decrement table is not only a useful means of summarizing mortality experience of a defined population subject to several risks of dying, but also a powerful analytical tool for the study of decrement data. The concept of multiple decrement originates in the investigation of component causes of death; however, it has applications in many research fields. In the actuarial sciences, for example, disability and mortality are distinct causes of claim, and the effects of exposure to both causes and their interaction must be analyzed in a meaningful way. Dissolution of a marriage may be because of death occurring to either one of the partners or because of divorce. Here there are three forces of decrement; death to the male, death to the female and divorce. Similarly, survival of an enterprise is subject to many forces of decrement and their interacting effects. In spite of numerous applications of this methodology, the most important use of multiple decrement tables still remains to be in the study of mortality.

The multiple decrement table is directly related to the theory of competing risks presented in Appendix III. The theory has been developed to evaluate the forces of mortality of competing risks under investigation. According to the theory, there are three types of probability of death with respect to a particular risk or risks.

1.1. Crude probability: The probability of death from a specific cause in the presence of competition of all other risks acting in a population.

1.2. Net probability: The probability of death if a specific risk is the only risk in effect in the population or, conversely, the probability of death if a specific risk is eliminated from the population.

1.3. Partial crude probability: The probability of death from a specific cause when another risk (or risks) is eliminated from the population.

Detailed discussion of these probabilities is presented in Appendix III. Clarification should be made of the terms "risk" and "cause." Both terms may refer to the same condition but are different on the time scale relative to the occurrence of death. Prior to death the condition in question is a risk; after death the condition is a cause (provided, of course, this is the condition from which an individual dies). We shall take up this point again in Appendix III.

An ordinary multiple decrement table contains only the crude probability of death for selected causes covering the entire life span of a well defined population. For easy comparison, the probability of death q_i without referring to causes is often included. Let $Q_{i\delta}$ be the probability that an individual alive at exact age x_i will die in interval (x_i, x_{i+1}) from cause R_δ in the presence of all other competing risks, for $i = 0,1,\ldots,w$; $\delta = 1,\ldots,r$. A typical decrement table is given in Table 1 on page 5-3.

There are two types of multiple decrement tables for the analysis of human mortality: The cohort multiple decrement table and the current multiple decrement table. As in the case of the life table, a cohort multiple decrement table records the mortality experience of a well-defined cohort of people from the birth of each person to the death of the last person of the group. When a cohort of people is subject to a number of risks of dying, there will be deaths from each of these risks within every age interval of life. The number of deaths from a specific cause (say R_δ) in an interval (x_i, x_{i+1}) divided by the number of individuals alive at the

Table 1

Multiple Decrement Table - The Crude Probability of Dying ($Q_{i\delta}$) from a specific cause (R_δ) in Age Interval (x_i, x_{i+1})

Age Interval (x_i, x_{i+1})	Probability of Dying in Interval (x_i, x_{i+1})	Causes of Death			
		R_1	R_2	...	R_r
0-1	q_0	Q_{01}	Q_{02}	...	Q_{0r}
1-5	q_1	Q_{11}	Q_{12}	...	Q_{1r}
⋮	⋮	⋮	⋮	⋮	⋮
$x_i - x_{i+1}$	q_i	Q_{i1}	Q_{i2}	...	Q_{ir}
⋮	⋮	⋮	⋮	⋮	⋮
x_w & over	q_w	Q_{w1}	Q_{w2}	...	Q_{wr}

beginning of the interval is an estimate of the (crude) probability of dying from cause R_δ, denoted by $\hat{Q}_{i\delta}$. This estimate is simply the proportion of individuals dying from a specific cause in a defined age interval. An aggregate of these proportions for different causes of death over all age intervals forms a cohort multiple decrement table. A detailed discussion and theoretical aspects of the table may be found in Appendix IV of this manual.

A current multiple decrement table, which is more useful for practical purposes, is the one derived from the mortality experience of a population of all ages over a short period of time, such as one year. The appearance of this table is exactly the same as the cohort multiple decrement table, but differs from the latter in the basic information from which the table is constructed. Specifically, the data for the current decrement table are the number of deaths from different causes and the corresponding mid-year population for each age group over the entire life span of a current population, from which age-and-cause specific death rates are computed. These rates in turn are then used to compute the estimate of the (crude) probability of dying ($\hat{Q}_{i\delta}$) from each cause R_δ. A brief description is presented below.

2. Computation of the Crude Probability, $\hat{Q}_{i\delta}$

Let us first reintroduce the symbols used in Chapter 3. For age interval (x_i, x_{i+1}) we let $n_i = x_{i+1}-x_i$ be the length of the interval, P_i the midyear population, D_i the number of deaths occurring during the calendar year, a_i the fraction of the last age interval lived by each of the D_i individuals, and N_i the number of people alive at x_i among whom D_i deaths occur. The age-specific death rate is defined by the ratio of D_i to the number of years lived by the N_i people in the interval (x_i, x_{i+1}), or

$$M_i = \frac{D_i}{(N_i - D_i)n_i + a_i\, n_i\, D_i} \tag{2.1}$$

When the denominator is estimated with the mid-year population, P_i,

$$(N_i - D_i)n_i + a_i\, n_i\, D_i = P_i \tag{2.2}$$

we have

$$M_i = \frac{D_i}{P_i} \quad . \tag{2.3}$$

The probability of dying in the interval (x_i, x_{i+1}) is estimated by

$$\hat{q}_i = \frac{D_i}{N_i} \quad , \tag{2.4}$$

where N_i can be derived from (2.2), or

$$N_i = \frac{1}{n_i}\, P_i[1 + (1-a_i)n_i\, M_i] \quad . \tag{2.2a}$$

Substituting (2.2a) in (2.4) gives

$$\hat{q}_i = \frac{n_i M_i}{1 + (1-a_i)n_i\, M_i} \quad . \tag{2.5}$$

The D_i deaths are now further divided according to cause with $D_{i\delta}$ dying from cause R_δ, $\delta = 1,\ldots,r$, and

$$D_i = D_{i1} + \ldots + D_{ir} \tag{2.6}$$

so that

$$M_{i\delta} = \frac{D_{i\delta}}{P_i} \qquad , \delta = 1,\ldots,r, \tag{2.7}$$

are age-cause-specific death rates. The estimate of the crude probability of dying from R_δ in the presence of competing risks is obviously

$$\hat{Q}_{i\delta} = \frac{D_{i\delta}}{N_i} \quad . \tag{2.8}$$

Substituting (2.2a) in (2.8) gives the formula for the crude probability

$$\hat{Q}_{i\delta} = \frac{n_i M_{i\delta}}{1+(1-a_i)n_i M_i} \quad . \tag{2.9}$$

We see from (2.4) and (2.8) that $Q_{i\delta}$ can be computed also from

$$\hat{Q}_{i\delta} = \frac{D_{i\delta}}{D_i} \hat{q}_i . \tag{2.9a}$$

It is easy to show that $\hat{Q}_{i\delta}$ in (2.9a) and $\hat{q}_i$ in (2.4) satisfy the relationship

$$\hat{Q}_{i1} + \dots + \hat{Q}_{ir} = \hat{q}_i . \tag{2.10}$$

The formula for the sample variance of the estimator $\hat{Q}_{i\delta}$ can be derived from

$$\mathrm{Var}(\hat{Q}_{i\delta}) = \frac{1}{N_i} Q_{i\delta}(1 - Q_{i\delta}) \tag{2.11}$$

by substituting $\hat{Q}_{i\delta}$ for $Q_{i\delta}$ and using (2.8):

$$\mathrm{Var}(\hat{Q}_{i\delta}) = \frac{1}{N_i} \hat{Q}_{i\delta}(1 - \hat{Q}_{i\delta}) = \frac{1}{D_{i\delta}} \hat{Q}^2_{i\delta}(1-\hat{Q}_{i\delta}). \tag{2.12}$$

The standard deviation of $\hat{Q}_{i\delta}$ is the square root of the variance in (2.12).

The steps involved in constructing a multiple decrement table may be summarized as follows:

2.1. Information needed from a current population.

(a) Number of deaths in each age interval (x_i, x_{i+1}) from each cause R_δ, $D_{i\delta}$, and the total number of deaths D_i, with

$$D_i = D_{i1} + \dots + D_{ir} . \tag{2.6}$$

(b) Mid-year population for each age interval (x_i, x_{i+1}), P_i

(c) The fraction of the last age interval of life, a_i, as given in Appendix V.

2.2. Computation of rates and probabilities

(a) Age-cause-specific death rate :

$$M_{i\delta} = \frac{D_{i\delta}}{P_i} \quad , \tag{2.7}$$

for each age and each cause; and the age-specific death rate:

$$M_i = \frac{D_i}{P_i} \quad . \tag{2.3}$$

(b) The probability of dying :

$$\hat{q}_i = \frac{n_i M_i}{1 + (1-a_i) n_i M_i} \quad , \tag{2.5}$$

and the crude probability of dying from R_δ for age interval (x_i, x_{i+1}) :

$$\hat{Q}_{i\delta} = \frac{D_{i\delta}}{D_i} \hat{q}_i \quad . \tag{2.9a}$$

2.3. Computation of the standard deviation:

$$S.D.(\hat{Q}_{i\delta}) = \sqrt{\frac{1}{D_{i\delta}} \hat{Q}^2_{i\delta} (1-\hat{Q}_{i\delta})} \tag{2.13}$$

and

$$S.D.(\hat{q}_i) = \sqrt{\frac{1}{D_i} \hat{q}^2_i (1-\hat{q}_i)} \quad . \tag{2.14}$$

To illustrate the computation, let us consider as an example the competing risks of death: cardiovascular diseases (R_1), cancer all forms (R_2), all accidents (R_3), infectious diseases (R_4), respiratory diseases (R_5), motor vehicle accidents (R_6), and all other causes (R_7) in the Sweden population age group (1,5), 1967 in Table 2. During 1967 there were a total of $D_1 = 250$ deaths occurring in the Sweden population between age one and five; and a mid-year population of $P_1 = 471{,}119$. The number of deaths, $D_1 = 250$, is entered in column (2) to the right of "all causes". Dividing D_1 by the mid-year population $P_1 = 471{,}119$ gives the age specific death rate $M_1 = .000531$ entered in column (3) on the same line. The fraction of last age interval of life is $a_1 = .43$ in this case and the age interval is $n_1 = 4$ years. With these values, we compute the probability of dying for this age interval as before from

$$\hat{q}_i = \frac{n_i M_i}{1+(1-a_i)n_i M_i}$$

$$\hat{q}_1 = \frac{4(.000531)}{1+(1-.43)4(.000531)} = .002121$$

which is recorded in column (4). The number of deaths are then identified by cause, with $D_{11} = 4$ deaths from cardiovascular diseases, etc. These numbers are entered in column (2). Dividing each of these values by the mid-year population P_1, we obtain the death rate specific for the cause in question as shown in column (3). The crude probability of dying from a specific cause when all other competing causes are acting may be computed from the corresponding death rate. But it is more convenient to use the relation

$$\hat{Q}_{i\delta} = \frac{D_{i\delta}}{D_i}\hat{q}_i\,, \qquad \text{for } \delta = 1,2,\ldots,r \quad . \tag{2.9a}$$

Thus, for R_2: cancer all forms, for example,

$$\hat{Q}_{12} = \frac{D_{12}}{D_1} \hat{q}_1 = \frac{46}{250} \; .002121 = .000390$$

These crude probabilities are shown in column (4).

The standard errors of these probabilities are computed from

$$S.D.(\hat{q}_i) = \sqrt{\frac{1}{D_i} \hat{q}_i^2(1-\hat{q}_i)} \quad = .0001340 \tag{2.14}$$

and

$$S.D.(\hat{Q}_{i\delta}) = \sqrt{\frac{1}{D_{i\delta}} \hat{Q}_{i\delta}^2(1-\hat{Q}_{i\delta})} \tag{2.13}$$

for $\delta = 1,2,\ldots,7$. The numerical values are recorded in column (6) of Table 2.

Such computations, which can be carried out easily with computers, are needed for each age group. The basic data, the mid-year population and the number of deaths by age interval and cause, for Sweden and Australia are given in Tables 3 and 4 respectively.

Table 2

Computation of the crude probability of dying from a specific cause and the corresponding standard error. Sweden population, age interval (1, 5), 1967

Cause of death R_δ (1)	Number of deaths[1] $D_{i\delta}$ (2)	Cause Specific Death Rate $M_{i\delta}$ (3)	Crude Probability of dying[2] $\hat{Q}_{i\delta}$ (4)	Standard deviation[3] S.D. $(\hat{Q}_{i\delta})$ (5)
All causes[4]	250	.000531	.002121	.0001340
R_1: Cardio-vascular diseases	4	.000008	.000034	.0000169
R_2: Cancer all forms	46	.000098	.000390	.0000575
R_3: All accidents	68	.000144	.000577	.0000700
R_4: Infectious diseases	14	.000030	.000119	.0000318
R_5: Respiratory diseases	37	.000079	.000314	.0000516
R_6: Motor vehicle accidents	19	.000040	.000161	.0000369
R_7: All other causes	81	.000172	.000687	.0000763

1/ $\hat{q}_i = \dfrac{n_i M_i}{1+(1-a_i)n_i M_i}$

2/ $\hat{Q}_{i\delta} = \dfrac{D_{i\delta}}{D_i}\hat{q}_i$

3/ $\text{S.D.} = \sqrt{\dfrac{1}{D_{i\delta}}\hat{Q}^2_{i\delta}(1-\hat{Q}_{i\delta})}$

4/ $D_1 = 250$
$\hat{q}_1 = 0.002121$

Table 3

Mid-year population and deaths by age and cause
Sweden, 1967

Age	Population	Total Deaths	Fraction of last age interval of life	Cause of Death: CVD R_1	Cancer R_2	All Accidents R_3	Infect. Diseases R_4	Respirat. Diseases R_5	Motor veh. Accidents R_6	All Other Causes R_7
(x_i, x_i+n_i)	P_i	D_i	a_i	D_{i1}	D_{i2}	D_{i3}	D_{i4}	D_{i5}	D_{i6}	D_{i7}
0-1	120905	1560	0.08	7	12	26	14	54	4	1447
1-5	471119	250	0.43	4	46	68	14	37	19	81
5-10	522261	171	0.45	5	31	80	3	7	39	45
10-15	534756	148	0.52	8	29	57	2	7	38	45
15-20	589158	318	0.56	22	21	187	8	11	135	69
20-25	656338	508	.050	23	53	226	6	10	136	190
25-30	510785	476	0.52	27	65	146	4	7	61	227
30-35	445412	517	0.53	47	89	128	6	8	52	239
35-40	462977	683	0.53	95	143	149	13	16	54	267
40-45	506480	1157	0.53	228	313	143	22	27	42	424
45-50	543670	1853	0.54	482	559	197	25	57	62	533
50-55	516154	2724	0.54	912	828	229	32	78	68	645
55-60	511489	4266	0.53	1742	1305	220	59	119	85	821
60-65	446800	6189	0.53	2905	1809	235	54	201	98	985
65-70	373773	8770	0.54	4625	2236	210	72	376	84	1251
70-75	286391	11339	0.53	6501	2463	198	75	646	71	1456
75-80	196498	13715	0.52	8225	2376	272	74	1019	70	1749
80-85	113212	12766	0.50	8042	1669	278	73	1197	28	1507
85+	59753	12373		8086	1036	351	48	1375	10	1477

Table 4

Mid-year population and deaths by age and cause - Australia, 1967

Age	Population	Total Deaths	Fraction of last age interval of life	Cause of Death						
				CVD	Cancer	All Accidents	Infect. Diseases	Respirat. Diseases	Motor veh. Accidents	All Other Causes
				R_1	R_2	R_3	R_4	R_5	R_6	R_7
(x_i, x_i+n_i)	P_i	D_i	a_i	D_{i1}	D_{i4}	D_{i3}	D_{i4}	D_{i5}	D_{i6}	D_{i7}
0-1	225600	4187	0.12	18	17	140	56	401	18	3555
1-5	925500	845	0.49	10	75	292	46	115	116	307
5-10	1198500	457	0.41	10	94	194	13	33	117	113
10-15	1110100	364	0.48	20	57	171	7	8	91	101
15-20	1051500	962	0.42	33	80	641	10	23	501	175
20-25	930500	1081	0.43	51	78	655	7	22	504	268
25-30	773000	871	0.47	90	84	379	8	24	256	286
30-35	705800	903	0.35	140	140	278	6	23	156	316
35-40	754200	1393	0.48	334	257	322	13	44	168	423
40-45	778300	2461	0.53	811	511	404	26	90	195	619
45-50	701300	3543	0.56	1570	753	357	24	125	183	714
50-55	648300	5184	0.52	2501	1271	359	36	186	186	831
55-60	560300	7239	0.54	3881	1622	378	59	333	176	966
60-65	446400	9086	0.54	5170	2011	321	56	486	157	1042
65-70	361400	11368	0.54	6753	2270	294	74	703	149	1274
70-75	277700	13495	0.54	8472	2295	315	96	902	122	1415
75-80	200700	15116	0.53	9843	2176	357	59	1119	134	1562
80-85	105300	12585	0.54	8556	1450	345	29	1010	74	1195
85+	55800	11544		7949	928	398	26	1153	31	1090

3. Multiple Decrement Tables for Sweden and Australia Populations

Two multiple decrement tables have been computed for the Australian population, 1967, and the Swedish population, 1967 [Tables 5 and 6]. The selected causes are cardiovascular diseases (R_1), cancer all forms (R_2), all accidents (R_3), infectious diseases (R_4), respiratory diseases (R_5), motor vehicle accidents (R_6) and all other causes (R_7). The crude probability of dying from each specific cause has been computed for every age interval. In addition, the probability of dying $\hat{q}_i$ without reference to cause of death is included in the tables so that the magnitude of the probability $\hat{Q}_{i\delta}$ for each risk R_δ relative to the total probability $\hat{q}_i$ can be determined. For example, the ratio $\hat{Q}_{i\delta}/\hat{q}_i$ will give the proportionate mortality due to a specific risk R_δ.

It may be noted that if $\hat{Q}_{i\delta}$ is computed for every risk, then the sum of the probabilities $\hat{Q}_{i\delta}$ overall possible risks R_δ will be equal to $\hat{q}_i$ [c.f., equation (2.10)]. **The sum of $\hat{Q}_{i\delta}$ over only selected risks is less than $\hat{q}_i$.**

For the purpose of testing for significance between the probabilities or making other statistical inferences, the standard deviations of the probabilities are also included in the tables.

Table 5

Multiple Decrement Table for Selected Causes of Death and the Standard Error of the Crude Probability of Dying

Sweden population, 1967

Age Interval (in years) (x_i, x_{i+1})	Probability of dying in interval (x_i, x_{i+1})		Crude Probability of Dying in Interval (x_i, x_{i+1}): Cardiovascular Diseases R_1		Cancer All Forms R_2		All Accidents R_3	
	$\hat{q}_i$	$S_{\hat{q}_i}$	$\hat{Q}_{i1}$	$S_{\hat{Q}_{i1}}$	$\hat{Q}_{i2}$	$S_{\hat{Q}_{i2}}$	$\hat{Q}_{i3}$	$S_{\hat{Q}_{i3}}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0- 1	.01275	.000321	.0001	.000022	.0001	.000028	.0002	.000042
1- 5	.00212	.000134	.0000	.000017	.0004	.000058	.0005	.000070
5-10	.00164	.000125	.0000	.000021	.0003	.000053	.0008	.000086
10-15	.00138	.000114	.0001	.000025	.0003	.000050	.0005	.000071
15-20	.00270	.000151	.0002	.000040	.0002	.000039	.0016	.000116
20-25	.00386	.000171	.0002	.000036	.0004	.000055	.0017	.000114
25-30	.00465	.000213	.0003	.000051	.0006	.000079	.0014	.000118
30-35	.00579	.000254	.0005	.000077	.0010	.000106	.0014	.000127
35-40	.00735	.000280	.0010	.000105	.0015	.000129	.0016	.000131
40-45	.01136	.000332	.0022	.000148	.0031	.000173	.0014	.000117
45-50	.01691	.000389	.0044	.000200	.0051	.000215	.0018	.000128
50-55	.02607	.000493	.0087	.000288	.0079	.000274	.0022	.000145
55-60	.04090	.000613	.0167	.000397	.0125	.000344	.0021	.000142
60-65	.06708	.000824	.0315	.000575	.0196	.000456	.0025	.000166
65-70	.11131	.001120	.0587	.000837	.0284	.000592	.0027	.000184
70-75	.18111	.001539	.1038	.001219	.0393	.000777	.0032	.000224
75-80	.29871	.002137	.1793	.001791	.0518	.001034	.0059	.000358
80-85	.43982	.002913	.2771	.002627	.0575	.001366	.0096	.000572

Table 5 (con't)

Multiple Decrement Table for Selected Causes of Death and the Standard Error of the Crude Probability of Dying

Sweden population, 1967

Age Interval (in years) (x_i, x_{i+1})	Crude Probability of Dying in Interval (x_i, x_{i+1})							
	Infectious Diseases R_4		Respiratory Diseases R_5		Motor Vehicle Accidents R_6		All Other Causes R_7	
	$\hat{Q}_{i4}$	$S_{\hat{Q}_{i4}}$	$\hat{Q}_{i5}$	$S_{\hat{Q}_{i5}}$	$\hat{Q}_{i6}$	$S_{\hat{Q}_{i6}}$	$\hat{Q}_{i7}$	$S_{\hat{Q}_{i7}}$
(1)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
0- 1	.0001	.000031	.0004	.000060	.0000	.000016	.0118	.000309
1- 5	.0001	.000032	.0003	.000052	.0002	.000037	.0007	.000076
5-10	.0000	.000017	.0001	.000025	.0004	.000060	.0004	.000064
10-15	.0000	.000013	.0001	.000025	.0004	.000058	.0004	.000063
15-20	.0001	.000024	.0001	.000028	.0011	.000098	.0006	.000070
20-25	.0000	.000019	.0001	.000024	.0010	.000089	.0014	.000105
25-30	.0000	.000020	.0001	.000026	.0006	.000076	.0022	.000147
30-35	.0001	.000027	.0001	.000032	.0006	.000081	.0027	.000173
35-40	.0001	.000039	.0002	.000043	.0006	.000079	.0029	.000176
40-45	.0002	.000046	.0003	.000051	.0004	.000064	.0042	.000202
45-50	.0002	.000046	.0005	.000069	.0006	.000072	.0049	.000210
50-55	.0003	.000054	.0007	.000084	.0007	.000079	.0052	.000242
55-60	.0005	.000074	.0011	.000105	.0008	.000088	.0079	.000274
60-65	.0006	.000080	.0022	.000153	.0011	.000107	.0107	.000338
65-70	.0009	.000108	.0048	.000246	.0011	.000116	.0159	.000445
70-75	.0012	.000138	.0103	.000404	.0011	.000135	.0233	.000502
75-80	.0015	.000187	.0222	.000688	.0015	.000182	.0381	.000894
80-85	.0025	.000294	.0412	.001167	.0010	.000182	.0519	.001302

Table 6

Multiple Decrement Table for Selected Causes of Death and the Standard Error of the Crude Probability of Dying

Australia Population, 1967

Age Interval (in years) (x_i, x_{i+1})	Probability of dying in interval (x_i, x_{i+1})		Crude Probability of Dying in Interval (x_i, x_{i+1}): Cardiovascular Diseases R_1		Cancer All Forms R_2		All Accidents R_3	
	$\hat{q}_i$	$S_{\hat{q}_i}$	$\hat{Q}_{i1}$	$S_{\hat{Q}_{i1}}$	$\hat{Q}_{i2}$	$S_{\hat{Q}_{i2}}$	$\hat{Q}_{i3}$	$S_{\hat{Q}_{i3}}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0- 1	.01826	.000280	.0001	.000019	.0001	.000018	.0006	.000052
1- 5	.00365	.000125	.0000	.000014	.0003	.000037	.0013	.000074
5-10	.00190	.000089	.0000	.000013	.0004	.000040	.0008	.000058
10-15	.00164	.000086	.0001	.000020	.0003	.000034	.0008	.000059
15-20	.00456	.000147	.0002	.000027	.0004	.000042	.0030	.000120
20-25	.00579	.000176	.0003	.000038	.0004	.000047	.0035	.000137
25-30	.00562	.000190	.0006	.000061	.0005	.000059	.0024	.000125
30-35	.00637	.000211	.0010	.000083	.0010	.000083	.0020	.000118
35-40	.00919	.000245	.0022	.000120	.0017	.000106	.0021	.000118
40-45	.01569	.000314	.0052	.000181	.0033	.000144	.0026	.000128
45-50	.02498	.000414	.0111	.000278	.0053	.000193	.0025	.000133
50-55	.03923	.000534	.0189	.000375	.0096	.000268	.0027	.000143
55-60	.06274	.000714	.0336	.000531	.0141	.000347	.0033	.000168
60-65	.09722	.000969	.0553	.000748	.0215	.000475	.0034	.000191
65-70	.14667	.001271	.0871	.001013	.0293	.000606	.0038	.000221
70-75	.21855	.001063	.1372	.001385	.0372	.000761	.0051	.000287
75-80	.31995	.002146	.2083	.001368	.0461	.000964	.0076	.000398
80-85	.40873	.003045	.3187	.002844	.0540	.001379	.0129	.000687

Table 6 (con't)

Multiple Decrement Table for Selected Causes of Death and the Standard Error of the Crude Probability of Dying

Australia Population, 1967

Age Interval (in years) (x_i, x_{i+1})	Crude Probability of Dying in Interval (x_i, x_{i+1}): Infectious Diseases R_4		Respiratory Diseases R_5		Motor Vehicle Accidents R_6		All Other Causes R_7	
	$\hat{Q}_{i4}$	$S_{\hat{Q}_{i4}}$	$\hat{Q}_{i5}$	$S_{\hat{Q}_{i5}}$	$\hat{Q}_{i6}$	$S_{\hat{Q}_{i6}}$	$\hat{Q}_{i7}$	$S_{\hat{Q}_{i7}}$
(1)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
0- 1	.0002	.000033	.0017	.000087	.0001	.000019	.0155	.000258
1- 5	.0002	.000029	.0005	.000046	.0005	.000046	.0013	.000076
5-10	.0001	.000015	.0001	.000024	.0005	.000045	.0005	.000044
10-15	.0000	.000012	.0000	.000013	.0004	.000043	.0005	.000045
15-20	.0000	.000015	.0001	.000023	.0024	.000106	.0008	.000063
20-25	.0000	.000014	.0001	.000025	.0027	.000120	.0014	.000088
25-30	.0001	.000018	.0002	.000032	.0017	.000103	.0018	.000109
30-35	.0000	.000017	.0002	.000034	.0011	.000088	.0022	.000125
35-40	.0001	.000024	.0003	.000044	.0011	.000085	.0028	.000136
40-45	.0002	.000033	.0006	.000060	.0012	.000089	.0039	.000158
45-50	.0002	.000035	.0009	.000079	.0013	.000095	.0050	.000188
50-55	.0003	.000045	.0014	.000103	.0014	.000103	.0063	.000217
55-60	.0005	.000067	.0029	.000158	.0015	.000115	.0084	.000268
60-65	.0006	.000080	.0052	.000235	.0017	.000134	.0111	.000343
65-70	.0010	.000111	.0091	.000341	.0019	.000157	.0164	.000457
70-75	.0016	.000159	.0146	.000483	.0020	.000179	.0229	.000602
75-80	.0012	.000162	.0237	.000700	.0028	.000245	.0331	.000823
80-85	.0011	.000200	.0376	.001161	.0028	.000320	.0445	.001259

4. Interpretation of a Multiple Decrement Table

A multiple decrement table, such as those presented in Tables 5 and 6, can serve many useful purposes. Significant points include the following:

1. Each probability $\hat{Q}_{i\delta}$ represents a measure of risk of dying from a specific cause to which a person is subject in a real population where other competing risks are also acting. For example, in Table 5, Column (4), age interval (60,65), we found $\hat{Q}_{60,1} = .0315$. This figure suggests that if the forces of mortality operating in the Swedish population, 1967, prevail, the probability is over three percent that a person of 60 years of age will die from cardiovascular disease within five years.

2. The entire array of probabilities $\hat{Q}_{i\delta}$ over all the age groups gives a profile of risk of dying from a specific cause during a person's life time. Thus the risk of dying from cardiovascular disease is negligible among young people, but increases with advancement of age. According to the Swedish, 1967, experiences, these diseases are the most serious cause of death for persons beyond age 50, and the chance is better than one in four (.2771) that a person of age 80 will die from cardiovascular disease in the following five years despite competition from other causes. Below age 40, however, cardiovascular disease is negligible as a cause of death, and between ages 1 and 10, the risk of death from these diseases is almost nonexistent. Alternatively, the risk of death from cancer all forms is more evenly spread over the age intervals but is less evenly spread than the risk of death from all accidents.

3. When viewed across several causes of death, the multiple decrement table shows relative risk of death either for a specific age group or for the entire life span. It is evident from these tables that among the leading causes of death, cardiovascular disease is a dominate cause, cancer all forms runs a poor second, while all accidents is a distant third. However, in age 40-50

cancer all forms is the most menacing disease in Sweden in 1967 and is a significant cause of death even below age 40. Within the latter age bracket, however, all accidents assumes the leading role as cause of death.

4. Each probability $\hat{Q}_{i\delta}$, when expressed in terms of the probability of dying $\hat{q}_i$, gives the proportionate mortality for each cause. This in turn provides the information as to what proportion of mortality in each age group may be attributed to specific causes.

5. A comprehensive comparison of cause specific mortality experience may be made among different countries, or of a country over time. Between Sweden and Australia, for example, the general mortality pattern is similar, but details vary. According to 1967 experience, the Australian population is subject to higher risks of death than the Swedish population in almost every age group and for each cause considered in this example. The only exceptions occur in the very old age brackets for a few causes. Beyond age 70 cancer all forms is a more eminent cause in Sweden than in Australia. A similar statement can be made for respiratory disease beyond age 80, and infectious diseases beyond age 75, although the magnitude of the probabilities for the latter case is quite small. It may also be noted that, while in Sweden, 1967, cardiovascular disease was the most serious cause of death from age 50 on, in Australia these diseases assume this role beginning in the early 30's.

6. Statistical inference about these rates can be readily made with the aid of the standard deviations listed. For example, hypotheses can be tested regarding the probability of dying from a specific cause between Sweden and Australia. Is the probability of dying from cardiovascular disease for age group (45,50) greater in Australia than Sweden? To answer this question, we compute the critical ratio (cf., Chapter 4).

$$Z = \frac{\hat{Q}_{45,1}(A) - \hat{Q}_{45,1}(S)}{\sqrt{S^2_{\hat{Q}_{45,1}(A)} + S^2_{\hat{Q}_{45,1}(S)}}} \tag{4.1}$$

which has a normal distribution with a mean zero and a variance one, if in fact $Q_{45,1}(A) = Q_{45,1}(S)$. The numerical value is

$$Z = \frac{.0111 - .0044}{\sqrt{(.000278)^2 + (.000200)^2}}$$
$$= 19.6$$

which is highly significant, as the corresponding probability is less than 1 in 10,000. In other words, if the probability of dying from cardiovascular diseases in Australia was equal to that in Sweden for a person of age 45-50, then the chances are less than 1 in 10,000 that a difference as great or greater than the one observed would occur. Based on the above findings we conclude that cardiovascular diseases werea more serious cause of death in Australia than it was in Sweden for the age group under consideration.

7. Caution should be observed in comparing the crude probabilities of dying from different causes in the same population and the same age group. In a particular age group, various causes are competing with one another for the life of an individual and the estimated probabilities are statistically dependent. Between any two probabilities (say, $\hat{Q}_{i1}$ and $\hat{Q}_{i2}$) there is a co-variance, which must be taken into account in making inference about these probabilities. The co-variance between $\hat{Q}_{i1}$ and $\hat{Q}_{i2}$, for example, is given by

$$\text{Cov}(\hat{Q}_{i1},\hat{Q}_{i2}) = -\frac{1}{N_i}\hat{Q}_{i1}\hat{Q}_{i2} = -\frac{1}{n_{i1}}\hat{Q}^2_{i1}\hat{Q}_{i2} \cdot \tag{4.2}$$

A similar formula holds for any two probabilities. How the covariance can be incorporated in statistical inference is demonstrated below. Suppose we want to compare two causes R_1: cardiovascular diseases and R_2: cancer all forms for people of age 55-60 in the Sweden population, 1967. From column (4) and (6), we found $\hat{Q}_{55,1} = .0167$ and $\hat{Q}_{55,2} = .0125$, with a difference

$$\hat{Q}_{55,1} - \hat{Q}_{55,2} = .0042 \quad . \tag{4.3}$$

Is this difference significantly greater than zero or can it be explained by chance? Here we are testing the hypothesis that $Q_{55,1} = Q_{55,2}$ against the alternative hypothesis that $Q_{55,1} > Q_{55,2}$. To test the hypothesis, we express the difference $\hat{Q}_{55,1} - \hat{Q}_{55,2}$ in terms of its standard deviation, which is given by

$$S.D.(\hat{Q}_{i1}-\hat{Q}_{i2}) = \sqrt{S^2_{\hat{Q}_{i1}} + S^2_{\hat{Q}_{i2}} - 2\,Cov(\hat{Q}_{i1},\hat{Q}_{i2})} \tag{4.4}$$

where the covariance can be computed as follows:

$$Cov(\hat{Q}_{i1},\hat{Q}_{i2}) = -\frac{1}{D_{i1}}\hat{Q}^2_{i1}\,\hat{Q}_{i2}$$

$$Cov(\hat{Q}_{55,1},\hat{Q}_{55,2}) = -.000000002$$

which is a very small number. Now the standard deviation can be computed

$$S.D.(\hat{Q}_{55,1}-\hat{Q}_{55,2}) = \sqrt{(.000397)^2 + (.000344)^2 + 2(.000000002)}$$

$$= 0.00053$$

The statistic used to test the hypothesis is again the normal deviate

$$Z = \frac{\hat{Q}_{55,1} - \hat{Q}_{55,2}}{S.D.(\hat{Q}_{55,1}-\hat{Q}_{55,2})} \tag{4.5}$$

and compare the numerical value of Z with the standard normal distribution. In this case we have

$$Z = \frac{.0167 - .0125}{.00053} = 7.9$$

which is highly significant. Thus according to Sweden 1967 experience, the probability of dying from cardiovascular disease is greater than cancer all forms for age interval (55,60).

In conclusion, the multiple decrement table presents a mortality profile over ages and causes of death for a population under study. It shows the relative, as well as the absolute, importance of various diseases and their variation over age and sex. With the information provided in the table, one can easily detect and determine the area of concern, the degree of seriousness of various diseases, and the type and amount of medical care and health services needed by persons in different age and sex categories.

CHAPTER 8

THE LIFE TABLE WHEN A PARTICULAR CAUSE IS ELIMINATED

1. Introduction

In Appendix III on competing risks, several types of probabilities of dying with respect to a particular cause have been discussed. Corresponding to each of these probabilities, a life table may be constructed to serve a specific purpose using the probability in question in place of $\hat{q}_i$. The procedure of construction is exactly the same as that of an ordinary life table described in Chapter 4, though the columns have different meanings. A life table derived from $\hat{q}_{i.1}$, the probability of dying when a cause R_1 (e.g., cardiovascular-renal disease) is eliminated as a cause of death, for example, may be used to evaluate the effect of the cardiovascular-renal diseases on the longevity of a human population in terms of the expectation of life or chance of survival. Generally, the event involved need not be survival or death and the subject is not limited to human beings. In a study of the effect of divorce on the longevity of marriage, for example, the event is the dissolution of marriage. If divorce (R_1) had been removed as a cause, how long is a marriage expected to last before death occurred to either one of the spouses? Application to other problems is possible wherever the concept of competing risks applies. We shall in this chapter describe two tables: (1) the life table when a specific cause is eliminated, and (2) the life table when a specific risk is the only risk operating. Empirical data will be used for illustration.

2. Computation of the Net Probability, $\hat{q}_{i.1}$

The life table in this chapter is derived from $q_{i.1}$, the net probability of dying when a particular cause R_1 is eliminated. This is one of the most important applications of the competing risks theory. Such a table may be constructed either for a cohort population or for a current population. In either case, the basic formula is (cf., Equation (2.29) in Appendix III)

$$q_{i.1} = (q_i - Q_{i1})(1 + \frac{1}{2} Q_{i1}) \tag{2.1}$$

where q_i is the probability of dying in interval (x_i, x_{i+1}) and Q_{i1} is the crude probability of dying from R_1 during the same interval in the presence of other competing risks. To avoid repetition, only the life table derived from mortality data of a current population will be discussed.

Let us consider, as an example, cardiovascular-renal diseases and the effect of their presence on the probability of dying and the expectation of life. For a typical age interval (x_i, x_{i+1}), with $x_{i+1} - x_i = n_i$ being the interval length, let D_i be the number of deaths occurring in age interval (x_i, x_{i+1}) during a calendar year, among them D_{i1} dying from R_1. Let P_i be the corresponding mid-year population, and a_i the fraction of last age interval of life. The age-specific death rate is computed as before from

$$M_i = \frac{D_i}{P_i} \tag{2.2}$$

and death rate specific for cause R_1, (CVR diseases) is computed from

$$M_{i1} = \frac{D_{i1}}{P_i} \quad . \tag{2.3}$$

Using the results in preceding chapters, we have the estimate of the probability [cf., Equation (4.3), Chapter 3]

$$\hat{q}_i = \frac{n_i M_i}{1 + (1-a_i) n_i M_i} \tag{2.4}$$

and of the crude probability [cf., equation (2.9), Chapter 7]

$$\hat{Q}_{i1} = \frac{n_i M_{i1}}{1 + (1-a_i)\, n_i M_i} \quad . \tag{2.5}$$

Substituting (2.4) and (2.5) in (2.1) yields the required estimate of the probability $q_{i.1}$:

$$\hat{q}_{i.1} = (\hat{q}_i - \hat{Q}_{i1})(1 + \frac{1}{2}\hat{Q}_{i1}) \quad . \tag{2.6}$$

With reference to formulas (2.1) through (2.6) above, computation of $\hat{q}_{i.1}$ is summarized in Table 1. For each age interval (x_i, x_{i+1}), the data required are midyear population P_i (Column 2), number of deaths from all causes D_i (Column 3), and number of deaths from the cause under study (in this case, cardiovascular-renal diseases) D_{i1} (Column 4). These figures, which are available in population and vital statistics publications, are used to compute death rate M_i (Column 5) and cause-specific death rate M_{i1} (Column 6). The fraction of last age interval of life, a_i, in Column (7) is given in Appendix V. Using formulas (2.4), (2.5) and (2.6) the probabilities $\hat{q}_i$, $\hat{Q}_{i1}$, and finally $\hat{q}_{i.1}$, are computed and recorded in Columns (8), (9), and (10), respectively.

For age interval (0, 1), for example, we have the midyear population P_0 = 1,794,784, the number of deaths D_0 = 48,063, deaths from cardiovascular diseases D_{01} = 228, and a_0 = .10. With these values, we compute the death rate from all causes using formula (2):

$$M_0 = \frac{D_0}{P_0} = \frac{48{,}063}{1{,}794{,}784} = .026779$$

$$\text{or } 26.78 \text{ per } 1{,}000, \tag{2.2a}$$

and the death rate from cardiovascular-renal diseases using formula (2.3):

$$M_{01} = \frac{D_{01}}{P_0} = \frac{228}{1,794,784} = .000127 \quad \text{or} \quad .13 \text{ per } 1,000 \ . \tag{2.3a}$$

The probability of dying, $\hat{q}_0$, is computed from formula (2.4) which is the same as in the ordinary life table in Chapter 3, namely

$$\hat{q}_0 = \frac{M_0}{1 + (1-a_0)\ M_0}$$

$$= \frac{.026779}{1 + (1-.10)\ .026779}$$

$$= .02615 \tag{2.4a}$$

and the crude probability of dying from cardiovascular-renal diseases is computed from

$$\hat{Q}_{01} = \frac{M_{01}}{1 + (1-a_0)\ M_0}$$

$$= \frac{.000127}{1 + (1-.10)\ .026779}$$

$$= .000124 \tag{2.5a}$$

and finally, the net probability $q_{0.1}$

$$\hat{q}_{0.1} = (\hat{q}_0 - \hat{Q}_{01})(1 + \frac{1}{2}\ \hat{Q}_{01})$$

$$= (.02615 - .000124)(1 + \frac{1}{2}\ .000124)$$

$$= .026028 \ . \tag{2.6a}$$

For age interval (1, 5), $n_1 = 5-1 = 4$, the rates and probabilities are successively computed as follows:

$$M_1 = \frac{D_1}{P_1} = \frac{7,409}{7,063,044}$$

$$= .001049 \quad \text{or} \quad 1.049 \text{ per } 1,000 \tag{2.2a}$$

Table 1.

Computation of the net probability of dying, $q_{i.1}$, when cardiovascular-renal (CVR) diseases (R_1) are eliminated as a cause of death, white males, United States, 1960.

Age Interval (in years) x_i to x_{i+1}	Mid-year Populations (a) P_i	Deaths from all causes (b) D_i	Deaths from Cardio Vascular renal diseases (c,d) D_{i1}	Death rate from all causes M_i (3)/(2)	Death rate from CVR M_{i1} (4)/(2)	Fraction of last Age Interval of Life a_i	Probability od dying $\hat{q}_i$	Crude Probability of dying from CVR $\hat{Q}_{i1}$	Net Probability of dying when CVR is eliminated q_{i1}
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0- 1	1794784	48063	228	.026779	.000127	.10	.02615	.000124	.02603
1- 5	7063044	7409	153	.001049	.000022	.39	.00419	.000088	.00410
5-10	8191158	4408	177	.000538	.000024	.46	.00269	.000108	.00258
10-15	7488562	3847	208	.000514	.000028	.54	.00257	.000139	.00243
15-20	5893946	7308	355	.001240	.000060	.57	.00618	.000300	.00588
20-25	4657470	7755	481	.001665	.000103	.49	.00829	.000514	.00778
25-30	4725480	7182	768	.001520	.000163	.50	.00757	.000810	.00676
30-35	5216424	9039	1808	.001733	.000347	.52	.00863	.001726	.00691
35-40	5461528	13803	4444	.002527	.000814	.54	.01256	.004045	.00854
40-45	5094821	21336	9125	.004188	.001791	.54	.02074	.008870	.01192
45-50	4850486	34247	16796	.007061	.003463	.54	.03474	.017037	.01785
50-55	4314976	50716	26812	.011753	.006214	.53	.05719	.030233	.02737
55-60	3774623	66540	36907	.017628	.009778	.52	.08456	.046904	.03858
60-65	3100045	85890	49649	.027706	.016016	.52	.12989	.075085	.05702
65-70	2637044	108726	65609	.041230	.024880	.52	.18759	.113198	.07908
70-75	1972947	119269	75371	.060452	.038202	.51	.26327	.166370	.10636
75-80	1214577	109193	73057	.089902	.060150	.51	.36837	.246465	.14106
80-85	591251	83885	58713	.141877	.099303	.48	.51822	.362716	.19679
85-90	235566	49502	36133	.210141	.153388	.45	.66589	.486055	.25627
90-95	56704	18253	13604	.321900	.239913	.41	.82555	.615285	.35901
95+	12333	4219	3136	.342090	.254277		1.00000		1.00000
TOTAL	78347769	860857[d]	473640[d]	.010988	.006045				

a. US Census of Population 1960, US Summary, Detailed Characteristics Table 156. Bureau of the Census, US Department of Commerce.

b. Vital Statistics of the US 1960, Vol. II, Part A, Table 5 - 11, National Centre for Health Statistics, US Department of Health, Education and Welfare.

c. Category Number 330 - 334, 400 - 468, and 592 - 594 of the 1955 Revision of the <u>International Classification of Diseases, Injuries, and Causes of Death</u>, World Health Organization, 1957.

d. Including those age not stated, 267 and 106, respectively.

$$M_{11} = \frac{D_{11}}{P_1} = \frac{153}{7,063,044}$$

$$= .000022 \quad \text{or} \quad .022 \text{ per } 1,000 \tag{2.3a}$$

$$\hat{q}_1 = \frac{4M_1}{1 + (1-a_1)\, 4M_1} = \frac{4(.001049)}{1 + (1-.39)\, 4(.001049)}$$

$$= .00419 \tag{2.4a}$$

$$\hat{Q}_{11} = \frac{4M_{11}}{1 + (1-a_1)\, 4M_1} = \frac{4(.000022)}{1 + (1-.39)\, 4(.001049)}$$

$$= .000088 \tag{2.5a}$$

and

$$\hat{q}_{1.1} = (\hat{q}_1 - \hat{Q}_{11})\,(1 + \tfrac{1}{2}\hat{Q}_{11})$$

$$= (.00419 - .000088)\,(1 + \tfrac{1}{2}\,.000088) = .004102. \tag{2.6a}$$

3. Construction of the Life Table

When all the values of $\hat{q}_{i.1}$ are computed, the columns in the life table can be obtained following the procedure described in Chapter 3. Beginning with a radix $\ell_{0.1} = 100,000$, we compute the number of deaths in (0, 1)*,

$$d_{0.1} = \ell_{0.1}\, q_{0.1}$$

$$= 100,000 \times .02603 = 2603 \tag{3.1}$$

the number living at age 1

$$\ell_{1.1} = \ell_{0.1} - d_{0.1}$$

$$= 100,000 - 2603 = 97397 \tag{3.2}$$

*A notation .1 is added in the subscript of $\ell_{i.1}$, $d_{i.1}$, $T_{i.1}$ and $\hat{e}_{i.1}$ to indicate that R_1 is eliminated.

and the number of years lived in (0, 1),

$$L_{0.1} = (\ell_{0.1} - d_{0.1}) + a_0\, d_{0.1}$$

$$= 97397 + .1 \times 2603 = 97657\,. \qquad (3.3)$$

Other figures in these columns for the subsequent age intervals (except for the last interval) can be computed in exactly the same way.

The computations for the last age interval (e.g., 95 and over) have been described in Chapter 3 [cf., Equations (3.10) to (3.12) in Chapter 3]. For easy reference, they are restated below. The number living at age 95, $\ell_{95\cdot1} = 21564$, is the survivors of interval (90, 95). The expectation of life $e_{95\cdot1}$ is computed directly from the death rate from causes other than cardiovascular-renal disease in the current population. Using the inverse relationship between the expectation and the death rate (cf. equation (3.7) in Chapter 3)

$$\hat{e}_{95.1} = \frac{P_{95}}{D_{95} - D_{95,1}} = \frac{12333}{4219 - 3136}$$

$$= 11.3878 \text{ years}\,. \qquad (3.4)$$

Since

$$\hat{e}_{95\cdot1} = \frac{T_{95\cdot1}}{\ell_{95\cdot1}}\,, \qquad (3.5)$$

we have

$$T_{95\cdot1} = \ell_{95\cdot1}\, \hat{e}_{95\cdot1}$$

$$= 21564 \times 11.3878$$

$$= 245567\,.$$

The remaining quantities in the last age interval may be derived from the obvious relationships, thus

$$L_{95\cdot 1} = T_{95\cdot 1} = 245,567 \tag{3.6}$$

$$d_{95\cdot 1} = \ell_{95\cdot 1} = 21,564 \tag{3.7}$$

and

$$\hat{q}_{95\cdot 1} = 1.00000 \quad . \tag{3.8}$$

With $L_{95\cdot 1}$ and all other $L_{i\cdot 1}$ determined, we proceed to compute $T_{i.1}$ from

$$T_{i.1} = L_{i.1} + \dots + L_{95\cdot 1} \quad . \tag{3.9}$$

For convenience, $T_{i.1}$ are computed successively from the highest age group, beginning with $T_{95.1} = 245,567$. For age 90, $T_{90.1}$ is computed from

$$T_{90.1} = L_{90.1} + T_{95.1}$$

$$= 132,580 + 245,567 = 378,147 \ ,$$

and $T_{85.1}$ from

$$T_{85.1} = L_{85.1} + T_{90.1}$$

and so on. In general,

$$T_{i.1} = L_{i.1} + T_{i+1.1} \quad . \tag{3.10}$$

The expectation of life (except for $\hat{e}_{95.1}$) is then obtained from

$$\hat{e}_{i.1} = \frac{T_{i.1}}{\ell_{i.1}} \tag{3.11}$$

for each i. For example,

Table 2.

Abridged Life Table when Cardiovascular Renal Diseases are eliminated as a cause of death for white males, United States, 1960.

