© Journal of the American Statistical Association
September 1972, Volume 67, Number 339
Applications Section

On Constructing Current Life Tables

CHIN LONG CHIANG*

The purpose of this note is to reintroduce a simple method of constructing current life tables proposed in [2], using the concept of the fraction of the last age interval of life, and to use it to construct life tables of the United States population.

1. INTRODUCTION

The main problem in developing a method of constructing current life tables is to derive a formula that expresses the probability of death (q_i) in terms of the corresponding death rate (m_i) for each age group. Several methods of obtaining such a formula have been proposed in the literature, including those of King [9], Greville [7], Reed and Merrell [10], Keyfitz [8], Sirken [11], Coale and Demeny [5] and Chiang [2]. However, most of these methods are so complex mathematically and conceptually that users in applied fields cannot appreciate the real value of life tables as a means of summarizing mortality experience, and theoretical statisticians find it difficult to consider the life table as an area of research. The method suggested in [2] was intended to resolve the problem. However, it was introduced informally and without a theoretical basis.

The purpose of the present note is to reintroduce the method from a theoretical viewpoint and to use it to construct life tables of the United States population. The outline of such a table is shown in Table 1. All the quantities

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 age 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	$\hat{e}_w$

in the life table, with the exception of ℓ_0 and a_i, are random variables. The probability distributions and statistical properties of these random variables have been discussed in detail in [1] and [3]; the problem of estimation has been treated in Grenander [6].

1.1. The method suggested in [2] is described here for easy reference. The observed data from a current population consist of the number of deaths (D_i) and the corresponding mid-year population (P_i) for each age group (x_i, x_i+n_i), where $n_i=x_{i+1}-x_i$ is the length of the interval. Using the total United States 1967 population in Table 2 as an example, for age interval (1, 5), $x_i=1$, $n_i=4$, the estimated population size on July 1, 1967 is $P_i=15{,}633{,}000$, and the number of deaths occurring during the year 1967 is $D_i=13{,}506$.

Suppose that there are N_i people of exact age x_i subject to the probability q_i of dying in the age interval (x_i, x_i+n_i), among whom D_i deaths actually occur. Then the estimate of q_i is given by

$$\hat{q}_i = \frac{D_i}{N_i} \cdot \tag{1.1}$$

The (observed) age specific death rate is defined as the ratio of D_i to the total number of years lived by N_i in the interval (x_i, x_i+n_i). Each of the D_1 people lives, on the average, a fraction a_i of the interval (x_i, x_i+n_i) before death occurring; as a group, the D_i people lived $a_in_iD_i$ years. The group of (N_i-D_i) survivors lived $n_i(N_i-D_i)$ years. Therefore, the death rate is

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

Conventionally, the denominator in (1.2) is estimated by the midyear population,

$$n_i(N_i - D_i) + a_in_iD_i = P_i, \tag{1.3}$$

to give the familiar formula

$$M_i = \frac{D_i}{P_i} \cdot \tag{1.4}$$

Solving (1.3) for N_i gives an estimate

$$N_i = \frac{1}{n_i}\left[P_i + (1 - a_i)n_iD_i\right]. \tag{1.5}$$

Substituting (1.5) in (1.1) and using (1.4) we obtain the basic relation between $\hat{q}_i$ and M_i,

$$\hat{q}_i = \frac{n_iM_i}{1 + (1 - a_i)n_iM_i} \cdot \tag{1.6}$$

Using a similar reasoning as in (1.3), we have

$$L_i = n_i(\ell_i - d_i) + a_in_id_i. \tag{1.7}$$

* Chin Long Chiang is professor of biostatistics; head, Division of Measurement Sciences; and co-chairman, Group in Biostatistics, University of California, Berkeley, Calif. 94720.

When $\hat{q}_i$ and ℓ_i for each age interval are obtained, the rest of the table can be determined as follows:

$$d_i = \ell_i q_i \text{ and } \ell_{i+1} = \ell_i - d_i, \quad \text{for } i = 0, 1, \cdots, w-1, \tag{1.8}$$

$$T_i = L_i + L_{i+1} + \cdots + L_w, \quad \text{for } i = 0, 1, \cdots, w-1, \tag{1.9}$$

and

$$\hat{e}_i = T_i / \ell_i, \quad \text{for } i = 0, 1, \cdots, w-1. \tag{1.10}$$

The quantities in the last age interval are computed from the equations:

$$\hat{e}_w = \frac{1}{M_w}, \quad T_w = L_w = \ell_w \hat{e}_w, \quad \text{and} \quad d_w = \ell_w. \tag{1.11}$$

2. CONSTRUCTION OF ABRIDGED LIFE TABLE FOR TOTAL UNITED STATES POPULATION, 1967

Age interval (in years)	Midyear population in interval (x_i, x_{i+1})[a]	Number of deaths in interval (x_i, x_{i+1})[b]	Death rate in interval (x_i, x_{i+1})	Fraction of last age interval of life	Proportion 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	3535000	79028	.022356	.09	.02191
1 - 5	15633000	13506	.000864	.39	.00345
5 - 10	20908000	8809	.000421	.46	.00210
10 - 15	19889000	8084	.000406	.54	.00203
15 - 20	17858000	18168	.001017	.57	.00507
20 - 25	15178000	19538	.001287	.49	.00641
25 - 30	12108000	16355	.001351	.50	.00673
30 - 35	10978000	18431	.001679	.52	.00836
35 - 40	11616000	28382	.002443	.54	.01215
40 - 45	12382000	45657	.003687	.54	.01828
45 - 50	11848000	68247	.005760	.54	.02842
50 - 55	10791000	96794	.008970	.53	.04392
55 - 60	9529000	130937	.013741	.52	.06651
60 - 65	8056000	163225	.020261	.52	.09660
65 - 70	6507000	199615	.030677	.52	.14286
70 - 75	5178000	238304	.046022	.51	.20679
75 - 80	3787000	250552	.066161	.51	.28466
80 - 85	2160000	219117	.101443	.48	.40135
85+	1173000	227987	.194362		1.00000

[a] U.S. Dept. of Commerce, Bureau of the Census, *Current Population Reports*, Series P-25, No. 441 (March 19, 1970), 14–5.

[b] U.S. Dept. of Health, Education and Welfare, Public Health Service, National Center for Health Statistics, *Vital Statistics of the U.S.*, 1967, Vol. II, Part B, Tables 7-5, 7–114, 7-115.

2. ABRIDGED LIFE TABLE FOR THE TOTAL UNITED STATES POPULATION, 1967

An important element of life table construction by the present method is a_i, the fraction of the last age interval of life; the practical usefulness of the method depends upon the invariant property of this fraction. Two extensive studies have been carried out on a_i. One study was based on some 135,000 California resident deaths in 1960 collected by the