Age Interval (in years)	Probability of dying in interval (x_i, x_{i+1})	Number living at age x_i	Number dying in interval (x_i, x_{i+1})	Fraction of last age interval of life	Number of years lived in interval (x_i, x_{i+1})	Total number of years lived beyond age x_i	Expectation of life at age x_i
x_i to x_{i+1}	$\hat{q}_{i.1}$	$l_{i.1}$	$d_{i.1}$	a_i	$L_{i.1}$	$T_{i.1}$	$\mathring{e}_{i.1}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0- 1	.02603	100000	2603	.10	97657	7894729	78.95
1- 5	.00410	97397	399	.39	388614	7797072	80.05
5-10	.00258	96998	250	.46	484315	7408458	76.38
10-15	.00243	96748	235	.54	483199	6924143	71.57
15-20	.00588	96513	567	.57	481346	6440944	66.74
20-25	.00778	95946	746	.49	477828	5959598	62.11
25-30	.00676	95200	644	.50	474390	5481770	57.58
30-35	.00691	94556	653	.52	471213	5007380	52.96
35-40	.00854	93903	802	.54	467670	4536167	48.31
40-45	.01192	93101	1110	.54	462952	4068497	43.70
45-50	.01785	91991	1643	.54	456178	3605545	39.19
50-55	.02737	90349	2473	.53	445933	3149367	34.86
55-60	.03858	87876	3390	.52	431244	2703434	30.76
60-65	.05702	84486	4817	.52	410869	2272190	26.89
65-70	.07908	79669	6300	.52	383225	1861321	23.36
70-75	.10636	73369	7804	.51	347725	1478096	20.15
75-80	.14106	65565	9249	.51	305165	1130371	17.24
80-85	.19679	56316	11082	.48	252767	825206	14.65
85-90	.25627	45234	11592	.45	194292	572439	12.66
90-95	.35901	33642	12078	.41	132580	378147	11.24
95+	1.00000	21564	21564		245567	245567	11.39

TABLE 3.

Abridged Life Table when Cardiovascular Renal Diseases are Eliminated as a cause of death for white females, United States, 1960.

Age Interval (in years)	Probability of dying in interval (x_i, x_{i+1})	Number living at age x_i	Number dying in interval (x_i, x_{i+1})	Fraction of last age interval of life	Number of years lived in interval	Total number of years lived beyond age x_i	Expectation of life at age x_i
x_i to x_{i+1}	$\hat{q}_{i.1}$	$\ell_{i.1}$	$d_{i.1}$	a_i	$L_{i.1}$	$T_{i.1}$	$\hat{e}_{i.1}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0 - 1	.01959	100000	1959	.10	98237	8676848	86.77
1 - 5	.00334	98041	327	.39	391366	8578611	87.50
5 - 10	.00184	97714	180	.46	488084	8187245	83.79
10 - 15	.00140	97534	137	.54	487355	7699161	78.94
15 - 20	.00227	97397	221	.57	486510	7211806	74.05
20 - 25	.00259	97176	252	.49	485237	6725296	69.21
25 - 30	.00295	96924	286	.50	483905	6240059	64.38
30 - 35	.00395	96638	382	.52	482273	5756154	59.56
35 - 40	.00583	96256	561	.54	479990	5273881	54.79
40 - 45	.00881	95695	843	.54	476536	4793891	50.10
45 - 50	.01282	94852	1216	.54	471463	4317355	45.52
50 - 55	.01775	93636	1662	.53	464274	3845892	41.07
55 - 60	.02287	91974	2103	.52	454823	3381618	36.77
60 - 65	.03241	89871	2913	.52	442364	2926795	32.57
65 - 70	.04416	86958	3840	.52	425574	2484431	28.57
70 - 75	.06179	83118	5136	.51	403007	2058857	24.77
75 - 80	.08920	77982	6956	.51	372868	1655850	21.23
80 - 85	.13492	71026	9583	.48	330214	1282982	18.06
85 - 90	.19040	61443	11699	.45	275043	952768	15.51
90 - 95	.29170	49744	14510	.41	205915	677725	13.62
95 +	1.00000	35234	35234		471810	471810	13.39

$$\hat{e}_{90.1} = \frac{T_{90.1}}{\ell_{90.1}}$$

$$= \frac{378,147}{33,642}$$

$$= 11.24 \quad .$$

This completes the procedure of constructing life tables.

We have used the 1960 US white male population as an example to illustrate the high force of mortality from cardiovascular-renal disease. For purposes of comparison, a table for the US white female 1960 population has also been constructed and is reproduced here.

4. Interpretation of Findings

Cardiovascular-renal (CVR) diseases have caused more deaths in the human population than any other disease. As a group, they are responsible for over 55 percent of all deaths in the United States in recent years. Equally impressive figures, but to a somewhat lesser extent, have been reported in the European countries. To evaluate the impact of these diseases on human longevity, we can compare the mortality and survival experience of the current population with the hypothetical experience of the same population under the condition that would exist if CVR diseases were removed as causes of death. The life table and the theory of competing risks provide the most convenient methods for the analysis of such a problem. In Tables 4 to 6, the probability of dying, the survival probability, and the expectation of life are given with and without the presence of CVR, each reflecting in a different way the effect these diseases have on the mortality of the population in question. A brief discussion on these findings follows.

Table 4.

Probability of dying and the effect of eliminating CVR diseases as a cause of death in each age interval, white males and females, U.S., 1960

Age interval (in years)	White males				White females			
	CVR present	CVR eliminated	Difference		CVR present	CVR eliminated	Difference	
$x_i - x_{i+1}$	$\hat{q}_i$	$\hat{q}_{i.1}$	$\hat{q}_i - \hat{q}_{i.1}$	$\frac{\hat{q}_i - \hat{q}_{i.1}}{\hat{q}_i}$	$\hat{q}_i$	$\hat{q}_{i.1}$	$\hat{q}_i - \hat{q}_{i.1}$	$\frac{\hat{q}_i - \hat{q}_{i.1}}{\hat{q}_i}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0- 1	.02615	.02603	.00012	0.5%	.01967	.01959	.00008	0.4%
1- 5	.00419	.00410	.00009	2.1	.00341	.00334	.00007	2.1
5-10	.00269	.00258	.00011	4.1	.00191	.00184	.00007	3.7
10-15	.00257	.00243	.00014	5.4	.00154	.00140	.00014	9.1
15-20	.00618	.00588	.00030	4.9	.00251	.00227	.00024	9.6
20-25	.00829	.00778	.00051	6.2	.00302	.00259	.00043	14.2
25-30	.00757	.00676	.00081	10.7	.00357	.00295	.00062	17.4
30-35	.00863	.00691	.00172	19.9	.00484	.00395	.00089	18.4
35-40	.01256	.00854	.00402	32.0	.00733	.00583	.00150	20.5
40-45	.02074	.01192	.00882	42.5	.01185	.00881	.00304	25.7
45-50	.03474	.01785	.01689	48.6	.01816	.01282	.00534	29.4
50-55	.05719	.02737	.02982	52.1	.02732	.01775	.00957	35.0
55-60	.08456	.03858	.04598	54.4	.03978	.02287	.01691	42.5
60-65	.12989	.05702	.07287	56.1	.06613	.03241	.03372	51.0
65-70	.18759	.07908	.10851	57.8	.10321	.04416	.05905	57.2
70-75	.26327	.10636	.15691	59.6	.16682	.06179	.10503	63.0
75-80	.36837	.14106	.22731	61.7	.26990	.08920	.18070	67.0
80-85	.51822	.19679	.32143	62.0	.42950	.13492	.29458	68.6
85-90	.66589	.25627	.40962	61.5	.59498	.19040	.40458	68.0
90-95	.82555	.35901	.46654	56.5	.78586	.29170	.49416	62.9
95+	1.00000	1.00000	0	0	1.00000	1.00000	0	0

Table 5

Probability of survival and the effect of eliminating CVR diseases as a cause of death, white males and females, U.S., 1960

Age interval (in years) $x_i - x_{i+1}$	White males				White females			
	CVR present	CVR eliminated	Difference		CVR present	CVR eliminated	Difference	
	p_{0i}	$\hat{p}_{0i.1}$	$\hat{p}_{0i.1}-\hat{p}_{0i}$	$\frac{\hat{p}_{0i.1}-\hat{p}_{0i}}{\hat{p}_{0i}}$	$\hat{p}_{0i}$	$\hat{p}_{0i.1}$	$\hat{p}_{0i.1}-\hat{p}_{0i}$	$\frac{\hat{p}_{0i.1}-\hat{p}_{0i}}{\hat{p}_{0i}}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0 - 1	1.00000	1.00000	.00000	0.0 %	1.00000	1.00000	1.00000	0.0 %
1 - 5	.97385	.97397	.00012	0.0	.98033	.98041	.00008	0.0
5 - 10	.96977	.96998	.00021	0.0	.97699	.97714	.00015	0.0
10 - 15	.96716	.96748	.00032	0.0	.97512	.97534	.00022	0.0
15 - 20	.96467	.96513	.00046	0.0	.97362	.97397	.00035	0.0
20 - 25	.95871	.95946	.00075	0.1	.97118	.97176	.00058	0.1
25 - 30	.95076	.95200	.00124	0.1	.96825	.96924	.00099	0.1
30 - 35	.94356	.94556	.00200	0.2	.96479	.96638	.00159	0.2
35 - 40	.93542	.93903	.00361	0.4	.96012	.96256	.00244	0.3
40 - 45	.92367	.93101	.00734	0.8	.95308	.95695	.00387	0.4
45 - 50	.90451	.91991	.01540	1.7	.94179	.94852	.00673	0.7
50 - 55	.87309	.90349	.03040	3.5	.92469	.93636	.01167	1.3
55 - 60	.82316	.87876	.05560	6.8	.89943	.91974	.02031	2.3
60 - 65	.75355	.84486	.09131	12.1	.86365	.89871	.03506	4.1
65 - 70	.65567	.79669	.14102	21.5	.80654	.86958	.06304	7.8
70 - 75	.53267	.73369	.20102	37.7	.72330	.83118	.10788	14.9
75 - 80	.39243	.65565	.26322	67.1	.60264	.77982	.17718	29.4
80 - 85	.24787	.56316	.31529	127.2	.43999	.71026	.27027	61.4
85 - 90	.11942	.45234	.33292	278.8	.25101	.61443	.36342	144.8
90 - 95	.03990	.33642	.29652	743.2	.10166	.49744	.39578	389.3
95+	.00696	.21564	.20868	2998.3	.02177	.35234	.33057	1518.5

Table 6

Expectation of life and the effect of eliminating CVR diseases as a cause of death, white males and females, U.S., 1960

Age interval	White males				White females			
	CVR present	CVR eliminated	Difference		CVR present	CVR eliminated	Difference	
$x_i - x_{i+1}$	$\hat{e}_i$	$\hat{e}_{i\cdot 1}$	$\hat{e}_{i\cdot 1} - \hat{e}_i$	$\frac{\hat{e}_{i\cdot 1} - \hat{e}_i}{\hat{e}_i}$	$\hat{e}_i$	$\hat{e}_{i\cdot 1}$	$\hat{e}_{i\cdot 1} - \hat{e}_i$	$\frac{\hat{e}_{i\cdot 1} - \hat{e}_i}{\hat{e}_i}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0 - 1	67.27	78.95	11.68	17.4 %	74.01	86.77	12.76	17.2 %
1 - 5	68.08	80.05	11.97	17.6	74.50	87.50	13.00	17.4
5 - 10	64.36	76.38	12.02	18.7	70.75	83.79	13.04	18.4
10 - 15	59.52	71.57	12.05	20.2	65.88	78.94	13.06	19.8
15 - 20	54.67	66.74	12.07	22.1	60.97	74.05	13.08	21.5
20 - 25	49.99	62.11	12.12	24.2	56.12	69.21	13.09	23.3
25 - 30	45.39	57.58	12.19	26.9	51.28	64.38	13.10	25.5
30 - 35	40.72	52.96	12.24	30.1	46.46	59.56	13.10	28.2
35 - 40	36.05	48.31	12.26	34.0	41.67	54.79	13.12	31.5
40 - 45	31.47	43.70	12.23	38.9	36.96	50.10	13.14	35.6
45 - 50	27.08	39.19	12.11	44.7	32.37	45.52	13.15	40.6
50 - 55	22.96	34.86	11.90	51.8	27.92	41.07	13.15	47.1
55 - 60	19.19	30.76	11.57	60.3	23.63	36.77	13.14	55.6
60 - 65	15.73	26.89	11.16	70.9	19.50	32.57	13.07	67.0
65 - 70	12.69	23.36	10.67	84.1	15.70	28.57	12.87	82.0
70 - 75	10.01	20.15	10.14	101.3	12.20	24.77	12.57	103.0
75 - 80	7.68	17.24	9.56	124.5	9.13	21.23	12.10	132.5
80 - 85	5.67	14.65	8.98	158.4	6.57	18.06	11.49	174.9
85 - 90	4.20	12.66	8.46	201.4	4.71	15.51	10.80	229.3
90 - 95	3.07	11.24	8.17	266.1	3.32	13.62	10.30	310.2
95+	2.92	11.39	8.47	290.1	2.98	13.39	10.41	349.3

Table 4 gives a comparison between $\hat{q}_i$ and $\hat{q}_{i.1}$. The difference, $\hat{q}_i - \hat{q}_{i.1}$, is the reduction in the probability of dying in age interval (x_i, x_{i+1}) if CVR diseases were eliminated as a risk of death, or, alternatively, the excess probability of dying due to the presence of these diseases. This difference, while not pronounced below age 30, advances with age at an accelerated rate: from .00012 (0.5%) for the first year of life to .46654 (56.5%) for age interval 90 to 95 in white males, U.S., 1960. If the effect of CVR diseases were removed, the reduction in the probability of dying for white males is 32% of the existing probability for age interval 35 to 40, over 50% for interval 50 to 55, and about 60% for interval 70 to 75. This general mortality pattern holds also for white females. The estimated probabilities $\hat{q}_i$ and $\hat{q}_{i.1}$, and their difference $\hat{q}_i - \hat{q}_{i.1}$ are lower for females than for males up to age 90. The relative reduction in the probability of dying is for females about 20% for age interval 35 to 40, 35% for interval 50 to 55, and 63% for interval 70 to 75. In fact from age 30 to age 70, the relative reduction in the probability of dying is lower for females than for males, but the reverse is true for other ages. Thus, relatively speaking, from age 70 on cardiovascular-renal diseases have a larger impact on white females than white males, although in absolute terms these diseases contribute more deaths in the white male population than in the white female population almost throughout life.

The impact of CVR diseases on the probability of survival is shown in Table 5, where $p_{0i} = \ell_i/\ell_0$ is taken from the life table of the entire white population for each sex, and $p_{0i.1} = \ell_{i.1}/\ell_{0.1}$ is from Tables 2 and 3. Although the impact on the probability of survival is less pronounced than the probability of dying in the younger ages, it is much more alarming in the older age groups. From age 30 on for males, and from age 40 on for females, the relative reduction in the survival probability due to the presence of CVR has been doubled over every 5 year age interval.

Table 6 shows that these diseases cause an average loss of 12 years in the expectation of life for white males under age 50 and 13 years for females under age 65. At older ages, the loss in the expectation of life decreases slightly in absolute value but increases spectacularly relative to the existing life expectancy. If CVR diseases were eliminated as a risk of death, a male could expect an 30% increase of length of life over the present life expectancy at 30 years of age, 50% at age 50 and 100% at age 70. Comparable percentages of increased lengths of life that could be expected at these ages (28% at age 30, 47% at age 50 and 103% at age 70) are found for a female.

4.1 Comparison of impact on human mortality of three major causes of deaths: All accidents, cancer all forms, and cardiovascular-renal diseases

Different diseases have definite effects on human mortality and longevity. Concerted efforts are being made through the World Health Organization and health programs of individual countries to reduce mortality due to specific diseases. Relative importance of diseases as causes of death play a significant role in determining the priority in overall health planning. The purpose of this section is to show how some leading causes of death may be compared using the life table and competing risk methodology.

Tables 7, 8 and 9, are the life tables of the Federal Republic of Germany 1970 population when cardiovascular-renal diseases (R_1), cancer all forms (R_2), and all accidents (R_3) respectively, are eliminated as causes of death. Each table shows a hypothetical pattern that would exist in the Federal Republic of Germany if the corresponding diseases were eliminated.

Table 7.

Life Table of the Federal Republic of Germany population, 1970 when cardiovascular diseases (R_1) are eliminated as a cause of death.

Age Interval (in years)	Probability of dying in interval (x_i, x_{i+1})	Number living at age x_i	Number dying in interval (x_i, x_{i+1})	Fraction of last age interval of life	Number of years lived in interval	Total number of years lived beyond age x_i	Observed Expectation of life at age x_i
x_i to x_{i+1}	$\hat{q}_{i.1}$	$\ell_{i.1}$	$d_{i.1}$	a_i	$L_{i.1}$	$T_{i.1}$	$\hat{e}_{i.1}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0- 1	.02117	100000	2117	.10	98095	7749467	77.49
1- 5	.00374	97883	366	.39	390639	7651372	78.17
5-10	.00257	97517	251	.46	486907	7260733	74.46
10-15	.00201	97266	196	.52	485860	6773826	69.64
15-20	.00505	97070	490	.57	484296	6287966	64.78
20-25	.00573	96580	553	.52	481573	5803670	60.09
25-30	.00511	96027	491	.51	478932	5322097	55.42
30-35	.00633	95536	605	.52	476228	4843165	50.69
35-40	.00832	94931	790	.54	472838	4366937	46.00
40-45	.01138	94141	1071	.53	468188	3894099	41.36
45-50	.01630	93070	1517	.51	461633	3425911	36.81
50-55	.02481	91553	2271	.58	452996	2964278	32.38
55-60	.03529	89282	3151	.54	439163	2511282	28.13
60-65	.05576	86131	4803	.54	419608	2072119	24.06
65-70	.08802	81328	7158	.52	389461	1652511	20.32
70-75	.12736	74170	9446	.52	348180	1263050	17.03
75-80	.18092	64724	11710	.51	294930	914870	14.13
80-85	.25626	53014	13585	.49	230428	619940	11.69
85+	1.00000	39429	39429		389512	389512	9.88

Table 8.

Life Table of the Federal Republic of Germany population, 1970
when cancer all forms (R_2) is eliminated as a cause of death.

Age Interval (in years) x_i to x_{i+1}	Probability of dying in interval (x_i, x_{i+1}) $\hat{q}_{i.2}$	Number living at age x_i $\ell_{i.2}$	Number dying in interval (x_i, x_{i+1}) $d_{i.2}$	Fraction of last age interval of life a_i	Number of years lived in interval $L_{i.2}$	Total number of years lived beyond age x_i $T_{i.2}$	Observed Expectation of life at age x_i $\hat{e}_{i.2}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0- 1	.02117	100000	2117	.10	98095	7323376	73.23
1- 5	.00343	97883	336	.39	390712	7225281	73.82
5-10	.00227	97547	221	.46	487138	6834569	70.06
10-15	.00182	97326	177	.52	486205	6347431	65.22
15-20	.00481	97149	467	.57	484741	5861226	60.33
20-25	.00553	96682	535	.52	482126	5376485	55.61
25-30	.00485	96147	466	.51	479593	4894359	50.90
30-35	.00603	95681	577	.52	477020	4414766	46.14
35-40	.00808	95104	768	.54	473754	3937746	41.40
40-45	.01116	94336	1053	.53	469205	3463992	36.72
45-50	.01595	93283	1488	.51	462769	2994787	32.10
50-55	.02433	91795	2233	.58	454286	2532018	27.58
55-60	.03681	89562	3297	.54	440227	2077732	23.20
60-65	.06501	86265	5608	.54	418427	1637505	18.98
65-70	.11328	80657	9137	.52	381356	1219078	15.11
70-75	.18512	71520	13240	.52	325824	837722	11.71
75-80	.29308	58280	17081	.51	249552	511898	8.78
80-85	.44942	41199	18516	.49	158779	262346	6.37
85+	1.00000	22683	22683		103567	103567	4.57

Table 9.

Life Table of the Federal Republic of Germany population, 1970 when all accidents (R_3) are eliminated as a cause of death.

Age Interval (in years)	Probability of dying in interval (x_i, x_{i+1})	Number living at age x_i	Number dying in interval (x_i, x_{i+1})	Fraction of last age interval of life	Number of years lived in interval	Total number of years lived beyond age x_i	Observed Expectation of life at age x_i
x_i to x_{i+1}	$q_{i.3}$	$\ell_{i.3}$	$d_{i.3}$	a_i	$L_{i.3}$	$T_{i.3}$	$\hat{e}_{i.3}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0- 1	.02054	100000	2054	.10	98151	7195696	71.96
1- 5	.00253	97946	248	.39	391179	7097545	72.46
5-10	.00123	97698	120	.46	488166	6706366	68.64
10-15	.00112	97578	109	.52	487628	6218200	63.73
15-20	.00203	97469	198	.57	486919	5730572	58.79
20-25	.00271	97271	264	.52	485721	5243653	53.91
25-30	.00334	97007	324	.51	484241	4757932	49.05
30-35	.00505	96683	488	.52	482244	4273691	44.20
35-40	.00790	96195	760	.54	479227	3791447	39.41
40-45	.01287	95435	1228	.53	474289	3312220	34.71
45-50	.02048	94207	1929	.51	466309	2837931	30.12
50-55	.03297	92278	3042	.58	455002	2371622	25.70
55-60	.05051	89236	4507	.54	435814	1916620	21.48
60-65	.08632	84729	7314	.54	406823	1480806	17.48
65-70	.14551	77415	11265	.52	360039	1073983	13.87
70-75	.22526	66150	14901	.52	294988	713944	10.79
75-80	.33585	51249	17212	.51	214076	418956	8.17
80-85	.48554	34037	16526	.49	128044	204880	6.02
85+	1.00000	17511	17511		76836	76836	4.39

For comparison, Tables 10a and 10b are produced from the life tables to show the differences $\hat{q}_i - \hat{q}_{i.1}$, $\hat{q}_i - \hat{q}_{i.2}$ and $\hat{q}_i - \hat{q}_{i.3}$, for each age group. These differences represent the increase in probability of dying due to the presence of the corresponding disease. Table 10b shows that the contribution of accidents to the probability of dying is quite uniform over most of the life span with the exception of very old ages where some increase has taken place. On the other hand, the differences in probability of dying for cancer all forms and cardiovascular-renal diseases are not significant for ages less than 30 years, but they increase rapidly with the advancement of age. The differences are higher for cancer than cardiovascular-renal diseases for age groups below 55, but the reverse is true for older ages. For age group 80 to 85, the difference for cardiovascular-renal diseases is more than four times as large as that for all cancers. Since the probability of dying in old age groups is much higher than in younger age groups, cardiovascular-renal diseases have a greater effect on human longevity than does cancer.

The relative impact of these three causes of death on human longevity becomes quite clear in Table 11. In this Table, the expectation of life $\hat{e}_i$ at each age when all risks are operating is being compared with the expectation of life ($\hat{e}_{i.1}$, $\hat{e}_{i.2}$, and $\hat{e}_{i.3}$) when one of the causes is eliminated. We see that cardiovascular diseases are by far the most important causes of death. Cancer all forms runs a distant second, and all accidents a poor third. The difference $\hat{e}_{i.1} - \hat{e}_i$ is quite constant throughout the life span. The figures show that the presence of cardio-vascular-renal diseases has, on the average, cost the people in the Federal Republic of Germany about 7 years loss of life. The corresponding difference for cancer decreases as age increases, with the largest difference of 2.58 years at age 1-5 and the smallest of .39 years

Table 10a.

Probability of dying when cardiovascular diseases (R_1), cancer all forms (R_2), or all accidents (R_3) are eliminated as a cause of death.

(The Federal Republic of Germany, 1970)

Age Interval (in years) (x_i, x_{i+1})	Probability of dying in interval (x_i, x_{i+1}) $\hat{q}_i$	Probability of Dying When A Cause is Eliminated: Cardiovascular (R_1) $\hat{q}_{i.1}$	Cancer All Forms (R_2) $\hat{q}_{i.2}$	All Accidents (R_3) $\hat{q}_{i.3}$
(1)	(2)	(3)	(4)	(5)
0- 1	.02123	.02117	.02117	.02054
1- 5	.00379	.00374	.00343	.00253
5-10	.00260	.00257	.00227	.00123
10-15	.00208	.00201	.00182	.00112
15-20	.00516	.00505	.00481	.00203
20-25	.00598	.00573	.00553	.00271
25-30	.00548	.00511	.00485	.00334
30-35	.00707	.00633	.00603	.00505
35-40	.00991	.00832	.00808	.00790
40-45	.01471	.01138	.01116	.01287
45-50	.02224	.01630	.01595	.02048
50-55	.03499	.02481	.02433	.03297
55-60	.05275	.03529	.03681	.05051
60-65	.08915	.05576	.06501	.08632
65-70	.14877	.08802	.11328	.14551
70-75	.23005	.12736	.18512	.22526
75-80	.34382	.18092	.29308	.33585
80-85	.49841	.25626	.44942	.48554
85+	1.00000	1.00000	1.00000	1.00000

TABLE 10b

Probability of dying and the effect of eliminating cardiovascular diseases (R_1), cancer all forms (R_2), or all accidents (R_3) as a cause of death in each age interval.

(The Federal Republic of Germany, 1970).

Interval x_i to x_{i+1}	Cardiovascular diseases, R_1.		Cancer all forms, R_2.		All accidents, R_3.	
	$\hat{q}_i-\hat{q}_{i.1}$	$\frac{\hat{q}_i-\hat{q}_{i.1}}{\hat{q}_i}$	$\hat{q}_i-\hat{q}_{i.2}$	$\frac{\hat{q}_i-\hat{q}_{i.2}}{\hat{q}_i}$	$\hat{q}_i-\hat{q}_{i.3}$	$\frac{\hat{q}_i-\hat{q}_{i.3}}{\hat{q}_i}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0 - 1	.00006	0.3%	.00006	0.3%	.00069	3.3%
1 - 5	.00005	1.3%	.00036	9.5%	.00126	33.2%
5 - 10	.00003	1.2%	.00033	12.7%	.00137	52.7%
10 - 15	.00005	2.4%	.00024	11.7%	.00094	45.6%
15 - 20	.00011	2.1%	.00035	6.8%	.00313	60.7%
20 - 25	.00025	4.2%	.00045	7.5%	.00327	54.7%
25 - 30	.00037	6.8%	.00063	11.5%	.00214	39.1%
30 - 35	.00074	10.5%	.00104	14.7%	.00202	28.6%
35 - 40	.00159	16.0%	.00183	18.5%	.00201	20.3%
40 - 45	.00333	22.6%	.00355	24.1%	.00184	12.5%
45 - 50	.00594	26.7%	.00629	28.3%	.00176	7.9%
50 - 55	.01018	29.1%	.01066	30.5%	.00202	5.8%
55 - 60	.01746	33.1%	.01594	30.2%	.00224	4.2%
60 - 65	.03339	37.5%	.02414	27.1%	.00283	3.2%
65 - 70	.06075	40.8%	.03549	23.9%	.00326	2.2%
70 - 75	.10269	44.6%	.04493	19.5%	.00479	2.1%
75 - 80	.16290	47.4%	.05074	14.8%	.00797	2.3%
80 - 85	.24215	48.6%	.04899	9.8%	.01287	2.6%
85 +		-		-		-

Table 11.

Expectation of life and the effect of elimination of cardiovascular diseases (R_1), Cancer all forms (R_2), or all accidents (R_3) as a cause of death in each age interval.

(The Federal Republic of Germany, 1970).

Age Interval (x_i to x_{i+1})	Observed Expectation of life $\hat{e}_i$	Expectation of life with elimination as cause of death					
		Cardiovascular Diseases 1.		Cancer all forms 2.		All accidents 3.	
		$\hat{e}_{i.1}$	$\hat{e}_{i.1}-\hat{e}_i$	$\hat{e}_{i.2}$	$\hat{e}_{i.2}-\hat{e}_i$	$\hat{e}_{i.3}$	$\hat{e}_{i.3}-\hat{e}_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0- 1	70.71	77.49	6.78	73.23	2.52	71.96	1.25
1- 5	71.24	78.17	6.93	73.82	2.58	72.46	1.22
5-10	67.51	74.46	6.95	70.06	2.55	68.64	1.13
10-15	62.68	69.64	6.96	65.22	2.54	63.73	1.05
15-20	57.80	64.78	6.98	60.33	2.53	58.79	0.99
20-25	53.09	60.09	7.00	55.61	2.52	53.91	0.82
25-30	48.39	55.42	7.03	50.90	2.51	49.05	0.66
30-35	43.64	50.69	7.05	46.14	2.50	44.20	C.56
35-40	38.94	46.00	7.06	41.40	2.46	39.41	0.47
40-45	34.30	41.36	7.06	36.72	2.42	34.71	0.41
45-50	29.77	36.81	7.04	32.10	2.33	30.12	0.35
50-55	25.39	32.38	6.99	27.58	2.19	25.70	0.31
55-60	21.21	28.13	6.92	23.20	1.99	21.48	0.27
60-65	17.24	24.06	6.82	18.98	1.74	17.48	0.24
65-70	13.66	20.32	6.66	15.11	1.45	13.87	0.21
70-75	10.59	17.03	6.44	11.71	1.12	10.79	0.20
75-80	7.98	14.13	6.15	8.78	0.80	8.17	0.19
80-85	5.82	11.69	5.87	6.37	0.55	6.02	0.20
85+	4.18	9.88	5.70	4.57	0.39	4.39	0.21

1. Cardiovascular diseases (A80-A88)
2. Cancer all forms (A45-A60)
3. All accidents (AE138-AE146)

Table 12.

Probability of dying and the effect of eliminating cancer all forms (R_2) as a cause of death

(Canada 1968 and France 1969)

Age Interval (in years) (x_i, x_{i+1})	Canada				France			
	Male		Female		Male		Female	
	$\hat{q}_i - \hat{q}_{i.2}$	$\frac{\hat{q}_i - \hat{q}_{i.2}}{\hat{q}_i}$	$\hat{q}_i - \hat{q}_{i.2}$	$\frac{\hat{q}_i - \hat{q}_{i.2}}{\hat{q}_i}$	$\hat{q}_i - \hat{q}_{i.2}$	$\frac{\hat{q}_i - \hat{q}_{i.2}}{\hat{q}_i}$	$\hat{q}_i - \hat{q}_{i.2}$	$\frac{\hat{q}_i - \hat{q}_{i.2}}{\hat{q}_i}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1-5	.00019	4.7%	.00022	6.8%	.00029	7.8%	.00025	8.2%
5-10	.00033	11.1	.00019	9.6	.00037	16.2	.00034	20.5
10-15	.00015	5.8	.00013	8.4	.00062	25.8	.00033	22.9
15-20	.00048	7.5	.00034	13.6	.00126	20.8	.00058	21.3
20-25	.00030	3.3	.00042	14.9	.00097	12.1	.00017	5.2
25-30	.00077	10.3	.00058	18.2	.00123	15.3	.00050	13.7
30-35	.00106	13.2	.00058	13.6	.00141	14.4	.00057	12.4
35-40	.00171	15.5	.00244	37.6	.00313	21.3	.00182	25.1
40-45	.00206	12.0	.00422	41.6	.00520	23.2	.00378	34.2
45-50	.00506	17.8	.00703	44.0	.00867	25.3	.00662	39.1
50-55	.00849	18.6	.01037	42.1	.01534	29.0	.00913	36.5
55-60	.01614	22.0	.01432	37.8	.02429	29.8	.01405	39.1
60-65	.02326	20.8	.01801	31.3	.03925	31.2	.01922	34.9
65-70	.03685	22.2	.02348	26.0	.05564	29.9	.02799	31.3
70-75	.04801	20.7	.03010	21.4	.07125	26.9	.04070	26.6
75-80	.04930	14.8	.03325	14.7	.08024	21.4	.05281	20.8
80-85	.06147	13.1	.03818	10.3	.07075	13.5	.05234	13.0

Table 13.

Expectation of life and the effect of eliminating cancer all forms (R_2) as a cause of death

(Canada 1968 and France 1969)

Age Interval (x_i, x_{i+1})	Canada				France			
	Male		Female		Male		Female	
	$\hat{e}_i$	$\hat{e}_{i.2}-\hat{e}_i$	$\hat{e}_i$	$\hat{e}_{i.2}-\hat{e}_i$	$\hat{e}_i$	$\hat{e}_{i.2}-\hat{e}_i$	$\hat{e}_i$	$\hat{e}_{i.2}-\hat{e}_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0-1	69.04	2.37	75.69	2.75	67.82	3.58	75.38	3.12
1-5	69.66	2.47	76.09	2.87	68.11	3.66	75.50	3.15
5-10	65.93	2.47	72.33	2.87	64.36	3.65	71.73	3.14
10-15	61.12	2.45	67.47	2.86	59.50	3.64	66.84	3.12
15-20	56.27	2.45	62.57	2.85	54.64	3.61	61.93	3.11
20-25	51.61	2.44	57.72	2.84	49.96	3.56	57.09	3.08
25-30	47.07	2.45	52.88	2.82	45.34	3.54	52.27	3.08
30-35	42.41	2.42	48.04	2.80	40.69	3.51	47.45	3.07
35-40	37.73	2.40	43.23	2.77	36.06	3.49	42.66	3.05
40-45	33.12	2.36	38.50	2.68	31.56	3.42	37.95	3.00
45-50	28.65	2.33	33.86	2.54	27.23	3.32	33.35	2.88
50-55	24.41	2.25	29.37	2.34	23.10	3.18	28.88	2.69
55-60	20.45	2.14	25.05	2.08	19.24	2.96	24.55	2.49
60-65	16.86	1.93	20.92	1.80	15.71	2.68	20.37	2.22
65-70	13.65	1.71	17.04	1.50	12.60	2.28	16.40	1.93
70-75	10.85	1.41	13.46	1.22	9.88	1.83	12.74	1.61
75-80	8.36	1.07	10.24	.94	7.53	1.37	9.56	1.28
80-85	6.25	.88	7.46	.78	5.54	1.01	6.93	.99
85+	4.61	.73	5.37	.76	4.01	.91	4.90	.93

at age 85. The average loss of length of life due to cancer is about 2.0 years. The length of life lost due to all accidents also decreases with the advancement of age. At age 0, the loss is 1.25, while at age 75, .19 years. On the average the loss due to all accidents is less than one year.

It may be noted that, in comparison with the findings in Table 6, cardiovascular-renal diseases are a more serious cause of death in the United States than they are in the Federal Republic of Germany.

4.2. Cancer all forms

Cancer all forms is next only to heart disease as a major cause of death. It claimed about 17 percent of all deaths in the United States in recent years. In spite of immeasurable amounts of scientific research effort, the cause of the disease is still unknown, and effective treatment is yet to be found. Concern has been expressed regarding the susceptibility to the disease as a function of age, sex, race, socio-economic status and others. To show how these diseases affect longevity of people of different ages, sex, and locality, we have computed the probability of dying $\hat{q}_{i.2}$ when cancer all forms is eliminated as a risk of death and the corresponding expectation of life $\hat{e}_{i.2}$ for the populations of Canada and France. The findings are recorded in Tables 12 and 13. For both the Canadian and French males (age 25-80) the difference $\hat{q}_i - \hat{q}_{i.2}$ increases as age advances, although the effect on French males is more pronounced. The reverse pattern holds for females between ages 20 and 60, where the Canadian females are more affected than the French females. In the age interval from 35 to 55 in Canada, the difference between the two probabilities is greater for females than for males. This may be attributable to the prevalence of breast cancer among women.

The number of years of life lost due to cancer all forms is greater for the French population than for the Canadian population for both sexes and all age categories. In France, the males would gain more years of life than females if cancer all forms was eliminated as a risk of death, while in Canada females would gain more years of life than males up to age 55.

5. The Life Table when a Particular Cause Alone is Operating in a Population

The procedure in constructing a life table when a particular risk is the only risk operating in a population is also the same as that described in Section 2 of this Chapter, except for the difference in the basic quantities. As an example, let us consider the net probability of dying, q_{i1}, when risk R_1 is the only risk acting. Since q_{i1} cannot be estimated directly, we make use of the result in the competing risks and estimate q_{i1} from the formula (cf., Equation (2.21a) in Appendix III),

$$q_{i1} = Q_{i1} \left(1 + \frac{1}{2} (q_i - Q_{i1})\right) \quad . \tag{5.1}$$

When a life table is for a current population, q_i and Q_{i1} are estimated, as in Section 2, from

$$\hat{q}_i = \frac{n_i M_i}{1 + (1-a_i)\, n_i M_i} \tag{2.4}$$

and

$$\hat{Q}_{i1} = \frac{n_i M_i}{1 + (1-a_i)\, n_i M_i} \tag{2.5}$$

and hence,

$$\hat{q}_{i1} = \hat{Q}_{i1} \left[1 + \frac{1}{2} (\hat{q}_i - \hat{Q}_{i1})\right] . \tag{5.2}$$

For the last age interval, e.g., 85 and over, $\hat{q}_{85,1} = 1$. When all the $\hat{q}_{i1}$ have been computed, we assume a radix $\ell_{01} = 100{,}000$ and proceed to construct the rest of the table in the same way as before. We shall not repeat the description.

CHAPTER 9

MEDICAL FOLLOW-UP STUDIES

1. Introduction

Statistical studies in the general category of medical follow-up and life testing have as their common immediate objective the estimation of life expectancy and survival rates for a defined population at risk. Such studies usually must be terminated before all survival information is complete and are therefore said to be truncated. The nature of the problem in an investigation concerned with the medical follow-up of patients is the same as in the life testing of electric bulbs, although differences in sample size may require different approaches. For illustration, we use cancer survival data of a large sample and therefore our terminology is the same as that of the medical follow-up study.

In a typical follow-up study, a group of individuals with some common morbidity experience is followed from a well-defined zero point, such as date of hospital admission. The purpose of the study might be to evaluate a certain therapeutic measure by comparing the expectation of life and survival rates of treated patients with those of untreated patients, or by comparing the expectation of life of treated and presumably cured patients with that of the general population. When the period of observation ends, there will usually remain a number of individuals for whom the mortality data is incomplete. First, some patients will still be alive at the close of the study.

Second, some patients will have died from causes other than those under study, so that the chance of dying from the specific cause cannot be determined directly. Finally, patients will be "lost" to the study because of follow-up failure. These three sources of incomplete information have created interesting statistical problems in the estimation of the expectation of life and survival rates. Many contributions have been made to methods of analysis of follow-up data. They include the studies of Greenwood [1925], Frost [1933], Berkson and Gage [1952], Fix and Neyman [1951], Boag [1949], Elveback [1958], Armitage [1959], Kaplan and Meier [1958], Dorn [1950], and Littell [1952]. For the material presented in this chapter, reference may be made to Chiang [1961a].

The purpose of this chapter is to adapt the life table methodology and competing risk theory, presented in Appendix II and III, to the special conditions of follow-up studies. Section 2 is concerned with the general type of study which investigates mortality experience without reference to cause of death. The maximum likelihood estimator of the probability of dying is derived, and a method is suggested for computing the observed expectation of life in such studies. Section 3 extends the discussion to follow-up studies with the consideration of competing risks, and presents formulas for the estimators of the net, crude, and partial crude probabilities. The problem of lost cases is treated in Section 4, where a patient's absence is considered as a competing risk. Application of the theoretical matter is illustrated with empirical data of a follow-up study of cervical cancer patients.

2. Estimation of Probability of Survival and Expectation of Life

Consider a follow-up program conducted over a period of y years. A total of N_0 patients are admitted to the program at any time during the study period and observed until death or until termination of the study, whichever comes first. The time of admission is taken as the common point of origin for all N_0 patients; thus N_0 is the number of patients with which the study begins, or the number alive at time zero. The time axis refers to the time of follow-up since admission, and x denotes the exact number of years of follow-up. A constant time interval of one year will be used for simplicity of notation, with the typical interval denoted by $(x, x+1)$, for $x=0,1,\cdots,y-1$. The symbol p_x will be used to denote the probability that a patient alive at time x will survive the interval $(x, x+1)$, and q_x the probability that he will die during the interval, with $p_x+q_x = 1$.

2.1. Basic random variables and likelihood functions. For each interval $(x, x+1)$ let N_x be the number of patients alive at the beginning of the interval. Clearly, N_x is also the number of survivors of those who entered the study at least x years before the closing date.[1/] The number N_x will decrease as x increases because of deaths and withdrawal of patients due to termination of the study. The decrease in N_x is systematically described below with reference to Table 1.

Table 1

Distribution of N_x patients according to withdrawal status and survival status in the interval $(x,x+1)$

Survival status	Withdrawal status in the interval		
	Total number of patients	Number to be observed for the entire interval*	Number due to withdraw during the interval**
Total	N_x	m_x	n_x
Survivors	s_x+w_x	s_x	w_x
Deaths	D_x	d_x	d'_x

* Survivors among those admitted to the study more than $(x+1)$ years before closing date for individual patients.

** Survivors among those admitted to the study less than $(x+1)$ years but more than x years before closing date for individual patients.

The N_x individuals who begin the interval $(x, x+1)$ comprise two mutually exclusive groups differentiated according to their date of entrance into the program. A group of m_x patients who entered the program more than $x+1$ years before the closing date will be observed for the entire interval; a second group of n_x patients who entered the program less than $x+1$ years before its termination is due to withdraw in the interval because the closing date precedes their $(x+1)$th anniversary date. Of the m_x patients d_x will die in the interval and s_x will survive to the end of the interval and become N_{x+1}; of the n_x patients d'_x will die before the closing date and w_x will survive to the closing date of the study. The sum $d_x+d'_x = D_x$ is the total number of deaths in the interval. Thus s_x, d_x, w_x, and d'_x are the basic random variables and will be used to estimate the probability p_x that a patient alive at x will survive the interval $(x, x+1)$, and its complement q_x.

Consider first the group of m_x individuals each of whom has a constant probability p_x of surviving and $q_x = 1-p_x$ of dying in the interval $(x, x+1)$. Thus, the random variable s_x has the binomial distribution:

$$C_1 \, p_x^{s_x} (1-p_x)^{d_x} \tag{2.1}$$

where C_1 is the binomial coefficient. The expected number of surviviors and the expected number of deaths are given by

$$E(s_x | m_x) = m_x p_x \qquad \text{and} \qquad E(d_x | m_x) = m_x(1-p_x) \tag{2.2}$$

respectively.

The distribution of the random variables in the group of n_x patients depends upon the time of withdrawal. A plausible assumption is that the withdrawals take place at random during the interval (x, x+1). Under this assumption the probability that a patient will survive to the closing date is

$$-(1-p_x)/\ln p_x \ , \tag{2.3}$$

which is approximately equal to $p_x^{1/2}$, or

$$-(1-p_x)/\ln p_x = p_x^{1/2} \ , \tag{2.4}$$

since the probability p_x of surviving the interval is almost always large. The quantities on both sides of (2.4) have been computed for selected values of p_x, and the results shown in Table 2 justify the approximation.

Table 2

Comparison between $p_x^{1/2}$ and $-(1-p_x)/\ln p_x$

p_x	$p_x^{1/2}$	$-(1-p_x)/\ln p_x$
.70	.837	.841
.75	.866	.869
.80	.894	.896
.85	.922	.923
.90	.949	.949
.95	.975	.975

Consequently, $p_x^{\frac{1}{2}}$ is taken as the probability of surviving to the closing date and $(1-p_x^{\frac{1}{2}})$ as the probability of dying before the time of withdrawal. Thus the probability distribution of the random variable w_x in the group of n_x patients due to withdraw is also binomial:

$$C_2 \; p_x^{\frac{1}{2}w_x}(1-p_x^{\frac{1}{2}})^{d'_x} \tag{2.5}$$

where C_2 is the binomial coefficient. The expected number of survivors and the expected number of deaths are given by

$$E(w_x|n_x) = n_x p_x^{\frac{1}{2}} \quad \text{and} \quad E(d'_x|n_x) = n_x(1-p_x^{\frac{1}{2}}) \;, \tag{2.6}$$

respectively.

Since the N_x individuals comprise two independent groups according to their withdrawal status, the likelihood function of all the random variables is the product of the two probability functions (2.1) and (2.5), or

$$L_x = C \; p_x^{(s_x+\frac{1}{2}w_x)}(1-p_x)^{d_x}(1-p_x^{\frac{1}{2}})^{d'_x} \;, \qquad x=0,1,\cdots,y-1. \tag{2.7}$$

where C stands for the product of the combinatorial factors in (2.1) and (2.5).

2.2. Maximum likelihood estimators of the probabilities p_x and q_x. The maximum likelihood estimators of the probability p_x is a value of p_x at which the function L_x in (2.7) attains a maximum. The estimator is given by

$$\hat{p}_x = \left[\frac{-\frac{1}{2}d'_x + \sqrt{\frac{1}{4}d'^2_x + 4(N_x - \frac{1}{2}n_x)(s_x + \frac{1}{2}w_x)}}{2(N_x - \frac{1}{2}n_x)} \right]^2 \tag{2.8}$$

with the complement

$$\hat{q}_x = 1 - \hat{p}_x \ , \qquad x = 0, 1, \cdots, y-1. \tag{2.9}$$

The maximum likelihood estimator (2.14) is not unbiased, but is consistent in the sense of Fisher. When the random variables s_x, w_x, and d'_x are replaced with their respective expectations as given by (2.5) and (2.9), the resulting expression is identical with the probability p_x.

The exact formula for the variance of the estimator $\hat{p}_x$ in (2.8) is unknown, but an approximate formula is stated below for practical applications.

$$S^2_{\hat{p}_x} = \frac{\hat{p}_x \, \hat{q}_x}{M_x} \tag{2.10}$$

where

$$M_x = m_x + n_x(1 + \hat{p}_x^{\frac{1}{2}})^{-1} \tag{2.11}$$

Formula (2.10) is quite similar to the variance of a binomial proportion except that M_x instead of N_x is in the denominator. However, M_x is the more logical choice, since a patient who is to be observed for a fraction of the period $(x, x+1)$ should be weighted less than one who is to be observed for the entire period. According to equation (2.11), the experience of each of the m_x patients is counted as a whole "trial," whereas the experience of each of the n_x patients due to withdraw is counted as a fraction $(1+p_x^{\frac{1}{2}})^{-1}$ of a "trial." The fraction is dependent upon the probability p_x of survival. The smaller the probability p_x, the larger will be the fraction. When $p_x=0$, $M_x = m_x+n_x$; when $p_x=1$, $M_x = m_x+\frac{1}{2}n_x$.

2.3. Estimation of survival probability. A life table for follow-up subjects can be readily constructed once $\hat{p}_x$ and $\hat{q}_x$ have been determined from (2.8) and (2.9) for each interval of the study. The procedure is the same as for the current life table. Because of their practical importance, we shall consider only the x-year survival rate and the expectation of life.

The x-year survival rate is an estimate of the probability that a patient will survive from the time of admission to the xth anniversary ; it is computed from

$$\hat{p}_{0x} = \hat{p}_0\hat{p}_1\cdots\hat{p}_{x-1} , \qquad x=1,2,\cdots,y. \qquad (2.12)$$

The sample variance of $\hat{p}_{0x}$ has the same form as that given in equation (2.7) of Chapter 3.

$$S^2_{\hat{p}_{0x}} = \hat{p}_{0x}^{\,2} \sum_{u=0}^{x-1} \hat{p}_u^{\,-2}\, S^{\,2}_{\hat{p}_u} \quad . \tag{2.13}$$

2.4. <u>Estimation of the expectation of life</u>. To avoid confusion in notation, let us denote by α a fixed number and by $\hat{e}_\alpha$ the observed expectation of life at time α computed from the following formula[2/]:

$$\hat{e}_\alpha = \tfrac{1}{2} + \hat{p}_\alpha + \hat{p}_\alpha \hat{p}_{\alpha+1} + \cdots + \hat{p}_\alpha \hat{p}_{\alpha+1} \cdots \hat{p}_{y-1} + \hat{p}_\alpha \hat{p}_{\alpha+1} \cdots \hat{p}_y + \cdots \; . \tag{2.14}$$

In a study covering a period of y years, if no survivors remain from the patients who entered the program in its first year, $\hat{p}_{y-1}$ will be zero, and $\hat{e}_\alpha$ can be computed from (2.14). However, usually there will be w_{y-1} survivors who were admitted in the first year of the program and are still living at the closing date. In such cases (2.8) shows that $\hat{p}_{y-1}$ is greater than zero, and the values of $\hat{p}_y, \hat{p}_{y+1}, \cdots$ are not observed within the time limits of the study. Consequently, $\hat{e}_\alpha$ cannot be obtained from equation (2. 14).

Nevertheless, $\hat{e}_\alpha$ may be computed with a certain degree of accuracy if w_{y-1} is small. Suppose we rewrite equation (2.14) in the form

$$\hat{e}_\alpha = \tfrac{1}{2} + \hat{p}_\alpha + \hat{p}_\alpha \hat{p}_{\alpha+1} + \cdots + \hat{p}_\alpha \hat{p}_{\alpha+1} \cdots \hat{p}_{y-1} + \hat{p}_{\alpha y}(\hat{p}_y + \hat{p}_y \hat{p}_{y+1} + \cdots), \tag{2.15}$$

where $\hat{p}_{\alpha y}$ is written for $\hat{p}_\alpha \hat{p}_{\alpha+1} \cdots \hat{p}_{y-1}$. The problem is to determine $\hat{p}_y, \hat{p}_{y+1}, \cdots$ in the last term, since the preceding terms can be computed from the data available.

Consider a typical interval $(z, z+1)$ beyond time y with the survival probability of p_z, for $z=y, y+1, \ldots$ If the force of mortality is constant beyond y, the probability of surviving the interval $(z, z+1)$ becomes independent of z, or

$$p_z = p, \qquad\qquad z = y, y+1, \ldots \quad . \tag{2.16}$$

Under this assumption, we may replace the last term of (2.15) with $\hat{p}_{\alpha y}(\hat{p}+\hat{p}^2+\ldots)$, which converges to $\hat{p}_{\alpha y}\hat{p}/(1-\hat{p})$, or

$$\hat{p}_{\alpha y}(\hat{p}+\hat{p}^2+\ldots) = \hat{p}_{\alpha y}\,\frac{\hat{p}}{1-\hat{p}} \quad . \tag{2.17}$$

As a result, we have

$$\hat{e}_\alpha = \tfrac{1}{2} + \hat{p}_\alpha + \hat{p}_\alpha\hat{p}_{\alpha+1} + \ldots + \hat{p}_\alpha\hat{p}_{\alpha+1}\cdots\hat{p}_{y-1} + \hat{p}_{\alpha y}\left(\frac{\hat{p}}{1-\hat{p}}\right) \quad . \tag{2.18}$$

Clearly, $\hat{p}$ may be set equal to $\hat{p}_{y-1}$ if the force of mortality is assumed to be constant beginning with time $(y-1)$ instead of time y. In order to have small sample variation, however, the estimate of $\hat{p}$ should be based on as large a sample as possible. Suppose there exists a time t, for $t<y$, such that $\hat{p}_t, \hat{p}_{t+1}, \ldots$ are approximately equal, thus indicating a constant force of mortality after time t. Then, $\hat{p}$ may be set equal to $\hat{p}_t$, and we have the formula for the observed expectation of life,

$$\hat{e}_\alpha = \tfrac{1}{2} + \hat{p}_\alpha + \hat{p}_\alpha\hat{p}_{\alpha+1} + \cdots + \hat{p}_\alpha\hat{p}_{\alpha+1}\cdots\hat{p}_{y-1} + \hat{p}_{\alpha y}\left(\frac{\hat{p}_t}{1-\hat{p}_t}\right) , \tag{2.19}$$

for $\alpha=0,\ldots,y-1$.

Although formula (2.19) holds for $\alpha=0,\cdots,y-1$, it is apparent that the smaller the value of α, the smaller the value of $\hat{p}_{\alpha y}$. When $\hat{p}_{\alpha y}$ is small, the error in assuming a constant force of mortality beyond y and in the choice of $\hat{p}_t$ will have but little effect on the value of $\hat{e}_\alpha$.

2.5. <u>Sample variance of the observed expectation of life</u>. In Appendix II we prove that the estimated probabilities of surviving any two non-overlapping intervals have a zero covariance; hence, the sample variance of the observed expectation of life may be computed from

$$s^2_{\hat{e}_\alpha} = \sum_{x\geq\alpha}\left\{\frac{\partial}{\partial\hat{p}_x}\hat{e}_\alpha\right\}^2 s^2_{\hat{p}_x} \quad . \tag{2.20}$$

The derivatives, taken at the observed point $\hat{p}_x$, $x\geq\alpha$, are given by

$$\left\{\frac{\partial}{\partial\hat{p}_x}\hat{e}_\alpha\right\} = \hat{p}_{\alpha x}\left[\hat{e}_{x+1} + \tfrac{1}{2}\right] , \qquad x\neq t \tag{2.21}$$

where

$$p_{\alpha x} = p_\alpha p_{\alpha+1} \cdots p_{\alpha-1} ,$$

and

$$\left\{\frac{\partial}{\partial\hat{p}_t}\hat{e}_\alpha\right\} = \hat{p}_{\alpha t}\left[\hat{e}_{t+1} + \tfrac{1}{2} + \frac{\hat{p}_{ty}}{(1-\hat{p}_t)^2}\right] , \qquad \alpha\leq t \tag{2.22}$$

For $t<\alpha$, the factors $\hat{p}_\alpha$, $\hat{p}_{\alpha+1}$, ..., $\hat{p}_{y-1}$ and $\hat{p}_{\alpha y}$ in (2.19) do not contain $\hat{p}_t$; hence the derivative

$$\frac{\partial}{\partial p_t}\hat{e}_\alpha = \frac{\partial}{\partial\hat{p}_t}\hat{p}_{\alpha y}\left(\frac{\hat{p}_t}{1-\hat{p}_t}\right) = \hat{p}_{\alpha y}\frac{1}{(1-\hat{p}_t)^2} \tag{2.22a}$$

Substituting (2.21), (2.22) and (2.22a) in (2.20) gives the sample variance of $\hat{e}_\alpha$,

$$S^2_{\hat{e}_\alpha} = \sum_{\substack{x=\alpha \\ x\neq t}}^{y-1} \hat{p}_{\alpha x}^{\,2} \left[\hat{e}_{x+1} + \tfrac{1}{2}\right]^2 S^2_{\hat{p}_x} + \hat{p}_{\alpha t}^{\,2} \left[\hat{e}_{t+1} + \tfrac{1}{2} + \frac{\hat{p}_{ty}}{(1-\hat{p}_t)^2}\right]^2 S^2_{\hat{p}_t}, \quad \alpha \leq t, \tag{2.23}$$

and

$$S^2_{\hat{e}_\alpha} = \sum_{x=\alpha}^{y-1} \hat{p}_{\alpha x}^{\,2} \left[\hat{e}_{x+1} + \tfrac{1}{2}\right]^2 S^2_{\hat{p}_x} + \frac{\hat{p}^2_{\alpha y}}{(1-\hat{p}_t)^4} S^2_{\hat{p}_t}, \qquad \alpha > t. \tag{2.24}$$

The value of $\hat{p}_x$ and the sample variance of $\hat{p}_x$ are obtained from formulas (2.8) and (2.10), respectively.

When the first term in formula (2.23) or (2.24) is taken out of the summation sign, we have a recursive equation

$$S^2_{\hat{e}_\alpha} = [\hat{e}_{\alpha+1} + \tfrac{1}{2}]^2 S^2_{\hat{p}_\alpha} + \hat{p}^2_\alpha S^2_{\hat{e}_{\alpha+1}}, \qquad \text{for } \alpha \neq t. \tag{2.25}$$

Therefore, the variance of $\hat{e}_\alpha$ may be computed successively beginning with the largest value of α.

2.6 <u>An example of life table construction for a follow-up population.</u>

Application of the methods developed in this section is illustrated with data collected by the Tumor Registry of the California State Department of Public Health . The material selected consists of 5,982 white female patients admitted to certain California hospitals and clinics between January 1, 1942, and December 31, 1954, with a diagnosis of cervical cancer. For the purpose of this illustration, the closing date is December 31, 1954, and the date of entrance to follow-up for each patient is the date of hospital admission. Each patient was observed until death or until the closing date, whichever came first.

The first step is to construct a table similar to Table 3, showing the survival experience of the patients grouped according to their withdrawal status for each time period of follow-up. The interval length selected (column 1) will depend upon the nature of the investigation; generally a fixed length of one year is used. The total number of patients admitted to the study is entered as N_0 in the first line of column 2, which is 5,982. Among them there were m_0=5,317 patients (column 3), observed for the entire interval (0,1). Of the m_0 patients, s_0 (4,030, column 4) survived to their first anniversary and d_0 (1,287, column 5) died during the first year of follow-up. In addition, there were n_0 (665, column 6) patients due to withdraw in the interval (0,1), of which w_0 (576, column 7) survived to the closing date and d'_0 (89, column 8) died before the closing date. The second interval began with the s_0 = 4,030 survivors from the first interval, which is entered as N_1 in column 2 of line 2. The N_1 patients were again divided successively by withdrawal and survival status. Of the N_1 patients, m_1 (3,489, column 3) were the survivors of those admitted prior to January 1, 1953, and hence were observed for the entire interval (1, 2); n_1 (541, column 8) were the survivors of those admitted during the year 1953 and hence were due to withdraw during the interval. At the beginning of the final interval (12, 13) there were N_{12} = 72 survivors of the patients admitted in 1942; all were due to withdraw during the last interval, or n_{12} = 72 (last line, column 8). Of the 72 patients, w_{12} = 72 (column 7) were alive at the closing date. This means that $\hat{p}_{12}$ is greater than zero, and $\hat{p}_z$ for $z \geq 12$ cannot be observed.

Table 3

SURVIVAL EXPERIENCE FOLLOWING DIAGNOSIS OF CANCER OF THE CERVIX UTERI
CASES INITIALLY DIAGNOSED 1942-1954
CALIFORNIA, U.S.A.

Interval since diagnosis (in years)	Number living at beginning of interval (x, x+1)	Number to be observed for entire interval (x, x+1)*			Number due for withdrawal in interval (x, x+1)**		
		Total number	Number surviving the interval	Number dying in the interval	Total due for withdrawal	Number living at time of withdrawal	Number dying before withdrawal
(x, x+1)	N_x	m_x	s_x	d_x	n_x	w_x	d'_x
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0-1	5982	5317	4030	1287	665	576	89
1-2	4030	3489	2845	644	541	501	40
2-3	2845	2367	2117	250	478	459	19
3-4	2117	1724	1573	151	393	379	14
4-5	1573	1263	1176	87	310	306	4
5-6	1176	918	861	57	258	254	4
6-7	861	692	660	32	169	167	2
7-8	660	496	474	22	164	161	3
8-9	474	356	344	12	118	116	2
9-10	344	256	245	11	88	85	3
10-11	245	164	158	6	81	78	3
11-12	158	76	72	4	82	80	2
12-13	72	0	0	0	72	72	0

*Survivors of those admitted more than x+1 years prior to closing date.
**Survivors of those admitted between x and x+1 years prior to closing date.

Source: California Tumor Registry, Department of Public Health, State of California

This material has been used to construct a life table for the cervical cancer patients. The steps involved are similar to those described in the construction of current life tables in Chapter 3. For easy reference, but at the expense of repetition, they are stated below:

(1) $\hat{p}_x$ and $\hat{q}_x$. For each interval (x, x+1), use formulas (2.8) and (2.9) of this chapter to compute $\hat{p}_x$ and $\hat{q}_x$.

(2) d_x and ℓ_x. Assume $\ell_0 = 100{,}000$, use $\hat{q}_0, \hat{q}_1, \cdots$ to obtain d_x and ℓ_x from

$$d_x = \ell_x \hat{q}_x \qquad \text{and} \qquad \ell_{x+1} = \ell_x - d_x$$

for $x = 0,1,\cdots,12$.

(3) a_x and L_x. The fraction of last year of life is assumed to be $a_x = .5$, which is quite appropriate for such studies. The quantity L_x is computed from

$$L_x = \ell_{x+1} + a_x d_x$$

or since $a_x = .5$ and $d_x = \ell_x - \ell_{x+1}$,

$$L_x = \tfrac{1}{2}(\ell_x + \ell_{x+1}), \qquad \text{for } x = 0,1,\cdots,12.$$

(4) T_x and $\hat{e}_x$ beyond the observation period. Information derived from a follow-up study is incomplete for the construction of a life table inasmuch as it is limited to the study period (13 years in this example). Therefore, some device needs to be developed for the computation of $\hat{e}_x$ beyond the last year of study, that is $\hat{e}_{13}$ in the present case. Here we make use of equation (2.19) of this chapter and

write

$$\hat{e}_{13} = \tfrac{1}{2} + \frac{\hat{p}_t}{1 - \hat{p}_t} \quad . \tag{2.19 a}$$

Estimating p_t with $\hat{p}_{11}$,

$$\hat{p}_t = \hat{p}_{11} = 1-\hat{q}_{11} = 1-.05106 = .94894$$

gives required value

$$\hat{e}_{13} = \tfrac{1}{2} + \frac{.94894}{1 - .94894} = 19.0848$$

Using this figure we compute

$$T_{13} = \ell_{13}\,\hat{e}_{13} = 34{,}277 \times 19.0848 = 654{,}170 \quad .$$

(5) T_x and $\hat{e}_x$. The quantities T_x and $\hat{e}_x$ for other intervals now can be obtained by simple computations. For example,

$$T_{12} = L_{12} + T_{13} \; .$$

In general

$$T_x = L_x + T_{x+1} \qquad \text{for } x = 0,1,\cdots,12,$$

and $\hat{e}_x$ (except for $\hat{e}_{13}$) is computed from[3/]

$$\hat{e}_x = \frac{T_x}{\ell_x} \qquad \text{for } x = 0,1,\cdots,12.$$

The results of the computations are given in Table 4.

For comparison between survival experience of different study groups or for making other statistical inferences, we computed the standard deviations of the survival rate [eq. (2.13)], of probability of death [eq. (2.10)], and of the expectation of life [eqs. (2.23), (2.24) and (2.25)] for each x. Numerical values of the standard errors and the main life table functions are shown in Table 5.

For example, at x=2, the calculations for $S_{\hat{q}_2}$ were as follows:

$$S^2_{\hat{q}_x} = \frac{\hat{q}_2(1-\hat{q}_x)}{M_x} \text{ where } M_x = m_x+n_x(1+\hat{p}_x^{1/2})^{-1}$$

$$M_2 = m_2+n_2+(1+\hat{p}_2^{1/2})^{-1} = 2367 + 478(1+.8967^{1/2})^{-1} = 2,612.495$$

$$S^2_{\hat{q}_2} = \frac{\hat{q}_2(1-\hat{q}_2)}{M_2} = \frac{.10303(.89697)}{2612.495} = .00003537$$

$$S_{\hat{q}_2} = \ .00003537 = .00595 = S_{p_2}$$

To calculate $S_{\hat{p}_{03}}$ from $S^2_{\hat{p}_{0x}} = \hat{p}^2_{0x} \sum_{u=1}^{x-1} p_u^{-2} S^2_{\hat{p}_u}$:

$$S^2_{\hat{p}_{03}} = p^2_{03} \sum_{u=0}^{2} p_u^{-2} S^2_{\hat{p}_u}$$

$$= (.55615)^2 \left[\frac{(.00569)^2}{(.75746)^2} + \frac{(.00626)^2}{(.81857)^2} + \frac{(.00595)^2}{(.98697)^2}\right]$$

$$= (.30930)[.0001589] = .00004915$$

$$S_{p_{03}} = \sqrt{.00004915} = .00701$$

To calculate $S_{\hat{e}_3}$ from $S^2_{\hat{e}_\alpha} = \hat{p}^2_\alpha S^2_{\hat{e}_{\alpha+1}} + [e_{\alpha+1} + \tfrac{1}{2}]^2 S^2_{\hat{p}_\alpha}$:

$$S^2_{\hat{e}_3} = \hat{p}^2_3 S^2_{\hat{e}_4} + [\hat{e}_4 + .5]^2 S^2_{\hat{p}_3}$$

$$= (1-.10303)^2 \ 5.09^2 + [19.31+.5]^2(.00595)^2$$

$$= 20.84450 + .01389 = 20.8584$$

$$S_{\hat{e}_3} = \sqrt{20.8584} = 4.567$$

Table 4

LIFE TABLE OF PATIENTS DIAGNOSED AS HAVING CANCER OF THE CERVIX UTERI

CASES INITIALLY DIAGNOSED 1942 - 1954
CALIFORNIA, U.S.A.

Interval since diagnosis (years)	Number living at time x	Probability of dying in interval (x, x+1)	Number dying in interval (x, x+1)	Fraction of last year of life	Number of years lived in interval (x, x+1)	Number of years lived beyond x	Observed Expectation of life at x
x, x+1	ℓ_x	$\hat{q}_x$	d_x	a_x	L_x	T_x	$\hat{e}_x$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0-1	100,000	.24254	24,254	.5	87,873	1,289,575	12.90
1-2	75,746	.18143	13,743	.5	68,875	1,201,702	15.86
2-3	62,003	.10303	6,388	.5	58,809	1,132,827	18.27
3-4	55,615	.08576	4,770	.5	53,230	1,074,018	19.31
4-5	50,845	.06413	3,261	.5	49,215	1,020,788	20.08
5-6	47,584	.05820	2,769	.5	46,200	971,573	20.42
6-7	44,815	.04376	1,961	.5	43,835	925,373	20.65
7-8	42,854	.04320	1,851	.5	41,929	881,538	20.57
8-9	41,003	.03369	1,381	.5	40,313	839,609	20.48
9-10	39,622	.04655	1,844	.5	38,700	799,296	20.17
10-11	37,778	.04385	1,657	.5	36,950	760,596	20.13
11-12	36,121	.05106	1,844	.5	35,199	723,646	20.03
12-13	34,277	.00000	0	.5	34,277	688,447	20.08
13	34,277					654,170	19.08*

* **For computation of e_{13}** and T_{13} see text (2.19a)

Table 5

SURVIVAL EXPERIENCE AFTER DIAGNOSIS OF CANCER OF THE CERVIX UTERI
CASES INITIALLY DIAGNOSED 1942-1954
CALIFORNIA, U.S.A.
THE MAIN LIFE TABLE FUNCTIONS AND THEIR STANDARD ERRORS

Interval since diagnoses (years)	x-year survival rate $\hat{p}_{0x}$		Estimated probability of death in interval (x, x+1)		Observed Expectation of life at x[a]	
(x, x+1)	$1000\ \hat{p}_{0x}$	$1000\ S_{\hat{p}_{0x}}$	$1000\ \hat{q}_x$	$1000\ S_{\hat{q}_x}$	$\hat{e}_x$	$S_{\hat{e}_x}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0-1	1000.00	0.00	242.54	5.69	12.90	2.83
1-2	757.46	5.80	181.43	6.26	15.86	3.74
2-3	620.03	6.65	103.03	5.95	18.27	4.57
3-4	556.15	7.01	85.76	6.38	19.31	5.09
4-5	508.45	7.33	64.13	6.50	20.08	5.56
5-6	475.84	7.61	58.20	7.23	20.42	5.94
6-7	448.15	7.95	43.76	7.34	20.65	6.31
7-8	428.54	8.29	43.20	8.45	20.57	6.60
8-9	410.03	8.71	33.69	8.85	20.48	6.89
9-10	396.22	9.17	46.55	12.15	20.17	7.13
10-11	377.78	9.98	43.85	14.30	20.13	7.47
11-12	361.21	10.97	51.06	20.30	20.03	7.81
12-13	342.77	12.73	00.00	00.00	20.08	7.79
13	342.77	12.73	-----	-----	19.08	7.79

Source: California Tumor Registry, Department of Public Health, State of California, U.S.A.

[a] t = 11

3. Consideration of Competing Risks

Most follow-up studies are conducted to determine the survival rates of patients affected with a specific disease. These patients are also exposed to other risks of death from which some of them may eventually die. In a study determining the effectiveness of radiation as a treatment for cancer, for example, some patients may die from heart disease. In such cases, the theory of competing risks is indispensible, and the crude, net, and partial crude probabilities all play important roles.

Let us assume, as in Appendix III, that r risks, denoted by $R_1,\cdots,R_r$, are acting simultaneously on each patient in the study. For risk R_δ there is a corresponding force of mortality $\mu(\tau;\delta)$, $\delta=1,\cdots,r$, and the sum

$$\mu(\tau;1) + \cdots + \mu(\tau;r) = \mu(\tau) \tag{3.1}$$

is the total force of mortality. Within the time interval $(x, x+1)$ we assume a constant force of mortality for each risk, $\mu(\tau;\delta) = \mu(x;\delta)$, which depends only on the interval $(x, x+1)$ and the risk R_δ; for all risks, $\mu(\tau) = \mu(x)$ for $x<\tau\leq x+1$.

Consider a subinterval $(x, x+t)$, and let $Q_{x\delta}(t)$ be the crude probability that an individual alive at time x will die prior to $x+t$, $0<t\leq 1$, from R_δ in the presence of all other risks in the population. It follows directly from equation (2.8) in Mathematical Appendix III that

$$Q_{x\delta}(t) = \frac{\mu(x;\delta)}{\mu(x)}\left[1 - p_x(t)\right], \quad 0<t\leq 1; \quad \delta=1,\cdots,r. \tag{3.2}$$

From (3.1) we see that the sum of the crude probabilities in (3.2) is equal to the complement of $p_x(t)$, or

$$Q_{x1}(t) + \cdots + Q_{xr}(t) + p_x(t) = 1\ , \qquad 0<t\leq 1. \qquad (3.3)$$

For $t=1$, we abbreviate $Q_{x\delta}(1)$ to $Q_{x\delta}$, etc. When $t = \frac{1}{2}$, we have the subinterval $(x, x+\frac{1}{2})$ and the corresponding crude probabilities

$$Q_{x\delta}(\tfrac{1}{2}) = \frac{\mu(x;\delta)}{\mu(x)}\left[1 - p_x^{\frac{1}{2}}\right] = Q_{x\delta}\left[1 + p_x^{\frac{1}{2}}\right]^{-1}, \quad \delta=1,\cdots,r. \qquad (3.4)$$

Equation (3.3) implies that

$$Q_{x1}\left[1+p_x^{\frac{1}{2}}\right]^{-1} + \cdots + Q_{xr}\left[1+p_x^{\frac{1}{2}}\right]^{-1} + p_x^{\frac{1}{2}} = 1, \qquad x=0,1,\cdots, y-1. \qquad (3.5)$$

The net and partial crude probabilities may be computed from the following approximate relations. The corresponding exact formulas are given in Section 2, Appendix III. The net probability of death in interval $(x,x+1)$ when R_δ is the only risk operating in a population is given by

$$q_{x\delta} = Q_{x\delta}[1+\tfrac{1}{2}(q_x-Q_{x\delta}) + \tfrac{1}{6}(q_x-Q_{x\delta})(2q_x-Q_{x\delta})] \quad ; \qquad (3.6)$$

the net probability of death if R_δ is eliminated as a risk of death is given by

$$q_{x\cdot\delta} = (q_x - Q_{x\delta})[1 + \tfrac{1}{2}Q_{x\delta} + \frac{1}{6}\,Q_{x\delta}(q_x + Q_{x\delta})] \quad , \ \delta=1,\ldots,r, \tag{3.7}$$

and the partial crude probability by

$$Q_{x\delta\cdot 1} = Q_{x\delta}[1 + \tfrac{1}{2}Q_{x\delta} + \frac{1}{6}\,Q_{x1}(q_x + Q_{x1})] \quad , \tag{3.8}$$

for $\delta = 2,\ldots,r$; $x = 0,1,\ldots,w-1$.

Our immediate problem is to estimate $Q_{x\delta}$, p_x, and q_x.

3.1. Basic random variables and likelihood functions. Identification of the random variables in the present case follows directly from the discussion in Section 2.1, except that deaths are further divided by cause, as shown in Table 6.

Table 6

Distribution of N_x patients according to withdrawal status, survival status, and cause of death in the interval $(x, x+1)$

	Withdrawal status in the interval		
	Total number of patients	Number to be observed for the entire interval*	Number due to withdraw during the interval**
Total	N_x	m_x	n_x
Survivors	s_x+w_x	s_x	w_x
Deaths, all causes	D_x	d_x	d'_x
Deaths due to cause			
R_1	D_{x1}	d_{x1}	d'_{x1}
.	.	.	.
.	.	.	.
.	.	.	.
R_r	D_{xr}	d_{xr}	d'_{xr}

* Survivors among those admitted to the study more than $(x+1)$ years before closing date.

** Survivors among those admitted to the study less than $(x+1)$ years but more than x years before closing date.

The m_x patients to be observed for the entire interval $(x, x+1)$ will be divided into $r+1$ mutually exclusive groups, with s_x surviving the interval and $d_{x\delta}$ dying from cause R_δ in the interval, $\delta=1,\cdots,r$. Since the sum of the corresponding probabilities is equal to unity (eq. (3.3) the random variables s_x, $d_{x1}, \cdots, d_{xr}$ have a multinomial distribution:

$$C_1 \, p_x^{s_x} \; Q_{x1}^{d_{x1}} \cdots Q_{xr}^{d_{xr}} \quad , \tag{3.10}$$

where $s_x + d_{x1} + \cdots + d_{xr} = m_x$, and C_1 is the combinatorial factor. The expected numbers are given by

$$E(s_x | m_x) = m_x p_x \qquad \text{and} \qquad E(d_{x\delta} | m_x) = m_x Q_{x\delta} \tag{3.11}$$

respectively.

In the group of n_x patients due to withdraw in interval (x, x+1), w_x will be alive at the closing date of the study and $d'_{x\delta}$ will die from R_δ before the closing date. Each of the n_x individuals has the survival probability $p_x^{\frac{1}{2}}$ [Cf. eq.(2.4)]and the probability of dying from risk R_δ before the closing date

$$Q_{x\delta}(\tfrac{1}{2}) = Q_{x\delta} \, (1 + p_x^{\frac{1}{2}})^{-1} \; , \qquad \delta=1,\cdots,r. \tag{3.12}$$

Since $p_x^{\frac{1}{2}}$ and the probabilities in (3.12) add to unity, as shown in (3.5), the random variables w_x, $d'_{x1}, \cdots, d'_{xr}$ also have a multinomial distribution:

$$C_2 \, p_x^{\frac{1}{2}w_x} \prod_{\delta=1}^{r} \left[Q_{x\delta} \; (1 + p_x^{\frac{1}{2}})^{-1} \right]^{d'_{x\delta}} \quad , \tag{3.13}$$

where $w_x + d'_{x1} + \cdots + d'_{xr} = n_x$, and C_2 is a combinatorial factor.

The expected numbers are

$$E(w_x|n_x) = n_x p_x^{\frac{1}{2}} \qquad \text{and} \qquad E(d'_{x\delta}|n_x) = n_x Q_{x\delta}(1+p_x^{\frac{1}{2}})^{-1} \tag{3.14}$$

respectively. Because of the independence of the two groups, the likelihood function of all the random variables in Table 6 is the product of (3.10) and (3.13):

$$L_x = c\ p_x^{s_x+\frac{1}{2}w_x} \prod_{\delta=1}^{r} Q_{x\delta}^{d_{x\delta}} \left[Q_{x\delta}(1+p_x^{\frac{1}{2}})^{-1}\right]^{d'_{x\delta}} , \tag{3.15}$$

where c stands for the product of the combinatorial factors in (3.10) and (3.13). Equation (3.15) may be simplified to give the final form of the likelihood function

$$L_x = c\ p_x^{s_x+\frac{1}{2}w_x} (1+p_x^{\frac{1}{2}})^{-d'_x} \prod_{\delta=1}^{r} Q_{x\delta}^{D_{x\delta}} . \tag{3.16}$$

3.2 <u>Estimation of crude, net, and partial crude probabiities</u>. We again use the maximum likelihood principle to obtain the estimators of the probabilities $p_x, Q_{x1}, \cdots, Q_{xr}$. The estimator of p_x is the same as that obtained in Section 2; namely

$$\hat{p}_x = \left[\frac{-\frac{1}{2}d'_x + \sqrt{\frac{1}{4}d'^2_x + 4(N_x-\frac{1}{2}n_x)(s_x+\frac{1}{2}w_x)}}{2(N_x-\frac{1}{2}n_x)}\right]^2 , \quad x=0,1,\cdots,y-1. \tag{3.17}$$

Therefore $\hat{q}_x$ $(=1-\hat{p}_x)$ also will have the same values as that in Section 2. The estimators of the crude probabilities are given by

$$\hat{Q}_{x\delta} = \frac{D_{x\delta}}{D_x}\hat{q}_x , \qquad \delta=1,2,\cdots,r, \quad x=0,1,\cdots,y-1. \tag{3.18}$$

We now use (3.17) and (3.18) in formulas (3.6) to (3.9) to obtain the following estimators of the net and partial crude probabilities:

$$\hat{q}_{x\delta} = \hat{Q}_{x\delta}[1 + \tfrac{1}{2}(\hat{q}_x - \hat{Q}_{x\delta}) + \frac{1}{6}(\hat{q}_x - \hat{Q}_{x\delta})(2\hat{q}_x - \hat{Q}_{x\delta})] \tag{3.19}$$

$$\hat{q}_{x\cdot\delta} = (\hat{q}_x - \hat{Q}_{x\delta})[1 + \tfrac{1}{2}\hat{Q}_{x\delta} + \frac{1}{6}\hat{Q}_{x\delta}(\hat{q}_x + \hat{Q}_{x\delta})] \quad , \; \delta = 1,\ldots,r, \tag{3.20}$$

and

$$\hat{Q}_{x\delta\cdot 1} = \hat{Q}_{x\delta}[1 + \tfrac{1}{2}\hat{Q}_{x1} + \frac{1}{6}\hat{Q}_{x1}(\hat{q}_x + \hat{Q}_{x1})] \quad , \; \delta = 2,\ldots,r;$$

$$x = 0,1,\ldots,v-1 \; . \tag{3.21}$$

These are also maximum likelihood estimators and consistent in Fisher's sense. Consider for example the estimator $\hat{Q}_{x\delta\cdot 1}$ in formula (3.21) of the partial crude probability. We have seen in Section 2 that $\hat{p}_x$ is consistent in Fisher's sense. When the other random variables are replaced with the corresponding expectations, the right side of (3.21) may be simplified to $Q_{x\delta\cdot 1}$ given in (3.8), proving the consistency.

3.3. An Example. The survival experience of cervical cancer patients presented in Table 3 in Section 2.6 is used once again to illustrate the application of the theory in this section. For easy reference, the cervical cancer patients data is reproduced in Table 7, except that the number of deaths, d_x and d'_x are further divided according to cause. In this example only two causes are considered, cancer of the cervix and other causes. Therefore, we have for each interval $(x, x+1)$,

$$d_x = d_{x1} + d_{x2} \qquad \text{and } d'_x = d'_{x1} + d'_{x2} \qquad .$$

During the first year of follow-up, for example, there were $d_0 = 1{,}287$ deaths occurring among those to be observed for the entire interval, and $d'_0 = 89$ deaths among those due for withdrawal in the interval. These numbers are divided by causes:

$$1{,}287 = 1{,}105 + 182 \qquad \text{and } 89 = 70 + 19 \qquad .$$

With the numerical values of the probability of survival $(\hat{p}_x)$ and the probability of dying $(\hat{q}_x)$ obtained in Section 2, simple application of formulas (3.18) and (3.19) yield the crude probability $\hat{Q}_{x\delta}$ and the net probability $\hat{q}_{x\delta}$. Since only two causes of death are studied, the probability $\hat{q}_{x2}$ is equal to $\hat{q}_{x\cdot 1}$; the probability of dying when cancer of the cervix uteri is eliminated is the same as the probability of dying from other causes when the other causes are the only causes acting. Table 8 shows the estimated probability of surviving each interval, and the crude and net probabilities of death from cancer of the cervix uteri (R_1) and all other causes of death (R_2).

Table 7

SURVIVAL EXPERIENCE FOLLOWING DIAGNOSIS OF CANCER OF THE CERVIX UTERI
CASES INITIALLY DIAGNOSED 1942-1954
CALIFORNIA, U.S.A.

Interval since diagnosis (in years)	Number living at beginning of interval (x, x+1)	Number to be observed during the entire interval (x, x+1)*					Number due for withdrawal in interval (x, x+1)				
		Total not due for withdrawal	Number surviving the interval	Number dying in the interval			Total due for withdrawal	Number living at time of withdrawal	Number dying before withdrawal		
				Total	Cancer of the cervix	Other causes			Total	Cancer of the cervix	Other causes
(x, x+1)	N_x	m_x	s_x	d_x	d_{x1}	d_{x2}	n_x	w_x	d'_x	d'_{x1}	d'_{x2}
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
0-1	5982	5317	4030	1287	1105	182	665	576	89	70	19
1-2	4030	3489	2845	644	557	87	541	501	40	31	9
2-3	2845	2367	2117	250	206	44	478	459	19	15	4
3-4	2117	1724	1573	151	113	38	393	379	14	8	6
4-5	1573	1263	1176	87	61	26	310	306	4	2	2
5-6	1176	918	861	57	24	33	258	254	4	3	1
6-7	861	692	660	32	16	16	169	167	2	2	0
7-8	660	496	474	22	11	11	164	161	3	2	1
8-9	474	356	344	12	5	7	118	116	2	1	1
9-10	344	256	245	11	7	4	88	85	3	2	1
10-11	245	164	158	6	4	2	81	78	3	1	2
11-12	158	76	72	4	1	3	82	80	2	1	1
12-13	72	0	0	0	0	0	72	72	0	0	0

* Survivors of those admitted more than x+1 years prior to closing date.
** Survivors of those admitted between x and x+1 years prior to closing date.
Source: California Tumor Registry, Department of Public Health, State of California, U.S.A.

Table 8

SURVIVAL EXPERIENCE AFTER DIAGNOSIS OF CANCER OF THE CERVIX UTERI
CASES INITIALLY DIAGNOSED 1942-1954
CALIFORNIA, U. S. A.

ESTIMATED CRUDE AND NET PROBABILITIES OF DEATH FROM CANCER
OF THE CERVIX UTERI AND FROM OTHER CAUSES

Interval since diagnosis (years)	Probability of surviving interval (x, x+1)	Crude probabilities of death in interval (x, x+1) from		Net probabilities of death in interval (x, x+1) when	
		Cervix cancer	Other causes	Cervix Cancer Acting Alone	Cervix Cancer Eliminated
(x, x+1)	1000 $\hat{p}_x$	1000 $\hat{Q}_{x1}$	1000 $\hat{Q}_{x2}$	1000 $\hat{q}_{x1}$	1000 $\hat{q}_{x2}$
(1)	(2)	(3)	(4)	(5)	(6)
0-1	757.46	207.11	35.43	211.17	39.77
1-2	818.57	155.97	25.46	158.11	27.71
2-3	896.97	84.65	18.38	85.46	19.22
3-4	914.24	62.89	22.87	63.63	23.63
4-5	935,87	44.40	19.73	44.85	20.19
5-6	941.80	25.76	32.44	26.19	32.87
6-7	956.24	23.17	20.59	23.41	20.84
7-8	956.80	22.47	20.73	22.70	20.97
8-9	966.31	14.44	19.25	14.58	19.39
9-10	953.45	29.93	16.62	30.18	16.88
10-11	956.15	24.36	19.49	24.60	19.73
11-12	948.94	17.02	34.04	17.32	34.34
12-13	1000.00	----	----	----	----

Source: California Tumor Registry, Department of Public Health, State of California, U.S.A.

4. Lost Cases

Every patient in a medical follow-up is exposed not only to the risk of dying, but also to the risk of being lost to the study because of follow-up failure. Untraceable patients have caused difficulties in determining survival rates, as have patients withdrawing due to the termination of a study. However, lost cases and withdrawals belong to entirely different categories. In a group of N_x patients beginning the interval $(x, x+1)$, for example, everyone is exposed to the risk of being lost, but only n_x patients are subject to withdrawal in the interval. Therefore, it is incorrect to treat lost cases and withdrawals equally in estimating probabilities of survival or death. For the purpose of determining the probability of dying from a specific cause, patients lost due to follow-up failure are not different from those dying of causes unrelated to the study. Being lost, therefore, should be considered as a competing risk, and the survival experience of lost cases should be evaluated by using the methods discussed in the preceding section. In this approach to the problem all formulas in Section 3 will remain intact, the solution requiring only a different interpretation of the symbols.

Suppose we let R_r denote the risk of being lost; for the time element $(\tau, \tau+\Delta)$ in the interval $(x, x+1)$ let

$$\mu(x;r)\Delta + o(\Delta) = \Pr\{\text{a patient will be lost to the study in } (\tau, \tau+\Delta) \text{ due to follow-up failure}\}, \quad x<\tau<x+1. \quad (4.1)$$

The following are a few examples of the new interpretation:

p_x = Pr{a patient alive at time x will remain alive and under observation at time $x+1$}, (4.2)

$q_x = 1-p_x$

= Pr{a patient alive at time x will either die or be lost to the study due to follow-up failure in interval $(x, x+1)$}, (4.3)

Q_{xr} = Pr{a patient alive at time x will be lost to the study in $(x, x+1)$}, (4.4)

$q_{x.r}$ = Pr{a patient alive at time x will die in interval $(x, x+1)$ if the risk R_r of being lost is eliminated}, (4.5)

$1-q_{x.r}$ = Pr{a patient alive at x will survive to time $x+1$ if the risk R_r of being lost is eliminated}, (4.6)

$Q_{x\delta.r}$ = Pr{a patient alive at x will die in $(x, x+1)$ from risk R_δ if the risk R_r of being lost is eliminated}. (4.7)

The probabilities in (4.5), (4.6), and (4.7) are equivalent to q_x, p_x, and $Q_{x\delta}$, respectively, if there is no risk of being lost.

The symbol d_{xr} in Table 6 now stands for the number of lost cases among the m_x patients and d'_{xr} for the number of lost cases among the n_x patients; the sum $D_{xr} = d_{xr} + d'_{xr}$ is the total number of cases lost in the interval. The probabilities in (4.2) through (4.7) can be estimated from formulas (3.17) through (3.21) in Section 3.2.

FOOTNOTES

1/ These methods are equally applicable to data based either on the date of last reporting for individual patients or on the common date.

2/ For simplicity, we assume for $n_x=1$ and $a_x=1$ for all x, then $c_x=1$ in the formula for $\hat{e}_\alpha$ [Chapter 4 (4.23)].

3/ To verify these computations find $\hat{e}_0$ with t=11 using formula (2.19) of this chapter:

$$\hat{e}_0 = \tfrac{1}{2} + \hat{p}_0 + \hat{p}_0\hat{p}_1 + \ldots + \hat{p}_0\hat{p}_1 \ldots \hat{p}_{y-1} + \hat{p}_{0y}\left(\frac{\hat{p}_t}{1-\hat{p}_t}\right)$$

$$= \tfrac{1}{2} + [\hat{p}_{01}+\hat{p}_{02}+ \ldots +\hat{p}_{0y}] + \hat{p}_{0y}\left(\frac{\hat{p}_{11}}{1-\hat{p}_{11}}\right)$$

$$= .5 + 6.02541 + .34277 \left(\frac{.94894}{.05106}\right)$$

$$= 12.8957 = 12.90$$

This serves as a check of $\hat{e}_0 = \frac{T_0}{\ell_0} = \frac{1,289,575}{100,000} = 12.89575$ and thus of all L_x and T_x.

APPENDIX I

Theoretical Justification of the Method of Life Table Construction in Chapter 3

Formulas (3.7) and (4.3) in Chapter 3, expressing the relation between the probability $\hat{q}_i$ and the corresponding death rate μ_i, were introduced as intuitive concepts for the purpose of application, but they can be derived from a theoretical viewpoint. Let $\mu(x)$ be the force of mortality (mortality intensity function) at age x. It is easy to see that the probability q_i, that an individual alive at exact age x_i will die in interval (x_i, x_i+n_i) is given by [cf., Appendix II, formula (2.7)]

$$q_i = 1-\exp\{-\int_0^{n_i} \mu(x_i+\xi)\, d\xi\} \quad . \tag{1}$$

For an individual at x_i, let I_i be the number of deaths in (x_i, x_{i+1}). Clearly, $I_i = 1$, if the individual dies in (x_i, x_{i+1}) with a probability q_i and $I_i = 0$, if the individual survives the interval, with a probability $1-q_i$. Therefore, the expected number of deaths in (x_i, x_{i+1}) is $E[I_i] = q_i$. The mortality rate m_i is the ratio of the expected number of deaths q_i to the number of years an individual expects to live in the interval, or

$$m_i = \frac{1-\exp\{-\int_0^{n_i} \mu(x_i+\xi)\, d\xi\}}{\int_0^{n_i} \exp\{-\int_0^{y} \mu(x_i+\xi)\, d\xi\}dy} \quad . \tag{2}$$

Let a random variable τ_i be the fraction of the interval (x_i, x_i+n_i) lived by an individual who dies at an age included in the interval, so that τ_i assumes values between 0 and 1. The expected value of τ_i is the fraction of the last age interval of life, denoted by a_i, i.e.,

$$E(\tau_i) = a_i \quad . \tag{3}$$

For each time $t, 0 \le t \le 1$, the probability density function of τ_i is

$$g(t)dt = \frac{[\exp\{-\int_0^{n_i t} \mu(x_i+\xi)d\xi\}]\ \mu(x_i+n_i t)n_i\ dt}{q_i} \tag{4}$$

$$0 \le t \le 1$$

The quantity on the right-hand side of (4) is the probability that an individual alive at x_i will die in interval $(x_i+n_i t,\ x_i+n_i t+dn_i t)$ providing that he dies in (x_i, x_i+n_i). According to the definition of τ_i, this is also the probability that τ_i will assume values in $(t,t+dt)$, which is the density function $g(t)dt$. The integral

$$\int_0^1 g(t)\ dt = \int_0^1 \frac{\exp\{-\int_0^{n_i t} \mu(x_i+\xi)\ d\xi\}}{q_i} \mu(x_i+n_i t)n_i dt = 1\ ; \tag{5}$$

thus τ_i is a proper random variable. The expected value of τ_i may be computed as follows.

$$a_i = E(\tau_i) = \int_0^1 tg(t)dt$$

$$= \int_0^1 \frac{t\ \exp\{-\int_0^{n_i t} \mu(x_i+\xi)\ d\xi\}}{q_i} \mu\,(x_i+n_i t)n_i dt\ . \tag{6}$$

Integrating the numerator in the last expression in (6) by parts gives the expression

$$a_i = \frac{-n_i \exp\{-\int_0^{n_i} \mu(x_i+\xi)d\xi\} + \int_0^{n_i} \exp\{-\int_0^{y} \mu(x_i+\xi)d\xi\}dy}{n_i q_i}\ . \tag{7}$$

Substituting (1), (2), and (3) in the resulting formula (7) yields

$$a_i = 1 - \frac{1}{q_i} + \frac{1}{n_i m_i}\ . \tag{8}$$

Solving (8) for q_i, we obtain the fundamental relationship between q_i and m_i

$$q_i = \frac{n_i m_i}{1+(1-a_i)n_i m_i} \tag{9}$$

For age interval $(x,x+1)$ of one year $(n=1)$, we write a'_x, q_x and m_x for a_i, q_i and m_i, respectively, and have from (9)

$$q_x = \frac{m_x}{1+(1-a'_x)m_x} \quad , \tag{10}$$

where

$$q_x = 1-\exp\{-\int_0^1 \mu(x+\xi)d\xi\} \quad \text{and} \quad m_x = \frac{q_x}{\int_0^1 \exp\{-\int_0^t \mu(x+\xi)d\xi\}dt} \tag{11}$$

and

$$a'_x = \int_0^1 \frac{t\exp\{-\int_0^t \mu(x+\xi)d\xi\}}{1-\exp\{-\int_0^1 \mu(x+\xi)d\xi\}}\mu(x+t)dt \quad . \tag{12}$$

Formulas (9) and (10) are completely analogous to formulas (4.3) and (3.7) in Chapter 3.

APPENDIX II

STATISTICAL THEORY OF LIFE TABLE FUNCTIONS

1. Introduction

The concept of the life table originated in longevity studies of man, where it was always presented as a subject peculiar to public health, demography, and actuarial science. As a result, its development has not received sufficient attention in the field of statistics. Actually, the problems of mortality studies are similar to those of reliability theory and life testing, and they may be described in terms familiar to the statistically oriented mind. From a statistical point of view, human life is subject to chance. The life table systematically records the outcomes of many such experiments for a large number of individuals over a period of time. Thus the quantities in the table are random variables. Theoretical studies of the subject from a purely statistical point of view have been made; the probability distributions of life table functions have been devised and some optimum properties of these functions when they are used as estimates of the corresponding unknown quantities have been explored. The reader may refer to [Chiang, (1968), Chapter 10] for detail. Estimation problems concerning life table functions have been discussed by Grenander [1965]. The purpose of this Appendix is to give a brief presentation of the theoretical aspects of the life table. A typical abridged life table is reproduced below.

Table 1

Life Table

Age interval (in years)	Number living at age x_i	Proportion dying in interval (x_i, x_{i+1})	Fraction of last interval of life	Number dying in interval (x_i, x_{i+1})	Number of years lived in interval (x_i, x_{i+1})	Total number of years lived beyond age x_i	Observed expectation of life at age x_i
x_i to x_{i+1}	ℓ_i	$\hat{q}_i$	a_i	d_i	L_i	T_i	$\hat{e}_i$
x_0 to x_1	ℓ_0	$\hat{q}_0$	a_0	d_0	L_0	T_0	$\hat{e}_0$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
x_w and over	ℓ_w	$\hat{q}_w$		d_w	L_w	T_w	$\hat{e}_w$

The following symbols are also used in the text:

$$p_{ij} = \Pr\{\text{an individual alive at age } x_i \text{ will survive to age } x_j\}, \quad i \le j;\ i, j=0,1,\cdots, \tag{1.1}$$

and

$$1 - p_{ij} = \Pr\{\text{an individual alive at age } x_i \text{ will die before age } x_j\}, \quad i \le j;\ i, j=0,1,\cdots. \tag{1.2}$$

When $x_j = x_{i+1}$, we drop the second subscript and write p_i for $p_{i,i+1}$. No particular symbol is introduced for the probability $1 - p_{ij}$ except when $x_j = x_{i+1}$, in which case we let $1 - p_i = q_i$.

Finally, the symbol e_i is used to denote the true, unknown expectation of life at age x_i, estimated by the "observed expectation of life," $\hat{e}_i$.

All the quantities in the life table, with the exception of ℓ_0 and a_i, are treated as random variables in this chapter. The radix, ℓ_0 is conventionally set equal to a convenient number, such as $\ell_0 = 100{,}000$, so that the value of ℓ_i clearly indicates the proportion of survivors to age x_i. We adopt the convention and consider ℓ_0 a constant in deriving the probability distributions of other life table functions. The distributions of the quantities in columns L_i and T_i are not discussed because of their limited use. One remark should be made regarding the final age interval (x_w and over): In a conventional table the last interval is usually an open interval, e.g., 95 and over ; statistically speaking, x_w is a random variable and is treated accordingly. However, discussion of this point, which is given in [Chiang, (1968), Chapter 10], will not be presented here. Throughout this appendix we shall assume a homogeneous population in which all individuals are subjected to the same force of mortality, and in which one individual's survival is independent of the survival of any other individual in the group.

2. Probability distribution of ℓ_x, the number of survivors at age x.

The various functions of the life table are usually given for integral ages or for other discrete intervals. In the derivation of the distribution of survivors, however, age is more conveniently treated as a continuous variable with formulas derived for ℓ_x, the number of individuals surviving the age interval (0,x), for all possible values of x.

The probability distribution of ℓ_x depends on the force of mortality, or intensity of risk of death, $\mu(x)$, defined as follows:

$$\mu(x)\Delta + o(\Delta) = \Pr\{\text{an individual alive at age } x \text{ will die in interval } (x,x+\Delta)\}. \tag{2.1}$$

Let the continuous random variable X be the life span of a person so that the distribution function

$$F_X(x) = \Pr\{X \le x\} \tag{2.2}$$

is the probability that the individual will die prior to (or at) age x. Consider now the interval $(0,x+\Delta)$ and the corresponding distribution function $F_X(x+\Delta) = \Pr\{X \le x+\Delta\}$. For an individual to die prior to $x+\Delta$ he must die prior to x or else he must survive to x and die during the interval $(x,x+\Delta)$. Therefore, the corresponding probabilities have the relation

$$F_X(x+\Delta) = F_X(x) + [1 - F_X(x)][\mu(x)\Delta + o(\Delta)] \tag{2.3}$$

or

$$\frac{F_X(x+\Delta) - F_X(x)}{\Delta} = [1 - F_X(x)]\,[\mu(x) + \frac{o(\Delta)}{\Delta}] \,. \tag{2.4}$$

Taking the limits of both sides of (2.4) as $\Delta \to 0$, we have the differential equation

$$\frac{d}{dt} F_X(x) = [1 - F_X(x)]\mu(x) \tag{2.5}$$

with the initial condition

$$F_X(0) = 0. \tag{2.6}$$

Integrating (2.5) and using (2.6) yields the solution

$$1 - F_X(x) = e^{-\int_0^x \mu(t)dt} = p_{0x} \tag{2.7}$$

Equation (2.7) gives the probability that one individual alive at age 0 will survive to age x. If there are ℓ_0 individuals alive at age 0 who are subject to the same force of mortality, the number ℓ_x of survivors at age x is clearly a binomial random variable with the probability p_{0x} of surviving to x and the probability distribution given by

$$\Pr\{\ell_x = k\} = \frac{\ell_0!}{k!(\ell_0-k)!} p_{0x}^k(1-p_{0x})^{\ell_0-k}, \quad k=0,1,\ldots,\ell_0. \tag{2.8}$$

For $x = x_i$, the probability that an individual will survive the age interval $(0,x_i)$ is

$$p_{0i} = \exp\{-\int_0^{x_i} \mu(\tau)d\tau\} \tag{2.9}$$

and the probability distribution of the number of survivors, ℓ_i, is

$$\Pr\{\ell_i = k_i | \ell_0\} = \binom{\ell_0}{k_i} p_{0i}^{k_i}(1-p_{0i})^{\ell_0-k_i}, \quad k_i = 0,1,\ldots,\ell_0 . \tag{2.10}$$

The expected value and variance of ℓ_i given ℓ_0 are

$$E(\ell_i|\ell_0) = \ell_0 p_{0i} \tag{2.11}$$

and

$$\sigma^2_{\ell_i|\ell_0} = \ell_0 p_{0i}(1-p_{0i}), \tag{2.12}$$

respectively.

In general, the probability of surviving the age interval (x_i,x_j) is

$$p_{ij} = \exp\left\{-\int_{x_i}^{x_j} \mu(\tau)d\tau\right\}, \quad \text{for } i\le j \tag{2.13}$$

with the obvious relation

$$p_{\alpha j} = p_{\alpha i}p_{ij}\,, \qquad \text{for } \alpha\le i\le j. \tag{2.14}$$

If we start with ℓ_i individuals at x_i, the number of survivors ℓ_j at x_j, for $i\le j$, is also a binomial random variable with the probability p_{ij} and

$$\Pr\{\ell_j = k_j|\ell_i\} = \frac{\ell_i!}{k_j!(\ell_i-k_j)!}\, p_{ij}^{k_j}\,(1-p_{ij})^{\ell_i-k_j}, \quad k_j=0,1,\ldots,\ell_i \tag{2.15}$$

with the expected value and variance given by

$$E(\ell_j|\ell_i) = \ell_i p_{ij} \tag{2.16}$$

and

$$\sigma^2_{\ell_j|\ell_i} = \ell_i p_{ij}(1-p_{ij}). \tag{2.17}$$

When $j=i+1$, (2.15) becomes

$$\Pr\{\ell_{i+1} = k_{i+1}|\ell_i\} = \frac{\ell_i!}{k_{i+1}!(\ell_i-k_{i+1})!}\, p_i^{k_{i+1}}\,(1-p_i)^{\ell_i-k_{i+1}}. \tag{2.18}$$

It is intuitively clear that given ℓ_i people alive at age x_i, the probability distribution of the number of people alive at x_j, for $x_j > x_i$, is independent of $\ell_0,\ell_1,\ldots,\ell_{i-1}$. This means that for each k_j

$$\Pr\{\ell_j = k_j \mid \ell_0, \ell_1, \ldots, \ell_i\} = \Pr\{\ell_j = k_j \mid \ell_i\} \quad . \tag{2.19}$$

Consequently,

$$E(\ell_j \mid \ell_0, \ldots, \ell_i) = E(\ell_j \mid \ell_i),$$

and

$$\sigma^2_{\ell_j \mid \ell_0, \ldots, \ell_i} = \sigma^2_{\ell_j \mid \ell_i} \; .$$

In other words, for each u the sequence $\ell_0, \ell_1, \ldots, \ell_u$ is a Markov process.

2.1. Mortality laws.

The survival probability in (2.7) has been known to life-table students for more than two hundred years. Unfortunately, it has not been given due recognition by investigators in statistics although differing forms of this function have appeared in various areas of research. We shall mention a few below in terms of the probability density function of X,

$$\begin{aligned} f_X(x) = \frac{dF_X(x)}{dx} &= \mu(x) e^{-\int_0^x \mu(t)dt} && x \geq 0 \\ &= 0 && x < 0. \end{aligned} \tag{2.20}$$

(i) Gompertz Distribution. In a celebrated paper on the law of human mortality, Benjamin Gompertz [1825] attributed death to two causes: chance, or the deterioration of the power to withstand destruction. In deriving his law of mortality, however, he considered only deterioration and assumed that man's power to resist death decreases at a rate proportional to the power itself. Since the force of mortality $\mu(t)$ is a measure of man's susceptibility to death, Gompertz used the reciprocal $1/\mu(t)$ as a measure of man's resistance to death and thus arrived at the formula

$$\frac{d}{dt}\left(\frac{1}{\mu(t)}\right) = -h\,\frac{1}{\mu(t)}\,, \tag{2.21}$$

where h is a positive constant. Integrating (2.21) gives

$$\ln\left(\frac{1}{\mu(t)}\right) = -ht+k \tag{2.22}$$

which when rearranged becomes the Gompertz law of mortality

$$\mu(t) = Bc^{t}. \tag{2.23}$$

The corresponding density function and distributions are given, respectively, by

$$f(x) = Bc^{x}e^{-B[c^{x}-1]/\ln c} \tag{2.24}$$

and

$$F_X(x) = 1 - \exp\{-\frac{B}{\ln c}(c^{x}-1)\}. \tag{2.25}$$

(ii) Makeham's distribution. In 1860 W. M. Makeham suggested the modification

$$\mu(t) = A + B\,c^{t} \tag{2.26}$$

which is a restoration of the missing component, "chance " to the Gompertz formula. In this case, we have

$$f(x) = [A+Bc^{x}]\exp\{-[Ax+B(c^{x}-1)/\ln c]\} \tag{2.27}$$

and

$$F_X(x) = 1 - \exp\{-[Ax+B(c^{x}-1)/\ln c]\}. \tag{2.28}$$

(iii) Weibull distribution. When the force of mortality (or failure rate) is assumed to be a power function of t, $\mu(t) = \mu a t^{a-1}$, we have

$$f(x) = \mu a x^{a-1}e^{-\mu x^{a}} \tag{2.29}$$

and

$$F_X(x) = 1-e^{-\mu x^{a}} \tag{2.30}$$

This distribution, recommended by W. Weibull [1939] for studies of the life span of materials, is used extensively in reliability theory.

(iv) <u>Exponential distribution.</u> If $\mu(t) = \mu$ is a constant, then

$$f(x) = \mu e^{-\mu x} \tag{2.31}$$

and

$$F_X(x) = 1-e^{-\mu x} \tag{2.32}$$

a formula that plays a central role in the problem of life testing (Epstein and Sobel [1953]).

3. Joint Probability Distribution of the Number of Survivors

Let us consider, for a given u, the joint probability distribution of $\ell_1, \ell_2, \ldots, \ell_u$ given ℓ_0,

$$\Pr\{\ell_1 = k_1, \ldots, \ell_u = k_u | \ell_0\}. \tag{3.1}$$

It follows from the Markovian property in (2.19) that

$$\Pr\{\ell_1=k_1, \ell_2=k_2, \ldots, \ell_u=k_u | \ell_0\}=\Pr\{\ell_1=k_1 | \ell_0\}\Pr\{\ell_2=k_2 | k_1\}\ldots\Pr\{\ell_u=k_u | k_{u-1}\}. \tag{3.2}$$

Substituting (2.15) in (3.2) yields a chain of binomial distributions:

$$\Pr\{\ell_1=k_1, \ell_2=k_2, \ldots, \ell_u=k_u | \ell_0\} = \prod_{i=0}^{u-1} \frac{k_i!}{k_{i+1}!(k_i-k_{i+1})!} p_i^{k_{i+1}}(1-p_i)^{k_i-k_{i+1}}$$

$$k_{i+1} = 0,1,\ldots,k_i, \quad \text{with } k_0=\ell_0. \tag{3.3}$$

Formula (3.3) shows that when a cohort of people is observed at regular points in time, the number of survivors, ℓ_{i+1}, to the end of the interval (x_i, x_{i+1}) has a binomial distribution depending solely on the number of individuals alive at the beginning of the interval $\ell_i=k_i$.

The covariance between ℓ_i and ℓ_j may be obtained directly from (3.3); a somewhat simpler approach is the following. By definition

$$\sigma_{\ell_i,\ell_j}=E(\ell_i\ell_j)-E(\ell_i)E(\ell_j)=E(\ell_i\ell_j)-(\ell_0 p_{0i})(\ell_0 p_{oj}) \tag{3.4}$$

where

$$E(\ell_i\ell_j)=E[\ell_i E(\ell_j | \ell_i)]=E[\ell_i^2 p_{ij}] = E[\ell_1^2]p_{ij} . \tag{3.5}$$

Since ℓ_i is a binomial random variable,

$$E[\ell_i^2] = \ell_0 p_{0i}(1-p_{0i}) + [\ell_0 p_{0i}]^2. \tag{3.6}$$

Substituting (3.5) and (3.6) successively in (3.4) and using the relationship $p_{0i}p_{ij}=p_{0j}$, we have the formula for the covariance

$$\sigma_{\ell_i,\ell_j} = \ell_0 p_{0j}(1-p_{0i}), \qquad i\le j;\ i,j=0,1,\ldots,u \ . \tag{3.7}$$

When $j=i$, (3.7) reduces to the variance of ℓ_i (equation (2.12)). The correlation coefficient ρ_{ℓ_i,ℓ_j} between ℓ_i and ℓ_j, therefore, is given by

$$\rho_{\ell_i,\ell_j} = \frac{p_{0j}(1-p_{0i})}{\sqrt{p_{0i}(1-p_{0i})p_{0j}(1-p_{0j})}} = \sqrt{\frac{p_{0j}(1-p_{0i})}{p_{0i}(1-p_{0j})}} \ , \tag{3.8}$$

which is always positive whatever may be $0<i<j$. For a given i, the correlation coefficient decreases as x_j increases. This means that the larger the number of individuals alive at x_i, the more survivors there are likely to be at x_j; but the effect of the former on the latter decreases when x_j becomes farther away from x_i. These results show that <u>for a given u, $\ell_1,\ldots,\ell_u$ in the life table form a chain of binomial distributions; the joint probability distribution, the expected values, covariances and correlation coefficients are</u> given in (3.3), (2.11), (3.7), <u>and</u> (3.8), <u>respectively</u>.

4. Joint Probability Distribution of the Numbers of Deaths

In a life table covering the entire life span of each individual in a given population, the sum of the deaths at all ages is equal to the size of the original cohort. Symbolically,

$$d_0 + d_1 + \dots + d_w = \ell_0, \tag{4.1}$$

where d_w is the number of deaths in the age interval (x_w and over). Each individual in the original cohort has a probability $p_{0i}q_i$ of dying in the interval (x_i, x_{i+1}), $i=0,1,\dots,w$. Since an individual dies once and only once in the span covered by the life table,

$$p_{00}q_0 + \dots + p_{0w}q_w = 1, \tag{4.2}$$

where $p_{00} = 1$ and $q_w = 1$. Thus we have the well-known results: The numbers of deaths, $d_0,\dots,d_w$, in a life table have a multinomial distribution with the joint probability distribution

$$\Pr\{d_0 = \delta_0,\cdots,d_w = \delta_w\} = \frac{\ell_0!}{\delta_0! \cdots \delta_w!}(p_{00}q_0)^{\delta_0} \cdots (p_{0w}q_w)^{\delta_w}; \tag{4.3}$$

the expectation, variance, and covariance are given, respectively, by

$$E(d_i|\ell_0) = \ell_0 p_{0i}q_i , \tag{4.4}$$

$$\sigma^2_{d_i} = \ell_0 p_{0i}q_i(1-p_{0i}q_i) , \tag{4.5}$$

and

$$\sigma_{d_i,d_j} = -\ell_0 p_{0i}q_i p_{0j}q_j \qquad \text{for } i\neq j;\ i,j=0,1,\cdots,w. \tag{4.6}$$

In the discussion above, age 0 was chosen only for simplicity. For any given age, say x_α, the probability that an individual alive at age x_α will die in the interval (x_i, x_{i+1}) subsequent to x_α is $p_{\alpha i} q_i$ and the sum

$$\sum_{i=\alpha}^{w} p_{\alpha i} q_i = 1 \, , \tag{4.7}$$

and thus the numbers of deaths in intervals beyond x_α also have a multinomial distribution.

5. Optimum Properties of $\hat{p}_j$ and $\hat{q}_j$

The quantity $\hat{q}_j$ (or $\hat{p}_j$) is an estimator of the probability that an individual alive at age x_j will die in (or survive) the interval (x_j, x_{j+1}), with

$$\hat{p}_j + \hat{q}_j = 1, \qquad j = 0, 1, \ldots . \tag{5.1}$$

Therefore, $\hat{p}_j$ and $\hat{q}_j$ have the same optimum properties. For convenience, we consider $\hat{p}_j$ in the following discussion.

5.1. Maximum likelihood estimator of p_j. The joint probability distribution (3.3), when expressed in terms of the random variables $\ell_1, \ldots, \ell_u$, may be rewritten as

$$L = \prod_{i=0}^{u-1} \frac{\ell_i!}{\ell_{i+1}!(\ell_i - \ell_{i+1})!} \, p_i^{\ell_{i+1}} (1-p_i)^{\ell_i - \ell_{i+1}} \tag{5.2}$$

which is known as the likelihood function of $\ell_1, \ldots, \ell_u$. When the right hand side of (5.2) is maximized with respect to p_j, we have the maximum likelihood estimators, say $\hat{p}_j$. In this case, the maximizing values $\hat{p}_j$ can be derived by differentiation. Letting

$$\log L = C + \sum_{i=0}^{u-1} \ell_{i+1} \log p_i + \sum_{i=0}^{u-1} (\ell_i - \ell_{i+1}) \log(1-p_i) \tag{5.3}$$

setting the derivative equal to zero,

$$\frac{\partial}{\partial p_j} \log L = \frac{\ell_{j+1}}{p_j} - \frac{\ell_j - \ell_{j+1}}{1-p_j} = 0, \tag{5.4}$$

and solving the equations (5.4), we have the maximum likelihood estimators

$$\hat{p}_j = \frac{\ell_{j+1}}{\ell_j} \qquad j = 0,1,\ldots,u-1 \quad . \tag{5.5}$$

It should be noted that if for some age x_w all the ℓ_w individuals alive at x_w die within the interval (x_w, x_{w+1}), then $\ell_i = 0$ for all $i > w$, so that there is no contribution to the likelihood function beyond the <u>w</u>th factor. Consequently, the maximum-likelihood estimator in (5.5) is defined only for $\ell_j > 0$. With this understanding, let us compute the first two moments.

We have shown in Section 2 that, given $\ell_j > 0$, the number ℓ_{j+1} has the binomial distribution, therefore

$$E[\hat{p}_j] = E\left(\frac{\ell_{j+1}}{\ell_j}\right) = E[\frac{1}{\ell_j} E(\ell_{j+1}|\ell_j)] = p_j, \tag{5.6}$$

and $\hat{p}_j$ and hence $\hat{q}_j$ are unbiased estimators of the corresponding probabilities. Direct computation gives also

$$E[\hat{p}_j^{\,2}] = E\left(\frac{1}{\ell_j}\right) p_j(1-p_j) + p_j^{\,2} \tag{5.7}$$

and the variance

$$\sigma^2_{\hat{p}_j} = E\left(\frac{1}{\ell_j}\right) p_j(1-p_j) = \sigma^2_{\hat{q}_j}. \tag{5.8}$$

When ℓ_0 is large, (5.8) may be approximated by

$$\sigma^2_{\hat{p}_j} = \frac{1}{E(\ell_j)} \; p_j(1-p_j). \tag{5.9}$$

Justification of (5.9) is left to the reader.

For the covariance between $\hat{p}_j$ and $\hat{p}_k$ for $j < k$, we require that ℓ_k and hence ℓ_j and ℓ_{j+1} be positive and compute the conditional expectation

$$E[\hat{p}_k|\hat{p}_j] = E\left[\left(\frac{\ell_{k+1}}{\ell_k}\right)\bigg|\hat{p}_j\right] = E\,\frac{1}{\ell_k}\,E(\ell_{k+1}|\ell_k)|\hat{p}_j = p_k = E(\hat{p}_k), \qquad (5.10)$$

from which it follows that

$$E[\hat{p}_j\hat{p}_k] = E[\hat{p}_j E(\hat{p}_k|\hat{p}_j)] = E[\hat{p}_j]\ E[\hat{p}_k]$$

and that

$$\sigma_{\hat{p}_j,\hat{p}_k} = 0. \qquad (5.11)$$

Observe that formula (5.11) of zero covariance holds only for proportions in two non-overlapping age intervals. If the two intervals considered both begin with age x_α but extend to ages x_j and x_k, respectively, the covariance between the proportions $\hat{p}_{\alpha j}$ and $\hat{p}_{\alpha k}$ is not equal to zero. Easy computation shows that

$$\sigma_{\hat{p}_{\alpha j},\hat{p}_{\alpha k}} = E\,\frac{1}{\ell_\alpha}\,p_{\alpha k}(1-p_{\alpha j}) \qquad \alpha < j \leq k. \qquad (5.12)$$

When $k = j$, (5.12) becomes the variance of $\hat{p}_{\alpha j}$.

<u>Although $\hat{p}_j$ and $\hat{p}_k$ have zero covariance, they are not independently distributed</u>. For example

$$\Pr\{\hat{p}_1 = .5|\hat{p}_0 = 1\} \neq \Pr\{\hat{p}_1 = .5|\hat{p}_0 = .8\}.$$

Thus we have shown that <u>the quantities $\hat{p}_j$ and $\hat{q}_j$ in the life table are the unbiased, maximum-likelihood estimators of the corresponding probabilities p_j and q_j</u>.

6. Distribution of $\hat{e}_\alpha$, the Observed Expectation of Life at Age x_α

The observed expectation of life summarizes the mortality experience of a population from a given age to the end of the life span. At age x_i the expectation expresses the average number of years remaining to each individual living at that age if all individuals are subjected to the estimated probabilities of death $\hat{q}_j$ for $j \geq i$. This is certainly the most useful column in the life table.

To avoid confusion in notation, let α denote a fixed number and x_α a particular age. We are interested in the distribution of $\hat{e}_\alpha$, the observed expectation of life at age x_α. Consider ℓ_α, the number of survivors to age x_α, and let Y_α denote the future lifetime beyond age x_α of a particular individual. Clearly, Y_α is a continuous random variable that can take on any non-negative real value. Let y_α be the value that the random variable Y_α assumes, then $x_\alpha+y_\alpha$ is the entire life span of the individual. Let $f(y_\alpha)$ be the probability density function of the random variable Y_α, and let dy_α be an infinitesimal time interval. Since Y_α can assume values between y_α and $y_\alpha+dy_\alpha$ if and only if the individual survives the age interval $(x_\alpha, x_\alpha+y_\alpha)$ and dies in the interval $(x_\alpha+y_\alpha, x_\alpha+y_\alpha+dy_\alpha)$, we have

$$f(y_\alpha)dy_\alpha = e^{-\int_{x_\alpha}^{x_\alpha+y_\alpha} \mu(\tau)d\tau} \mu(x_\alpha+y_\alpha)dy_\alpha \qquad y_\alpha \geq 0. \tag{6.1}$$

Function $f(y_\alpha)$ in (6.1) is a proper probability density function since it is never negative and since the integral of the function from $y_\alpha = 0$ to $y_\alpha = \infty$ is equal to unity. Clearly, $f(y_\alpha)$ can never be negative whatever the value of y_α. To evaluate the integral

$$\int_0^\infty f(y_\alpha)dy_\alpha = \int_0^\infty e^{-\int_{x_\alpha}^{x_\alpha+y_\alpha}\mu(\tau)d\tau}\mu(x_\alpha+y_\alpha)dy_\alpha \tag{6.2}$$

we define a quantity Φ

$$\Phi = \int_{x_\alpha}^{x_\alpha+y_\alpha}\mu(\tau)d\tau = \int_0^{y_\alpha}\mu(x_\alpha+t)dt \tag{6.3}$$

and substitute the differential

$$d\Phi = \mu(x_\alpha+y_\alpha)dy_\alpha \tag{6.4}$$

in the integral to give the solution

$$\int_0^\infty f(y_\alpha)dy_\alpha = \int_0^\infty e^{-\Phi}d\Phi = 1. \tag{6.5}$$

The mathematical expectation of the random variable Y_α is the expected length of life beyond age x_α, and thus is the true expectation of life at age x_α. In accordance with the definition given the symbol e_α, we may write

$$e_\alpha = \int_0^\infty y_\alpha f(y_\alpha)dy_\alpha = \int_0^\infty y_\alpha e^{-\int_{x_\alpha}^{x_\alpha+y_\alpha}\mu(\tau)d\tau}\mu(x_\alpha+y_\alpha)dy_\alpha. \tag{6.6}$$

Thus the expectation e_α and the variance

$$\sigma^2_{Y_\alpha} = \int_0^\infty (y_\alpha-e_\alpha)^2 f(y_\alpha)dy_\alpha \tag{6.7}$$

both depend on the intensity of risk of death $\mu(\tau)$.

The expectation of life at age x_α is conventionally defined as

$$e_\alpha = \int_0^\infty e^{-\int_{x_\alpha}^{x_\alpha+y_\alpha}\mu(\tau)d\tau} dy_\alpha \; . \tag{6.8}$$

It is instructive to prove that the two alternative definitions (6.6) and (6.8) are identical. Let $u = y_\alpha$, $du = dy_\alpha$,

$$v = -e^{-\int_{x_\alpha}^{x_\alpha+y_\alpha}\mu(\tau)d\tau} \tag{6.9}$$

and

$$dv = e^{-\int_{x_\alpha}^{x_\alpha+y_\alpha}\mu(\tau)d\tau} \mu(x_\alpha+y_\alpha)dy_\alpha \; . \tag{6.10}$$

Integrating (6.6) by parts gives

$$\int_0^\infty y_\alpha \, e^{-\int_{x_\alpha}^{x_\alpha+y_\alpha}\mu(\tau)d\tau} \mu(x_\alpha+y_\alpha)dy_\alpha$$

$$= -y_\alpha \, e^{-\int_{x_\alpha}^{x_\alpha+y_\alpha}\mu(\tau)d\tau} \Bigg|_0^\infty + \int_0^\infty e^{-\int_{x_\alpha}^{x_\alpha+y_\alpha}\mu(\tau)d\tau} dy_\alpha \; . \tag{6.11}$$

The first term on the right vanishes and the second term is the same as (6.8), proving the identity.

6.1. The variance of the expectation of life, $\hat{e}_\alpha$. The future lifetimes of ℓ_α survivors may be regarded as a sample of ℓ_α independent and identically distributed random variables, $Y_{\alpha k}$, $k=1, \ldots, \ell_\alpha$, each of which has the probability density function (6.1), the expectation (6.6), and the variance (6.7). According to the central limit theorem, for large ℓ_α the distribution of the sample mean

$$\bar{Y}_\alpha = \frac{1}{\ell_\alpha} \sum_{k=1}^{\ell_\alpha} Y_{\alpha k} \tag{6.12}$$

is approximately normal with an expectation as given in (6.6) and a variance $\sigma^2_{Y_\alpha} / \ell_\alpha$. It has been shown in Chapter 4, Section 3, that the sample mean $\bar{Y}_\alpha$ is equal to the observed expectation of life $\hat{e}_\alpha$, or

$$\bar{Y}_\alpha = \hat{e}_\alpha \,. \tag{6.13}$$

Therefore, the variance of $\hat{e}_\alpha$ is also $\sigma^2_{Y_\alpha} / \ell_\alpha$. For practical purposes, we need to have a formula for the variance of $\hat{e}_\alpha$ which can be estimated for the cohort and the current life tables. The formula of $\hat{e}_\alpha$ is given by

$$\hat{e}_\alpha = \frac{1}{\ell_\alpha} \left[\sum_{i=\alpha}^{w-1} \{ n_i(\ell_i - d_i) + a_i n_i d_i \} + a_w n_w d_w \right] \tag{6.14}$$

Using the relation $d_i = \ell_i - \ell_{i+1}$; $i=\alpha,\ldots,w-1$, we rewrite (6.14) as

$$\hat{e}_\alpha = a_\alpha n_\alpha + \sum_{i=\alpha+1}^{w} c_i \frac{\ell_i}{\ell_\alpha} = a_\alpha n_\alpha + \sum_{i=\alpha+1}^{w} c_i \hat{p}_{\alpha i} \tag{6.15}$$

where $c_i = (1-a_{i-1})n_{i-1} + a_i n_i$. Because the proportion $\hat{p}_{\alpha i}$ in (6.15) is an unbiased estimate of $p_{\alpha i}$, the expectation of $\hat{e}_\alpha$ as given by (6.6) is simply

$$e_\alpha = a_\alpha n_\alpha + \sum_{i=\alpha+1}^{w} c_i p_{\alpha i}, \qquad \alpha=0,1,\cdots,w. \tag{6.16}$$

The observed expectation of life as given in (6.15) is a linear function of $\hat{p}_{\alpha j}$, which in the current life table is computed from

$$\hat{p}_{\alpha j} = \hat{p}_\alpha \hat{p}_{\alpha+1} \cdots \hat{p}_{j-1}, \qquad j=\alpha+1,\cdots,w. \tag{6.17}$$

Clearly, the derivatives taken at the true point $(p_\alpha, p_{\alpha+1}, \cdots, p_{j-1})$ are

$$\frac{\partial}{\partial \hat{p}_i} \hat{p}_{\alpha j} = p_{\alpha i}\, p_{i+1,j}\,, \qquad \text{for} \quad \alpha \le i < j;$$

$$= 0\,, \qquad \text{for} \quad i \ge j. \tag{6.18}$$

Hence

$$\frac{\partial}{\partial \hat{p}_i} \hat{e}_\alpha = \sum_{j=i+1}^{w} c_j\, p_{\alpha i}\, p_{i+1,j}$$

$$= p_{\alpha i}\left[c_{i+1} + \sum_{j=i+2}^{w} c_j\, p_{i+1,j}\right]$$

$$= p_{\alpha i}\left[e_{i+1} + (1-a_i)n_i\right] \quad . \tag{6.19}$$

Because of (6.18), the derivative (6.19) vanishes when $i = w$. Since it has been shown in Section 5 [cf. equation (5.11)] that the covariance between proportions of survivors of two non-overlapping age intervals is zero, the variance of the observed expectation of life may be computed from the following

$$\sigma^2_{\hat{e}_\alpha} = \sum_{i=\alpha}^{w-1} \left\{\frac{\partial}{\partial \hat{p}_i} \hat{e}_\alpha\right\}^2 \sigma^2_{\hat{p}_i} \quad . \tag{6.20}$$

Substitution of (6.19) in (6.20) gives the formula

$$\sigma^2_{\hat{e}_\alpha} = \sum_{i=\alpha}^{w-1} p_{\alpha i}^2 \{e_{i+1} + (1-a_i)n_i\}^2\, \sigma^2_{\hat{q}_i}\,, \qquad \alpha = 0, 1, \cdots, w-1. \tag{6.21}$$

Thus we have

<u>Theorem</u>: <u>If the distribution of deaths in the age interval</u> (x_i, x_{i+1}) <u>is such that, on the average, each of the</u> d_i <u>individuals lives</u> $a_i n_i$ <u>years in the interval, for</u> $i = \alpha, \alpha+1, \ldots, w$, <u>then as</u> ℓ_α <u>approaches infinity, the probability distribution of the observed expectation of life at age</u> x_α, <u>as given by (6.15), is asymptotically normal and has the mean and variance as given by (6.16) and (6.21), respectively</u>.

APPENDIX III

THE THEORY OF COMPETING RISKS

A Historical Note - Daniel Bernoulli's Work

The concept of competing risks is not new; it seems to have originated in a controversy over the value of vaccination. The first systematic discussion of the problem was by Daniel Bernoulli in 1760 in his article entitled, "Essai d'une nouvelle analyse de la mortalite causee par la petite verole et des avantages de l'inoculation pour la prevenir." The main objective of the memoir was to determine the mortality caused by smallpox at various ages. Since his work created much discussion in his time and opened up a new area of study in competing risks, it may be appropriate to review briefly Daniel Bernoulli's approach.

Let ℓ_x denote the number who survive to age x; among them s_x have not had smallpox. Assume that in a year smallpox attacks one out of n individuals who have not had the disease, and one out of every m individuals who contract the disease dies. Both n and m are assumed to be constant. Within the time element dx, the number who die is $-d\ell_x$, and the number who die from smallpox is

$$\frac{s_x\,dx}{mn} \tag{1}$$

and therefore the number who die from other causes is

$$-d\ell_x - \frac{s_x\,dx}{mn} \quad . \tag{2}$$

Now the number of those who have not had smallpox will decrease during the time element dx through contracting smallpox (s_x dx/n) and through death (a proportion s_x/ℓ_x of that in (2)). Denoting this reduction by $-ds_x$, we have the equation,

$$-ds_x = \frac{s_x\,dx}{n} - \frac{s_x}{\ell_x}\left(d\ell_x + \frac{s_x\,dx}{mn}\right) \quad . \tag{3}$$

Equation (3) may be rearranged to yield

$$\frac{s_x\,d\ell_x - \ell_x\,ds_x}{s_x^2} = \left(m\,\frac{\ell_x}{s_x} - 1\right)\frac{dx}{mn} \tag{4}$$

or

$$n\ d\ \ln\left(m\,\frac{\ell_x}{s_x} - 1\right) = dx \quad . \tag{5}$$

Integrating both sides of (5) gives

$$\left(m\,\frac{\ell_x}{s_x} - 1\right)^n = e^{x+c}$$

or

$$s_x = \frac{m\ell_x}{1+e^{(x+c)/n}} \quad . \tag{6}$$

To determine the constant of integration c, we observe that at x=0, $s_0 = \ell_0$ so that $e^{c/n} = (m-1)$, and thus

$$s_x = \frac{m\ell_x}{1+(m-1)e^{x/n}} \quad . \tag{7}$$

Using formula (7) and assuming m = n = 8, Bernoulli calculated ℓ_x and s_x on the basis of Halley's table of Breslow.

Bernoulli also derived a formula for the number of survivors had there been no smallpox. Using a similar approach, he showed that this number of survivors, denoted by z_x, is given by

$$z_x = \frac{m\ell_x \, e^{x/n}}{1+(m-1)e^{x/n}} \tag{8}$$

The right-hand side of (8) increases as either m or n decreases and approaches the limit $m\,\ell_x/(m-1)$ as x increases indefinitely.

After discussing the subject of the mortality from smallpox, Daniel Bernoulli proceeded to the discussion of inoculation. He admitted that there was some danger in inoculation against smallpox, but he found that on the whole it was advantageous. Based on his calculations, he concludes that inoculation would increase the average length of life by three years.

An important assumption in Daniel Bernoulli's solution of the problem was the constant incidence rate (1/n) and constant case fatality rate (1/m). D'Alembert (1717-1783), Trembly (1749-1811), and Laplace (1749-1827) all had considered the case when n and m both are functions of age x. It was D'Alembert who was the most critical of Bernoulli's solution. Although he too recognized the value of inoculation, he felt that Bernoulli had over-estimated its advantage. While he failed to provide a neat solution to the problem, D'Alembert brought up a significant distinction between the physical life and the real life of an individual. By the physical life, he meant life in the ordinary sense; by the real life he meant the portion of existence during which the individual enjoys life in a disease-free state. Theoretical pursuit of this aspect of the problem, however, was not in evidence either in D'Alembert's work or in that of others. A detailed account of the controversy may be found in Todhunter [1949]. Thus, Bernoulli, D'Alembert, Trembley and Laplace each derived a method of determining the change in population composition that would take place if smallpox were eliminated as a cause of death. It was Makeham [1874], however, who first formulated a theory of decremental forces and explored its practical applications.

Actuarial mathematicians have applied Makeham's work to develop multiple-decrement theory in the study of life contingencies. For a detailed account of the theory, reference should be made to C. W. Jordan [1952]. In the last thiry years, the theory of competing risks have attracted much attention in the field of health and statistics. Greville [1948] discussed deterministically multiple decrement tables, Fix and Neyman [1951] studied the problem of competing risks for cancer patients; and Chiang [1961a] approached the problem from a stochastic viewpoint. Other interesting studies include those by Berkson and Elveback [1960], Berman [1963], Cornfield [1957], and Kimball [1958], [1969], David [1970], Pike [1970], Mantel and Bailar [1970] and Chiang [1970].

1. Introduction

Although the basic characteristics of mortality studies are that death is not a repetitive event and that usually death is attributed to a single cause, in cause-specific mortality studies the various risks competing for the life of an individual must be considered as well. For example, in an investigation of congenital malformation as a cause of infant death, some subjects would die from other causes such as tuberculosis. These infants have no chance either of dying from congenital malformation or of surviving the first year of life. What then would be the contribution of their survival experience to such a mortality study and what adjustment would have to be made for the competing effect of tuberculosis in the assessment of congenital malformation as a cause of death? Competing risks must also be taken into account in studies of the relative susceptibility of individuals in different illness states to other diseases. For instance, would people suffering from arteriosclerotic heart disease be more likely to die from pneumonia than those without a heart condition? Any meaningful comparison between the two groups with respect to their susceptibility to pneumonia would have to evaluate the effect of arteriosclerosis as a competing risk. The following three types of probability of death from a specific cause are necessary for an understanding of the study of survival as well as the application of life table methodology to such problems as those above.

1. Crude probability: The probability of death from a specific cause in the presence of all other risks acting in a population, or

$Q_{i\delta}$ = Pr{an individual alive at time x_i will die in the interval (x_i, x_{i+1}) from cause R_δ in the presence of all other risks in the population}.

2. Net probability: The probability of death if a specific risk is the only risk in effect in the population or, conversely, the probability of death if a specific risk is eliminated from the population.

$q_{i\delta}$ = Pr{an individual alive at x_i will die in the interval (x_i, x_{i+1}) if R_δ is the only risk acting on the population};

$q_{i\cdot\delta}$ = Pr{an individual alive at x_i will die in the interval (x_i, x_{i+1}) if R_δ is eliminated as a risk of death}.

3. Partial crude probability: The probability of death from a specific cause when another risk (or risks) is eliminated from the population.

$Q_{i\delta\cdot 1}$ = Pr{an individual alive at x_i will die in the interval (x_i, x_{i+1}) from R_δ if R_1 is eliminated from the population};

$Q_{i\delta\cdot 12}$ = Pr{an individual alive at x_i will die in the interval (x_i, x_{i+1}) from R_δ if R_1 and R_2 are eliminated from the population}.

When the cause of death is not specified, we have the probabilities

p_i = Pr{an individual alive at x_i will survive the interval (x_i, x_{i+1})}

and

q_i = Pr{an individual alive at x_i will die in the interval (x_i, x_{i+1})},

with $p_i+q_i = 1$.

The use of the terms "risk" and "cause" needs clarification. Both terms may refer to the same condition but are distinguished here by their position in time relative to the occurrence of death. Prior to death the condition referred to is a risk; after death the same condition is the cause. For example, tuberculosis is a risk of dying to which an individual is exposed, but tuberculosis is also the cause of death if it is the disease from which the individual eventually dies.

In the human population the net and partial crude probabilities cannot be estimated directly, but only through their relations with the crude probability. The study of these relations is part of the problem of "competing risks," or "multiple decrement." Formulas expressing relations between net and crude probabilities have been developed by assuming either a constant intensity of risk of death (force of mortality) or a uniform distribution of deaths. We will review these formulas in this Appendix assuming a constant relative intensity. Partial crude probabilities have not received due attention in view of their often indispensable role in studies of cause-specific mortality. Their relations with the corresponding crude probabilities will also be discussed.

2. Relations between Crude, Net and Partial Crude Probabilities

Suppose that r risks of death are acting simultaneously on each individual in a population, and let these risks be denoted by $R_1,\cdots,R_r$. For each risk, R_δ, there is a corresponding intensity function (or force of mortality) $\mu(t;\delta)$ such that,

$$\begin{aligned}\mu(t;\delta)\Delta + o(\Delta) = \Pr\{&\text{an individual alive at time } t \text{ will die in interval}\\ &(t,\ t+\Delta) \text{ from risk } R_\delta\},\quad \delta=1,\cdots,r,\end{aligned} \tag{2.1}$$

and the sum

$$\mu(t;1) + \cdots + \mu(t;r) = \mu(t) \tag{2.2}$$

is the total intensity (or the total force of mortality). Although for each risk R_δ the intensity $\mu(t;\delta)$ is a function of time t, we assume that within the time interval $(x_i,\ x_{i+1})$ the ratio

$$\frac{\mu(t;\delta)}{\mu(t)} = c_{i\delta} \tag{2.3}$$

is independent of time t, but is a function of the interval $(x_i,\ x_{i+1})$ and risk R_δ. Assumption (2.3), which is known as the proportionality assumption, permits the risk-specific intensity $\mu(t;\delta)$ to vary in absolute magnitude, but requires that it remain a constant proportion of the total intensity in an interval.

Consider death without specification of cause. The probability that an individual alive at x_i will survive the interval $(x_i,\ x_{i+1})$ is

$$p_i = e^{-\int_{x_i}^{x_{i+1}} \mu(t)\,dt}\ , \qquad i=0,1,\cdots \tag{2.4}$$

and the probability of his dying in the interval is $q_i = 1-p_i$ (see formula (2.9) in Appendix II).

To derive the crude probability of dying from risk R_δ, we consider a point t within the interval (x_i, x_{i+1}). The probability that an individual alive at x_i will die from R_δ in interval (t, t+dt) is

$$e^{-\int_{x_i}^{t} \mu(\tau)d\tau} \mu(t;\delta)dt \tag{2.5}$$

where the exponential function is the probability of surviving from x_i to t when all risks are acting, and the factor $\mu(t;\delta)dt$ is the instantaneous death probability from risk R_δ in time interval (t, t+dt). Summing (2.5) over all possible values of t, for $x_i \le t \le x_{i+1}$, gives the crude probability

$$Q_{i\delta} = \int_{x_i}^{x_{i+1}} e^{-\int_{x_i}^{t} \mu(\tau)d\tau} \mu(t;\delta)dt \ . \tag{2.6}$$

Under the proportionality assumption (2.3), (2.6) may be rewritten as

$$Q_{i\delta} = \frac{\mu(t;\delta)}{\mu(t)} \int_{x_i}^{x_{i+1}} e^{-\int_{x_i}^{t} \mu(\tau)d\tau} \mu(t)\ dt. \tag{2.7}$$

Integrating gives

$$Q_{i\delta} = \frac{\mu(t;\delta)}{\mu(t)} \left[1 - e^{-\int_{x_i}^{x_{i+1}} \mu(t)dt} \right] = \frac{\mu(t;\delta)}{\mu(t)}\ q_i \ ; \tag{2.8}$$

hence

$$\frac{\mu(t;\delta)}{\mu(t)} = \frac{Q_{i\delta}}{q_i} \ , \qquad x_i \le t < x_{i+1}; \quad \delta = 1, \cdots, r \ . \tag{2.9}$$

Equation (2.9) is obvious, for if the ratio of the risk-specific intensity to the total intensity is constant throughout an interval, this constant should also be equal to the ratio of the corresponding probabilities of dying over the entire interval. Equations (2.2) and (2.9) imply a trivial equality

$$Q_{i1} + \cdots + Q_{ir} = q_i, \qquad i=0,1,\cdots. \tag{2.10}$$

Remark 1. Equation (2.9) suggests a similarity between the intensity functions $\mu(t;\delta)$ and the probability $Q_{i\delta}$. For example, from (2.9) we have

$$\frac{\mu(t;\delta)}{\mu(t;\varepsilon)} = \frac{Q_{i\delta}}{Q_{i\varepsilon}} \quad , \tag{2.11}$$

so that the relative magnitude between any two probabilities is equal to the relative magnitude between the corresponding intensity functions. However, when several sets of values are considered, the variation among $Q_{i\delta}$ may be quite different from the variation among $\mu(t;\delta)$. To illustrate, let $\mu(t;\delta) = \mu(i;\delta)$ for $x_i \leq t \leq x_{i+1}$ and $\delta = 1,\cdots,r$; so that $\mu(t) = \mu(i)$. Then (2.9) implies that

$$\mu(i;\delta) = -\frac{Q_{i\delta}}{q_i} \ln (1-q_i) \ . \tag{2.12}$$

Suppose we let Q_{i1} increase but keep $Q_{i2},\cdots,Q_{ir}$ unchanged. The intensity functions $\mu(i;2),\cdots,\mu(i;r)$ will not remain constant, but rather they increase with the increasing values of Q_{i1} (or with increasing values of q_i, since $q_i = Q_{i1}+\cdots+Q_{ir}$). In other words, the function

$$h(q_i) = -\frac{1}{q_i} \ln (1-q_i) \tag{2.13}$$

on the right-hand side of (2.12) is a monotonically increasing function of q_i. Taking the derivative of $h(q_i)$ with respect to q_i yields

$$h'(q_i) = \frac{1}{q_i^2}\left[\ln(1-q_i) + \frac{q_i}{1-q_i}\right] = \sum_{n=2}^{\infty} \frac{n-1}{n} q_i^{n-2} \tag{2.13}$$

since q_i is between 0 and 1. The last expression in (2.13) is always positive for positive values of q_i. Hence the function $h(q_i)$ increases with q_i and $\mu(i;\delta)$ increases with Q_{i1}, as required to be shown.

A numerical example for r = 3 risks is given below. It is also easy to see that

$$\frac{Q_{i1}}{\mu(i,1)} = \frac{Q_{i2}}{\mu(i,2)} = \frac{Q_{i3}}{\mu(i,3)} \tag{2.14}$$

Table 1. Probabilities and Intensity Functions

Q_{i1}	Q_{i2}	Q_{i3}	q_i	$\mu(i;1)$	$\mu(i;2)$	$\mu(i;3)$
.01	.01	.30	.32	.0121	.0121	.3615
.05	.01	.30	.36	.0620	.0124	.3719
.10	.01	.30	.41	.1287	.0129	.3860
.25	.01	.30	.56	.3665	.0147	.4398
.50	.01	.30	.81	1.0251	.0205	.6151

2.1. Relations between crude and net probabilities.

The net probability of death in the interval (x_i, x_{i+1}) when R_δ is the only operating risk is obviously

$$q_{i\delta} = 1 - e^{-\int_{x_i}^{x_{i+1}} \mu(t;\delta)dt} \tag{2.15}$$

which, in view of (2.3), can be written as

$$q_{i\delta} = 1 - e^{-\frac{\mu(t;\delta)}{\mu(t)} \int_{x_i}^{x_{i+1}} \mu(t)dt} = 1 - p_i^{\mu(t;\delta)/\mu(t)} \quad . \tag{2.16}$$

With equation (2.9), (2.16) gives the relation between the net and the crude probabilities,

$$q_{i\delta} = 1 - p_i^{Q_{i\delta}/q_i}, \qquad \delta = 1,\cdots,r. \tag{2.17}$$

Formula (2.17) may be simplified. Using the $p_i = 1-q_i$, the second term on the right-hand side is expanded in terms of Newton's binomial series,

$$p_i^{Q_{i\delta}/q_i} = (1-q_i)^{Q_{i\delta}/q_i}$$

$$= 1 + \frac{Q_{i\delta}}{q_i}(-q_i) + \ldots + \binom{Q_{i\delta}/q_i}{k}(-q_i)^k + \ldots \tag{2.18}$$

where the binomial coefficient is defined as follows:

$$\binom{Q_{i\delta}/q_i}{k} = \frac{1}{k!} \frac{Q_{i\delta}}{q_i} \left(\frac{Q_{i\delta}}{q_i} - 1\right) \cdots \left(\frac{Q_{i\delta}}{q_i} - k + 1\right) \tag{2.19}$$

for $k=0,1,2,\ldots$.

Because of small values of q_i, the first four terms of the infinite series in (2.18) give a good approximation. As a result, we have

$$p_i^{Q_{i\delta}/q_i} \doteq 1 - Q_{i\delta} - \frac{1}{2} Q_{i\delta}(q_i - Q_{i\delta}) - \frac{1}{6} Q_{i\delta}(q_i - Q_{i\delta})(2q_i - Q_{i\delta}) \ . \tag{2.20}$$

Substituting (2.20) in (2.17) yields the relationship

$$q_{i\delta} = Q_{i\delta}\left[1 + \frac{1}{2}(q_i - Q_{i\delta}) + \frac{1}{6}(q_i - Q_{i\delta})(2q_i - Q_{i\delta})\right] \tag{2.21}$$

When the first three terms of the binomial series are taken, we have

$$q_{i\delta} = Q_{i\delta}\left[1 + \frac{1}{2}(q_i - Q_{i\delta})\right], \tag{2.21 a}$$

which may be used when q_i is extremely small.

The net probability of death when risk R_δ is eliminated can be derived in a similar way. When R_δ is eliminated as a cause of death, the force of mortality is $\mu(t) - \mu(t;\delta)$. In this case, the probability that an individual alive at x_i will die in $(t, t+dt)$ is

$$e^{-\int_{x_i}^{t}[\mu(\tau)-\mu(\tau;\delta)]d\tau}[\mu(t) - \mu(t;\delta)]dt\,, \quad \text{for } x_i<t\le x_{i+1}\,, \tag{2.22}$$

where the exponential function is the probability of his surviving from x_i to t and $[\mu(t) - \mu(t;\delta)]dt$ is the probability that he will die in the time element $(t, t+dt)$. For different values of t, the corresponding events associated with the probability in (2.22) are mutually exclusive. Using the addition theorem we have the net probability that the individual will die in the interval (x_i, x_{i+1})

$$q_{i\cdot\delta} = \int_{x_i}^{x_{i+1}} e^{-\int_{x_i}^{t}[\mu(\tau)-\mu(\tau;\delta)]d\tau}[\mu(t) - \mu(t;\delta)]dt\,. \tag{2.23}$$

Since (2.9) imples that the ratio

$$\frac{\mu(t) - \mu(t;\delta)}{\mu(t)} = \frac{q_i - Q_{i\delta}}{q_i}$$

is independent of time t, we write

$$\begin{aligned}\mu(t) - \mu(t;\delta) &= \frac{\mu(t) - \mu(t;\delta)}{\mu(t)}\,\mu(t)\\ &= \frac{q_i - Q_{i\delta}}{q_i}\,\mu(t)\end{aligned} \tag{2.24}$$

and

$$\mu(\tau) - \mu(\tau;\delta) = \frac{q_i - Q_{i\delta}}{q_i}\,\mu(\tau)\quad. \tag{2.25}$$

Substituting (2.24) and (2.25) in (2.23) gives

$$q_{i.\delta} = \int_{x_i}^{x_{i+1}} \left[e^{-\int_{x_i}^{x_{i+1}} \mu(\tau)d\tau} \right]^{(q_i-Q_{i\delta})/q_i} \left(\frac{q_i-Q_{i\delta}}{q_i}\right)\mu(t)dt, \tag{2.26}$$

and integrating the right-hand side of (2.26) yields the relationship

$$q_{i\cdot\delta} = 1 - p_i^{(q_i-Q_{i\delta})/q_i} \quad . \tag{2.27}$$

Formula (2.27) also can be simplified using Newton's binomial expression as was formula (2.17). Again taking the first four terms of the series, we have

$$p_i^{(q_i-Q_{i\delta})/q_i} = (1-q_i)^{(q_i-Q_{i\delta})/q_i}$$

$$= 1 - (q_i-Q_{i\delta}) - \frac{1}{2}(q_i-Q_{i\delta})Q_{i\delta} - \frac{1}{6}(q_i-Q_{i\delta})Q_{i\delta}(q_i+Q_{i\delta}) \tag{2.28}$$

Substituting (2.28) in (2.27) and simplifying the resulting expression gives the desired formula[1/]

$$q_{i\cdot\delta} = (q_i-Q_{i\delta}) \left[1 + \frac{1}{2} Q_{i\delta} + \frac{1}{6} Q_{i\delta}(q_i+Q_{i\delta})\right] \ . \tag{2.29}$$

Because of the absence of competing risks, the net probability is always greater than the corresponding crude probability, or

$$q_{i\delta} > Q_{i\delta}. \tag{2.30a}$$

Further, if two risks R_δ and R_ε are such that

$$Q_{i\delta} > Q_{i\varepsilon},$$

then

$$q_{i\delta} > q_{i\varepsilon} \qquad \text{and} \qquad q_{i\cdot\delta} < q_{i\cdot\varepsilon} \ . \tag{2.30b}$$

Verification of (2.30a) and (2.30b) is left to the reader.

2.2. Relation between crude and partial crude probabilities.

Suppose now that risk R_1 is eliminated from the population. In the presence of all other risks, let $Q_{i\delta\cdot 1}$ be the partial crude probability that an individual alive at time x_i will die in the interval (x_i, x_{i+1}) from cause R_δ for $\delta=2,\cdots,r$. We shall express $Q_{i\delta\cdot 1}$ in terms of the probabilities p_i and q_i and of the crude probabilities Q_{i1} and $Q_{i\delta}$. Using the multiplication and addition theorems as in (2.22) we have

$$Q_{i\delta\cdot 1} = \int_{x_i}^{x_{i+1}} e^{-\int_{x_i}^{t} [\mu(\tau)-\mu(\tau;1)]d\tau} \mu(t;\delta)dt \quad . \tag{2.31}$$

To simplify (2.31), we note from (2.9) that the ratio $\mu(t;\delta)/[\mu(t)-\mu(t;1)]$ is equal to $Q_{i\delta}/(q_i-Q_{i1})$ and is independent of time t. The partial crude probability may then be rewritten as

$$Q_{i\delta\cdot 1} = \frac{\mu(t;\delta)}{\mu(t)-\mu(t;1)} \int_{x_i}^{x_{i+1}} e^{-\int_{x_i}^{t} [\mu(\tau)-\mu(\tau;1)]d\tau} [\mu(t)-\mu(t;1)]dt$$

$$= \frac{Q_{i\delta}}{q_i-Q_{i1}} \left[1 - e^{-\int_{x_i}^{x_{i+1}} [\mu(t)-\mu(t;1)]dt} \right] \tag{2.32}$$

$$= \frac{Q_{i\delta}}{q_i-Q_{i1}} \, q_{i\cdot 1} \quad .$$

Substituting (2.29) for $\delta=1$ in (2.32) gives the final formula[2/]

$$Q_{i\delta\cdot 1} = Q_{i\delta}[1 + \frac{1}{2} Q_{i1} + \frac{1}{6} Q_{i1}(q_i+Q_{i1})] \; , \qquad \delta =2,\ldots,r. \tag{2.33}$$

The sum of $Q_{i\delta\cdot 1}$ for $\delta=2,\cdots,r$ is equal to the net probability of death when risk R_1 is eliminated from the population. That is,

$$\sum_{\delta=2}^{r} Q_{i\delta\cdot 1} = \sum_{\delta=2}^{r} Q_{i\delta}[1 + \frac{1}{2} Q_{i1} + \frac{1}{6} Q_{i1}(q_i+Q_{i1})] = (q_i-Q_{i1})[1+ \frac{1}{2}Q_{i1} + \frac{1}{6}Q_{i1}(q_i+Q_{i1})]$$

$$= q_{i.1} \quad , \tag{2.34}$$

as might have been anticipated.

Formula (2.33) can be easily generalized to other cases where more than one risk is eliminated. If risks R_1 and R_2 are eliminated, the partial crude probability that an individual alive at time x_i will die from cause R_δ in the interval (x_i, x_{i+1}) is[3/]

$$Q_{i\delta\cdot 12} = Q_{i\delta}[1 + \frac{1}{2} (Q_{i1} + Q_{i2}) + \frac{1}{6}(Q_{i1}+ Q_{i2})(q_i+Q_{i1}+Q_{i2})] \quad . \tag{2.35}$$

In the discussion of these three types of probability, both q_i and p_i are assumed to be greater than zero but less than unity. If q_i were zero (p_i=1), then $Q_{i\delta}$ would also be zero for $\delta=1,\cdots,r$. Then the ratios $Q_{i\delta}/q_i$, $Q_{i\delta}/(q_i-Q_{i1})$, and $(q_i-Q_{i1})/q_i$ and formulas (2.17), (2.26), (2.33) and (2.35) would become meaningless. In other words, if an individual were certain to survive an interval, it would be meaningless to speak of his chance of dying from a specific risk. On the other hand, if p_i were zero (q_i=1), formula (2.4) shows that the integral

$$\int_{x_i}^{x_{i+1}} \mu(t)\, dt$$

would approach infinity; this fortunately is extremely unrealistic.

3. Competing Risks with Interaction

The theory of competing risks presented in Section 2 was based on the independence assumption in (2.2). Under this assumption, the risks act independently of one another and the presence or elimination of one risk has no effect on the intensity functions (forces of mortality) of other risks. The validity of the assumption depends upon the diseases under consideration. One can certainly visualize a situation where the independence assumption does not hold. The presence of tuberculosis (R_1), for example, may affect the chance of dying from pneumonia, R_2. Once tuberculosis is eliminated as a risk of death, how does one evaluate the probability of dying from pneumonia? The problem can be resolved by creating another risk, R_{12}, the interaction between tuberculosis and pneumonia, with the intensity function $\mu(t;1,2)$. When tuberculosis is eliminated as a risk of death, the interaction vanishes also. The purpose of this section is to study the theory of competing risks with the consideration of interactions.

For any two risks R_δ and R_ε, we denote by $R_{\delta\varepsilon}$ their interaction and by $\mu(t;\delta,\varepsilon)$ the corresponding intensity function, with

$$\mu(t;\delta,\varepsilon) \geqq 0 \qquad (3.2)$$

When two risks have positive interaction, the intensity function $\mu(t;\delta,\varepsilon)$ is positive and has the following probabilistic meaning:

$$\begin{aligned}\mu(t;\delta,\varepsilon)\Delta + o(\Delta) = \Pr\{&\text{an individual alive at time } t \text{ will die} \\ &\text{in interval } (t,t+\Delta) \text{ from } R_{\delta\varepsilon}\}\end{aligned} \qquad (3.3)$$

If two risks have no interaction, $\mu(t;\delta,\varepsilon) = 0$. It is conceivable that, for two particular risks, presence of one decreases the probability of dying from the other, so that $\mu(t;\delta,\varepsilon) < 0$. In such cases proper interpretation is the following

$[\mu(t;\delta) + \mu(t;\varepsilon) + \mu(t;\delta,\varepsilon)]\Delta + o(\Delta) = \Pr\{$an individual alive at t will die in interval $(t,t+\Delta)$ from either R_δ or $R_\varepsilon\}$. (3.4)

For convenience of our discussion, we assume that all $\mu(t;\delta,\varepsilon)$ are non-negative.

Under the present framework, the intensity functions satisfy the relation

$$\sum_{\delta=1}^{r} \mu(t;\delta) + \sum_{\delta=1}^{r-1} \sum_{\varepsilon=\delta+1}^{r} \mu(t;\delta,\varepsilon) = \mu(t). \tag{3.5}$$

We shall assume, as in Section 2, that the intensity functions $\mu(t;\delta)$ satisfy condition (2.3) and that the ratio

$$\frac{\mu(t;\delta,\varepsilon)}{\mu(t)} = c_{i\delta\varepsilon}, \text{ for } x_i \leq t < x_{i+1}, \tag{3.6}$$

is independent of time t, but is a function of the interval (x_i,x_{i+1}) and risks R_δ and R_ε. In this case the formulas for the probability $p_i(q_i)$, the crude probability $Q_{i\delta}$ and the net probability $q_{i\delta}$ all remain the same as those in Section 2. Namely,

$$p_i = e^{-\int_{x_i}^{x_{i+1}} \mu(t)dt}, \qquad q_i = 1 - p_i \tag{2.4}$$

$$Q_{i\delta} = \int_{x_i}^{x_{i+1}} e^{-\int_{x_i}^{t} \mu(\tau)d\tau} \mu(t;\delta)dt \tag{2.6}$$

and

$$q_{i\delta} = \int_{x_i}^{x_{i+1}} e^{-\int_{x_i}^{t} \mu(\tau;\delta)d\tau} \mu(t;\delta)dt \quad . \tag{2.15}$$

so that the relation,

$$q_{i\delta} = 1-p_i^{Q_{i\delta}/q_i} \quad , \tag{2.17}$$

and

$$q_{i\delta} = Q_{i\delta}[1 + \frac{1}{2}(q_i-Q_{i\delta}) + \frac{1}{6}(q_i-Q_{i\delta})(2q_i-Q_{i\delta})] \tag{2.21}$$

also holds. Corresponding to the interaction $R_{\delta\varepsilon}$, there is the crude probability of dying from $R_{\delta\varepsilon}$ in the presence of all risks,

$$Q_{i\delta\varepsilon} = \int_{x_i}^{x_{i+1}} e^{-\int_{x_i}^{t} \mu(\tau)d\tau} \mu(t;\delta,\varepsilon)dt \ . \tag{3.7}$$

In view of the proportionality assumption in (3.6), we may rewrite (3.7) as follows

$$\begin{aligned} Q_{i\,\delta\varepsilon} &= \frac{\mu(t;\delta,\varepsilon)}{\mu(t)} \int_{x_i}^{x_{i+1}} e^{-\int_{x_i}^{t} \mu(\tau)d\tau} \mu(t)dt \\ &= \frac{\mu(t;\delta,\varepsilon)}{\mu(t)}\, q_i \end{aligned} \tag{3.7a}$$

When risk R_1 is removed as a cause of death, its interactions with all other causes, $R_{12},\cdots,R_{1r}$, will all vanish, and the net probability of dying in (x_i,x_{i+1}) is given by

$$q_{i\cdot 1} = 1 - e^{-\int_{x_i}^{x_{i+1}} [\mu(t)-\mu(t,1)-\sum_{\varepsilon=2}^{r}\mu(t;1,\varepsilon)]dt} \tag{3.8}$$

Using the proportionality assumption (2.3) and (3.6), the exponent in (3.8) may be rewritten

$$\int_{x_i}^{x_{i+1}} [\mu(t)-\mu(t;1)-\sum_{\varepsilon=2}^{r}\mu(t;1,\varepsilon)]dt = [\int_{x_i}^{x_{i+1}} \mu(t)dt]\frac{\mu(t)-\mu(t;1) - \sum_{\varepsilon=2}^{r} \mu(t;1,\varepsilon)}{\mu(t)}$$

which, because of (2.9) and (3.7a), becomes

$$[\int_{x_i}^{x_{i+1}} \mu(t)dt]\frac{q_i-Q_{i1}-\sum Q_{i1\varepsilon}}{q_i} \quad . \tag{3.9}$$

Substituting (3.9) in (3.8) yields a relationship between the net probability $q_{i\cdot 1}$ and q_i and the crude probabilities:

$$q_{i\cdot 1} = 1-p_i^{(q_i-Q_{i1}-\sum Q_{i1\varepsilon})/q_i} \tag{3.10}$$

where the summation $\Sigma Q_{i1\varepsilon}$ in the exponent is taken over $\varepsilon = 2,\ldots,r$.

Applying the binomial expression to the last term in (3.10) and taking the first three terms of the infinite series as in (2.17), we have

$$\begin{aligned} p_i^{(q_i-Q_{i1}\,-\,\Sigma\, Q_{i1\varepsilon})/q_i} &= (1-q_i)^{(q_i-Q_{i1}\,-\,\Sigma\, Q_{i1\varepsilon})/q_i} \\ &= 1 - (q_i-Q_{i1}-\Sigma\, Q_{i1\varepsilon}) - \tfrac{1}{2}(q_i-Q_{i1}-\Sigma Q_{i1\varepsilon})(Q_{i1}+\Sigma Q_{i1\varepsilon}) \\ &\quad - \tfrac{1}{6}(q_i-Q_{i1}-\Sigma Q_{i1\varepsilon})(Q_{i1}+\Sigma Q_{i1\varepsilon})(q_i+Q_{i1}+\Sigma Q_{i1\varepsilon}) \end{aligned} \tag{3.11}$$

and the approximate formula [4]

$$q_{i\cdot 1} = (q_i-Q_{i1} - \sum_{\varepsilon=2}^{r} Q_{i1\varepsilon})\ [1 + \tfrac{1}{2}\,(Q_{i1} + \sum_{\varepsilon=2}^{r} Q_{i1\varepsilon}) + \tfrac{1}{6}(Q_{i1}+\sum_{2}^{r}Q_{i1\varepsilon})(q_i+Q_{i1}+\sum_{2}^{r}Q_{i1\varepsilon})] \tag{3.12}$$

The partial crude probability of dying from R_δ when R_1 is eliminated can be obtained in a similar manner. Corresponding to formula (2.31) we now have

$$\begin{aligned} Q_{i\delta\cdot 1} &= \int_{x_i}^{x_{i+1}} e^{-\int_{x_i}^{t}[\mu(\tau)-\mu(\tau;1)-\sum\mu(\tau;1,\varepsilon)]d\tau}\ \mu(t;\delta)dt \\ &= \frac{\mu(t;\delta)}{\mu(t)-\mu(t;1)-\sum\mu(t;1,\varepsilon)} \int_{x_i}^{x_{i+1}} e^{-\int_{x_i}^{t}[\mu(\tau)-\mu(\tau;1)-\sum\mu(\tau;1,\varepsilon)]d\tau}\ [\mu(t)-\mu(t;1)-\Sigma\mu(t;1,\varepsilon)]dt\ , \end{aligned} \tag{3.13}$$

where

$$\frac{\mu(t;\delta)}{\mu(t)-\mu(t;1)-\sum\mu(t;1,\varepsilon)} = \frac{Q_{i\delta}}{q_i-Q_{i1}-\sum_{\varepsilon=2}^{r}Q_{i1\varepsilon}} \tag{3.14}$$

and

$$\int_{x_i}^{x_{i+1}} e^{-\int_{x_i}^{t}[\mu(\tau)-\mu(\tau;1)-\sum\mu(\tau;1,\varepsilon)]d\tau}[\mu(t)-\mu(t;1)-\sum\mu(t;1\varepsilon)]dt = 1-p_i^{(q_i-Q_{i1}-\sum Q_{i1\varepsilon})/q_i} \tag{3.15}$$

$$= q_{i\cdot 1} \; .$$

Therefore

$$Q_{i\delta\cdot 1} = \frac{Q_{i\delta}}{q_i-Q_{i1}-\sum_{\varepsilon=2}^{r}Q_{i1\varepsilon}}\, q_{i\cdot 1} \; . \tag{3.16}$$

In the following **appendix** we shall present the problem of estimation and the multiple decrement tables without **considering the interaction.** The case where the interactions are present is completely analogous.

FOOTNOTES

1/ When the first three terms of the binomial series are taken, we have

$$q_{i.\delta} = (q_i - Q_{i\delta})[1 + \frac{1}{2} Q_{i\delta}] \quad . \tag{2.29a}$$

2/ When formula (2.29a) in footnote 1 is used, we have

$$Q_{i\delta.1} = Q_{i\delta}[1 + \frac{1}{2} Q_{i1}] \quad . \tag{2.33a}$$

3/ Corresponding to formula (2.33a) in footnote 2, we have

$$Q_{i\delta.12} = Q_{i\delta}[1 + \frac{1}{2}(Q_{i1} + Q_{i2})] \quad . \tag{2.35a}$$

4/ Corresponding to formula (2.29a) in footnote 1, we have

$$q_{i.1} = (q_i - Q_{i1} - \sum_{2}^{r} Q_{i1\varepsilon})[1 + \frac{1}{2}(Q_{i1} + \sum_{2}^{r} Q_{i1\varepsilon})]. \tag{3.12a}$$

APPENDIX IV

MULTIPLE DECREMENT TABLES

1. Introduction

In studies of competing risks in a given population, deaths are classified according to cause. The number of deaths from each specific cause is the basic random variable for estimating the corresponding probability and for making inference about the population in question. The statistical theory involves is the multinomial distribution, which is described as follows.

Suppose that ℓ_i individuals alive at the beginning of an age interval (x_i, x_{i+1}) are subject to r risks of death, $R_1, \cdots, R_r$, with the corresponding probabilities $Q_{i1}, \cdots, Q_{ir}$, respectively. Let $d_{i\delta}$ be the number of deaths occurring in the interval from R_δ so that the sum

$$d_{i1} + \cdots + d_{ir} = d_i \tag{1.1}$$

is the total number of deaths, and the difference

$$\ell_i - d_i = \ell_{i+1} \tag{1.2}$$

is the number of survivors at the end of the interval. This means that

$$\ell_i = d_{i1} + \cdots + d_{ir} + \ell_{i+1} \ . \tag{1.3}$$

The corresponding probabilities have a similar relationship (cf. equation (2.10), Appendix III)

$$Q_{i1} + \cdots + Q_{ir} = q_i \ ; \tag{1.4}$$

and the difference

$$1 - q_i = p_i \tag{1.5}$$

is the probability of survival, so that

$$1 = Q_{i1}+\cdots+Q_{ir}+p_i \ . \tag{1.6}$$

Equations (1.3) and (1.6) define a multinomial distribution, where the random variables $d_{i1},\cdots,d_{ir}$ and ℓ_{i+1} have the joint probability[1]

$$\frac{\ell_i!}{d_{i1}!\cdots d_{ir}!\ell_{i+1}!} \quad Q_{i1}^{\ d_{i1}}\cdots Q_{ir}^{\ d_{ir}}\ p_i^{\ \ell_{i+1}} \tag{1.7}$$

where $d_{i1}+\ldots+d_{ir}+\ell_{i+1} = \ell_i$.

In formula (1.7) the quantity $Q_{i1}^{\ d_{i1}}\cdots Q_{ir}^{\ d_{ir}}\ p_i^{\ \ell_{i+1}}$ is the probability that $d_{i\delta}$ specified individual will die from R_δ, for $\delta=1,\cdots,r$, and the remaining ℓ_{i+1} individuals will survive the interval (x_i,x_{i+1}). The combinational factor

$$\frac{\ell_i!}{d_{i1}!\cdots d_{ir}!\ell_{i+1}!}$$

is the number of possibilities that $d_{i\delta}$ people among ℓ_i will die from R_δ and ℓ_{i+1} will survive. The expected values, the variances and the covariances of the distribution can be obtained from (1.7) directly. However, the following approach is somwhat simpler.

Each $d_{i\delta}$ is in effect a binomial random variables in ℓ_i "trials" with the binomial probability $Q_{i\delta}$. Therefore, the expected value and the variance of $d_{i\delta}$ are given by

$$E(d_{i\delta}|\ell_i) = \ell_i Q_{i\delta} \tag{1.8}$$

and

$$\mathrm{Var}(d_{i\delta}|\ell_i) = \ell_i Q_{i\delta}(1-Q_{i\delta}),\ \delta=1,\cdots,r. \tag{1.9}$$

---

(1) For simplicity of formulas, no symbols for the values that the random variables ℓ_i and $d_{i\delta}$ assume are introduced in this **Appendix.**

The covariance between any two random variables $d_{i\delta}$ and $d_{i\varepsilon}$ is

$$\operatorname{Cov}(d_{i\delta}, d_{i\varepsilon} | \ell_i) = -\ell_i Q_{i\delta} Q_{i\varepsilon} \quad , \quad \delta \neq \varepsilon;\ \delta, \varepsilon = 1, \ldots, r. \tag{1.10}$$

Therefore the correlation coefficient between $d_{i\delta}$ and $d_{i\varepsilon}$ is negative.

$$\rho_{d_{i\delta}, d_{i\varepsilon} | \ell_i} = - \frac{\ell_i Q_{i\delta} Q_{i\varepsilon}}{\sqrt{\ell_i Q_{i\delta}(1-Q_{i\delta}) \ell_i Q_{i\varepsilon}(1-Q_{i\varepsilon})}}$$

$$= - \sqrt{\frac{Q_{i\delta} Q_{i\varepsilon}}{(1-Q_{i\delta})(1-Q_{i\varepsilon})}} \quad . \tag{1.11}$$

Formula (1.11) shows that the larger the probabilities $Q_{i\delta}$ and $Q_{i\varepsilon}$, the greater will be the correlation coefficient in absolute value. Thus the greater the number of deaths from one cause, the smaller will be the number of deaths from another cause; and the two risks, R_δ and R_ε, are indeed competing risks.

The covariance between the number of deaths from a specific cause $d_{i\delta}$ and the number of survivors ℓ_{i+1} can be obtained in a similar manner. The formula is

$$\operatorname{Cov}(d_{i\delta}, \ell_{i+1} | \ell_i) = -\ell_i Q_{i\delta} p_i \tag{1.12}$$

and hence the correlation coefficient

$$\rho_{d_{i\delta} \ell_{i+1}} = - \sqrt{\frac{Q_{i\delta}\, p_i}{(1-Q_{i\delta})(1-p_i)}} \tag{1.13}$$

which also increases in absolute value with $Q_{i\delta}$ and p_i. This means that the greater the number of deaths from one cause, the fewer will be the survivors.

When ℓ_i is a given fixed number, the variance of ℓ_i is zero. Applying the formula of the variance of a sum of random variables to formula (1.3) we have

$$\sum_{\delta=1}^{r} \mathrm{Var}(d_{i\delta}|\ell_i) + \mathrm{Var}(\ell_{i+1}|\ell_i) + 2\sum_{\delta=1}^{r-1}\sum_{\varepsilon=\delta+1}^{r} \mathrm{Cov}(d_{i\delta},d_{i\varepsilon}|\ell_i)$$

$$+ 2\sum_{\delta=1}^{r} \mathrm{Cov}(d_{i\delta},\ell_{i+1}|\ell_i) = 0 . \tag{1.14}$$

Substituting formulas (1.9), (1.10), (1.12) and the conditional variance of ℓ_{i+1} given ℓ_i,

$$\mathrm{Var}(\ell_{i+1}|\ell_i) = \ell_i p_i q_i , \tag{1.15}$$

in (1.14) yields a relationship

$$\sum_{\delta=1}^{r} \ell_i\, Q_{i\delta}(1-Q_{i\delta}) + \ell_i p_i q_i - 2\sum_{\delta=1}^{r-1}\sum_{\varepsilon=\delta+1}^{r} \ell_i\, Q_{i\delta}\, Q_{i\varepsilon}$$

$$- 2\sum_{\delta=1}^{r} \ell_i\, Q_{i\delta}\, P_i = 0 \tag{1.16}$$

Verification of (1.16) is left to the reader.

2. The Chain Multinomial Distributions

In the preceding section we were concerned with the probability distribution of the number of deaths occurring in interval (x_i, x_{i+1}) based on ℓ_i people alive at age x_i. When we start at age x_0 with ℓ_0 individuals, the number of survivors $\ell_1, \ell_2, \ldots$ at ages $x_1, x_2 \ldots$ are themselves random variables. The probability distribution of each ℓ_i is dependent upon the number of survivors of the preceding interval, for i=1,2... As a result, we have a chain of multinomial distributions. In other words, for any positive integer, u, the joint probability distribution of all the random variables $d_{i1}, \ldots, d_{ir}, \ell_{i+1}$, for $i=0,1,\ldots,u$ is

$$\prod_{i=0}^{u} \frac{\ell_i!}{d_{i1}!\ldots d_{ir}!\ \ell_{i+1}!} Q_{i1}^{d_{i1}} \ldots Q_{ir}^{d_{ir}} p_i^{\ell_{i+1}} \tag{2.1}$$

with $d_{i\delta}$ and ℓ_{i+1} being non-negative integers and satisfying **the restriction** $d_{i1}+\ldots+d_{ir}+\ell_{i+1} = \ell_i$.

The expected values and variances of the random variables $d_{i\delta}$ may be derived from those obtained in Section 1 by using the rule on conditional expectation and conditional variance. The expectation of $d_{i\delta}$ is the expectation of the condition expection of $d_{i\delta}$ given ℓ_i

$$\begin{aligned} E(d_{i\delta}) &= E[E(d_{i\delta}|\ell_i)] = E[\ell_i\ Q_{i\delta}] \\ &= E(\ell_i)Q_{i\delta} = \ell_0 p_{0i} Q_{i\delta} \end{aligned} \tag{2.2}$$

where

$$p_{0i} = \exp\{-\int_0^{x_i} \mu(t)dt\}$$

is the probability of surviving from x_0 to x_i.

The rule on the variances is a little more complex. When ℓ_i is a random variable, the conditional variance of $d_{i\delta}$ given ℓ_i is also a random variable and has an expectation

$$E[\mathrm{Var}(d_{i\delta}|\ell_i)] = E[\ell_i Q_{i\delta}(1-Q_{i\delta})]$$

$$= E(\ell_i)Q_{i\delta}(1-Q_{i\delta}) = \ell_0 p_{0i} Q_{i\delta}(1-Q_{i\delta}) \quad . \tag{2.3}$$

On the other hand, the conditional expectation $E(d_{i\delta}|\ell_i)$, being a random variable, has the variance

$$\mathrm{Var}[E(d_{i\delta}|\ell_i)] = \mathrm{Var}[\ell_i Q_{i\delta}] = Q_{i\delta}^2 \ \mathrm{Var}(\ell_i)$$

$$= Q_{i\delta}^2 \ \ell_0 p_{0i}(1-p_{0i}) \quad . \tag{2.4}$$

According to the rule, the variance of $d_{i\delta}$ is given by

$$\mathrm{Var}(d_{i\delta}) = E[\mathrm{Var}(d_{i\delta}|\ell_i)] + \mathrm{Var}[E(d_{i\delta}|\ell_i)] \tag{2.5}$$

Substituting (2.3) and (2.4) in (2.5) and simplifying the resulting expression yield the formula

$$\mathrm{Var}(d_{i\delta}) = \ell_0 \, p_{0i} \, Q_{i\delta}(1-p_{0i}Q_{i\delta}), \quad \begin{array}{l}\delta=1,\ldots,r;\\ i=0,\ldots,u.\end{array} \tag{2.6}$$

Regarding the covariance of $d_{i\delta}$ and $d_{i\varepsilon}$, the rule is

$$\mathrm{Cov}(d_{i\delta},d_{i\varepsilon}) = E[\mathrm{Cov}(d_{i\delta},d_{i\varepsilon}|\ell_i)] + \mathrm{Cov}[E(d_{i\delta}|\ell_i), E(d_{i\varepsilon}|\ell_i)]$$

and hence the formula for the covariance is

$$\mathrm{Cov}(d_{i\delta},d_{i\varepsilon}) = -\ell_0 p_{0i} Q_{i\delta} p_{0i} Q_{i\varepsilon}, \quad \begin{array}{l}\delta\neq\varepsilon;\ \delta,\ \varepsilon=1,\ldots,r;\\ i=0,\ldots,u.\end{array} \tag{2.7}$$

Formulas (2.6) and (2.7) can be justified intuitively. An individual alive at x_0 has a probability $p_{0i}Q_{i\delta}$ of dying from R_δ in interval (x_i, x_{i+1}). The number of individuals dying from R_δ , $d_{i\delta}$, has a binomial distribution in ℓ_0 "trials" with the probability $p_{0i}Q_{i\delta}$. It follows from the binomial theory that the variance of $d_{i\delta}$ is given by (2.6) [Cf. formula (1.9)]. Similarly, the random variables $d_{i\delta}$ and $d_{i\varepsilon}$ have a joint multinomial distribution with the corresponding probabilities $p_{0i}Q_{i\delta}$ and $p_{0i}Q_{i\varepsilon}$, respectively. Therefore, their covariance is given by

$$\operatorname{Cov}(d_{i\delta},d_{i\varepsilon}) = -\ell_0 p_{0i}Q_{i\delta}p_{0i}Q_{i\varepsilon} \tag{2.8}$$

and their correlation coefficient is

$$\rho_{d_{i\delta},d_{i\varepsilon}} = -p_{0i}\sqrt{\frac{Q_{i\delta}}{1-p_{0i}Q_{i\delta}}}\sqrt{\frac{Q_{i\varepsilon}}{1-p_{0i}Q_{i\varepsilon}}} . \tag{2.9}$$

The negative correlation coefficient again indicates the competition between two risks, and the negative correlation is more pronounced when the corresponding probabilities of death, $Q_{i\delta}$ and $Q_{i\varepsilon}$, are large. Also the correlation coefficient increases in absolute value with p_{0i}, the probability of surviving the interval (x_0, x_i). Since p_{0i} decreases with x_i, the competition between two risks is more acute at young ages or when the probability of dying $Q_{i\delta}$ $(Q_{i\varepsilon})$ is large.

For the numbers of deaths occurring in two different age intervals (x_i, x_{i+1}) and (x_j, x_{j+1}), the corresponding covariance is obtained by using once again the fact that the random variables $d_{i\delta}$ and $d_{j\varepsilon}$ have a joint multinomial distribution in ℓ_0 "trials"with the corresponding probabilities $p_{0i}Q_i$ and $p_{0j}Q_j$ so that

$$\operatorname{Cov}(d_{i\delta},d_{j\varepsilon}) = -\ell_0 p_{0i}Q_{i\delta}\, p_{0j}Q_{j\varepsilon} \tag{2.10}$$

and the correlation coefficient

$$\rho_{d_{i\delta},d_{j\varepsilon}} = -\sqrt{\frac{p_{0i}\ Q_{i\delta}}{1-p_{0i}Q_{i\delta}}}\sqrt{\frac{p_{0j}\ Q_{j\varepsilon}}{1-p_{0j}Q_{j\varepsilon}}} \tag{2.11}$$

is again negative. Without loss of generality, we may assume that i < j and use the relationship $p_{0j} = p_{0i}\, p_{ij}$ and write

$$\rho_{d_{i\delta},d_{j\varepsilon}} = - p_{0i} \sqrt{p_{ij}} \sqrt{\frac{Q_{i\delta}}{1 - p_{0i}Q_{i\delta}}} \sqrt{\frac{Q_{j\varepsilon}}{1 - p_{0j}Q_{j\varepsilon}}} \tag{2.12}$$

which differs from the correlation coefficient in (2.9) by the factor $\sqrt{p_{ij}}$, and equals (2.9) when $j=i$. Thus the correlation coefficient between $d_{i\delta}$ and $d_{j\varepsilon}$ is generally smaller in absolute value than the correlation coefficient between $d_{i\delta}$ and $d_{i\varepsilon}$. For a fixed x_i, the probability p_{ij} of surviving from x_i to x_j decreases as x_j increases. This means that the competition between risks at two different ages diminishes as the two ages become more distant.

The covariance between the number dying and the number surviving can be derived in a similar way:

$$\mathrm{Cov}(d_{i\delta},\ell_j) = - \ell_0 p_{0i} Q_{i\delta} p_{0j} \tag{2.13}$$

and

$$\mathrm{Cov}(\ell_i,\ d_{j\delta}) = \ell_0(1-p_{0i})p_{0j}Q_{j\delta}\ , \qquad \begin{array}{l}\delta= 1,\ldots,r;\\ i<j;\ i,j=0,1,\ldots.\end{array} \tag{2.14}$$

It is interesting to note that the covariance between ℓ_i and $d_{j\delta}$ in (2.14) is the only one carrying a positive sign. The positive covariance in (2.14) indicates that the larger the number of survivors at age x_i, the greater the probability that a larger number of deaths from R_δ will occur in a subsequent interval $(x_j,\ x_{j+1})$. The covariance between ℓ_i and ℓ_j

$$\sigma_{\ell_i,\ell_j} = \ell_0(1-p_{0i})p_{0j}\ , \qquad i < j, \tag{2.15}$$

has been given in (3.7) in **Appendix II.** These results show that, for each u, the random variables $d_{i\delta}$ and ℓ_{i+1}, for $i=0,1,\ldots,u$; $\delta=1,\ldots,r$, have a chain of multinomial distributions with the probability distribution given in (2.1) and the expectations and covariances given in (2.2) through (2.15).

3. Estimation of the Crude Probabilities

The estimators of the crude probabilities Q_{ik} and p_i can be derived directly from the joint probability function

$$L = \prod_{i=0}^{u} \frac{\ell_i!}{d_{i1}! \ldots d_{ir}! \; \ell_{i+1}!} \; Q_{i1}^{d_{i1}} \ldots Q_{ir}^{d_{ir}} \; p_i^{\ell_{i+1}} \qquad (3.1)$$

by using the maximum likelihood principle.

The logarithm of the likelihood functions is

$$\log L = k + \sum_{i=0}^{u} [d_{i1} \log Q_{i1} + \ldots + d_{ir} \log Q_{ir} + \ell_{i+1} \log p_i] \quad (3.2)$$

where k is constant and the probabilities are not all independent but satisfy the relationship for each i

$$Q_{i1} + \ldots + Q_{ir} + p_i = 1 \; . \qquad (3.3)$$

The estimators $\hat{Q}_{i1}$, ..., $\hat{Q}_{ir}$, and $\hat{p}_i$ are the maximizing values of log L subject to condition (3.3). Using the Lagrange method we maximize

$$\Phi = k + \sum_{i=0}^{u} [\sum_{\delta=1}^{r} d_{i\delta} \log Q_{i\delta} + \ell_{i+1} \log p_i - \lambda_i (\sum_{\delta=1}^{r} Q_{i\delta} + p_i - 1)] \; .$$

Differentiating Φ with respect to $Q_{i1},\ldots,Q_{ir}$, p_i and setting the derivatives equal to zero, we have the following simultaneous equations

$$\frac{\partial}{\partial Q_{i\delta}} \Phi = \frac{d_{i\delta}}{\hat{Q}_{i\delta}} - \lambda_i = 0 \qquad \text{or} \qquad \hat{Q}_{i\delta} = \frac{d_{i\delta}}{\lambda_i} \, , \qquad \delta=1,\ldots,r \qquad (3.4)$$

$$\frac{\partial}{\partial p_i} \Phi = \frac{\ell_{i+1}}{\hat{p}_i} - \lambda_i = 0 \qquad \text{or} \qquad \hat{p}_i = \frac{\ell_{i+1}}{\lambda_i} \, , \qquad (3.5)$$

For each i, there are r+2 equations in (3.3), (3.4), and (3.5) with r+2 unknowns: $\hat{Q}_{i1},\ldots,\hat{Q}_{ir}$, $\hat{p}_i$ and λ_i, where λ_i is known as the Lagrange coefficient. To solve these equations simultaneously we substitute (3.4) and (3.5) in (3.3),

$$\frac{d_{i1}}{\lambda_i} + \ldots + \frac{d_{ir}}{\lambda_i} + \frac{\ell_{i+1}}{\lambda_i} = 1$$

or

$$\lambda_i = d_{i1} + \ldots + d_{ir} + \ell_{i+1} = \ell_i \tag{3.6}$$

and use (3.6) in (3.4) and (3.5) to obtain the maximum likelihood estimators

$$\hat{Q}_{i\delta} = \frac{d_{i\delta}}{\ell_i}, \qquad \begin{matrix} \delta=1,\ldots,r \\ i=0,\ldots,u \end{matrix} \tag{3.7}$$

and

$$\hat{p}_i = \frac{\ell_{i+1}}{\ell_i}, \qquad i=0,\ldots,u \quad . \tag{3.8}$$

On the right-hand side of (3.7) and (3.8) are the proportions dying in the interval (x_i, x_{i+1}) from R_δ and the proportion surviving the interval, respectively. Thus the maximum likelihood estimators derived (3.7) and (3.8) are consistent with our intuition. Further, they are unbiased estimators, since their expected values are equal to the corresponding probabilities. This is demonstrated below.

$$E[\hat{Q}_{i\delta}] = E[\frac{d_{i\delta}}{\ell_i}] = E[E(d_{i\delta}|\ell_i)\ \frac{1}{\ell_i}], \tag{3.9}$$

where the conditional expectation is given in Section 1,

$$E(d_{i\delta}|\ell_i) = \ell_i Q_{i\delta} \quad , \tag{1.8}$$

so that

$$E[\hat{Q}_{i\delta}] = Q_{i\delta} \tag{3.10}$$

as required to be shown.

The variances and covariances of the estimators can be computed directly. For the estimator $\hat{Q}_{i\delta}$,

$$\operatorname{Var}(\hat{Q}_{i\delta}) = E[\hat{Q}_{i\delta}^{\,2}] - Q_{i\delta}^{\,2} , \tag{3.11}$$

where

$$E[\hat{Q}_{i\delta}^{\,2}] = E[\frac{d_{i\delta}^{\,2}}{\ell_i^{\,2}}] = E[E(d_{i\delta}^{\,2}|\ell_i)\,\frac{1}{\ell_i^{\,2}}] \quad . \tag{3.12}$$

We recall from Section 1 that, given ℓ_i, the number of deaths $d_{i\delta}$ is a binomial random variable having the variance

$$\operatorname{Var}(d_{i\delta}|\ell_i) = \ell_i Q_{i\delta}(1-Q_{i\delta}), \tag{1.9}$$

and the expectation of the square

$$E(d_{i\delta}^{\,2}|\ell_i) = \ell_i Q_{i\delta}(1-Q_{i\delta}) + \ell_i^{\,2} Q_{i\delta}^{\,2} \quad . \tag{3.13}$$

Consequently, the expectation in (3.12) may be rewritten

$$E[\hat{Q}_{i\delta}^{\,2}] = E(\frac{1}{\ell_i})\; Q_{i\delta}(1-Q_{i\delta}) + Q_{i\delta}^{\,2} \quad . \tag{3.14}$$

Substituting (3.14) in (3.11) gives the formula

$$\operatorname{Var}(\hat{Q}_{i\delta}) = E(\frac{1}{\ell_i})\; Q_{i\delta}(1-Q_{i\delta}), \qquad \begin{matrix}\delta=1,\ldots,r\\ i=0,\ldots,u\end{matrix} \quad . \tag{3.15}$$

Using a similar approach, we obtain

$$\operatorname{Var}(\hat{p}_i) = \operatorname{Var}(\hat{q}_i) = E(\frac{1}{\ell_i})p_i q_i \tag{3.16}$$

$$\operatorname{Cov}(\hat{Q}_{i\delta},\, \hat{Q}_{i\varepsilon}) = -E(\frac{1}{\ell_i})\; Q_{i\delta}\; Q_{i\varepsilon} \tag{3.17}$$

and

$$\operatorname{Cov}(\hat{Q}_{i\delta},\, \hat{p}_i) = -E(\frac{1}{\ell_i})\; p_i\; Q_{i\delta} \quad . \tag{3.18}$$

When the original cohort ℓ_0 is large, the expectation of the reciprocal of ℓ_i may be approximated by the reciprocal of the expectation. That is

$$E\left(\frac{1}{\ell_i}\right) \doteq \frac{1}{E(\ell_i)} = \frac{1}{\ell_0 p_{0i}} \quad . \tag{3.19}$$

In order to make inferences concerning the crude probabilities, it is necessary to find the sample estimates of the standard errors of $\hat{Q}_{i\delta}$. This may be done by substituting $\hat{Q}_{i\delta}$, $\hat{p}_i$, and ℓ_i for the corresponding unknown parameters in (3.15) and (3.16) and taking the square root of the resulting formulas. Thus,

$$S_{\hat{Q}_{i\delta}} = \sqrt{\frac{1}{\ell_i}\hat{Q}_{i\delta}(1-\hat{Q}_{i\delta})} \quad , \qquad \delta=1,\ldots,r \tag{3.20}$$

and

$$S_{\hat{q}_i} = \sqrt{\frac{1}{\ell_i}\hat{q}_i(1-\hat{q}_i)} \quad . \qquad i=1,\ldots,u \quad . \tag{3.21}$$

The main results obtained in this Appendix may be summarized in the following table.

Table 1. Multiple Decremental Table

Age Interval (years)	Number Living at Age x_i	Proportion Dying in Interval (x_i,x_{i+1})	Proportions Dying in (x_i,x_{i+1}) by Causes		
			R_1	. . .	R_r
$x_i - x_{i+1}$	ℓ_i	$\hat{q}_i \quad S_{\hat{q}_i}$	$\hat{Q}_{i1} \quad S_{\hat{Q}_{i1}}$	. . .	$\hat{Q}_{ir} \quad S_{\hat{Q}_{ir}}$
$x_0 - x_1$	ℓ_0	$\hat{q}_0 \quad S_{\hat{q}_0}$	$\hat{Q}_{01} \quad S_{\hat{Q}_{01}}$		$\hat{Q}_{0r} \quad S_{\hat{Q}_{0r}}$
.	.	. .	. .	. . .	. .
.	.	. .	. .	. . .	. .
.	.	. .	. .	. . .	. .

APPENDIX V

Fraction of Last Age Interval of Life a_i

Table 1

Austria, 1969

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i		
	Both Sexes	Male	Female
0-1	.12	.12	.12
1-5	.37	.37	.37
5-10	.47	.47	.47
10-15	.51	.51	.49
15-20	.58	.58	.55
20-25	.48	.49	.48
25-30	.51	.50	.54
30-35	.53	.53	.53
35-40	.53	.52	.53
40-45	.52	.51	.54
45-50	.54	.54	.53
50-55	.52	.53	.52
55-60	.53	.54	.53
60-65	.54	.53	.54
65-70	.53	.52	.54
70-75	.52	.50	.53
75-80	.51	.50	.51
80-85	.48	.47	.49
85-90	.45	.44	.45
90-95	.40	.40	.40

Table 2

Fraction of Last Age Interval of Life, a_i

California, 1960

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i
0-1	.10
1-5	.39
5-10	.46
10-15	.57
15-20	.57
20-25	.49
25-30	.50
30-35	.53
35-40	.54
40-45	.54
45-50	.54
50-55	.53
55-60	.51
60-65	.53
65-70	.52
70-75	.52
75-80	.51
80-85	.49
85-90	.46
90-95	.40

Table 3

Fraction of Last Age Interval of Life, a_i

Canada, 1968

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i Both Sexes	Male	Female
0-1	.11	.11	.12
1-5	.41	.42	.40
5-10	.45	.45	.44
10-15	.54	.54	.53
15-20	.57	.59	.53
20-25	.48	.47	.51
25-30	.50	.49	.53
30-35	.52	.52	.52
35-40	.53	.53	.53
40-45	.54	.54	.54
45-50	.53	.53	.53
50-55	.54	.54	.54
55-60	.54	.53	.54
60-65	.53	.53	.53
65-70	.53	.52	.53
70-75	.52	.51	.53
75-80	.52	.51	.53
80-85	.50	.49	.51
85-90	.47	.46	.48

Table 4

Fraction of Last Age Interval of Life, a_i

Costa Rica, 1963

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i Both Sexes	Male	Female
0-1	.28	.27	.28
1-5	.29	.29	.28
5-10	.40	.42	.38
10-15	.49	.50	.50
15-20	.55	.55	.55
20-25	.53	.53	.54
25-30	.53	.51	.55
30-35	.51	.51	.51
35-40	.49	.51	.48
40-45	.53	.54	.52
45-50	.53	.51	.55
50-55	.53	.53	.52
55-60	.54	.55	.53
60-65	.53	.55	.51
65-70	.54	.56	.52
70-75	.53	.52	.54
75-80	.51	.52	.51
80-85	.50	.50	.50

Table 5

Fraction of Last Age Interval of Life, a_i

Finland, 1968

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i Both Sexes	Male	Female
0-1	.09	.08	.09
1-5	.38	.41	.34
5-10	.49	.48	.49
10-15	.52	.53	.50
15-20	.53	.53	.54
20-25	.51	.52	.51
25-30	.51	.52	.48
30-35	.52	.51	.52
35-40	.54	.54	.53
40-45	.55	.54	.55
45-50	.53	.52	.54
50-55	.54	.54	.53
55-60	.53	.53	.54
60-65	.53	.53	.54
65-70	.52	.51	.53
70-75	.52	.51	.53
75-80	.51	.49	.52
80-85	.47	.47	.48

Table 6

Fraction of Last Age Interval of Life, a_i

France, 1969

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i		
	Both Sexes	Male	Female
0-1	.16	.15	.17
1-5	.38	.39	.36
5-10	.46	.47	.45
10-15	.54	.55	.52
15-20	.56	.56	.55
20-25	.51	.50	.51
25-30	.51	.51	.52
30-35	.53	.53	.54
35-40	.53	.53	.52
40-45	.53	.53	.53
45-50	.54	.54	.54
50-55	.52	.52	.52
55-60	.53	.53	.53
60-65	.53	.52	.53
65-70	.53	.52	.54
70-75	.52	.51	.53
75-80	.51	.50	.52
80-85	.49	.48	.50
85-90	.46	.45	.47
90-95	.41	.39	.42

Table 7

Fraction of Last Age Interval of Life, a_i

East Germany, 1967

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i Both Sexes	Male	Female
0-1			
1-5	.38	.38	.38
5-10	.46	.46	.46
10-15	.52	.53	.51
15-20	.56	.58	.54
20-25	.50	.50	.51
25-30	.52	.51	.53
30-35	.52	.52	.53
35-40	.52	.52	.52
40-45	.54	.54	.54
45-50	.54	.55	.54
50-55	.52	.53	.52
55-60	.54	.54	.54
60-65	.54	.53	.54
65-70	.53	.53	.54
70-75	.52	.51	.53
75-80	.51	.49	.52
80-85	.48	.47	.49
85-90	.43	.43	.43
90-95	.39	.39	.39

Table 8

Fraction of Last Age Interval of Life, a_i

West Germany, 1969

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i: Both Sexes	Male	Female
0-1	.10	.10	.11
1-5	.39	.39	.38
5-10	.46	.46	.46
10-15	.52	.51	.52
15-20	.57	.58	.54
20-25	.52	.51	.53
25-30	.51	.51	.51
30-35	.52	.52	.53
35-40	.54	.54	.55
40-45	.53	.53	.53
45-50	.51	.51	.51
50-55	.58	.58	.57
55-60	.54	.54	.54
60-65	.54	.53	.54
65-70	.52	.52	.53
70-75	.52	.51	.53
75-80	.51	.49	.52
80-85	.49	.47	.49
85-90	.44	.43	.45
90-95	.39	.38	.40

Table 9

Fraction of Last Age Interval of Life, a_i

Hungary, 1967

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i Both Sexes	Male	Female
0-1	.10	.10	.11
1-5	.35	.35	.33
5-10	.45	.47	.42
10-15	.52	.51	.54
15-20	.55	.57	.52
20-25	.51	.52	.50
25-30	.52	.52	.53
30-35	.52	.51	.52
35-40	.53	.52	.55
40-45	.53	.52	.53
45-50	.54	.54	.53
50-55	.53	.53	.52
55-60	.54	.54	.54
60-65	.53	.53	.54
65-70	.53	.52	.54
70-75	.52	.51	.53
75-80	.50	.50	.51
80-85	.48	.47	.48

Table 10

Fraction of Last Age Interval of Life, a_i

Ireland, 1966

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i: Both Sexes	Male	Female
0-1	.13	.12	.13
1-5	.38	.39	.37
5-10	.47	.47	.46
10-15	.48	.48	.46
15-20	.55	.56	.54
20-25	.51	.50	.53
25-30	.51	.50	.53
30-35	.52	.52	.51
35-40	.55	.56	.54
40-45	.54	.55	.54
45-50	.50	.50	.50
50-55	.53	.53	.52
55-60	.52	.53	.52
60-65	.52	.52	.53
65-70	.52	.51	.53
70-75	.52	.52	.53
75-80	.49	.49	.50
80-85	.48	.48	.48
85-90	.45	.44	.46
90-95	.39	.38	.40

Table 11

Fraction of Last Age Interval of Life, a_i

North Ireland, 1966

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i: Both Sexes	Male	Female
0-1	.13	.13	.14
1-5	.36	.38	.35
5-10	.45	.47	.41
10-15	.50	.49	.52
15-20	.58	.59	.56
20-25	.52	.54	.48
25-30	.51	.53	.49
30-35	.52	.50	.56
35-40	.53	.51	.55
40-45	.53	.54	.53
45-50	.56	.57	.55
50-55	.54	.54	.54
55-60	.55	.54	.55
60-65	.54	.53	.55
65-70	.52	.52	.53
70-75	.52	.51	.53
75-80	.50	.49	.51
80-85	.50	.49	.51

Table 12

Fraction of Last Age Interval of Life, a_i

Italy, 1966

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i		
	Both Sexes	Male	Female
0-1	.16	.15	.17
1-5	.35	.36	.35
5-10	.46	.47	.45
10-15	.53	.54	.53
15-20	.53	.53	.52
20-25	.51	.51	.50
25-30	.52	.51	.53
30-35	.53	.52	.54
35-40	.53	.53	.53
40-45	.53	.53	.53
45-50	.54	.54	.54
50-55	.54	.54	.53
55-60	.54	.54	.54
60-65	.53	.53	.54
65-70	.52	.52	.53
70-75	.52	.51	.53

Table 13

Fraction of Last Age Interval of Life, a_i

The Netherlands, 1968

Age Interval x_i-x_{i+1}	Fraction of Last Age Interval of Life a_i		
	Both Sexes	Male	Female
0-1	.11	.11	.11
1-5	.41	.43	.39
5-10	.47	.47	.45
10-15	.51	.50	.53
15-20	.54	.55	.52
20-25	.49	.48	.51
25-30	.51	.50	.53
30-35	.51	.51	.51
35-40	.54	.54	.54
40-45	.53	.53	.53
45-50	.55	.55	.54
50-55	.54	.54	.53
55-60	.54	.54	.53
60-65	.53	.52	.54
65-70	.53	.52	.54
70-75	.52	.51	.53
75-80	.51	.50	.52
80-85	.49	.49	.50
85-90	.46	.46	.47
90-95	.42	.42	.42

Table 14

Fraction of Last Age Interval of Life, a_i

Norway, 1968

Age Interval x_i-x_{i+1}	Fraction of Last Age Interval of Life a_i Both Sexes	Male	Female
0-1	.12	.10	.14
1-5	.44	.46	.42
5-10	.45	.46	.42
10-15	.56	.55	.60
15-20	.55	.56	.52
20-25	.51	.50	.52
25-30	.48	.48	.50
30-35	.54	.55	.55
35-40	.54	.55	.54
40-45	.56	.56	.56
45-50	.54	.53	.54
50-55	.53	.54	.53
55-60	.53	.53	.54
60-65	.54	.54	.53
65-70	.54	.53	.55
70-75	.53	.52	.54
75-80	.51	.50	.52
80-85	.50	.49	.50
85-90	.47	.46	.47
90-95	.42	.41	.43

Table 15

Fraction of Last Age Interval of Life, a_i

Okinawa, 1960

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i		
	Both Sexes	Male	Female
0-1	.32	.32	.31
1-5	.38	.37	.40
5-10	.45	.47	.45
10-15	.50	.51	.48
15-20	.50	.51	.49
20-25	.51	.53	.49
25-30	.52	.51	.53
30-35	.51	.52	.50
35-40	.50	.48	.52
40-45	.52	.51	.52
45-50	.53	.53	.54
50-55	.52	.52	.52
55-60	.52	.52	.52
60-65	.53	.52	.54
65-70	.53	.53	.53
70-75	.52	.52	.53
75-80	.52	.52	.52
80-85	.50	.50	.50

Table 16

Fraction of Last Age Interval of Life, a_i

Panama, 1968

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i: Both Sexes	Male	Female
0-1	.23	.23	.24
1-5	.33	.33	.33
5-10	.44	.44	.44
10-15	.49	.49	.50
15-20	.54	.53	.56
20-25	.53	.52	.54
25-30	.49	.49	.49
30-35	.48	.48	.49
35-40	.48	.47	.50
40-45	.49	.49	.50
45-50	.51	.51	.52
50-55	.53	.53	.53
55-60	.52	.52	.52
60-65	.52	.52	.52
65-70	.52	.53	.52
70-75	.44	.44	.45

Table 17

Fraction of Last Age Interval of Life, a_i

Portugal, 1960

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i		
	Both Sexes	Male	Female
0-1	.26	.25	.27
1-5	.27	.27	.27
5-10	.42	.44	.41
10-15	.50	.50	.50
15-20	.53	.54	.52
20-25	.53	.52	.54
25-30	.52	.52	.52
30-35	.52	.52	.52
35-40	.52	.53	.52
40-45	.53	.53	.53
45-50	.53	.53	.53
50-55	.53	.54	.53
55-60	.54	.53	.54
60-65	.54	.53	.54
65-70	.54	.53	.55
70-75	.53	.52	.54
75-80	.52	.51	.53
80-85	.48	.47	.49
85-90	.45	.44	.46
90-95	.39	.38	.40

Table 18

Fraction of Last Age Interval of Life, a_i

Romania, 1965

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i Both Sexes	Male	Female
0-1	.23	.22	.24
1-5	.33	.34	.32
5-10	.46	.47	.43
10-15	.51	.51	.51
15-20	.56	.56	.54
20-25	.51	.51	.51
25-30	.51	.51	.52
30-35	.51	.51	.50
35-40	.53	.52	.53
40-45	.52	.52	.53
45-50	.54	.55	.53
50-55	.53	.53	.53
55-60	.54	.54	.54
60-65	.53	.53	.53
65-70	.54	.52	.55
70-75	.51	.51	.52

Table 19

Fraction of Last Age Interval of Life, a_i

Scotland, 1968

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i Both Sexes	Male	Female
0-1	.13	.13	.23
1-5	.40	.42	.38
5-10	.44	.44	.43
10-15	.53	.53	.53
15-20	.56	.57	.55
20-25	.49	.48	.52
25-30	.51	.51	.52
30-35	.53	.53	.53
35-40	.54	.53	.54
40-45	.54	.54	.54
45-50	.54	.55	.54
50-55	.53	.54	.52
55-60	.54	.54	.53
60-65	.53	.53	.54
65-70	.52	.52	.53
70-75	.51	.50	.52
75-80	.50	.49	.51
80-85	.49	.47	.50

Table 20

Fraction of Last Age Interval of Life, a_i

Spain, 1965

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i Both Sexes	Male	Female
0-1			
1-5	.38	.39	.37
5-10	.46	.47	.46
10-15	.53	.53	.52
15-20	.55	.56	.53
20-25	.54	.53	.55
25-30	.51	.50	.52
30-35	.52	.52	.52
35-40	.53	.53	.53
40-45	.54	.53	.54
45-50	.54	.54	.54
50-55	.54	.54	.54
55-60	.54	.54	.54
60-65	.54	.53	.55
65-70	.54	.53	.55
70-75	.53	.52	.54
75-80	.52	.51	.53

Table 21

Fraction of Last Age Interval of Life, a_i

Sri Lanka, 1952

Age Interval x_i-x_{i+1}	Fraction of Last Age Interval of Life a_i Male	Female
0-1*	.28	.35
1-5	.46	.45
5-10	.53	.53
10-15	.55	.42
15-20	.49	.55
20-25	.51	.54
25-30	.52	.53
30-35	.53	.54
35-40	.53	.54
40-45	.54	.53
45-50	.54	.53
50-55	.53	.53
55-60	.53	.53
60-65	.53	.53
65-70	.53	.53
70-75	.52	.52
75-80	.50	.45
80-85	.42	.35
85-90	.35	

* a_0 values are estimated from the experience of the India 1941-50 populations

Table 22

Fraction of Last Age Interval of Life, a_i

Sweden, 1966

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i Both Sexes	Male	Female
0-1	.08	.08	.08
1-5	.44	.44	.45
5-10	.45	.44	.48
10-15	.53	.52	.55
15-20	.56	.57	.53
20-25	.51	.50	.53
25-30	.52	.53	.51
30-35	.53	.52	.55
35-40	.52	.53	.51
40-45	.53	.53	.54
45-50	.54	.55	.53
50-55	.54	.55	.53
55-60	.54	.54	.53
60-65	.53	.53	.54
65-70	.54	.53	.55
70-75	.53	.52	.54
75-80	.52	.51	.53
80-85	.50	.49	.50
85-90	.46	.45	.47
90-95	.42	.41	.42

Table 23

Fraction of Last Age Interval of Life, a_i

Switzerland, 1968

Age Interval x_i-x_{i+1}	Fraction of Last Age Interval of Life a_i		
	Both Sexes	Male	Female
0-1	.10	.10	.11
1-5	.36	.37	.36
5-10	.45	.45	.45
10-15	.52	.54	.47
15-20	.57	.58	.52
20-25	.49	.48	.49
25-30	.49	.50	.48
30-35	.51	.53	.49
35-40	.54	.54	.53
40-45	.53	.53	.54
45-50	.55	.55	.55
50-55	.54	.54	.53
55-60	.54	.55	.53
60-65	.54	.53	.54
65-70	.53	.53	.54
70-75	.52	.52	.53
75-80	.51	.50	.52
80-85	.50	.49	.51
85-90	.47	.45	.48
90-95	.41	.39	.42

Table 24

Fraction of Last Age Interval of Life, a_i

United States, 1970

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i		
	Both Sexes	Male	Female
0-1	.09	.09	.09
1-5	.40	.40	.39
5-10	.46	.47	.45
10-15	.55	.56	.53
15-20	.54	.55	.53
20-25	.51	.51	.52
25-30	.51	.50	.52
30-35	.52	.52	.53
35-40	.53	.53	.53
40-45	.54	.54	.53
45-50	.54	.54	.53
50-55	.53	.53	.53
55-60	.53	.53	.53
60-65	.52	.52	.53
65-70	.52	.51	.53
70-75	.51	.51	.53
75-80	.51	.50	.52
80-85	.49	.48	.50

Table 25

Fraction of Last Age Interval of Life, a_i

Yugoslavia, 1968

Age Interval $x_i - x_{i+1}$	Fraction of Last Age Interval of Life a_i		
	Both Sexes	Male	Female
0-1	.23	.22	.24
1-5	.29	.31	.28
5-10	.45	.46	.43
10-15	.51	.51	.52
15-20	.53	.54	.53
20-25	.51	.52	.50
25-30	.51	.52	.50
30-35	.52	.53	.52
35-40	.53	.53	.53
40-45	.53	.52	.53
45-50	.54	.54	.54
50-55	.52	.52	.52
55-60	.54	.54	.55
60-65	.53	.53	.54
65-70	.54	.53	.55
70-75	.52	.51	.53
75-80	.49	.49	.50
80-85	.49	.48	.49
85-90	.45	.45	.46
90-95	.38	.38	.38

APPENDIX VI-A

COMPUTER PROGRAM FOR ABRIDGED LIFE TABLE CONSTRUCTION

Identification

Program name	ABRIDGE
Author	Patrick Wong Based on original work by Linda Kwok. Program was further modified by Carol Langhauser to handle WHO data (1969-70) in August, 1974
Department	Biostatistics Program School of Public Health University of California Berkeley, California
Date	February, 1973
Environment	Machine = CDC 6400 Operating System = Calidoscope (SCM) version 01.2-A Coding Language = FORTRAN

Purpose

This program constructs abridged life tables for a series of countries based on the method developed by Chin Long Chiang.

Input card preparation

All input data are assumed to be broken down into 5 year age intervals (except the first year of life) up to age 85 with the last interval being age 85 and over as follows: 0-1, 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, 60-65, 65-70, 70-75, 75-80, 80-85, 85+. Details are given below.

1. Fractions of year lived by those dying in the interval are punched in F3.2 format consecutively starting from column one. Columns 61-80 can be used for optional population ID.

2. Title for the population date in columns 1-80. Standard format:

{'TOTAL'/'MALE'/'FEMALE'} 'POPULATION', (country), (year)

e.g., TOTAL POPULATION, CALIFORNIA, 1970
MALE POPULATION, CANADA, 1968

3. Midyear populations of that country in each age interval in 10I8 format. Two cards are required to accommodate the data for 19 age intervals.

4. Title for the death date in columns 1-80. Standard format: same as 2, except substituting the word 'DEATHS' for 'POPULATION'

5. Number of deaths from all causes in each age interval in 10I8 format. Two cards are required.

Cards in 1-5 can be repeated for as many countries as one desires. The program is terminated if a_0 read in columns 1-3 of card 1 is greater than or equal to 0.80.

Output

For each country, the following quantities are printed out:

1. Raw input data

   a_i = fractions of year lived by those dying in each age interval
   P_i = mid-year populations in each age interval
   D_i = number of deaths in each age interval

2. Construction of abridged life table

   $x_i - x_{i+1}$ = age interval
   P_i
   D_i
   M_i = age specific death rate
   a_i
   $\hat{q}_i$ = proportion dying in interval

3. The abridged life table

   $x_i - x_{i+1}$ = age interval
   $\hat{q}_i$
   ℓ_i = number living at age x_i (ℓ_0 = 100,000)
   d_i = number dying in interval (x_i, x_{i+1})
   a_i
   L_i = number of years lived in interval
   T_i = total number of years lived beyond age x_i
   $\hat{e}_i$ = observed expectation of life at age x_i

Program limitations

1. The current version of the program only handles 19 age intervals, broken down as described in the section, Input card preparation.

2. The maximum population size and number of deaths in any age interval has to be less than a hundred million. However, the input data format card can easily be changed to handle larger or smaller limits.

Computational procedure

1. All input data of a country (a_i, P_i, D_i) are read in.

2. The age specific death rates are computed for each age interval:

$$M_i = D_i/P_i$$

3. Proportions dying in interval:

$$\hat{q}_i = \frac{n_i M_i}{1+(a_i)*M_i*n_i}$$

where n_i = length of age interval = $x_{i+1}-x_i$.

4. Number alive at age x_i:

$$\ell_i = \ell_i-1 - d_i$$

In the program, the radix ℓ_0 is set to be 100,000 for convenience.

5. Number of life table deaths in the interval:

$$d_i = \ell_i \hat{q}_i$$

Note that d_i's are dependent on the radix ℓ_0

6. Number of years lived in interval:

$$L_i = n_i*(\ell_i-d_i) + a_i*n_i*d_i$$

7. Total number of years lived beyond age x_i:

$$T_i = L_i+L_{i+1} + \dots + L_w$$

$$= T_{i+1} + L_i$$

where x_w = last age interval, i.e., 85 and over.

8. Observed expectation of life at age x_i:

$$\hat{e}_i = \frac{T_i}{\ell_i}$$

9. In the final age interval x_w,

$$d_w = \ell_w$$

$$L_w = d_w/M_w$$

$$T_w = L_w$$

$$\hat{e}_w = T_w/\ell_w$$

Reference

Introduction to Stochastic Processes in Biostatistics. Chapter 9: "The life table and its construction," Chin Long Chiang, John Wiley and Sons, Inc., 1968.

Program listing and sample input deck setup

Caution: modifications of certain statements in the program might be required if used on machines other than CDC 6400. The obvious modifications are:

1. The first statement in the program, the Program Statement, might not be required by other machines.

2. The syntax of the read/write statements might be slightly different for different machines.

3. The input and output unit numbers for card reader and printer are probably different in different computer installations.

4. The format statements can be changed if the input data are in a different format than what this program assumes.

5. A format of A10 is used in the program for the input and output of all data titles since a maximum of ten characters can be stored in one word on a CDC machine. A different A format width and corresponding changes in the dimensions of the title arrays are called for on machines with different word structure. For example, IBM 360/370 machines only handle four characters in a word and a format of A4 has to be used when reading in or printing out character data.

```
      PROGRAM ABRIDGE (INPUT, OUTPUT)
C  PROGRAM CARD IS REQUIRED FOR CDC 6400 RUN COMPILER
C
      DIMENSION TITLE (8), TITLE 2(8)
C  LARGER DIMENSIONS SHOULD BE USED FOR ARRAYS TITLE AND TITLE2 FOR
C  MACHINES THAT HANDLE LESS THAN 10 CHARACTERS PER WORD
C
      REAL AI(20), QI(19), E(19)
      INTEGER SL(19), PI(19), CL(19), DI(19), T(20)
      REAL M(19)
C ***********************************************************************
C  CONSTRUCTION OF ABRIDGED LIFE TABLES                                  *
C                                                                        *
C THIS PROGRAM WAS WRITTEN AND DEBUGGED BY PATRICK WONG IN FEB., 1973    *
C BASED ON THE PRELIMINARY WORK OF LINDA WONG.                           *
C                                                                        *
C THIS PROGRAM WAS FURTHER MODIFIED BY CAROL LANGHAUSER                  *
C TO HANDLE W.H.O. DATA IN AUGUST, 1974.                                 *
C                                                                        *
C ASSUME ALL INPUT DATA TO BE BROKEN DOWN INTO THE FOLLOWING 19 AGE      *
C INTERVALS - 0-1,1-5,5-10,10-15,15-20,20-25,25-30,30-35,35-40,40-45,    *
C 45-50,50-55,55-60,60-65,65-70,70-75,75-80,80-85,85+                    *
C                                                                        *
C INPUT DATA.                                                            *
C AI( )=FRACTION OF LAST AGE INTERVAL OF LIFE                            *
C PI( )=MID-YEAR POPULATION IN THE AGE INTERVAL                          *
C DI( )=NUMBER OF DEATHS IN THE AGE INTERVAL                             *
C ***********************************************************************
C
C READ AND PRINT DATA..
  KCT=0
C
C READ A(I) S WITH OPTIONAL TITLE IN COL. 61-80
 500 READ 1, (AI(I), I=1,20), TITLE(1), TITLE(2)
  KCT=KCT+1
  PRINT 102, KCT
  DO 272 I=1,19
  NN=19-I
  IF(AI(NN).NE.0.) GO TO 274
 272 CONTINUE
 274 CONTINUE
     PRINT 4, (AI(I), I=1,NN)
     PRINT 99, (TITLE (I), I=1,2)
     IF((0.8-AI(1)).LE.0.) GO TO 600
C
C READ TITLE FOR POPULATION DATA
     READ 108, (TITLE (I), I=1,8)
     PRINT 109,(TITLE (I), I=1,8)
C
C READ MIDYEAR POPULATIONS IN EACH AGE INTERVALS
     READ 2, (PI(I), I=1,19)
     PRINT 3, (PI(I),I =1,19)
C
C READ TITLE FOR DEATH DATA
     READ 108, (TITLE2(I), I=1,8)
     PRINT 109,(TITLE2(I), I=1,8)
C
C READ NUMBERS OF DEATHS IN EACH AGE INTERVALS
     READ 2, (DI(I), I=1,19)
     PRINT3, (DI(I), I=1,19)
```

```
      PRINT 119,KCT
C
 CHECK DATA DECK..
      LAST=-5
      DO 300 I.1,18
      IF (AI(I).GT.O.) LAST=LAST +5
  300 CONTINUE
      LSAT=-10
      DO 301  I=1,19
      IF(PI(I).GT.O) LSAT=LSAT+5
  301 CONTINUE
      LTSA=-10
      DO 302 I=1,19
      IF(DI(I).GT.O) LTSA=LTSA+5
  302 CONTINUE
      IF(LAST.EQ.LSAT.AND.LSAT.EQ.LTSA) GO TO 305
      PRINT 304, JJ
      GO TO 500
C
  305 J=LAST/5+1
      QI(1)=DI(1)/(PI(1) + (1.-AI(1))*DI(1))
      QI(J+1)=1.
      SL(1)=100000
      D(1)=SL(1)*QI(1)+0.5
      CL(1)=(SL(1)-D(1))+AI(1)*D(1)+0.5
      JLAST=J+1
      DO 306 I=1,JLAST
      F=DI(I)
      G=PI(I)
      M(I)=F/G
  306 CONTINUE
C
C  COMPUTE Q(I) D(I) SL(I) CL(I) AND ROUND OFF Q(I)
      N=4
      DO 307 I=2,J
      QI(I)=N*M(I)/(1.+(1.-AI(I))*M(I)*N)
      TEMP=QI(I)*100000.+0.5
      ITEMP=TEMP
      TEMP=ITEMP
      QI(I)=TEMP/100000.
      SL(I)=SL(I-1)-D(I-1)
      D(I)=SL(I)*QI(I)+0.5
      CL(I)=N*(SL(I)-D(I))+AI(I)*N*D(I)+0.5
      N=5
  307 CONTINUE
      SL(J+1)=SL(J)-D(J)
      D(J+1)=SL(J+1)
      CL(J+1)=SL(J+1)/M(J+1)+0.5
C
C  COMPUTE  E(I)  AND T (I)
      T(J+2)=0
      I=J+1
  308 T(I)=T(I+1)+CL(I)
      F=T(I)
      G=SL(I)
      E(I)=F/G
      I=I-1
      IF(I.GT.O) GO TO 308
C
      PRINT 8, TITLE
```

```
      PRINT 14
      K=1
      KK=0
      DO 77 I=1,J
      PRINT 7,KK,K,PI(I),DI(I),M(I),AI(I),QI(I)
      KK=K
      K=KK+5
      IF(KK.EQ.1) K=KK+4
   77 CONTINUE
      PRINT 5,LAST,PI(J+1),DI(J+1),M(J+1)
      PRINT 100,TITLE
      PRINT 9
      K=1
      KK=0
      DO 78 I=1,J
      PRINT 6,KK,K,QI(I),SL(I),D(I),AI(I),CL(I),T(I),E(I)
      KK=K
      K=KK+5
      IF(KK.EQ.1) K=KK+4
   78 CONTINUE
      PRINT 13,LAST,SL(J+1),D(J+1),CL(J+1),T(J+1),E(J+1)
      GO TO 500
  600 PRINT 200
C
C FORMAT STATEMENTS
C CAUTION - ALL A10 FORMATS SHOULD BE CHANGED TO APPROPRIATE
C WIDTH FOR NON CDC MACHINES
C
C OUTPUT DATA FORMATS
  102 FORMAT(1H6,/,*11XXXXXX8XXXXXX16XXXXXX24XXX DATA DECK NO.*,12,* PRI
     INT OUT XXXXXXXX64XXXXXX72XXXXXX80*)
    4 FORMAT (1X,20F3.2,2A10)
   99 FORMAT (1H+,59X,2A10)
  109 FORMAT (1X,8A10)
    3 FORMAT (1X,10I8)
  119 FORMAT (  * 1XXXXXX8XXXXXX16XXXXXX24XXX END OF DATA DECK NO.*,12,
    1   * XXX56XXXXXX64XXXXXX72XXXXXX80*)
  304 FORMAT(1H1,1X,*INPUT DATA DECK NO.*,13,2X,*IN ERROR*)
    8 FORMAT (1H1,/ * CONSTRUCTION OF ABRIDGED LIFE TABLE FOR *,8A10,//)
   14 FORMAT (3X,*AGE*,13X,*MIDYEAR*,8X,*NUMBER OF *,6X,*DEATHS*,9X,*FRACT
     1ION*,7X,*PROPORTION*/3X,*INTERVAL*,8X,*POPULATION*,5X,*DEATHS*,9X,
     2*RATE*,11X,*OF LAST*,8X,*DYING IN*,/3X,*(IN YEARS)*,6X,*IN INTERVA
     3L*,4X,*IN INTERVAL*,4X,*IN INTERVAL*,4X,*AGE INTERVAL*,3X,*INTERVA
     4L*/3X,*(X(I) TO X(I+1))*,1X,*(X(I),X(I+1))*,1X,*(X(I),X(I+1))*,2X,
     5*(X(I),X(I+1))*,2X,*OF LIFE*,9X,*(X(I),X(I+1))*/19X,*P(I)*,11X,*D(
     6I)*,11X,*M(I)*,11X,*A(I)*,13X,*Q(I)*/)
    7 FORMAT(/3X,I2,1H-,12,7X,110,7X,17,8X,F9.6,9X,F4.2,10X,F9.5)
    5 FORMAT(/3X,I2,1H+,9X,I10,7X,17,8X,F9.6,25X,7H1.00000)
  100 FORMAT(1H1,IX,*ABRIDGED LIFE TABLE FOR*, 8A10/)
    9 FORMAT(/3X,*AGE*,12X,*PROPORTION*,4X,*NUMBER*,2X,*NUMBER*,8X,*FRAC
     1TION*,5X,*NUMBER*,9X,*TOTAL*,11X,*OBSERVED*/3X,*INTERVAL*  ,7X,*DY
     2ING IN*,6X,*LIVING*,2X,*DYING IN*,6X,*OF LAST*,6X,*OF YEARS*,7X,*N
     3UMBER OF*,7X,*EXPECTATION*/3X,*(IN YEARS)*,5X,*INTERVAL*,6X,*AT AG
     4E*,2X,*INTERVAL*,6X,*AGE INTERVAL*,1X,*LIVED IN*,7X,*YEARS LIVED*,
     55X,*OF LIFE AT*/3X,*X(I) TO X(I+1)*,1X,*(X(I),X(I+1))*,1X,*X(I)*,4
     66X,*(X(I),X(I+1))*,1X,*OF LIFE*,6X,*INTERVAL*,7X,*BEYOND AGE X(I)*,
     71X,*AGE X (I)*/67X,*(X(I),X(I+1))*,/20X,*Q(I)*, 8X,*SL(I)*,7X,*D(I)
     8*, 8X,*A(I)*,9X,*CL(I)*,11X,*T(I)*,11X,*E(I)*/)
    6 FORMAT(/3XI2,1H-,I2,8XF10.5,6XI6,2XI6,8XF5.2,8X18,7XI10,6XF7.2)
   13 FORMAT(/3X,I2,1H+,13X,7H1.00000,6X,I6,2X,I6,21X,18,7X,I10,6X,F7.2)
```

```
   200 FORMAT(81H 1XXXXXX8XXXXXX16XXXXXX24XXX END OF ALL DATA DECKS XXXX5
      16XXXXXX64XXXXXX72XXXXXX80 )
C
C   INPUT DATA FORMATS - CAN BE MODIFIED IF NEEDED
     1 FORMAT(   20F3.2,2A10)
     2 FORMAT (  10I8)
   108 FORMAT(   8A10)
       STOP
       END
.10.42.44.55.60.49.50.52.54.54.53.54.52.52.51.51.50.49       CALIF.MALE, 1970
MALE POPULATION, CALIFORNIA, 1970
  173822  663481  975971  998536  930884  872256  726974   611232  575226  592330
  607160  529935  451259  363840  278585  202534  134280    78528   49842
MALE DEATHS, CALIFORNIA, 1970
    3574     607     462     452    1432    1996    1412     1251    1596    2486
    4052    5580    7596    9222   10667   11022   11042     9255    8406
 08.40.45.54.56.49.53.51.52.54.53.53.52.53.52.53.52.51       FEMALE, CAL. 1970
FEMALE POPULATION, CALIFORNIA, 1970
  166661  638717  942146  965145  886495  868710  730640   608157  574773  616220
  638743  553917  481985  406930  342220  281897  207817   132425   92849
FEMALE DEATHS, CALIFORNIA, 1970
    2660     442     261     283     622     706     659      713     992    1628
    2670    3368    4346    5087    6421    8127   10283    10874   14077
.09.41.44.54.59.49.51.52.53.54.53.53.52.52.51.52.51.50         CALIFORNIA, 1970
TOTAL POPULATION, CALIFORNIA, 1970
  340483 1302198 1918117 1963681 1817379 1740966 1457614  1219389 1149999 1208550
 1245903 1083852  933244  770770  620805  484431  342097   210953  142691
TOTAL DEATHS, CALIFORNIA, 1970
    6234    1049     723     735    2054    2702    2071     1964    2588    4114
    6722    8948   11942   14309   17088   19149   21325    20129   22483
```

APPENDIX VI-B

COMPUTER PROGRAM FOR LIFE TABLE CONSTRUCTION WHEN A PARTICULAR CAUSE OF DEATH IS ELIMINATED

Identification

Program name	SPCELT
Author	Patrick Wong Based on original work by Linda Kwok. Program was further modified by Carol Langhauser to handle WHO data.
Department	Biostatistics Program School of Public Health University of California Berkeley, California
Date	February 25, 1973
Environment	Machine = CDC 6400 Operating System = Calidoscope (SCM) version 01.2-A Coding Language = FORTRAN

Purpose

This program constructs abridged life tables when a specific cause is eliminated as a cause of death based on the method developed by Chin Long Chiang.

Input card preparation

All input data are assumed to be broken down into most five year age intervals as follows: 0-1, 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, 60-65, 65-70, 70-75, 75-80, 85+. Details are given below.

1. Number of specific causes of death for the following country in columns 1-2 (maximum is 25).

2. Fractions of year lived by those dying in the interval are punched consecutively in F3.2 format beginning at column 1. Columns 61-80 can be used for optimal population identification.

3. Title for population data in columns 1-80. Standard format:

'TOTAL'/'MALE'/'FEMALE' 'POPULATION', (country), (year)

e.g., TOTAL POPULATION, CANADA, 1970
FEMALE POPULATION, AUSTRIA, 1969

4. Midyear population of that country is each age interval in 10I8 format. Two cards are required.

5. Title for death data in columns 1-80. Standard format:

Col. 1-50 'DEATH FROM ALL CAUSES'
Col. 51-80 same as title for population data described in 3.

e.g., Col. 1 - DEATH FROM ALL CAUSES, Col. 51 - MALE POPULATION, CANADA, 1970.

6. Number of deaths from all causes in each age interval in 10I8 format. Two cards are required.

7. Title for a specific cause of death in columns 1-80. Standard format:

Col. 1-10 'DEATH FROM'
Col. 20-40 (specific cause of death)
Col. 51-80 same as title for population data described in 3.

e.g., Col. 1 - DEATH FROM INFECTIOUS DISEASES, Col. 51 - TOTAL POPULATION, USA, 1970

8. Number of deaths from that specific cause in each age interval in 10I8 format. Two cards are required.

Cards in 7,8 are to be repeated for each specific cause of death for the number of times as specified in card 1.

Cards 1-8 can then be repeated with data from another country. The program is terminated if the number specified in card 1 is greater than 25.

Output

For each country, the following output are produced:

1. Raw input data

r = number of specific causes of death

a_i = fractions of year lived by those dying in each age interval

P_i = midyear population in each age interval

D_i = total number of deaths in each age interval

$D_{i\delta}$ = number of deaths in each age interval from a specific cause, δ; $\delta = 1,\ldots,r$

2. Abridged life tables when each specific cause (R_δ) is eliminated as a cause of death:

$x_i - x_i + 1$ = age interval

$q_{i.\delta}$ = probablity that an individual alive at x_i will die in the interval (x_i, x_{i+1}) if cause R_δ is eliminated as a risk of death

$\ell_{1.\delta}$ = number living at age x_i if cause R_δ is eliminated as a risk of death ($\ell_{0.\delta}$ = 100,000)

$d_{i.\delta}$ = number dying in interval (x_i, x_{i+1}) if cause R_δ is eliminated as a risk of death.

a_i = fraction of year lived by those dying in age interval (x_i, x_{i+1})

$L_{i.\delta}$ = number of years lived in interval if R_δ is eliminated as a risk of death

$T_{i.\delta}$ = total number of years lived beyond age x_i if R_δ is eliminated as a risk of death

$\hat{e}_{i.\delta}$ = observed expectation of life at age x_i if R_δ is eliminated as a risk of death.

Program limiations

1. The current version of the program only handles 19 age intervals broken down as described in the section Input card preparation.

2. The maximum population size and number of deaths in any age interval has to be less than a hundred million. However, the input data format card can easily be changed to handle larger or smaller data fields.

3. The maximum number of specific causes of death to be specified in columns 1-2 of card 1 is currently 25. This number can also be changed to handle a larger limit.

Computational procedure

1. All input data of a country (r, a_i, P_i, D_i, $D_{i\delta}$) are read in.

2. Compute proportions dying in interval (q_i)

$$q_i = \frac{n_i * D_i}{(P_i + (1 - a_i) * n_i * D_i)}$$

where n_i = length of interval = $x_{i+1} - x_i$.

3. Compute the probabilities of dying in interval ($q_{i.\delta}$) when a specific cause (R_δ) is eliminated as a cause of death

$$q_{i.\delta} = 1-(1-q_i)^{(1-D_{i\delta}/D_i)}$$

4. Abridged life table for cause R_δ is then constructed using $q_{i.}$ instead of q_i following the same procedure as described in the program writeup for program ABRIDGE.

Reference

Introduction to Stochastic Processes in Biostatistics. Chapter 11: "Competing risks," Chin Long Chiang, John Wiley and Sons, Inc., 1968.

Program listing and sample deck setup

Caution: See same section in program writeup for ABRIDGE.

```
          PROGRAM SPCELT(INPUT,OUTPUT)
C       PROGRAM CARD IS REQUIRED FOR CDC 6400 RUN COMPILER
          DIMENSION AI(20), QI(19), QQ(19)
          INTEGER PI(19), DI(19), DC(19,25)
C  ***********************************************************************
C  CONSTRUCTION OF ABRIDGED LIFE TABLE WHEN A SPECIFIC CAUSE             *
C  IS ELIMINATED AS A CAUSE OF DEATH..                                   *
C  THIS PROGRAM WAS WRITTEN AND DEBUGGED  BY PATRICK WONG                *
C  BASED ON THE PRELIMINARY WORK OF LINDA WONG.                          *
C                                                                        *
C  THIS PROGRAM WAS FURTHER MODIFIED BY CAROL LANGHAUSER TO HANDLE       *
C  W.H.O. DATA                                                           *
C                                                                        *
C  INPUT DATA..                                                          *
C  ASSUME ALL INPUT DATA TO BE BROKEN DOWN INTO 19 AGE INTERVALS -       *
C  LT 1,1-5,5-10,10-15,15-20,20-25,25-30,30-35,35-40,40-45,45-50,        *
C  50-55,55-60,60-65,65-70,70-75,75-80,80-85,85+                         *
C  JCAUSE=NUMBER OF SPECIFIC CAUSES OF DEATH IN THE DATA DECK CONCERNED. *
C        MAX.=25                                                         *
C  AI( )=FRACTIONS OF LAST AGE INTERVAL OF LIFE.                         *
C  PI( )=MID-YEAR POPULATION IN THE AGE INTERVAL.                        *
C  DI( )=DEATH BY ALL CAUSES IN THE AGE INTERVAL                         *
C  DC( )=DEATH BY A SPECIFIC CAUSE                                       *
C                                                                        *
C  WORKING VARIABLES..                                                   *
C  QI( )=LIFE TABLE PROPORTION OF DEATHS BY ALL CAUSES                   *
C  QQ( )=LIFE TABLE PROPORTION OF DEATHS WHEN A SPECIFIC CAUSE IS        *
C        ELIMINATED AS A CAUSE OF DEATH                                  *
C  ***********************************************************************
C
C  LARGER DIMENSIONS SHOULD BE USED FOR THE FOLLOWING TITLE ARRAYS IN
C  MACHINES THAT HANDLE LESS THAN 10 CHARACTERS PER WORD
      REAL B,TITLE(8),TITLE1(8,25),TITLE2(8)
      K=0
C
C  READ NUMBER OF SPECIFIC CAUSE OF DEATH
  500 READ 13,JCAUSE
      K=K+1
      PRINT 102,K
      PRINT 15,JCAUSE
C
C  PROGRAM TERMINATES IF JCAUSE GT 25
      IF((26-JCAUSE).LE.0) GO TO 600
C
C  READ A(I)S WITH OPTIONAL TITLE IN COL. 61-80
      READ 1,(AI(I), I=1,20),TITLE2(1),TITLE2(2)
C
C  CALCULATE WORKING INDEX..
      DO 202 I=1,19
      NN=19-I
      IF(AI(NN).NE.0.) GO TO 204
  202 CONTINUE
  204 NM=NN+1
      MM=5*NN-5
C
      PRINT 4,(AI(I), I=1,NN)
      PRINT 99,(TITLE2(I), I=1,2)
C
C  READ TITLE FOR POPULATION DATA
      READ 108,(TITLE(I), I=1,8)
```

```
      PRINT 109,(TITLE(I), I=1,8)
C
C  READ MIDYEAR POPULATIONS IN EACH AGE INTERVAL
      READ 2,(PI(I), I=1,19)
      PRINT 3,(PI(I), I=1,19)
C
C  READ TITLE FOR DEATH DATA
      READ  108,(TITLE2(I), I=1,8)
      PRINT 109,(TITLE2(I), I=1,8)
C
C  READ NUMBERS OF DEATHS IN EACH AGE INTERVAL
      READ  2,(DI(I), I=1,19)
      PRINT 3,(DI(I), I=1,19)
C
   DO 170 J=1,JCAUSE
C
C  READ TITLE FOR A SPECIFIC CAUSE OF DEATH
      READ 108,B,(TITLE1(I,J), I=1,7)
      PRINT 109,B,(TITLE1(I,J), I=1,7)
C
C  READ NUMBERS OF DEATHS FROM THAT SPECIFIC CAUSE OF DEATH IN EACH AGE
C  INTERVAL
      READ 2,(DC(I,J), I=1,19)
      PRINT 3,(DC(I,J), I=1,19)
  170 CONTINUE
      PRINT 119,K
C
C  COMPUTE QI( )..
      DO 112 I=1,NN
      N=5
      IF(I.EQ.1) N=1
      IF(I.EQ.2) N=4
      QI(I)=N*DI(I)/(PI(I)+(1.-AI(I))*NDI(I))
  112 CONTINUE
C
C  COMPUTE THE PROBABILITY OF DYING QQ ( ) WHEN A SPECIFIC CAUSE
C  IS ELIMINATED AS A CAUSE OF DEATH
      DO 700 J=1,JCAUSE
      DO 110 I=1,NN
      F=DC(I,J)
      G=DI(I)
      EE=1.-F/G
      QQ(I)=1.-(1.-QI(I))**EE
      TEMP=QQ(I)*100000.+0.5
      ITEMP = TEMP
      TEMP=ITEMP
      QQ(I)=TEMP/100000.
  110 CONTINUE
      QQ(NM)=1.
C
      F=DI(NM)-DC(NM,J)
      G=PI(NM)
      WM=F/G
      PRINT 7,(TITLE(I), I=1,5)
      PRINT 10,(TITLE1(I,J), I=1,3)
      PRINT 9
      CALL ABRLIF(AI,QQ,WM,NN)
  700 CONTINUE
      GO TO 500
  600 PRINT 100
```

```
C
C  FORMAT STATEMENTS
C  CAUTION - ALL A10 FORMATS SHOULD BE CHANGED TO APPROPRIATE WIDTH
C  FOR NON CDC MACHINES
C
C  INPUT DATA FORMATS - CAN BE CHANGED IF NEEDED
      1 FORMAT(   20F3.2,2A10)
      2 FORMAT(   10I8)
     13 FORMAT(I2)
    108 FORMAT(   8A10)
C
C  OUTPUT DATA FORMATS
C
      3 FORMAT(1X,10I8)
      4 FORMAT(1X,20F3.2,2A10)
      7 FORMAT( 1H1,//,25H ABRIDGED LIFE TABLE FOR ,5A10)
      9 FORMAT(/3X,*AGE*,12X,*PROPORTION*,4X,*NUMBER*,2X,*NUMBER*,8X,*FRAC
       1TION*,5X,*NUMBER*,9X,*TOTAL*,11X,*OBSERVED*/3X,*INTERVAL*   ,7X,*DY
       2ING IN*,6X,*LIVING*,2X,*DYING IN *,6X,*OF LAST*,6X,*OF YEARS*,7X,*N
       3UMBER OF*,7X,*EXPECTATION*/3X,*(IN YEARS)*,5X,*INTERVAL*,6X,*AT AG
       4E*,2X,*INTERVAL*,6X,*AGE INTERVAL*,1X,*LIVED IN*,7X,*YEARS LIVED*,
       55X,*OF LIFE AT*/3X,*X(I) TO X(I+1)*,1X,*(X(I),X(I+1))*,1X,*X(I)*,4
       6X,*(X(I),X(I+1))*,1X,*OF LIFE*,6X,*INTERVAL*,7X,*BEYOND AGE X(I)*,
       71X,*AGE X(I)*/67X,*(X(I),X(I+1))*,/20X,*Q(I.1)*6X,*SL(I)*,7X,*D(I)
       8*, 8X,*A(I)*,9X,*CL(I)*,11X,*T(I)*,11X,*E(I.1)*/)
     10 FORMAT(5H WHEN 1X,R9, 2A10/35H IS ELIMINATED AS A CAUSE OF DEATH )
     15 FORMAT(1X, I2)
     99 FORMAT(1H+,59X,2A10)
    100 FORMAT (81H 1 XXXXXX8XXXXXX16XXXXXX24XXX END OF ALL DATA DECKS XXXX5
       16XXXXXX64XXXXXX72XXXXXX80 )
    102 FORMAT(1H7,/,*11XXXXXX8XXXXXX16XXXXXX24XXX DATA DECK NO.*,12,* PRI
       1NT OUT XXXXXXXX64XXXXXX72XXXXXX80*)
    109 FORMAT(1X,8A10)
    119 FORMAT(  * 1XXXXXX8XXXXXX16XXXXXX24XXX END OF DATA DECK No.*,12,
       1  * XXX56XXXXXX64XXXXXX72XXXXXX80*)
        STOP
        END
        SUBROUTINE ABRLIF(AI,QI,WM,J)
C
C  CONSTRUCTION OF ABRIDGED LIFE TABLE..
        DIMENSION AI(20),QI(19),E(19)
        INTEGER SL(19),CL(19),D(19),T(20)
C
C  COMPUTE D( ),SL( ),CL( )..
        SL(1)=100000
        D(1)=SL(1)*QI(1)+0.5
        CL(1)=(SL(1)-D(1))+AI(1)*D(1)+0.5
        N=4
        DO 307 I=2,J
        SL(I)=SL(I-1)-D(I-1)
        D(I)=SL(I)*QI(I)+0.5
        CL(I)=N*(SL(I)-D(I)+AI(I)*N*D(I)+0.5
        N=5
    307 CONTINUE
        SL(J+1)=SL(J)-D(J)
        D(J+1)=SL(J+1)
        CL(J+1)=SL(J+1)/WM+0.5
C
C  COMPUTE E( ) AND T( )..
        T(J+2)=0
```

```
      I=J+1
  308 T(I)=T(I+1)+CL(I)
      F=T(I)
      G=SL(I)
      E(I)=F/G
      I=I-1
      IF(I.GT.0) GO TO 308
      K=1
      KK=0
      DO 78 I=1,J
      PRINT 6,KK,K,QI(I),SL(I),D(I),AI(I),CL(I),T(I),E(I)
      KK=K
      K=KK+5
      IF(KK.EQ.1) K=KK+4
   78 CONTINUE
      LAST=(J-1)*5
      PRINT 13,LAST,SL(J+1),D(J+1),CL(J+1),T(J+1),E(J+1)
    6 FORMAT(/3X12,1H-,12,8XF10.5,6XI6,4XI6,6XF5.2,8XI8,7XI10,6XF7.2)
   13 FORMAT(/3X,I2,1H+,13X,7H1.00000,6X,I6,4X,I6,19X,I8,7X,I10,6X,F7.2)
      RETURN
      END
12
 .11.41.45.54.57.48.50.52.53.54.53.54.54.53.53.52.52.52      CANADA, TOTAL, 1968
TOTAL POPULATION, CANADA, 1970
  365000 1503300 2301400 2297100 2068200 1851800 1508000 1289800 1276900 1295100
 1222800 1045900  930100  745500  585000  448100  329600  194300  119100
DEATHS FROM ALL CAUSES                                 TOTAL POPULATION, CANADA, 1970
    7001    1263    1102     988    1948    2105    1564    1636    2217    3566
    5327    7362   10419   13060   16054   17989   19915   19773   22653
DEATHS FROM CAUSE 1                                    TOTAL POPULATION, CANADA, 1970
     190     102      46      27      20      14      12      15      27      35
      61      72      92     126      98     112     114      92      93
DEATHS FROM CAUSE 2                                    TOTAL POPULATION, CANADA, 1970
      25     125     185     147     136     159     167     234     455     895
    1515    2239    3066    3678    4152    4171    3832    2969    2332
DEATHS FROM CAUSE 3                                    TOTAL POPULATION, CANADA, 1970
       0       0       0       0       0       3       3      15      31      48
      86     159     204     308     362     411     400     337     272
DEATHS FROM CAUSE 4                                    TOTAL POPULATION, CANADA, 1970
       0       0       1       1       1       2       3      11      52     117
     256     439     712     822     876     800     514     293     162
DEATHS FROM CAUSE 5                                    TOTAL POPULATION, CANADA, 1970
      28      15       9      26      48      71     116     214     456    1024
    1889    3026    4772    6496    8579   10154   12137   12838   15478
DEATHS FROM CAUSE 6                                    TOTAL POPULATION, CANADA, 1970
      20       9       7      21      31      51      73     157     355     852
    1637    2644    4159    5602    7171    8177    9301    9639   11712
DEATHS FROM CAUSE 7                                    TOTAL POPULATION, CANADA, 1970
     735     171      58      53      55      48      55      53      90     132
     228     352     545     768    1016    1322    1436    1523    2058
DEATHS FROM CAUSE 8                                    TOTAL POPULATION, CANADA, 1970
    5199     183      77      50      46      36      22      24      12      20
      16      36      28      11      14      10       2       5       3
DEATHS FROM CAUSE 9                                    TOTAL POPULATION, CANADA, 1970
     186      13       2       4       5      15      10      12      17      26
      27      31      41      47      55      68      80     141     336
DEATHS FROM CAUSE 10                                   TOTAL POPULATION, CANADA, 1970
     328     490     579     528    1294    1272     758     594     545     610
     623     543     543     466     405     361     408     417     604
```

DEATHS FROM CAUSE 11 TOTAL POPULATION, CANADA, 1970

20	168	342	260	881	864	425	312	259	265
245	231	232	204	186	143	124	85	62	

DEATHS FROM CAUSE 12 TOTAL POPULATION, CANADA, 1970

0	0	0	17	145	256	207	195	231	250
258	221	211	150	117	70	42	32	11	

REFERENCES

Aalen, O. [1976]. Nonparametric inference in connection with multiple decrement models. Scand. J. Stat., 3, 15-27.

Adlakha, A. [1972]. Model life tables: an empirical test of their applicability to less developed countries. Demography, 9, 589-601.

Armitage, P. [1951]. The statistical theory of bacterial populations subject to mutation. J. Royal Statist. Soc., B14, 1-33.

Armitage, P. [1959]. The comparison of survival curves. J. Royal Statist. Soc., A122, 279-300.

Arriaga, E. E. [1968]. New Life Tables for Latin American Populations in the Ninteenth and Twentieth Centuries. Monograph Series No. 3. Institute of International Studies, Univ. of Ca., Berkeley.

Bailar, B. A. [1976]. Some sources of error and their effect on census statistics. Demography, 13, 273-286.

Balakrishnan, T. R., J. D. Allingham, and J. F. Kanter [1970]. Analysis of oral contraceptive use through multiple decrement life table techniques. Demography, 7, 459-65.

Barclay, G. W. [1958]. Techniques of Population Analysis. Wiley, New York.

Barlow, R. E. and F. Proschan [1965]. Mathematical Theory of Reliability. Wiley, New York.

Bartholomew, D. J. [1963]. The sampling distribution of an estimate arising in life testing. Technometrics, 5, 361-74.

Bartlett, M. S. [1956]. An Introduction to Stochastic Processes. Cambridge University Press.

Bayo, F. [1968]. United States Life Tables by Causes of Death, 1959-61. Health Service Publication No. 1252, 1.

Berkson, J. and L. Elveback [1960]. Competing exponential risks, with particular reference to smoking and lung cancer. J. Amer. Statist. Assoc., 55, 415-428.

Berkson, J. and R. P. Gage [1950]. Calculation of survival rates for cancer. Proc. Staff Meetings Mayo Clinic 25, 270-86.

Berkson J. and R. P. Gage [1952]. Survival curve for cancer patients following treatment. J. Amer. Statist. Assoc., 47, 501-515.

Berman, S. M. [1963]. A note on extreme values, competing risks and semi-Markov processes. Ann. Math. Statist., 34, 1104-1106.

Berry, G. [1970]. Parametric analysis of disease incidences in multiway tables. Biometrics, 26, 572-579.

Birbaum, Z. W., J. D. Esary, and S. C. Saunders [1961]. Multi-component systems and structures and their reliability. Technometrics, 3, 55-77.

Boag, J. W. [1949]. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. J. Royal Statist. Soc., 11, 15-53.

Broardman, T. J. and P. J. Kendell [1970]. Estimation in compound exponential failure models. Technometrics, 12, 391-900.

Bolander, A. M. [1971]. A Comparative Study of Mortality by Cause in Four Nordic Countries, 1966-1968, with Special Reference to Male Excess Mortality. Statistical Reports 1971, 9. National Central Bureau of Statistics, Stockholm.

Breslow, N. [1974]. Covariance analysis of censored survival data. Biometrics, 30, 89-99.

Breslow, N. and J. Crowley [1974]. A large sample study of the life table and product limit estimates under random censorship. Ann. Stat., 2, 437-453.

Brindley, Jr., E. C. and W. A.Thompson, Jr. [1972]. Dependence and aging aspects of multivariate survival. J. Am. Stat. Assoc., 67, 822-830.

Brockmayer, E., H. L. Halstrom, and A. Jensen [1948]. The Life Work of A. K. Erlang. Copenhagen Telephone Co., Copenhagen

Brownlee, J. [1913]. The relationship between 'corrected' death rates and life table death rates. Journal of Hygiene, XIII, 178-190.

Brownlee, J. [1922]. The use of death rates as a measure of hygienic conditions. Medical Research Council Special Report Series, 60. H. M. Stationery Office, London.

Campbell, H. [1965]. Changes in mortality trends in England and Wales, 1931-1961. Vital and Health Statistics, Series 3, No. 3.

Canadian Department of National Health and Welfare and the Dominion Bureau of Statistics [1960]. Illness and Health Care in Canada, Canadian Sickness Survey, 1950-1951. The Queen's Printer and Controller of Stationery, Ottawa.

Chase, C. C. [1967]. International comparison of perinatal and infant mortality. Vital and Health Statistics, Series 3, No. 6, 1-97. U. S. Dept. of H. E. W., Public Health Services.

Chiang, C. L. [1960]. A stochastic study of the life table and its applications: 1. Probability distributions of the biometric functions. Biometrics, 16, 618-635.

Chiang, C. L. [1960b]. A stochastic study of the life table and its applications: II. Sample variance of the observed expectation of life and other biometric functions. Human Biol., 32, 221-238.

Chiang, C. L. [1961a]. A stochastic study of the life table and its applications: III. The follow-up study with the consideration of competing risks, 17, 57-78.

Chiang, C. L. [1961b]. Standard error of the age-adjusted death ratio. Vital Statistics, Special Reports Selected Studies, 47, 275-285. National Center for Health Statistics.

Chiang, C. L. [1961c]. On the probability of death from specific causes in the presence of competing risks. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press.

Chiang, C. L. [1965]. A stochastic model of competing risks of illness and competing risks of death. Stochastic Models in Medicine and Biology, J. Gurland, Ed. University of Wisconsin Press, Madison, pp 323-354.

Chiang, C. L. [1966]. On the expectation of the reciprocal of a random variable. The Amer. Statist., 20, 28.

Chiang, C. L. [1967]. Variance and covariance of life table functions estimated from a sample of deaths. Vital and Health Statistics, Ser. 2, No. 20, 1-8. National Center for Health Statistics.

Chiang, C. L. [1968]. Introduction to Stochastic Processes in Biostatistics. Wiley, New York.

Chiang, C. L. [1970]. Competing risks and conditional probabilities. Biometrics, 26, 767-76.

Chiang, C. L. [1972]. On constructing current life tables. J. Am. Stat. Assoc., 67, 538-541.

Chiang, C. L., F. D. Norris, F. Olsen, P. W. Shipley, and H. E. Supplee [1961]. Determination of the fraction of last year of life and its variation by age, race, and sex. (An unpublished manuscript presented at the Annual Meeting of the American Public Health Assoc., Detroit, Michigan, November 1961.)

Chu, G., C. Langhauser, J. Fortman, P. Schoenfeld, and F. Belzer [1973]. A survival analysis of patients undergoing dialysis or renal transplantation. Trans. Amer. Soc. Artificial Internal Organs, 19, 126-129.

Coale, A. J. and P. Demeny [1966]. Regional Life Tables and Stable Populations. Princeton University Press, Princeton, N.J.

Cochran, W. G. [1963]. Sampling Techniques (2nd ed.). John Wiley and Sons, New York.

Cochran, W. G. [1968]. The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics, 24, 295-313.

Cohen, J. [1965]. Routine morbidity statistics as a tool for defining public health priorities. Isr. J. Med. Sci., 1, 457.

Cohen, J. E. [1975]. Livelihood benefits of small improvements in the life table. Health Serv. Res., 1, 82-96.

Cornfield, J. [1951]. A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast and cervix. J. Natl. Cancer Inst., 11, 1269-1275.

Cornfield, J. [1957]. The estimation of the probability of developing a disease in the presence of competing risks. J. Amer. Public Health Assoc., 47, 601-607.

Cox, D. R. [1959]. The analysis of exponentially distributed life-times with two types of failure. J. R. Statist. Soc. B, 411-21.

Cox, D. R. [1972]. Regression models and life-tables. J. R. Stat. Soc., Series B, 34, 187-220.

Cramer, H. [1946]. Mathematical Methods of Statistics. Princeton University Press.

Cutler, S. J. and F. Ederer [1958]. Maximum utilization of the life table method in analyzing survival. J. Chronicle Diseases, 8, 699-712.

Dad Gupta, A. S. [1954]. Accuracy index of census age distributions. Proceedings of the World Population Conference, 4. United Nations, New York.

Das Gupta, A., S. Chotechanapibal, T. Chalothorn, and W. Siripak [1965]. Population Perspective of Thailand. Sankhya, Series B, 27, 1-46.

David, H. A. [1970]. On Chiang's proportionality assumption in the theory of competing risks. Biometrics, 26, 336-9.

David, H. A. [1974]. Parametric approaches to the theory of competing risks. Reliability and Biometry (Proschan and Serfling Eds.) SIAM, 275-290. Philadelphia.

Demeny, P. [1965]. Estimation of vital rates for populations in the process of destabilization. Demography, 2, 516-530.

Densen, P. M. [1950]. Long-time follow-up in morbidity studies: the definition of the group to be followed. Human Biol.. 22, 233.

Derksen, J. B. D. [1948]. The calculation of mortality rates in the construction of life tables. A mathematical statistical study. Population Studies, 1, 457.

Devroede, G. and W. F. Taylor [1976]. On calculating cancer risk and survival of ulcerative colitis patients with the life table method. Gastroenterology, 71, 505-9.

Doering, C. R. and A. L. Forbes [1939]. Adjusted death rates. Proceedings of the National Academy of Science, 25, 461-467.

Doll, R. and P. Cook [1967]. Summarizing indices for comparison of cancer incidence data. Internat. J. Cancer, 2, 269-279.

Dorn, H. [1950]. Methods of analysis for follow-up studies. Human Biol., 22, 238-248.

Drolette, M. E. [1975]. The effect of incomplete follow-up. Biometrics, 31, 135-144.

Dublin, L. I. and A. J. Lotka [1937]. Uses of the life table in vital statistics. Journal of the American Public Health Assoc.

Dublin, L. I., A. J. Lotka, and M. Spiegelman [1949]. Length of Life: A Study of the Life Table. Ronald Press, New York.

Duda, R. and G. Duda [1967]. Life tables of the population of the city of Jassy. Rev. Med. Chir. Soc. Med. Nat. Iasi., 73, 673-80.

Du Pasquier, L. G. [1913]. Mathematische theorie der invaliditatsversicherung. Mitteil. Verein. Schweizer Versicherungsmath, 8, 1-153.

Du Pasquier, L. G. [1912]. Mathematische theorie der invaliditatsversicherung. Mitteil. Verein. Schweizer Versicherungsmath, 7, 1-7.

Eckler, A. R. and W. N. Hurwitz [1958]. Response variance and bias in censuses and surveys. Bulletin of the International Statistical Institute 36, Part 2, 12-35. Stockholm.

Efron, B. [1965]. The two sample problem with censored data. Proc. 5th Berkeley Symp. Math. Statist. Probl. IV, 831-53.

Elandt-Johnson, R. C. [1973]. Age-at-onset distribution in chronic diseases. A life table approach to analysis of family data. J. Chronic Dis., 26, 529-45.

El-Badry, M. A. [1969]. Higher female than male mortality in some countries of South Asia: A digest. J. Amer. Statist. Assoc., 64, 1234-1244.

Elveback, L. [1958]. Estimation of survivorship in chronic disease: The actuarial method. J. Amer. Statist. Assoc., 53, 420-440.

Elveback, L. R. [1966]. Discussion of "Indices of mortality and tests of their statistical significance." Hum. Biol., 38, 322-324.

Epstein, B. and M. Sobel [1953]. Life testing. J. Amer. Stat. Assoc., 48, 486-502.

Fabia, J. and M. Drolette [1970]. Life tables up to age 10 for mongols with and without congenital heart defect. J. Ment. Defic. Res., 14, 235-42.

Faulkner, J. E. and R. B. McHugh [1972]. Bias in observable cancer age and life-time of mice subject to spontaneous mammary carcinomas. Biometrics, 28, 489-98.

Fisher, L. and P. Kanarek [1974]. Presenting censored survival data when censoring and survival times may not be independent. Reliability and Biometry (Proschan and Serfling, Eds.) SIAM, 291-302. Philadelphia.

Fix, E. and J. Neyman [1951]. A simple stochastic model of recovery, relapse, death, and loss of patients. Human Biol, 23, 205-241.

Fleiss, J. L. [1973]. Statistical Methods for Rates and Proportions. Wiley, New York.

Fleiss, J. L., D. L. Dunner, F. Stallone, and R. R. Fieve [1976]. The life table. A method for analyzing longitudinal studies. Arch. Gen. Psychiatry, 33, 107-12.

Flinn, M. W. [1970]. British Population Growth, 1700-1850. Macmillan, London.

Fréchet, M. [1947]. Sur les expressions analytique de la mortalité variables pour la vie entiere. Journal de la Société de Statistique de Paris, 88, 261-285.

Frost, W. H. [1933]. Risk of persons in familiar contact with pulmonary tuberculosis. Amer. J. Public Health, 23, 426-432.

Frumkin, G. [1954]. Estimation de la qualité des statistiques démographiques. Proceedings of the World Population Conference, 4. United Nations, New York.

Fukushima, N. Z. [1974]. A study on the method of constructing abridged life tables and the interpolation for individual years of life. J. Med. Sci., 20, 11-48.

Gail, M. [1975]. A review and critique of some models used in competing risk analysis. Biometrics, 31, 209-22.

Gehan, E. A. [1969]. Estimating survival functions from the life table. J. Chronic Dis., 21, 629-44.

George, L. and A. P. Norman [1971]. Life tables for cystic fibrosis. Arch. Dis. Child, 46, 139-43.

Gershenson, H. [1961]. Measurement of Mortality. Society of Actuaries, Chicago.

Gille, H. [1949]. The demographic history of the Northern European coutries in the eighteenth century. Population Studies, 3, 3-65.

Glass, D. V. and D. E. C. Eversley (eds.) [1965]. Population in History. Edward Arnold, London.

Glover, J. W. [1921]. United States Life Tables, 1890, 1901, 1910, and 1901-1910. Bureau of the Census, Washington.

Gompertz, B. [1825]. On the nature of the function expressive of the law of human mortality. Phil. Trans. Royal Soc., 155, 513-583. London

Graunt, J. [1662]. Natural and Political Observations Made upon the Bills of Mortality. Reprinted by the John Hopkins Press, Baltimore, 1939.

Greenwood, M., et al [1922]. Discussion on the value of life-tables in statistical research. Journal of the Royal Statistical Society, 85, 537-560.

Greenwood, M. [1925]. A report on the natural duration of cancer. Rep. Public Health Med. Sub., 33, 1-26.

Greenwood, M. [1926]. A report on the natural duration of cancer. Reports on Public Health and Medical Subjects, 33, 1-26. His Majesty's Stationery Office.

Greenwood, M. [1942]. Medical statistics from Graunt to Farr. Biometrika, 32, 101-127, 203-225.

Grenander, U. [1956]. On the theory of mortality measurement. Skandinavisk Aktuarietidskrift, 39, 1-55.

Greville, T. N. E. [1943]. Short methods of constructing abridged life tables. Record Amer. Inst. Actuaries, 32, 29-43.

Greville, T. N. E. [1946]. United States Life Tables and Actuarial Tables, 1939-41. National Office of Vital Statistics.

Greville, T. N. E. [1948]. Mortality tables analyzed by cause of death. Record Amer. Inst. Actuaries, 37, 283-294.

Greville, T. N. E. (ed.) [1972]. Population Dynamics. Academic Press, New York.

Gupta, R. B. [1975]. Impact of partial elimination of specific causes of death on the human survivorship: An Indian situation. Indian J. Pub. Health, 19, 11-33.

Gupta, R. B. and G. R. Rao [1973]. Effect of elimination of different causes of death on expectation of life: Bombay, 1960-61. Indian J. Med. Res., 61, 950-61.

Gurunanjappa, B. S. [1969]. Life tables for Alaska natives. Public Health Rep., 84, 65-9.

Haenszel, W. [1950]. A standardized rate for mortality defined in units of lost years of life. J. Amer. Public Health Assoc., 40, 17-26.

Haenszel, W. (ed.) [1966]. Epidemiological approaches to the study of cancer and other chronic diseases. National Cancer Institute Monograph No. 19.

Hajnal. J. [1950]. Rates of dissolution of mairrages in England and Wales, 1938-39. Papers of the Royal Commission on Population, 2, 178-187. H. M. Stationery Office, London.

Hakama, M. [1970]. Age adjustment of incidence rates in cancer epidemiology. Acta. Path. Microb. Scand., Suppl. 213.

Hakulinan, R. and L. Teppo [1976]. The increase in working years due to elimination of cancer as a cause of death. Int. J. Cancer, 17, 429-35.

Haldane, J. B. S. [1927]. A mathematical theory of natural and artificial selection, Part V: Selection and Mutation. Proc. Royal Soc., 23, 838-844. Edinburgh.

Halley, E. [1693]. An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslau. Philos. Trans. Royal Soc., 17, 596-610. London

Hanse, M. H., W. N. Hurwitz, and M. A. Bershad [1961]. Measurement errors in censuses and surveys. Bulletin of the International Statistical Institute 38, Part 2, 359-374. Tokyo.

Harris, T. E., P. Meier, and J. W. Tukey [1950]. Timing of the distribution of events between observations. Human Biol., 22, 249-270,

Heasman, M. A. and L. Lipworth [1966]. Studies on Medical and Population Subjects No. 20. Accuracy of Certification of Cause of Death.General Register Office of England and Wales.

Hemminki, E., K. Hemmink, T. Hakulinen, and M. Hakana [1976]. Increase in years of life after eliminating causes of death: significance for health priorities. Scand. J. Soc. Med., 4, 1-6.

Hemminki, E., K. Hemminki, and T. Hakulinen [1974]. Life expectancy of the population and the influence of most occuring causes of death. Duodecim, 90, 1167-79.

Herman, R. J. and Patell, Rusi, K. N. [1971]. Maximum likelihood estimation for multi-risk model. Technometrics, 13, 385-96

Hill, A. B. [1966]. Principles of Medical Statistics. Oxford Univ. Press, New York.

Hoel, D. G. [1972]. A representation of mortality data by competing risks. Biometrics, 28, 475-88.

Hoem, J. M. [1970]. A probabilistic approach to nuptiality. Biométrie-Praximétrie, 11, 3-19.

Hoem, J. M. [1970]. Probabilistic fertility models of the life table type. Theor. Popul. Biol., 1, 12-38.

Hoem, J. M. [1971]. On the interpretation of certain vital rates as averages of underlying forces of transition. Theor. Popul. Biol., 2, 454-458.

Hoem, J. M. [1975]. The construction of increment-decrement life tables: a comment on articles by R. Schoen and V. Nelson. Demography, 12, 661-4.

Hogg, R. V. and A. T. Craig [1965]. Introduction to Mathematical Statistics (2nd ed.). Macmillan, New York.

Holford, T. R. [1976]. Life tables with concomitant information. Biometrics, 32, 587-97.

Hyrenius, H. and J. Quist [1970]. Life table technique for the working ages. Demography, 7, 393-9.

Irwin, A. C. [1976]. Life tables as "predictors" of average longevity. Can. Med. Assoc. J., 114, 539-41.

Irwin, J. O. [1949]. The standard error of an estimate of expectation of life. J. Hygiene, 47, 188-189.

Jacobson, P. H. [1964]. Cohort survival for generations since 1840. Milbank Memorial Fund Quarterly.

James, G., R. E. Patton, and S. Heslin. [1955]. Accuracy of cause-of-death statements on death certificates Public Health Reports, 7, 39-51. United States Department of Health, Education, and Welfare.

Jordan, C. W., Jr. [1967]. Life Contingencies (2nd ed.). Society of Actuaries, Chicago.

Kalton, G. [1968]. Standardization: A technique to control for extraneous variables. Appl. Statist., 17, 118-136.

Kaplan, E. L. and P. Meier [1958]. Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc., 53, 457-481.

Karn, M. N. [1931]. An inquiry into various death rates and the comparative influence of certain diseases on the duration of life. Ann. Eugen., 4, 279.

Karn, M. N. [1933]. A further study of methods of constructing life tables when certain causes of death are eliminated. Biometrika, 25, 91-101.

Kendall, M. G. [1946]. The Advanced Theory of Statistics, Vol 1. Charles Griffin, London.

Kendall, M. G. and A. Stuart [1961]. The Advanced Theory of Statistics, Vol. 2. Griffin, London.

Keyfitz, N. [1966]. Sampling variance of standardized mortality rates. Hum. Biol., 38, 309-317.

Keyfitz, N. [1966]. A life table that agrees with the data. J. of Amer. Statist. Assoc., 61, 305-312.

Keyfitz, N. [1968]. Introduction to the Mathematics of Population. Addison-Wesley, Reading, Mass.

Keyfitz, N. [1970]. Finding probabilities from observed rates, or how to make a life table. The American Statistician, **24**, 28-33.

Keyfitz, N. and W. Flieger. [1968]. World Population: An Analysis of Vital Data. Univ. of Chicago Press, Chicago.

Keyfitz, N. and Frauenthal [1975]. An improved life table method. Biometrics, **31**, 889-99.

Keyfitz, N., S. H. Preston, and R. Schoen [1972]. Inferring probabilities from rates: extension to multiple decrement. Skand. Aktuarietidskr, 1-13.

Kilpatrick, S. J. [1963]. Mortality comparisons in socio-economic groups. Appl. Statist., **12**, 65-86.

Kimball, A. W. [1958]. Disease incidence estimation in populations subject to multiple causes of death. Bull. Internat. Statist. Inst., **36**, 193-204.

Kimball, A. W. [1969]. Models for the estimation of competing risks from grouped data. Biometrics, **25**, 329-37.

King, G. [1914]. On a short method of constructing an abridged mortality table. J. Inst. Actuaries, **48**, 294-303.

Kitagawa, E. M. [1955]. Components of a difference between two rates. J. Amer. Statist. Assoc., **50**, 1168-1194.

Kitagawa, E. M. [1964]. Standardized comparisons in population research. Demography, **1**, 296-315.

Kitagawa, E. M. [1966]. Theoretical considerations in the selection of a mortality index, and some empirical comparisons. Hum. Biol., **38**, 293-308.

Kloetzel, K. and J. C. Dias [1968]. Mortality in Chagas' disease: life-table for the period 1949-1967 in an unselected population. Rev. Inst. Med. Trop. Sao Paulo, **10**, 5-8.

Krall, J. M. and J. C. Hickman [1971]. Adjusting multiple-decrement tables. Transactions of the Society of Actuaries, **22**, 163-179.

Krall, J. M., V. A. Uthoff and J. B. Harley [1975]. A step-up procedure for selecting variables associated with survival. Biometrics, **31**, 49-57.

Krishnan, P. [1971]. Divorce tables for females in the United States: 1960. Journal of Marriage and the Family, **33**, 318-320.

Kruegel, D. L. and J. M. Peck [1974]. Maryland abridged life tables by color and sex: 1969-1971. Md. State Med. J., **23**, 49-55.

Kuzma, J. W. [1967]. A comparison of two life table methods. Biometrics, **23**, 51-64.

Lehmann, E. [1959]. Testing Statistical Hypotheses. Wiley, New York.

Lew, E. A. and F. Seltzer [1970]. Uses of the life table in public health. Milbank Mem. Fund Q., 48, Suppl., 15-37.

Linder, F. E. and R. D. Grove [1959]. Techniques of vital statistics, Reprint of Chapters I-IV. Vital Statistics Rates in the United States, 1900-1940. U. S. Government Printing Office, Washington, D. C.

Littell, A. S. [1952]. Estimation of the T-year survival rate from follow-up studies over a limited period of time. Human Biol., 24, 87-116.

Lopez, A. [1961]. Problems in Stable Population Theory. Office of Population Research, Princeton University, Princeton, New Jersey.

Makeham, W. M. [1860]. On the law of mortality and the construction of annuity tables. J. Inst. Actuaries, 8.

Makeham, W. M. [1974]. On an application of the theory of the composition of decremental forces. J. Inst. Actuaries, 18, 317-322.

Mantel, N. [1974]. Branching experiments: a generalized application of the life-table method. Proceedings of SIMS Conference on Epidemiology (D. Ludwig, K. L. Cooke, Ed.), 69-74. Society for Industrial and Applied Mathematics, Philadelphia.

Mantel, N. and J. C. Bailar, III [1970]. Model I or Model II in competing risk analysis? Biometrics, 26, 861-3.

Mantel, N. and W. Haenszel [1959]. Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute, 22.

Mantel, N. and C. R. Stark [1968]. Computation of indirect-adjusted rates in the presence of confounding. Biometrics, 24, 997-1005.

Marshall, A. W. and I. Olkin [1967]. A multivariate exponential distribution. J. Amer. Statist. Soc., 62, 30-44.

Mattila, A. and K. Rosendahl [1969]. Factors affecting life expectancy. Acta. Socio-Med. Scand. 1, Suppl. 1, 51.

McCann, J. C. [1976]. A technique for estimating life expectancy with crude vital rates. Demography, 13, 259-72.

Medsger, T. A., Jr., A. T. Masi, G. P. Rodnan, T. G. Benedek, and H. Robinson [1971]. Survival with systemic sclerosis (scleroderma). A life-table analysis of clinical and demographic factors in 309 patients. Ann. Intern. Med., 75, 369-76.

Medsger, T. A., Jr. and A. T. Masi [1973]. Survival with scleroderma. II. A life-table analysis of clinical and demographic factors in 358 male U.S. veteran patients. J. Chronic Dis., 26, 647-60.

Medsger, T. A., Jr., H. Robinson, and A. T. Masi [1971]. Factors affecting survivorship in polymyositis. A life-table study of 124 patients. Arthritis Rheum., 14, 249-58.

Merrell, M. and L. E. Shulman [1955]. Determination of prognosis in chronic disease, illustrated by systemic lupus erythematseus. J. Chron. Dis., 1, 12-32.

Miller, R. S. and J. L. Thomas [1958]. The effect of larval crowding and body size on the longevity of adult Drosophila Melangaster. Ecology, 39, 118-125.

Mitra, S. [1973]. On the efficiency of the estimates of life table functions. Demography, 10, 421-6.

Mode, C. J. and G. S. Littman [1975]. Methodological Issues Underlying Multiple Decrement Life Table Analysis. Institute for Popul. Stud., Drexel Univ., Philadelphia.

Mode, C. H. [1976]. A large sample investigation of a multiple decrement life table estimator. Mathematical Biosciences, 32, 111-23.

Moeschberger, M. L. and H. A. David [1971]. Life tests under competing causes of failure and the theory of competing risks. Biometrics, 27, 909-33.

Moeschberger, M. A. [1974]. Life tests under competing causes of failure. Technometrics, 16, 39-47.

Mood, A. M. and F. A. Graybill [1963]. Introduction to the Theory of Statistics (2nd ed.). McGraw-Hill, New York.

Moriyama, I. M. [1956]. Development of the present concept of cause of death. American Journal of Public Health, 46, 436-441.

Mould, R. F. [1976]. Calculation of survival rates by the life table and other methods. Clin. Radiol., 27, 33-8.

Mukherjee, S. P. and S. K. Das [1975]. Abridged life tables for rural West Bengal, 1969. Indian J. Public Health, 19, 3-9.

Myers, R. J. [1964]. Analysis of Mortality in the Soviet Union According to 1958-59 life tables. Transactions, Society of Actuaries.

Nanjo, Z. [1967]. Life tables for deaths from vascular lesions affecting central nervous system by prefecture, 1959-61. Fukushima J. Med. Sci., 14, 1-20.

Nanjo, Z. [1967]. A study on deaths from malignant neoplasms including neoplasms of lymphatic and haematopoietic tissues by prefecture using the life table, 1959-61. Fukushima J. Med. Sci., 14, 21-43.

Nanjo, Z. [1974]. A study on the method of constructing abridged life tables and the interpolation for individual years of life. Fukushima J. Med. Sci., 20, 11-48.

National Center for Health Statistics [1974]. Vital Statistics of the United States, 1970, Volume 2, Mortality. U. S. Government Printing Office, Washington, D. C.

National Office of Vital Statistics [1959]. Method of constructing the 1949-51 National, Divisional, and state life tables. Vital Statistics -- Special Reports, 41, 1959.

Neidhardt, F. O. [1971]. Estimation of survival in life table methods, applied to a sample of prostatectomies. Nord. Med., 85, 129-30.

Neumann, H. G. [1970]. Evaluation of the results of intrauterine contraception. 2 year analysis of the Rostock studies by means of the life table method of Tietze and Potter. Geburtshilfe Frauenheilkd, 30, 537-47.

Neyman, J. [1950]. First Course in Probability and Statistics. Holt, New York.

Oechsli, F. W. [1975]. A population model based on a life table that includes marriage and parity. Theor. Popul. Biol., 2, 229-45.

Oleinick, A. and N. Mantel [1970]. Family studies in systemic lupus erythematosus. II. Mortality among siblings and offspring of index cases with a statistical appendix concerning life table analysis. J. Chronic Dis., 22, 617-24.

Oster, J., M. Mikkelsen, and A. Nielsen [1975]. Mortality and life-table in Down's syndrome. Acta. Paediatr. Scand., 64, 322-6.

Oficina Sanitaria Panamericana [1940]. Consultas. Boletín de la Oficina Sanitaria Panamericana, 19, 283-285.

Pachal, T. K. [1975]. A note on the relation between expectation of life at birth and life table mortality rate for the age group 0-4 years. Indian J. Public Health, 19, 9-10.

Padley, R. [1959]. Cause of death statements in Ceylon: A study in levels of diagnostic reporting. Bulletin of the World Health Organization, 20, 677-695.

Park, C. B. [1955]. Statistical observations on death rates and causes of death in Korea. Bulletin of the World Health Organization, 13, 69-108.

Park, C. B. and Y. S. Matsumoto [1971]. Hawaii life table values by causes of death: 1959-61. Hawaii Med. J., 30, 184-90.

Pascua, M. [1952]. Evolution of mortality in Europe during the twentieth century. Epidemiological and Vital Statistics Report (World Health Organization), 5, 4-8.

Pike, M. C. and F. J. C. Roe [1963]. An actuarial method of analysis in an experiment in two-stage carcinogenesis. Brit. J. Cancer, 17, 605-10.

Pike, M. C. [1970]. A note on Kimball's paper 'Models for the estimation of competing risks from grouped data'. Biometrics, 26, 579-91.

Preston, S. H., N. Keyfitz, and R. Schoen [1972]. Causes of Death: Life Tables for National Populations. Seminar Press, New York.

Rao, C. R. [1945]. Information and accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc., 37, 81-91.

Reed, L. H. and M. Merrill [1939]. A short method for constructing an abridged life table. Amer. J. Hygiene, 30, 33062.

Reid, D. D. and G. A. Rose [1964]. Assessing the comparability of mortality statistics. British Medical Journal, 2, 1437-1439.

The Registrar-General's Decennial Supplement, England and Wales, 1921 [1933]. Part III, Estimates of Population, Statistics of Marriages, Births, and Deaths, 1911-1920, xxxiii-lxix. H. M. Stationery Office, London.

Registrar General's Statistical Review of Engalnd and Wales for the Year 1934 [1936]. New Annual Series, No. 14. H. M. Stationery Office, London.

Registrar-General's Statistical Review of England and Wales for the year 1937 [1940]. New Annual Series, No. 17. H. M. Stationery Office, London.

Robinson, M. J. and A. P. Norman [1975]. Life tables for cystic fibrosis. Arch. Dis. Child., 50, 962-5.

Sadie, J. L. [1970]. An evaluation of demographic data pertaining to the non-White population of South Africa. South African Journal of Economics, 38, 1-34.

Santas, J. L. [1972]. Evaluation of the mortality ratio and life expectancy levels in the State of Sao Paulo, Brazil in 1970. Rev. Saude Publica, 6, 269-72.

Schwartz, D. and P. Lazar [1963]. Taux de mortalité par une cause donnée de décès en tenant compte des autres causes de décès ou de disparition. Unité de Recherches Statistiques de l'Institut National d'Hygiène, Ministère de la Santé Publique, France.

Saveland, W. and P. C. Glick [1969]. First marriage decrement tables by color and sex for the United States in 1958-60. Demography, 6, 243-260.

Schoen, R. [1975]. Constructing increment-decrement life tables. Demography, 12, 313-324.

Schoen, R. and V. E. Nelson [1974]. Marriage, divorce and mortality: A life table analysis. Demography, 11, 267-290.

Seigel, D. G. [1975]. Life table rates and person month ratios as summary statistics for contraceptive trials. J. Steroid Biochem., 6, 933-6.

Sewell, W. H. [1972]. Life table analysis of the results of coronary surgery. Chest, 61, 481.

Shapiro, S. E. R. Schlesinger, and R. E. L. Nesbitt [1968]. Infant, Perinatal, Maternal and Childhood Mortality in the United States. American Public Health Association, Vital and Health Statistics Monographs, Harvard Univ. Press, Cambridge, Mass.

Sheps, M. C. [1958]. Shall we count the living or the dead? New England J. Med., 259, 1210-14.

Sheps, M. C. [1959]. An examination of some methods of comparing several rates or proportions. Biometrics, 15, 87-97.

Sheps, M. [1961]. Marriage and mortality. American Journal of Public Health, 51, 547-555.

Sirken, M. G. [1964]. Comparison of two methods of constructing abridged life tables by reference to a 'standard' table. Vital and Health Statistics, Series 2, No. 4, 1-11. National Center for Health Statistics, U. S. Dept. of Health, Education and Welfare.

Snow, E. C. [1920]. An elementary rapid method of constructing an abridged life table. Supplement to the 75th Annual Report of the Registrar-General of Births, Deaths, and Marriages in England and Wales. Part II. Abridged Life Tables.

Spiegelman, M. [1968]. Introduction to Demography (revised edition). Cambridge, Harvard University Press.

Spiegelman, M. and H. H. Marks [1966]. Empirical testing of standards for the age adjustment of death rates by the direct method. Hum Biol., 38, 280-292.

Stark, C. R. and N. Mantel [1966]. Effects of maternal age and birth order on the risk of mongolism and leukemia. J. Natl. Cancer Inst., 37, 687-698.

Stolnitz, G. J. [1956]. Life Tables from Limited Data: A Demographic Approach. Office of Population Research, Princeton University, Princeton, New Jersey.

Sullivan, J. M. [1971]. A review of Taiwanese infant and child mortality statistics 1961-1968. Taiwan Population Studies Working Paper No. 10. Ann Arbor, Univ. of Michigan Population Studies Center.

Swartout, H. O. and R. G. Webster [1940]. To what degree are mortality statistics dependable? American Journal of Public Health, 30, 811-815.

Terris, M. and M. Glasser [1974]. A life table analysis of the relation of prenatal care to prematurity. Am. J. Public Health, 64, 869-75.

Thylstrup, A. and I. Rolling [1975]. The life table method in clinical dental research. Community Dent, Oral Epidemiol, 3, 5-10.

Todhunter, I. [1949]. A History of the Mathematical Theory of Probability. Chelsea.

United Nations, Department of Economic and Social Affairs [1955]. Manuels on methods of estimating population, ST/SOA/Ser. A. Population Studies No. 23, Manuel II. Methods of appraisal of quality of basic data for population estimates. United Nations, N. Y.

United Nations, Department of Economic and Social Affairs [1958]. Handbook of population census methods. Ser. E. No. 5, Rev. 1, Vol. I. General aspects of a population census. United Nations, N. Y.

United Nations, Department of Economic and Social Affairs [1967]. Manuels on methods of estimating population, ST/SOA/Ser. A. 42. Manuel IV. Methods of estimating basic demographic measures from incomplete data.

United Nations Statistical Office. Department of Economic and Social Affairs [1955]. Handbook of vital statistics methods. Ser. F. No. 7. United Nations, New York.

United Nations Statistical Office, Department of Economic and Social Affairs [1971]. Demographic Yearbook, 1970, 22nd Edition, United Nations, New York.

United States Bureau of the Census [1975]. Census of population and Housing, 1970. Evaluation and Research Program, Accuracy of Data for Selected Ppulation Characteristics as Measured by the 1970 CPS-Census Match, Series PHC(E)-11. Washington, D.C., U.S. Government Printing Office.

United States Department of Health, Education and Welfare [1963]. Comparability of statistics of cause of death according to the Fifth and Sixth Revisions of the international list. Vital Statistics-Special Reports, Selected Studies 51. U.S. Government Printing Office, Washington, D. C.

United States Public Health Service, National Center for Health Statistics [1964]. Recent mortality rates in Chile, Ser. 3, No. 2. U. S. Government Printing Office, Washington D. C.

Walker, A. E., H.K. Leichs, H. Lechtape-Gruter, W. F. Caveness, and C. Kretschmann [1971]. The life expectancy of head injured men with and without epilepsy. Zentralbl Neurochir 32, 3-9.

Weibull, W. [1939]. A statistical theory of the strength of material. Ing. Vetenskaps Akad. Handl. 151

Weiss, K. M. [1973]. A method for approximating age-specific fertility in the construction of life tables for anthropological populations. Human Biology, 45, 195-210.

Westergaard, H. [1916]. Scope and method of statistics. J. of the American Statistical Association. 15, 260-264.

Wiesler, H. [1954]. Une méthods simple pour la construction de table de mortalite abregees. World Population Conference, Vol. IV. United Nations, N.Y.

Wilson, E. B. [1938]. The standard deviation of sampling for life expectancy. J. of the American Statistical Association, 33, 705-708.

World Health Organization [1967]. The accuracy and comparability of death statistics. Chronicle . 21, 11-17.

World Health Organization [1952]. Comparability of statistics of causes of death according to the Fifth and Sixth Revisions of the International List. Bulletin of the World Health Organization, Suppl. 4.

Yerushalmy, J. [1951]. A mortality index for use in place of the age-adejusted death rate. American J. of Public Health, 41, 907-922.

Yule, G. U. [1924], A methematical theory of evolution based on the conclusion of Dr. J. C. Willis, F.R.S. Phil Trans. Royal Soc. (London), B213, 21-87.

Yule G. U. [1934]. On some points relating to vital statistics, more especially statistics of occupational mortality. J. of the Royal Statistical Society. 97, 1-84.

Zahl, S. [1955]. A Markov process model for follow-up studies. Human Biol 27, 90-120.

GLOSSARY

CHAPTER 1

Formulas

Probability of A:

$$\Pr\{A\} = \frac{n(A)}{n} \quad . \tag{2.1}$$

$$0 \leq \Pr\{A\} \leq 1 \quad . \tag{2.2}$$

$$\Pr\{\bar{A}\} = 1 - \Pr\{A\} \tag{2.5a}$$

$$\Pr\{AB\} = \frac{n(AB)}{n} \tag{2.6}$$

Conditional probability:

$$\Pr\{B|A\} = \frac{\Pr\{AB\}}{\Pr\{A\}} \tag{2.9}$$

Multiplication theorem:

$$\Pr\{AB\} = \Pr\{A\} \times \Pr\{B|A\} \tag{2.13}$$

$$\Pr\{ABC\} = \Pr\{A\} \times \Pr\{B|A\} \times \Pr\{C|AB\} \tag{2.15}$$

$$\Pr\{ABCD\} = \Pr\{A\} \times \Pr\{B|A\} \times \Pr\{C|AB\} \times \Pr\{D|ABC\} \quad . \tag{2.16}$$

Addition theorem:

$$\Pr\{A \text{ or } B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{AB\} \quad . \tag{2.21}$$

$$\Pr\{A \text{ or } B \text{ or } C\} = \Pr\{A\} + \Pr\{B\} + \Pr\{C\}$$

$$- \Pr\{AB\} - \Pr\{BC\} - \Pr\{CA\} + \Pr\{ABC\} \tag{2.22}$$

$$\Pr\{A \text{ or } B \text{ or } C \text{ or } D\} = \Pr\{A\} + \Pr\{B\} + \Pr\{C\} + \Pr\{D\}$$

$$- \Pr\{AB\} - \Pr\{AC\} - \Pr\{AD\} - \Pr\{BC\} - \Pr\{BD\} - \Pr\{CD\}$$

$$+ \Pr\{ABC\} + \Pr\{ABD\} + \Pr\{ACD\} + \Pr\{BCD\} - \Pr\{ABCD\} \tag{2.23}$$

Distributive law:

$$\Pr\{A(B \text{ or } C)\} = \Pr\{AB \text{ or } AC\} \tag{2.27}$$

$$\Pr\{(A \text{ or } B)(C \text{ or } D)\} = \Pr\{AC \text{ or } AD \text{ or } BC \text{ or } BD\} \tag{2.28}$$

CHAPTER 2

		page no.
a_i	- Fraction of the last age interval of life. It is the expected fraction of the interval (x_i,x_{i+1}) lived by an individual who dies at an age in the interval (x_i,x_{i+1}).	19
d_i	- Number of life table deaths in the age interval (x_i,x_{i+1}).	19
D	- Total number of deaths in a current population.	22
D_δ	- Number of deaths from cause R_δ in a current population.	22
D_i	- Number of deaths in the age group (x_i,x_{i+1}) in a current population.	20
$D_{i\delta}$	- Number of deaths from cause R_δ in age group (x_i,x_{i+1}) in a current population.	23
D_s	- Total number of deaths in a standard population.	29
D_{si}	- Number of deaths in the age interval (x_i,x_{i+1}) in the standard population.	29
D_u	- Total number of deaths in community u.	28
D_{ui}	- Number of deaths in the age interval (x_i,x_{i+1}) in community u.	28
$\hat{e}_0$	- Observed expectation of life at age zero.	38
ℓ_i	- Number alive at exact age x_i in the life table population.	19
L_i	- Number of years lived in (x_i,x_{i+1}) by ℓ_i individuals.	37
n_i	- Length of the age interval (x_i,x_{i+1}); $n_i = x_{i+1}-x_i$.	19
N_i	- (Hypothetical) number of individuals alive at exact age x_i.	20
P	- Total midyear population.	22
P_i	- Midyear population in age interval (x_i,x_{i+1}).	23
P_s	- Total midyear standard population.	29
P_{si}	- Midyear population in the age interval (x_i,x_{i+1}) of the standard population.	29
P_u	- Total midyear population of community u.	29
P_{ui}	- Midyear population in age interval (x_i,x_{i+1}) of community u.	28

		page no.
T_o	- Total number of years lived by the life table population beyond x_0.	37
x_i	- Exact age in years at the lower limit of the i-th interval.	13
x_{i+1}	- Exact age in years at the upper limit of the i-th interval.	13
Age-specific death rate	- $M_i = \dfrac{\text{Number dying in } (x_i, x_{i+1})}{\text{Number of years lived in } (x_i, x_{i+1}) \text{ by those alive at } x_i}$	19
Fetal death rate	- (alias "stillbirth rate). Two definitions are available:	
	$\dfrac{\text{Number of fetal deaths or 28 or more weeks of gestation}}{\text{Number of live births + fetal deaths of 28 or more weeks of gestation}} \times 1000$	24
	$\dfrac{\text{Number of fetal deaths of 20 weeks or more of gestation}}{\text{Number of live births + fetal deaths of 20 or more weeks of gestation}} \times 1000$	24
Neonatal mortality rate	- $\dfrac{\text{Number of deaths under 28 days of age}}{\text{Number of live births}} \times 1000$	24
Perinatal mortality rate	- There are two definitions in common use:	
	$\dfrac{\text{Number of deaths under 7 days + fetal deaths of 28 weeks or more of gestation}}{\text{Number of live births + fetal deaths of 28 weeks or more of gestation}}$	24
	$\dfrac{\text{Number of deaths under 28 days of life + fetal deaths of 20 or more weeks of gestation}}{\text{Number of live births + fetal deaths of 20 or more weeks of gestation}} \times 1000$	25
Post neonatal mortality rate	- $\dfrac{\text{Number of deaths at age 28 days through one year}}{\text{Number of live births - neonatal deaths}} \times 1000$	25
Infant mortality rate	- $\dfrac{\text{Number of deaths under one year of age}}{\text{Number of live births}} \times 1000$	25
Fetal death ratio	- $\dfrac{\text{Number of fetal deaths of 20 or more weeks of gestation}}{\text{Number of live births}}$	25
Maternal mortality rate	- $\dfrac{\text{Number of maternal deaths}}{\text{Number of live births}} \times 1000$	25

Probability of dying for age interval (x_i, x_{i+1}):

$$\hat{q}_i = \frac{d_i}{\ell_i} \tag{1.3}$$

$$\hat{q}_i = \frac{D_i}{N_i} \quad . \tag{1.3a}$$

Age-specific death rate for age interval (x_i, x_{i+1}):

$$M_i = \frac{d_i}{n_i(\ell_i - d_i) + a_i n_i d_i} \quad , \tag{1.2}$$

$$M_i = \frac{D_i}{n_i(N_i - D_i) + a_i n_i D_i} \tag{1.2a}$$

$$M_i = \frac{D_i}{P_i} \quad . \tag{1.5}$$

Relationship between q_i and M_i for age interval (x_i, x_{i+1}):

$$\hat{q}_i = \frac{n_i M_i}{1+(1-a_i)n_i M_i} \quad . \tag{1.4}$$

Cause-specific death rate for cause R_δ:

$$M_\delta = \frac{D_\delta}{P} \times 100{,}000 \tag{1.9}$$

Age-cause-specific death rate for cause R_δ age interval (x_i, x_{i+1}):

$$M_{i\delta} = \frac{D_{i\delta}}{P_i} \times 100{,}000 \tag{1.10}$$

Crude death rate:

$$M = \frac{D}{P} \times 1000 \qquad (1.6)$$

Crude death rate for community u:

$$C.D.R. = D_u/P_u \quad . \qquad (3.3)$$

Age-specific death rate for community u and age interval (x_i, x_{i+1}):

$$M_{ui} = D_{ui}/P_{ui} \ , \qquad (3.5)$$

Crude death rate (as a weighted average of M_{ui}):

$$C.D.R. = \sum_i \frac{P_{ui}}{P_u} M_{ui} \quad , \qquad (3.8)$$

Direct method of adjustment:

$$D.M.D.R. = \frac{\sum_i P_{si} M_{ui}}{P_s} \quad , \qquad (3.10)$$

Comparative mortality ratio:

$$C.M.R. = \tfrac{1}{2} \sum_i \left(\frac{P_{ui}}{P_u} + \frac{P_{si}}{P_s} \right) M_{ui} \qquad (3.11)$$

Indirect method of adjustment:

$$I.M.D.R. = \frac{D_s/P_s}{\sum_i P_{ui} M_{si}/P_u} \ \frac{D_u}{P_u} \quad . \qquad (3.12)$$

Life table death rate:

$$L.T.D.R. = \sum_i \frac{L_i}{T_0} \ M_{ui} \quad , \qquad (3.15)$$

$$L.T.D.R. = \frac{1}{\hat{e}_0} \quad . \qquad (3.21)$$

Equivalent average death rate:

$$\text{E.A.D.R.} = \sum_i \frac{n_i}{\sum_i n_i} M_{ui} \tag{3.22}$$

Relative mortality index

$$\text{R.M.I.} = \sum_i \frac{P_{ui}}{P_u} \frac{M_{ui}}{M_{si}} \tag{3.23}$$

Mortality index:

$$\text{M.I.} = \frac{\Sigma n_i \dfrac{M_{ui}}{M_{si}}}{\Sigma n_i} \tag{3.24}$$

Standardized mortality ratio:

$$\text{S.M.R.} = \frac{\Sigma P_{ui} M_{ui}}{\Sigma P_{ui} M_{si}} \tag{3.25}$$

CHAPTER 3

		page no.
a_i	- Fraction of the last age interval of life.	47
D_i	- The number of deaths in the age group (x_i, x_{i+1}) in a current population.	43
$E(D_i)$	- Expected number of deaths in the interval (x_i, x_{i+1})	43
D_s	- The total number of deaths in the standard population.	53
D_u	- The total number of deaths in community u.	53
$E(\dot{q}_i \mid N_i)$	- The conditional expectation of q_i given N_i.	49
L_i	- Number of years lived in (x_i, x_{i+1}) by N_i individuals.	52
M_i	- The age specific death rate for the interval (x_i, x_{i+1})	46
M_{si}	- The age-specific death rate for interval (x_i, x_{i+1}) in the standard population.	53
M_{ui}	- The age-specific death rate for interval (x_i, x_{i+1}) in community u.	53
n_i	- The length of the age interval (x_i, x_{i+1}); $n_i = x_{i+1} - x_i$.	47
N_i	- (Hypothetical) number of individuals alive at exact age x_i.	43
P_i	- Midyear population in age interval (x_i, x_{i+1}).	47
P_s	- Total midyear standard population.	52
P_{si}	- Midyear population in the age interval (x_i, x_{i+1}) of the standard population.	52
P_u	- Total midyear population of community u.	52
P_{ui}	- Midyear population in age interval (x_i, x_{i+1}) of community u.	52
q_i	- Probability of death in the interval (x_i, x_{i+1}).	43
$\hat{q}_i$	- Estimate of the probability of death in the interval (x_i, x_{i+1}).	43
$\hat{q}_{ui}$	- Estimate of the probability of death in the interval (x_i, x_{i+1}) of community u.	59
R	- General symbol for an adjusted rate or mortality index.	54
$S^2_{D_i}$	- The sample variance of D_i.	48

		page no.
$S^2_{q_i}$	- The sample variance of q_i.	44
S^2_R	- Sample variance of an adjusted rate or mortality index, R.	55
$\sigma^2_{D_i}$	- Variance of D_i.	43
$\sigma^2_{q_i}$	- Variance of q_i.	44
$\sigma^2_{q_i \mid N_i}$	- Conditional variance of q_i given N_i.	49
w_i	- Weight of M_{ui} used to calculate adjusted rates and mortality indices.	53
T_0	- Total number of years lived by the life table population beyond x_0.	52
x_i	- Exact age in years at the lower limit of the age interval.	43
x_{i+1}	- Exact age in years at the upper limit of the age interval.	43
Binomial Distribution	- If an event has a constant probability q of occurring in any one trial, then the number of times (D) that the event will occur in N independent trials has a binomial distribution, with the expected value E(D) = Nq and variance $\sigma^2_D = N\, q(1-q)$.	43
Coefficient of Variation of R	- A measure of the magnitude of the standard deviation of an adjusted rate, R, relative to R itself.	57

Formulas

Expectations and variances of the number of deaths and probability of dying in (x_i, x_{i+1}):

$$E(D_i) = N_i q_i \tag{2.1}$$

$$\hat{q}_i = \frac{D_i}{N_i} \tag{2.3}$$

$$\sigma^2_{D_i} = N_i q_i (1-q_i). \tag{2.2}$$

$$E(\hat{q}_i) = E\left(\frac{D_i}{N_i}\right) = \frac{1}{N_i} E(D_i) = \frac{1}{N_i} N_i q_i = q_i \, . \tag{2.4}$$

$$\sigma^2_{\hat{q}_i} = \frac{1}{N_i} q_i (1-q_i). \tag{2.5}$$

$$S^2_{\hat{q}_i} = \frac{1}{N_i} \hat{q}_i (1-\hat{q}_i) \, . \tag{2.6}$$

95% confidence interval for q_i:

$$Z = \frac{\hat{q}_i - q_i}{\sqrt{q_i(1-q_i)/N_i}} \tag{2.7}$$

$$\hat{q}_i - 1.96\, S_{\hat{q}_i} < q_i < \hat{q}_i + 1.96\, S_{\hat{q}_i} \tag{2.10}$$

Sample variances of M_i, $\hat{q}_i$, and $\hat{q}_x$:

$$S^2_{D_i} = N_i \hat{q}_i(1-\hat{q}_i) = D_i(1-\hat{q}_i) \qquad (3.7)$$

$$\hat{q}_i = \frac{D_i}{N_i} \qquad (2.3)$$

$$M_i = \frac{D_i}{P_i} \qquad (3.4)$$

$$S^2_{\hat{q}_i} = \frac{1}{D_i} \hat{q}_i^2 (1-\hat{q}_i) \quad . \qquad (3.5)$$

$$S^2_{\hat{q}_x} = \frac{1}{D_x} \hat{q}_x^2 (1-\hat{q}_x) \qquad (8.5)$$

$$S^2_{M_i} = \frac{1}{P_i} M_i(1-\hat{q}_i) \quad . \qquad (3.8)$$

Sample variance of age adjusted death rates:

$$R = \sum_i w_i M_{ui} \qquad (6.1)$$

$$S^2_R = \sum_i w_i^2 S^2_{M_{ui}} \quad . \qquad (6.3)$$

$$S^2_R = \sum_i w_i^2 \frac{M_{ui}}{P_{ui}} (1-\hat{q}_{ui}) \quad , \qquad (6.4)$$

Sample variance of direct method of age adjustment:

$$S^2_{DMDR} = \sum_i \left(\frac{P_{si}}{P_s}\right)^2 \frac{M_{ui}}{P_{ui}} [1-\hat{q}_{ui}] \tag{7.1}$$

$$\text{Coeff. of variation of } R = \frac{S_R}{R} \tag{7.5}$$

Sample variance for life table death rate:

$$LTDR = \frac{\Sigma L_x M_{ux}}{\Sigma L_x} = \frac{\Sigma d_x}{\Sigma L_x} = \frac{\ell_0}{T_0} = \frac{1}{\hat{e}_0} . \tag{8.1}$$

$$S^2_{LTDR} = \frac{1}{\hat{e}_0^4} S^2_{\hat{e}_0} . \tag{8.2}$$

$$S^2_{\hat{e}_0} = \frac{1}{\hat{e}_0^4} \sum_{x \geq 0} \hat{p}^2_{0x} [(1-a_x)n_x + \hat{e}_{x+n_x}]^2 S^2_{\hat{q}_x} \tag{8.4}$$

Summary of adjusted death rates and indices (Table 1):

Crude death rate (C.D R.)
$$\frac{\sum_i P_{ui} M_{ui}}{P_u}$$

Direct method of adjustment (D.M.D.R.)
$$\frac{\sum_i P_{si} M_{ui}}{P_s}$$

Comparative mortality rate (C.M.R.)
$$\tfrac{1}{2} \sum_i \left(\frac{P_{ui}}{P_u} + \frac{P_{si}}{P_s} \right) M_{ui}$$

Indirect method of adjustment (I.M.D.R.)
$$\frac{(D_s/P_s)(D_u/P_u)}{\sum_i P_{ui} M_{si} \; / P_u}$$

Life table death rate (L.T.D.R.)
$$\frac{\sum_i L_i M_{ui}}{\sum_i L_i}$$

Equivalent average death rate (E.A.D.R.)
$$\frac{\sum_i n_i M_{ui}}{\sum_i n_i}$$

Relative mortality index (R.M.I.)
$$\frac{\sum_i P_{ui} \dfrac{M_{ui}}{M_{si}}}{P_u}$$

Mortality index (M.I.)
$$\frac{\sum_i n_i \dfrac{M_{ui}}{M_{si}}}{\Sigma n_i}$$

Standardized mortality ratio (S.M.R.)
$$\sum_i \frac{P_{ui}}{\Sigma P_{ui} M_{si}} M_{ui}$$

CHAPTER 4

		page no.
a'_x	- Fraction of the last year of life.	66
d_x	- Number of life table deaths in the age interval (x,x+1).	66
D_x	- Number of deaths in the age interval (x,x+1) in a current population.	71
e_x	- The expectation of life at age x.	68
$\hat{e}_x$	- Observed expectation of life at age x.	67
ℓ_x	- Number alive at exact age x in a life table population.	65
L_x	- Number of years lived in (x,x+1) by the ℓ_x individuals.	67
M_x	- Age-specific death rate in the interval (x,x+1).	71
N_x	- (Hypothetical) number of individuals alive at exact age x.	71
$\hat{p}_x$	- Proportion of those alive at age x surviving the interval (x,x+1).	68
$\hat{p}_{xy}$	- Proportion of those alive at age x surviving to age y.	68
P_x	- Midyear population in age interval (x,x+1).	71
$\hat{q}_x$	- Estimate of the probability of dying in (x,x+1).	65

		page no.
T_x	- Total number of years lived by the life table population beyond age x.	67
x	- Lower limit of age interval (x,x+1).	65
x+1	- Upper limit of age interval (x,x+1).	65
x_w	- Lower limit of the final age interval in a life table.	65
Abridged Life Table	- A life table with age intervals greater than one year (beyond age 1).	64
Complete Life Table	- A life table with single year age intervals.	64
Current Life Table	- A life table based on current mortality and population data.	63
Cohort Life Table	- A life table based on the mortality experience of a single group of individuals.	62

Formulas

Relationship between life table functions in the complete life table:

$$d_x = \ell_x \hat{q}_x, \qquad x=0,1,\ldots,w \quad . \tag{2.1}$$

$$\ell_{x+1} = \ell_x - d_x, \qquad x=0,1,\ldots,w-1 \quad . \tag{2.2}$$

$$L_x = (\ell_x - d_x) + a_x' d_x \qquad x=0,1,\ldots,w-1 \quad . \tag{2.3}$$

$$T_x = L_x + L_{x+1} + \ldots + L_w , \qquad x=0,1,\ldots,w \quad . \tag{2.5}$$

$$T_x = L_x + T_{x+1} \; . \tag{2.6}$$

$$\hat{e}_x = \frac{T_x}{\ell_x} , \qquad x=0,1,\ldots,w. \tag{2.7}$$

$$\hat{p}_x = 1 - \hat{q}_x , \tag{2.8}$$

$$\hat{p}_{xy} = \hat{p}_x \hat{p}_{x+1} \cdots \hat{p}_{y-1} = \frac{\ell_y}{\ell_x} , \tag{2.9}$$

Computation of $\hat{q}_x$, L_w, T_w, and $\hat{e}_w$:

$$\hat{q}_x = \frac{D_x}{N_x} \quad . \tag{3.1}$$

$$M_x = \frac{D_x}{(N_x - D_x) + a_x' D_x} \quad . \tag{3.2}$$

$$M_x = \frac{D_x}{P_x} \quad . \tag{3.4}$$

$$\hat{q}_x = \frac{M_x}{1 + (1 - a_x')M_x} \quad . \tag{3.7}$$

$$L_w = \frac{\ell_w}{M_w} \quad . \tag{3.11}$$

$$T_w = L_w \qquad \text{and} \qquad \hat{e}_w = \frac{T_w}{\ell_w} = \frac{L_w}{d_w} = \frac{1}{M_w} \quad . \tag{3.12}$$

CHAPTER 5

		page no.
a_i	- Fraction of the last age interval of life.	69
d_i	- Number of life table deaths in the age interval (x_i,x_{i+1})	69
D_i	- Number of deaths in the age interval (x_i,x_{i+1}) in a current population	93
$\hat{e}_i$	- Observed expectation of life at age x_i.	68
ℓ_i	- Number alive at exact age x_i in a life table population.	94
L_i	- Number of years lived in the interval (x_i,x_{i+1}) by the ℓ_i individuals.	94
M_i	- Age-specific death rate in the interval (x_i,x_{i+1}).	93
n_i	- Length of the age interval (x_i,x_{i+1}); $n_i = x_{i+1}-x_i$.	93
N_i	- (Hypothetical) number of individuals alive at exact age x_i.	93
P_i	- Midyear population in age interval (x_i,x_{i+1}).	94
$\hat{q}_i$	- Estimate of the probability of dying in interval (x_i,x_{i+1}).	94
T_i	- Total number of years lived by the life table population beyond age x_i.	95
x_i	- Lower limit of age interval (x_i,x_{i+1}).	92
x_{i+1}	- Upper limit of age interval (x_i,x_{i+1}).	92

Formulas

Construction of abridged life table:

$$\hat{q}_i = \frac{D_i}{N_i} \quad . \tag{2.1}$$

$$M_i = \frac{D_i}{(N_i-D_i)n_i + a_i n_i D_i} \quad . \tag{2.2}$$

$$\hat{q}_i = \frac{n_i M_i}{1 + (1-a_i)n_i M_i} \quad . \tag{2.3}$$

$$M_i = \frac{D_i}{P_i} \tag{2.4}$$

$$d_i = \ell_i \hat{q}_i, \qquad i=0,1,\ldots,w-1, \tag{2.5}$$

$$\ell_{i+1} = \ell_i - d_i, \qquad i=0,1,\ldots,w-1, \tag{2.6}$$

$$L_i = n_i(\ell_i - d_i) + a_i n_i d_i, \qquad i=0,1,\ldots,w-1\ . \tag{2.7}$$

$$L_w = \frac{\ell_w}{M_w}\ , \tag{2.8}$$

$$\hat{e}_i = \frac{L_i + L_{i+1} + \ldots + L_w}{\ell_i}\ , \quad i=0,\ldots w\ . \tag{2.9}$$

Computation of the fraction of last age interval of life, a_i:

$$a_1 = \frac{q_1 a_1' + p_1 q_1 (1+a_2') + p_1 p_2 q_3 (2+a_3') + p_1 p_2 p_3 q_4 (3+a_4')}{4(1 - p_1 p_2 p_3 p_4)} . \quad (3.1)$$

$$a_2 = \frac{.5q_5 + (1+.5)p_5 q_6 + (2+.5)p_5 p_6 q_7 + (3+.5)p_5 p_6 p_7 q_8 + (4+.5)p_5 p_6 p_7 p_8 q_9}{5(1 - p_5 p_6 p_7 p_8 p_9)}$$

$$= \frac{p_5 q_6 + 2p_5 p_6 q_7 + 3p_5 p_6 p_7 q_8 + 4p_5 p_6 p_7 p_8 q_9}{5(1 - p_5 p_6 p_7 p_8 p_9)} + .1 \quad , \quad (3.3)$$

$$q_5 + p_5 q_6 + p_5 p_6 q_7 + p_6 p_6 p_7 q_8 + p_5 p_6 p_7 p_8 q_9 = 1 - p_5 p_6 p_7 p_8 p_9 . \quad (3.4)$$

Computation of cohort life table functions:

$$\ell_x - \ell_{x+n} = d_x \tag{5.1}$$

$$\ell_{x+n} = \ell_x - d_x \quad . \tag{5.2}$$

$$\hat{q}_x = \frac{d_x}{\ell_x} \quad . \tag{5.3}$$

$$L_x = n\ell_{x+n} + (1-a_x)nd_x \quad . \tag{5.4}$$

$$T_x = L_x + \quad \ldots + L_w \tag{5.5}$$

$$\hat{e}_x = \frac{T_x}{\ell_x} \;, \quad x = 0,1,\ldots,w \quad . \tag{5.6}$$

CHAPTER 6

		page no.
a_i	- Fraction of the last age interval of life.	131
d_i	- Number of life table deaths in the age interval (x_i, x_{i+1}).	122
D_i	- Number of deaths in the age interval (x_i, x_{i+1}) in a current population.	119
e_α	- Expectation of life at age x_α.	119
$\hat{e}_\alpha$	- Observed expectation of life at age x_α.	131
ℓ_0	- Life table population at age x_0. It is an arbitrarily assigned number and is referred to as the radix.	134
ℓ_i	- Number alive at exact age x_i in the life table population.	122
L_i	- Number of years lived in the interval (x_i, x_{i+1}) by the ℓ_i individuals.	125
n_i	- Length of the age interval (x_i, x_{i+1}); $n_i = x_{i+1} - x_i$.	131
p_i	- Probability of surviving the interval (x_i, x_{i+1}).	119
$\hat{p}_i$	- Estimate of the probability of surviving the interval (x_i, x_{i+1}).	119
p_{ij}	- Probability of surviving from age x_i to age x_j.	119
$\hat{p}_{ij}$	- Estimate of the probability of surviving from age x_i to age x_j.	123
p_{0i}	- Probability of surviving from age x_0 to age x_i.	123
$\hat{p}_{0j}$	- Estimate of the probability of surviving from age x_0 to age x_j.	123
q_i	- Probability of death in the interval (x_i, x_{i+1}).	119
$\hat{q}_i$	- Estimate of the probability of dying in age interval (x_i, x_{i+1}).	119
S.E.	- Standard error.	121
S.E.(diff.)	- Standard error of a difference.	128
$S_{\hat{e}_i}$	- Sample standard error of $\hat{e}_i$.	139
$S_{\hat{q}_i}$	- Sample standard error of $\hat{q}_i$.	120

		page no.
$S_{\hat{p}_{0i}}$	- Sample standard deviation of $\hat{p}_{0i}$.	126
S^2	- Sample variance.	122
$S^2_{\hat{e}_\alpha}$	- Sample variance of $\hat{e}_\alpha$.	134
$S^2_{\hat{q}_i}$	- Sample variance of $\hat{q}_i$.	119
$S^2_{\hat{p}_i}$	- Sample variance of $\hat{p}_i$.	119
$S^2_{\hat{p}_{0i}}$	- Sample variance of $\hat{p}_{0i}$.	124
$S^2_{Y_\alpha}$	- Sample variance of Y_α.	134
$S^2_{\bar{Y}_\alpha}$	- Sample variance of $\bar{Y}_\alpha$.	134
T_i	- Total number of years lived by the life table population beyond age x_i.	125
x_w	- Lower limit of the final age interval in a life table.	119
x_i	- Lower limit of age interval (x_i, x_{i+1}).	119
x_{i+1}	- Upper limit of age interval (x_i, x_{i+1}).	119
$Y_{\alpha k}$	- Length of life beyond age x_α of the k-th individual in the group of ℓ_α, for $k=1,2,\ldots,\ell_\alpha$.	131
$\bar{Y}_\alpha$	- Mean length of life beyond age x_α.	131

Formulas

Estimation and hypothesis testing concerning probability q_i:

$$S^2_{\hat{q}_i} = S^2_{\hat{p}_i} \tag{2.1}$$

$$S^2_{\hat{q}_i} = \frac{1}{D_i}\, \hat{q}_i^2\, (1-\hat{q}_i) \tag{2.2}$$

$$\Pr\{\hat{q}_i - 1.96\, S_{\hat{q}_i} < q_i < \hat{q}_i + 1.96\, S_{\hat{q}_i}\} = .95 \tag{2.3}$$

$$Z = \frac{\hat{q}_0(1960) - \hat{q}_0(1970)}{\text{S.E.}[\hat{q}_0(1960) - \hat{q}_0(1970)]} \tag{2.4}$$

Hypothesis testing concerning survival probaiblity p_{ij}:

$$p_{ij} = p_i \, p_{i+1} \cdots p_{j-1} \tag{3.1}$$

$$p_{ij} = (1-q_i)(1-q_{i+1}) \ \cdots \ (1-q_{j-1}) \ . \tag{3.2}$$

$$S^2_{\hat{p}_{ij}} = \hat{p}_{ij}^{\,2} \sum_{h=i}^{j-1} \hat{p}_h^{-2} \, S^2_{\hat{p}_h} \tag{3.7}$$

$$Z = \frac{\hat{p}_{0,20}(\text{U.S.}) - \hat{p}_{0,20}(\text{Cal.})}{\text{S.E.(diff.)}} \ . \tag{3.7a}$$

NOTE: For the cohort life table, $\hat{p}_{ij}$ is computed directly from

$$\hat{p}_{ij} = \frac{\ell_j}{\ell_i} \quad , \tag{3.5}$$

with the variance given by:

$$S^2_{\hat{p}_{ij}} = \frac{1}{\ell_i} \hat{p}_{ij}(1-\hat{p}_{ij}) \ . \tag{3.10}$$

Mean life time and expectation of life:

$$\overline{Y}_\alpha = \frac{1}{\ell_\alpha} \sum_{k=1}^{\ell_\alpha} Y_{\alpha k} \tag{4.1}$$

$$\overline{Y} = \frac{L_\alpha + L_{\alpha+L} + \ldots + L_w}{\ell_\alpha} \tag{4.13}$$

$$\overline{Y}_\alpha = \hat{e}_\alpha \tag{4.2}$$

Variance of the observed expectation of life:

$$S^2_{Y_\alpha} = \frac{1}{\ell_\alpha} \sum_{i=\alpha}^{w} [(x_i - x_\alpha + a_i n_i) - \hat{e}_\alpha]^2 \, d_i \;. \tag{4.14}$$

$$S^2_{\hat{e}_\alpha} = S^2_{\bar{Y}_\alpha} = \frac{1}{\ell_\alpha} \, S^2_{Y_\alpha} \;, \tag{4.15}$$

$$S^2_{\hat{e}_\alpha} = \sum_{i=\alpha}^{w-1} \hat{p}^2_{\alpha i} \, [(1-a_i)n_i + \hat{e}_{i+1}]^2 \, S^2_{\hat{p}_i} \tag{4.27}$$

CHAPTER 7

		page no.
Crude probability	- The probability of death from a specific cause in the presence of competition of all other risks acting in a population.	141
Net probability	- The probability of death if a specific risk is the only risk in effect in a population, or conversely, the probability of death if a specific risk is eliminated from a population.	142
Partial crude probability	- The probability of death from a specific cause when another risk is (or risk are) eliminated from a population.	142
Risk and cause	- Both terms may refer to the same condition but are different on the time scale relative to the occurrence of death. Prior to death the condition in question is a risk; after death the condition is a cause (provided, of course, this is the condition from which an individual dies).	142
Cohort multiple decrement table	- A cohort multiple decrement table records the mortality experience by cause of a well defined cohort of people from birth to the death of the last person of the group.	142
Current multiple decrement table	- A current multiple decrement table is the one derived from the mortality experience by cause of a population of all ages during a current year.	144

Formulas

Age specific death rate:

$$M_i = \frac{D_i}{(N_i - D_i)n_i + a_i\, n_i\, D_i}, \qquad (2.1)$$

$$M_i = \frac{D_i}{P_i}\,. \qquad (2.3)$$

Estimate of the probability of dying:

$$\hat{q}_i = \frac{D_i}{N_i}\,, \qquad (2.4)$$

Relationship between M_i and q_i:

$$\hat{q}_i = \frac{n_i M_i}{1 + (1-a_i)n_i M_i} \quad . \tag{2.5}$$

Age-cause-specific death rate:

$$M_{i\delta} = \frac{D_{i\delta}}{P_i} \qquad , \ \delta = 1,\ldots,r. \tag{2.7}$$

Estimate of the crude probability of dying from risk R_δ:

$$\hat{Q}_{i\delta} = \frac{n_i M_{i\delta}}{1+(1-a_i)n_i M_i} \quad , \tag{2.9}$$

$$\hat{Q}_{i\delta} = \frac{D_{i\delta}}{D_i} \hat{q}_i \ . \tag{2.9a}$$

Relationship between the probability of dying q_i and the crude probabilities, $Q_{i\delta}$:

$$\hat{Q}_{i1} + \ldots + \hat{Q}_{ir} = \hat{q}_i \ . \tag{2.10}$$

Variance of the estimate of crude probability $\hat{Q}_{i\delta}$:

$$\text{Var}(\hat{Q}_{i\delta}) = \frac{1}{N_i} Q_{i\delta}(1 - Q_{i\delta}) \quad . \tag{2.11}$$

$$\text{Var}(\hat{Q}_{i\delta}) = \frac{1}{N_i} \hat{Q}_{i\delta}(1 - \hat{Q}_{i\delta}) = \frac{1}{D_{i\delta}} \hat{Q}^2_{i\delta}(1-\hat{Q}_{i\delta}). \tag{2.12}$$

Standard deviation (standard error) of the estimate $\hat{Q}_{i\delta}$:

$$S.D.(\hat{Q}_{i\delta}) = \sqrt{\frac{1}{D_{i\delta}} \hat{Q}_{i\delta}^2 (1-\hat{Q}_{i\delta})} \quad . \tag{2.13}$$

Standard deviation (standard error) of the estimate $\hat{q}_i$:

$$S.D.(\hat{q}_i) = \sqrt{\frac{1}{D_i} \hat{q}_i^2 (1-\hat{q}_i)} \quad . \tag{2.14}$$

Covariance between $\hat{Q}_{i1}$ and $\hat{Q}_{i2}$ from the same population:

$$Cov(\hat{Q}_{i1},\hat{Q}_{i2}) = -\frac{1}{N_i} \hat{Q}_{i1}\hat{Q}_{i2} = -\frac{1}{D_{i1}} \hat{Q}_{i1}^2 \hat{Q}_{i2} \quad . \tag{4.2}$$

Standard deviation (standard error) of the difference $\hat{Q}_{i1}-\hat{Q}_{i2}$:

$$S.D.(\hat{Q}_{i1}-\hat{Q}_{i2}) = \sqrt{S^2_{\hat{Q}_{i1}} + S^2_{\hat{Q}_{i2}} - 2\ Cov(\hat{Q}_{i1},\hat{Q}_{i2})} \tag{4.4}$$

Critical ratio for comparing $\hat{Q}_{i1}$ between two populations:

$$z = \frac{\hat{Q}_{45,1}(A) - \hat{Q}_{45,1}(S)}{\sqrt{S^2_{\hat{Q}_{45,1}(A)} + S^2_{\hat{Q}_{45,1}(S)}}} \quad . \tag{4.1}$$

CHAPTER 8

page no.

Net probability:

$\hat{q}_{i.1}$ – The probability of dying in (x_i, x_{i+1}) when R_1 is eliminated as a risk of death. 163

q_{i1} – The probability of dying in (x_i, x_{i+1}) when R_1 is the only risk acting in a population. 190

$\hat{q}_i - \hat{q}_{i.1}$ – Reduction in the probability of dying in interval (x_i, x_{i+1}) due to the presence of risk R_1. 178

Expectation of life:

$\hat{e}_{i.1}$ – The expectation of life at age x_i when R_1 is eliminated as a risk of death. 183

$\hat{e}_{i.1} - \hat{e}_i$ – Reduction in the expectation of life at age x_i due to the presence of risk R_1. 183

Formulas

The net probabilities:

$$q_{i.1} = (q_i - Q_{i1})(1 + \frac{1}{2} Q_{i1}) \tag{2.1}$$

$$q_{i1} = Q_{i1} (1 + \frac{1}{2} (q_i - Q_{i1})) \quad . \tag{5.1}$$

Computation of the estimate $\hat{q}_{i.1}$:

$$M_{i1} = \frac{D_{i1}}{P_i} \quad . \tag{2.3}$$

$$\hat{q}_i = \frac{n_i M_i}{1 + (1 - a_i) n_i M_i} \tag{2.4}$$

$$\hat{Q}_{i1} = \frac{n_i M_{i1}}{1 + (1 - a_i) n_i M_i} \quad . \tag{2.5}$$

$$\hat{q}_{i.1} = (\hat{q}_i - \hat{Q}_{i1})(1 + \frac{1}{2} \hat{Q}_{i1}) \quad . \tag{2.6}$$

Formulas used in the construction of life tables when R_1 is eliminated:

$$d_{0.1} = \ell_{0.1}\, q_{0.1} \tag{3.1}$$

$$\ell_{1.1} = \ell_{0.1} - d_{0.1}$$

(3.2)

$$L_{0.1} = (\ell_{0.1} - d_{0.1}) + a_0\, d_{0.1}$$

(3.3)

$$\hat{e}_{95.1} = \frac{P_{95}}{D_{95} - D_{95,1}}$$

(3.4)

$$T_{95\cdot 1} = \ell_{95\cdot 1}\, \hat{e}_{95\cdot 1}$$

$$L_{95\cdot 1} = T_{95\cdot 1} \tag{3.6}$$

$$d_{95\cdot 1} = \ell_{95\cdot 1} \tag{3.7}$$

$$\hat{q}_{95\cdot 1} = 1.00000 \tag{3.8}$$

$$T_{i.1} = L_{i.1} + \dots + L_{95\cdot 1}\ . \tag{3.9}$$

$$\hat{e}_{i.1} = \frac{T_{i.1}}{\ell_{i.1}} \tag{3.11}$$

CHAPTER 9

		page no.
x	- The number of years since admission to a follow-up study.	195
N_x	- The number of patients alive at the beginning of the interval $(x,x+1)$, $N_x = m_x + n$.	195
m_x	- The number of patients who entered a follow-up study more than $x+1$ years before the closing date who will be observed for the entire interval $(x,x+1)$.	196
n_x	- The number of patients who entered the study less than $x+1$ years before the closing date and are due to withdraw in the interval $(x,x+1)$.	196
d_x	- The number of patients among m_x dying in the interval $(x,x+1)$.	196
s_x	- The number of patients among m_x surviving to the end of the interval $(x,x+1)$.	197
d'_x	- The number of patients among n_x who will die before the time of withdrawal.	197
w_x	- The number of patients who survive to the time of withdrawal.	197
$\hat{p}_{0x}$	- Estimate of the probability of surviving from admission to the interval $(0,x)$.	201
$\hat{e}_\alpha$	- Estimate of the expectation of life at $x=\alpha$.	202
$\mu(t)$	- Total force of mortality at time t.	213
$\mu(t;\delta)$	- The force of mortality for risk R_δ at time t.	213
$d_{x\delta}$	- The number of patients among d_x dying from risk R_δ.	216
$d'_{x\delta}$	- The number among d'_x dying from risk R_δ.	217

<u>Formulas</u>

Binomial distribution among those not due for withdrawal in $(x,x+1)$:

$$C_1 \, p_x^{s_x} (1-p_x)^{d_x} \tag{2.1}$$

$$E(s_x | m_x) = m_x p_x \quad \text{and} \quad E(d_x | m_x) = m_x(1-p_x) \tag{2.2}$$

Binomial distribution among those due for withdrawal in (x,x+1):

$$c_2 \; p_x^{\frac{1}{2}w_x}(1-p_x^{\frac{1}{2}})^{d'_x} \tag{2.5}$$

$$E(w_x|n_x) = n_x p_x^{\frac{1}{2}} \qquad \text{and} \qquad E(d'_x|n_x) = n_x(1-p_x^{\frac{1}{2}}) \quad , \tag{2.6}$$

Estimate of the probability of survival p_x and its sample variance:

$$\hat{p}_x = \left[\frac{-\frac{1}{2}d'_x + \sqrt{\frac{1}{4}d'^2_x + 4(N_x-\frac{1}{2}m_x)(s_x + \frac{1}{2}w_x)}}{2(N_x - \frac{1}{2}m_x)}\right]^2 \tag{2.8}$$

$$\hat{q}_x = 1-\hat{p}_x \quad , \qquad x=0,1,\cdots,y-1. \tag{2.9}$$

$$S^2_{\hat{p}_x} = \frac{\hat{p}_x \, \hat{q}_x}{M_x} \tag{2.10}$$

where

$$M_x = m_x + n_x(1+\hat{p}_x^{\frac{1}{2}})^{-1} \tag{2.11}$$

Estimate of survival probability p_{0x} and its sample variance:

$$\hat{p}_{0x} = \hat{p}_0\hat{p}_1\cdots\hat{p}_{x-1} \quad , \qquad\qquad x=1,2,\cdots,y. \qquad (2.12)$$

$$S^2_{\hat{p}_{0x}} = \hat{p}_{0x}^{\ 2} \sum_{u=0}^{x-1} \hat{p}_u^{\ -2}\, S_{\hat{p}_u}^{\ 2} \quad . \qquad (2.13)$$

Estimate of expectation of life and its sample variance:

$$\hat{e}_\alpha = \tfrac{1}{2} + \hat{p}_\alpha + \hat{p}_\alpha\hat{p}_{\alpha+1} + \cdots + \hat{p}_\alpha\hat{p}_{\alpha+1}\cdots\hat{p}_{y-1} + \hat{p}_{\alpha y}\left(\frac{\hat{p}_t}{1-\hat{p}_t}\right) \quad , \qquad (2.19)$$

$$S^2_{\hat{e}_\alpha} = \sum_{\substack{x=\alpha \\ x\neq t}}^{y-1} \hat{p}_{\alpha x}^{\ 2}\left[\hat{e}_{x+1} + \tfrac{1}{2}\right]^2 S^2_{\hat{p}_x} + \hat{p}_{\alpha t}^{\ 2}\left[\hat{e}_{t+1} + \tfrac{1}{2} + \frac{\hat{p}_{ty}}{(1-\hat{p}_t)^2}\right]^2 S^2_{\hat{p}_t}, \quad \alpha\leq t, \qquad (2.23)$$

$$S^2_{\hat{e}_\alpha} = \sum_{x=\alpha}^{y-1} \hat{p}_{\alpha x}^{\ 2}\left[\hat{e}_{x+1} + \tfrac{1}{2}\right]^2 S^2_{\hat{p}_x} + \frac{\hat{p}^2_{\ \alpha y}}{(1-\hat{p}_t)^4} S^2_{\hat{p}_t} \quad , \qquad \alpha>t. \qquad (2.24)$$

Forces of mortality:

$$\mu(\tau;1) + \cdots + \mu(\tau;r) = \mu(\tau) \qquad (3.1)$$

Crude probability of dying:

$$Q_{x\delta}(t) = \frac{\mu(x;\delta)}{\mu(x)}\Big[1 - p_x(t)\Big], \quad 0<t\leq 1; \quad \delta=1,\cdots,r. \tag{3.2}$$

$$Q_{x1}(t) + \cdots + Q_{xr}(t) + p_x(t) = 1\,, \qquad 0<t\leq 1. \tag{3.3}$$

$$Q_{x\delta}(\tfrac{1}{2}) = \frac{\mu(x;\delta)}{\mu(x)}\Big[1 - p_x^{\frac{1}{2}}\Big] = Q_{x\delta}\Big[1 + p_x^{\frac{1}{2}}\Big]^{-1}, \quad \delta=1,\cdots,r. \tag{3.4}$$

Net probabilities of dying:

$$q_{x\delta} = Q_{x\delta}[1+\tfrac{1}{2}(q_x-Q_{x\delta}) + \frac{1}{6}(q_x-Q_{x\delta})(2q_x-Q_{x\delta})] \quad ; \tag{3.6}$$

$$q_{x\cdot\delta} = (q_x-Q_{x\delta})[1 + \tfrac{1}{2}Q_{x\delta} + \frac{1}{6}Q_{x\delta}(q_x+Q_{x\delta})] \quad , \delta=1,\ldots,r, \tag{3.7}$$

Partial crude probability of dying:

$$Q_{x\delta\cdot 1} = Q_{x\delta}[1 + \tfrac{1}{2}Q_{x\delta} + \frac{1}{6}Q_{x1}(q_x+Q_{x1})] \quad , \tag{3.8}$$

Multinomial distribution among those not due for withdrawal in (x,x+1):

$$C_1 \, p_x^{s_x} \; Q_{x1}^{d_{x1}} \cdots Q_{xr}^{d_{xr}} \quad , \tag{3.10}$$

$$E(s_x \mid m_x) = m_x p_x \qquad \text{and} \qquad E(d_{x\delta} \mid m_x) = m_x Q_{x\delta} \tag{3.11}$$

Multinomial distribution among those due for withdrawal in (x,x+1):

$$C_2 \, p_x^{\frac{1}{2} w_x} \prod_{\delta=1}^{r} \left[Q_{x\delta} \;\; (1 + p_x^{\frac{1}{2}})^{-1} \right]^{d'_{x\delta}} \quad , \tag{3.13}$$

$$E(w_x \mid n_x) = n_x p_x^{\frac{1}{2}} \qquad \text{and} \qquad E(d'_{x\delta} \mid n_x) = n_x Q_{x\delta} (1 + p_x^{\frac{1}{2}})^{-1} \tag{3.14}$$

Estimates of probabilities of dying:

$$\hat{Q}_{x\delta} = \frac{D_{x\delta}}{D_x} \hat{q}_x \quad , \qquad \begin{array}{l} \delta = 1,2,\cdots,r, \\ x = 0,1,\cdots,y-1. \end{array} \tag{3.18}$$

$$\hat{q}_{x\delta} = \hat{Q}_{x\delta} [1 + \tfrac{1}{2}(\hat{q}_x - \hat{Q}_{x\delta}) + \frac{1}{6} [\hat{q}_x - \hat{Q}_{x\delta})(2\hat{q}_x - \hat{Q}_{x\delta})] \tag{3.19}$$

$$\hat{q}_{x \cdot \delta} = (\hat{q}_x - \hat{Q}_{x\delta}) [1 + \tfrac{1}{2} \hat{Q}_{x\delta} + \frac{1}{6} \hat{Q}_{x\delta} (\hat{q}_x + \hat{Q}_{x\delta})] \quad , \; \delta = 1, \ldots, r, \tag{3.20}$$

and

$$\hat{Q}_{x\delta \cdot 1} = \hat{Q}_{x\delta} [1 + \tfrac{1}{2} \hat{Q}_{x\delta} + \frac{1}{6} \hat{Q}_{x1} (\hat{q}_x + \hat{Q}_{x1})] \quad , \; \delta = 2, \ldots, r; \qquad x = 0, 1, \ldots, y-1 \; . \tag{3.21}$$

Lost cases:

$\mu(x;r)\Delta + o(\Delta)$ = Pr{a patient will be lost to the study in $(\tau, \tau+\Delta)$ due to follow-up failure},

$$x < \tau < x+1 \ . \tag{4.1}$$

p_x = Pr{ a patient alive at time x will remain alive and under observation at time x+1}. (4.2)

$q_x = 1-p_x$ = Pr{a patient alive at time x will either die or be lost to the study due to follow-up failure in interval (x,x+1)}. (4.3)

Q_{xr} = Pr{a patient alive at time x will be lost to the study in (x,x+1)}. (4.4)

$q_{x.r}$ = Pr{a patient alive at time x will die in interval (x,x+1) if the risk R_r of being lost is eliminated}. (4.5)

$1-q_{x.r}$ = Pr{a patient alive at x will survive to time x+1 if the risk R_r of being lost is eliminated}. (4.6)

$Q_{x\delta.r}$ = Pr{a patient alive at x will die in (x,x+1) from risk R_δ if the risk R_r of being lost is eliminated}. (4.7)

APPENDIX I

page no.

a_i - The fraction of last age interval of life. The expected fraction of the interval (x_i, x_i+n_i) lived by an individual who dies at an age included in the interval 227

τ_i - The fraction of the interval (x_i, x_i+n_i) lived by an individual who dies at an age included in the interval. τ_i is a random variable whose expectation is a_i, or $E(\tau_i) = a_i$. 227

M_i - Age specific death rate. The ratio of the observed number of deaths (D_i) to the total number of years lived in the interval (x_i, x_i+n_i) by those who are alive at x_i. M_i is a random variable. 227

m_i - (Theoretical) age specific death rate. The ratio of the expected number of deaths to the expected number of years lived in the interval (x_i, x_i+n_i) by those who are alive at x_i. m_i is an unknown theoretical value. 227

q_i - Probability of dying in interval (x_i, x_i+n_i). 227

$\mu(x)$ - Force of mortality (mortality intensity function) age age x. 227

Formulas

Relationship between q_i and m_i:

$$q_i = 1-\exp\{-\int_0^{n_i} \mu(x_i+\xi)\, d\xi\} \tag{1}$$

$$m_i = \frac{1-\exp\{-\int_0^{n_i} \mu(x_i+\xi)\, d\xi\}}{\int_0^{n_i} \exp\{-\int_0^{y} \mu(x_i+\xi)\, d\xi\} dy} \tag{2}$$

$$g(t)dt = \frac{[\exp\{-\int_0^{n_i t} \mu(x_i+\xi) d\xi\}]\ \mu(x_i+n_i t) n_i\ dt}{q_i} \tag{4}$$

$$0 \leq t \leq 1$$

$$a_i = E(\tau_i) = \int_0^1 tg(t)dt \qquad (6)$$

$$q_i = \frac{n_i m_i}{1+(1-a_i)n_i m_i} \qquad (9)$$

$$q_x = \frac{m_x}{1+(1-a_x')m_x} \qquad (10)$$

APPENDIX II

		page no.
Y_α	- The future life time beyond age x_α. This is a random variable.	247
e_α	- The true expectation of life byond age x_α. Y_α is a random variable whose expectation is e_α, or $E(Y_\alpha) = e_\alpha$.	248
$f(y_\alpha)$	- Probability density function of Y_α. The product $f(y_\alpha)dy_\alpha$ is the probability that an individual alive at x_α will survive the period $(x_\alpha, x_\alpha+y_\alpha)$ and then die in the interval $(x_\alpha+y_\alpha, (x_\alpha+y_\alpha+dy_\alpha)$	247
X	- The life span of an individual. It is a continuous random variable.	234
$F_X(x)$	- Distribution function of the length of life X. It is the probability of dying prior to, or at, age x.	234
ℓ_x	- The number of individuals surviving to age x.	234
p_{ij}	- Probability that an individual alive at age x_i will survive to age x_j, for $i \leq j$.	232
$1-p_{ij}$	- Probability that an individual alive at age x_i will die before age x_j.	232
p_{0x}	- Probability that one individual alive at age 0 will survive to age x.	235
$p_{0i}q_i$	- Probability that an individual alive at age x_0 will die in the interval (x_i, x_{i+1}), $i=0,1,\ldots,w$.	242
$p_{\alpha i}q_i$	- Probability that an individual alive at age x_α will die in the interval (x_i, x_{i+1}) subsequent to x_α.	243
ρ_{ℓ_i,ℓ_j}	- Correlation between ℓ_i and ℓ_j.	241
σ_{ℓ_i,ℓ_j}	- Covariance between ℓ_i, ℓ_j	241

Formulas

Distribution of the length of life X and the number of survivors ℓ_x:

$$\mu(x)\Delta + o(\Delta) = \Pr\{\text{an individual alive at age } x \text{ will die in interval } (x, x+\Delta)\} \tag{2.1}$$

$$F_X(x) = \Pr\{X \leq x\} \tag{2.2}$$

$$1 - F_X(x) = e^{-\int_0^x \mu(t)dt} = p_{0x} \tag{2.7}$$

$$\Pr\{\ell_x = k\} = \frac{\ell_0!}{k!(\ell_0-k)!} p_{0x}^k (1-p_{0x})^{\ell_0-k}, \quad k=0,1,\ldots,\ell_0 \tag{2.8}$$

$$E(\ell_i|\ell_0) = \ell_0 p_{0i} \tag{2.11}$$

$$\sigma^2_{\ell_i|\ell_0} = \ell_0 p_{0i}(1-p_{0i}) \tag{2.12}$$

$$p_{ij} = \exp\left\{-\int_{x_i}^{x_j} \mu(\tau)d\tau\right\}, \quad \text{for } i\leq j \tag{2.13}$$

$$f_X(x) = \frac{dF_X(x)}{dx} = \mu(x)e^{-\int_0^x \mu(t)dt} \qquad x\geq 0 \tag{2.20}$$

$$= 0 \qquad x<0$$

Gompertz distribution:

$$\mu(t) = Bc^t \tag{2.23}$$

$$f(x) = Bc^x e^{-B[c^x-1]/\ln c} \tag{2.24}$$

$$F_X(x) = 1 - \exp\{-\frac{B}{\ln c}(c^x-1)\} \tag{2.25}$$

Makeham distribution:

$$\mu(t) = A + B\ c^t \qquad (2.26)$$

$$f(x) = [A+Bc^x]\ \exp\{-[Ax+B(c^x-1)/\ln\ c]\} \qquad (2.27)$$

$$F_X(x) = 1 - \exp\{-[Ax+B(c^x-1)/\ln\ c]\} \qquad (2.28)$$

Weibull distribution:

$$\mu(t) = \mu a t^{a-1} \qquad (2.29)$$

$$f(x) = \mu a x^{a-1} e^{-\mu x^a} \qquad (2.29a)$$

$$F_X(x) = 1-e^{-\mu x^a} \qquad (2.30)$$

Exponential distribution:

$\mu(t) = \mu$ is a constant

$$f(x) = \mu e^{-\mu x} \qquad (2.31)$$

$$F_X(x) = 1-e^{-\mu x} \qquad (2.32)$$

Joint distribution of $\ell_1, \ell_2, \ldots, \ell_u$, and their correlation:

$$\Pr\{\ell_1=k_1, \ell_2=k_2, \ldots, \ell_u=k_u \mid \ell_0\} = \prod_{i=0}^{u-1} \frac{k_i!}{k_{i+1}!(k_i-k_{i+1})!} p_i^{k_{i+1}} (1-p_i)^{k_i-k_{i+1}}$$

$$k_{i+1} = 0,1,\ldots,k_i, \quad \text{with } k_0=\ell_0 \tag{3.3}$$

$$\rho_{\ell_i,\ell_j} = \frac{p_{0j}(1-p_{0i})}{\sqrt{p_{0i}(1-p_{0i})p_{0j}(1-p_{0j})}} = \sqrt{\frac{p_{0j}(1-p_{0i})}{p_{0i}(1-p_{0j})}} \tag{3.8}$$

$$p_{00}q_0 + \cdots + p_{0w}q_w = 1, \tag{4.2}$$

Joint distribution of $d_0, d_1, \ldots, d_w$, and their correlation:

$$\Pr\{d_0 = \delta_0, \cdots, d_w = \delta_w\} = \frac{\ell_0!}{\delta_0! \cdots \delta_w!} (p_{00}q_0)^{\delta_0} \cdots (p_{0w}q_w)^{\delta_w} \tag{4.3}$$

$$E(d_i \mid \ell_0) = \ell_0 p_{0i} q_i \tag{4.4}$$

$$\sigma^2_{d_i} = \ell_0 p_{0i} q_i (1-p_{0i}q_i) \tag{4.5}$$

$$\sigma_{d_i,d_j} = -\ell_0 p_{0i} q_i p_{0j} q_j \qquad \underline{\text{for}} \quad i \neq j;\ i,j=0,1,\cdots,\underline{w}. \tag{4.6}$$

Maximum likelihood estimates of $p_0, p_1, \ldots, p_u$:

$$L = \prod_{i=0}^{u-1} \frac{\ell_i!}{\ell_{i+1}!(\ell_i-\ell_{i+1})!} p_i^{\ell_{i+1}} (1-p_i)^{\ell_i-\ell_{i+1}} \tag{5.2}$$

$$\hat{p}_j = \frac{\ell_{j+1}}{\ell_j} \qquad j = 0,1,\ldots,u-1 \tag{5.5}$$

$$E[\hat{p}_j] = E\left(\frac{\ell_{j+1}}{\ell_j}\right) = E\left[\frac{1}{\ell_j} E(\ell_{j+1} \mid \ell_j)\right] = p_j \tag{5.6}$$

$$\sigma^2_{\hat{p}_j} = E\left(\frac{1}{\ell_j}\right) p_j(1-p_j) = \sigma^2_{\hat{q}_j} \tag{5.8}$$

When ℓ_0 is large, $$\sigma^2_{\hat{p}_j} = \frac{1}{E(\ell_j)} \; p_j(1-p_j) \tag{5.9}$$

$$\sigma_{\hat{p}_j,\hat{p}_k} = 0 \tag{5.11}$$

$$\sigma_{\hat{p}_{\alpha j},\hat{p}_{\alpha k}} = E\left(\frac{1}{\ell_\alpha}\right) p_{\alpha k}(1-p_{\alpha j}) \qquad \alpha < j \le k \tag{5.12}$$

Expectation of life and its estimate:

$$e_\alpha = \int_0^\infty y_\alpha f(y_\alpha)dy_\alpha = \int_0^\infty y_\alpha e^{-\int_{x_\alpha}^{x_\alpha+y_\alpha}\mu(\tau)d\tau} \mu(x_\alpha+y_\alpha)dy_\alpha \tag{6.6}$$

$$\sigma^2_{Y_\alpha} = \int_0^\infty (y_\alpha-e_\alpha)^2 \; f(y_\alpha)dy_\alpha \tag{6.7}$$

$$\bar{Y}_\alpha = \hat{e}_\alpha \tag{6.13}$$

$$\hat{e}_\alpha = a_\alpha n_\alpha + \sum_{i=\alpha+1}^{w} c_i \frac{\ell_i}{\ell_\alpha} = a_\alpha n_\alpha + \sum_{i=\alpha+1}^{w} c_i \, \hat{p}_{\alpha i} \tag{6.15}$$

$$e_\alpha = a_\alpha n_\alpha + \sum_{i=\alpha+1}^{w} c_i \, p_{\alpha i} \, , \qquad \alpha=0,1,\cdots,w. \tag{6.16}$$

$$\sigma^2_{\hat{e}_\alpha} = \sum_{i=\alpha}^{w-1} p_{\alpha i}^2\{e_{i+1} + (1-a_i)n_i\}^2 \, \sigma^2_{\hat{q}_i} \qquad \alpha=0,1,\cdots,w-1. \tag{6.21}$$

APPENDIX III

		page no.
$Q_{i\delta}$	– Crude probability of dying from risk R_δ. $Q_{i\delta}$ = Pr{an individual alive at age x_i will die in the interval (x_i, x_{i+1}) from R_δ in the presence of other competing risks operating in the population}	259
$q_{i\delta}$	– Net probability of dying from risk R_δ. $q_{i\delta}$ = Pr{an individual alive at age x_i will die in the interval (x_i, x_{i+1}) when R_δ is the only risk operating in the population}	259
$q_{i.\delta}$	– Net probability of dying when R_δ is eliminated as a risk of death. $q_{i.\delta}$ = Pr{an individual alive at age x_i will die in the interval (x_i, x_{i+1}) when R_δ is eliminated as a risk of death}	259
$Q_{i\delta.1}$	– Partial crude probability of dying $Q_{i\delta.1}$ = Pr{an individual alive at age x_i will die in the interval (x_i, x_{i+1}) from R_δ when R_1 is eliminated as a risk of death}	259
$Q_{i\delta.12}$	– Partial crude probability of dying $Q_{i\delta.12}$ = Pr{an individual alive at age x_i will die in the interval (x_i, x_{i+1}) from R_δ when R_1 and R_2 are eliminated as risks of death}	259
p_i	– Probability of surviving the interval (x_i, x_{i+1})	259
q_i	– Probability of dying in the interval (x_i, x_{i+1})	259
$R_{\delta\varepsilon}$	– Interaction between risks R_δ and R_ε.	270
$\mu(t;\delta)$	– Force of mortality associated with risk R_δ, $\delta = 1,\ldots,r$.	261
$\mu(t)$	– Total force of mortality. $\mu(t) = \mu(t;1) + \ldots + \mu(t;r)$	261
$\mu(t;\delta,\varepsilon)$	– Force of mortality associated with the interaction $R_{\delta\varepsilon}$.	270

Formulas

Relationship between three types of probabilities:

$$\mu(t;\delta)\Delta + o(\Delta) = \Pr\{\text{an individual alive at time } t \text{ will die in interval } (t, t+\Delta) \text{ from risk } R_\delta\},\ \delta=1,\cdots,r \tag{2.1}$$

$$\mu(t;1) + \cdots + \mu(t;r) = \mu(t) \tag{2.2}$$

Proportionality Assumption: $$\frac{\mu(t;\delta)}{\mu(t)} = c_{i\delta} \tag{2.3}$$

$$Q_{i1} + \cdots + Q_{ir} = q_i \qquad i=0,1,\cdots. \tag{2.10}$$

$$q_{i\delta} = 1 - p_i^{Q_{i\delta}/q_i} \qquad \delta=1,\cdots,r. \tag{2.17}$$

When q_i is extremely small, $$q_{i\delta} = Q_{i\delta}\left[1 + \frac{1}{2}(q_i - Q_{i\delta})\right] \tag{2.21 a}$$

$$q_{i\cdot\delta} = 1 - p_i^{(q_i-Q_{i\delta})/q_i} \quad . \tag{2.27}$$

$$q_{i\cdot\delta} = (q_i-Q_{i\delta})\left[1 + \frac{1}{2}Q_{i\delta} + \frac{1}{6}Q_{i\delta}(q_i+Q_{i\delta})\right] \tag{2.29}$$

$$Q_{i\delta\cdot 1} = Q_{i\delta}\left[1 + \frac{1}{2}Q_{i1} + \frac{1}{6}Q_{i1}(q_i+Q_{i1})\right] \qquad \delta=2,\ldots,r. \tag{2.33}$$

$$\mu(t;\delta,\varepsilon)\Delta + o(\Delta) = \Pr\{\text{an individual alive at time } t \text{ will die in interval } (t,t+\Delta) \text{ from } R_{\delta\varepsilon}\} \tag{3.3}$$

$$\sum_{\delta=1}^{r}\mu(t;\delta) + \sum_{\delta=1}^{r-1}\sum_{\varepsilon=\delta+1}^{r}\mu(t;\delta,\varepsilon) = \mu(t) \tag{3.5}$$

$$q_{i\cdot 1} = 1 - p_i^{(q_i - Q_{i1} - \sum_{2}^{r} Q_{i1\varepsilon})/q_i} \tag{3.10}$$

$$q_{i\cdot 1} = (q_i - Q_{i1} - \sum_{\varepsilon=2}^{r} Q_{i1\varepsilon}) \; [1 + \frac{1}{2} (Q_{i1} + \sum_{\varepsilon=2}^{r} Q_{i1\varepsilon}) + \frac{1}{6}(Q_{i1} + \sum_{2}^{r} Q_{i1\varepsilon})(q_i + Q_{i1} + \sum_{2}^{r} Q_{i1\varepsilon})] \tag{3.12}$$

$$Q_{i\delta\cdot 1} = \frac{Q_{i\delta}}{q_i - Q_{i1} - \sum_{\varepsilon=2}^{r} Q_{i1\varepsilon}} \, q_{i\cdot 1} \tag{3.16}$$

APPENDIX IV

		page no.
$d_{i\delta}$	- Number of deaths from cause R_δ in the interval (x_i, x_{i+1}).	277
$\rho_{d_{i\delta}, d_{i\varepsilon} \mid \ell_i}$	- Correlation coefficient between $d_{i\delta} d_{i\varepsilon}$, given ℓ_i.	279
$\rho_{d_{i\delta}, d_{i\varepsilon}}$	- Correlation coefficient between $d_{i\delta}$ and $d_{i\varepsilon}$.	283
$S_{\hat{Q}_{i\delta}}$	- Standard error of $\hat{Q}_{i\delta}$.	288

Formulas

$$d_{i1}+\cdots+d_{ir} = d_i \tag{1.1}$$

$$\ell_i = d_{i1}+\cdots+d_{ir}+\ell_{i+1} \tag{1.3}$$

$$Q_{i1}+\cdots+Q_{ir} = q_i \tag{1.4}$$

$$1 = Q_{i1}+\cdots+Q_{ir}+p_i \tag{1.6}$$

Joint probability distribution of $d_{i1},\ldots,d_{ir},\ell_{i+1}$ given ℓ_i:

$$\frac{\ell_i!}{d_{i1}!\cdots d_{ir}!\,\ell_{i+1}!} \; Q_{i1}^{\;d_{i1}}\cdots Q_{ir}^{\;d_{ir}} \; p_i^{\;\ell_{i+1}} \tag{1.7}$$

where $d_{i1}+\ldots+d_{ir}+\ell_{i+1} = \ell_i$.

$$E(d_{i\delta}|\ell_i) = \ell_i Q_{i\delta} \tag{1.8}$$

$$\mathrm{Var}(d_{i\delta}|\ell_i) = \ell_i Q_{i\delta}(1-Q_{i\delta}), \quad \delta=1,\cdots,r. \tag{1.9}$$

$$\rho_{d_{i\delta},d_{i\varepsilon}|\ell_i} = -\sqrt{\frac{Q_{i\delta}p_i}{(1-Q_{i\delta})(1-Q_{i\varepsilon})}} \tag{1.11}$$

$$\rho_{d_{i\delta},\ell_{i+1}|\ell_i} = -\sqrt{\frac{Q_{i\delta}p_i}{(1-Q_{i\delta})(1-p_i)}} \tag{1.13}$$

$$\mathrm{Var}(\ell_{i+1}|\ell_i) = \ell_i p_i q_i \tag{1.15}$$

Joint probability distribution of all the random variables $d_{i1},\ldots,d_{ir},\ell_{i+1}$, for $i=0,1,\ldots,u$, given ℓ_0:

$$\prod_{i=0}^{u} \frac{\ell_i!}{d_{i1}!\ldots d_{ir}!\ \ell_{i+1}!} \cdot Q_{i1}^{\ d_{i1}} \ldots Q_{ir}^{\ d_{ir}}\ p_i^{\ \ell_{i+1}} \tag{2.1}$$

$$E(d_{i\delta}) = E(\ell_i)Q_{i\delta} = \ell_0 p_{0i} Q_{i\delta} \tag{2.2}$$

$$\mathrm{Var}(d_{i\delta}) = \ell_0\ p_{0i}\ Q_{i\delta}(1-p_{0i}Q_{i\delta}), \quad \begin{array}{l}\delta=1,\ldots,r;\\ i=0,\ldots,u.\end{array} \tag{2.6}$$

$$\mathrm{Cov}(d_{i\delta},d_{i\varepsilon}) = -\ell_0 p_{0i}Q_{i\delta}p_{0i}Q_{i\varepsilon}, \quad \begin{array}{l}\delta\neq\varepsilon;\ \delta,\ \varepsilon=1,\ldots,r;\\ i=0,\ldots,u.\end{array} \tag{2.7}$$

$$\rho_{d_{i\delta},d_{i\varepsilon}} = -\ p_{0i}\sqrt{\frac{Q_{i\delta}}{1-p_{0i}Q_{i\delta}}}\sqrt{\frac{Q_{i\varepsilon}}{1-p_{0i}Q_{i\varepsilon}}} \tag{2.9}$$

$$\rho_{d_{i\delta},d_{j\varepsilon}} = -\ p_{0i}\ \sqrt{p_{ij}}\sqrt{\frac{Q_{i\delta}}{1-p_{0i}Q_{i\delta}}}\sqrt{\frac{Q_{j\varepsilon}}{1-p_{0j}Q_{j\varepsilon}}}\ ,\ \text{for } i<j \tag{2.12}$$

$$\mathrm{Cov}(d_{i\delta},\ell_j) = -\ \ell_0 p_{0i} Q_{i\delta} p_{0j} \tag{2.13}$$

$$\mathrm{Cov}(\ell_i,\ d_{j\delta}) = \ell_0 (1-p_{0i}) p_{0j} Q_{j\delta} \qquad \begin{matrix} \delta = 1,\ldots,r; \\ i<j;\ i,j=0,1,\ldots. \end{matrix} \tag{2.14}$$

$$\sigma_{\ell_i,\ell_j} = \ell_0 (1-p_{0i}) p_{0j} \qquad i < j, \tag{2.15}$$

Maximum likelihood estimates of p_i, $Q_{i\delta}$:

$$L = \prod_{i=0}^{u} \frac{\ell_i!}{d_{i1}! \ldots d_{ir}!\ \ell_{i+1}!}\ Q_{i1}^{\ d_{i1}} \ldots Q_{ir}^{\ d_{ir}}\ p_i^{\ \ell_{i+1}} \tag{3.1}$$

$$\hat{Q}_{i\delta} = \frac{d_{i\delta}}{\ell_i}, \qquad \begin{matrix} \delta=1,\ldots,r \\ i=0,\ldots,u \end{matrix} \tag{3.7}$$

$$\hat{p}_i = \frac{\ell_{i+1}}{\ell_i}, \qquad i=0,\ldots,u \tag{3.8}$$

$$E[\hat{Q}_{i\delta}] = Q_{i\delta} \tag{3.10}$$

$$\mathrm{Var}(\hat{Q}_{i\delta}) = E(\frac{1}{\ell_i})\ Q_{i\delta}(1-Q_{i\delta}), \qquad \begin{matrix} \delta=1,\ldots,r \\ i=0,\ldots,u \end{matrix} \tag{3.15}$$

$$\mathrm{Var}(\hat{p}_i) = \mathrm{Var}(\hat{q}_i) = E(\frac{1}{\ell_i}) p_i q_i \tag{3.16}$$

$$\mathrm{Cov}(\hat{Q}_{i\delta},\ \hat{Q}_{i\varepsilon}) = -E(\frac{1}{\ell_i})\ Q_{i\delta}\ Q_{i\varepsilon} \tag{3.17}$$

$$\mathrm{Cov}(\hat{Q}_{i\delta},\ \hat{p}_i) = -E(\frac{1}{\ell_i})\ p_i\ Q_{i\delta} \tag{3.18}$$

$$S_{\hat{Q}_{i\delta}} = \sqrt{\frac{1}{\ell_i}\ \hat{Q}_{i\delta}(1-\hat{Q}_{i\delta})} \qquad \delta=1,\ldots,r \tag{3.20}$$

$$S_{\hat{q}_i} = \sqrt{\frac{1}{\ell_i}\ \hat{q}_i(1-\hat{q}_i)} \qquad \ell=0,\ldots,u\ . \tag{3.21}$$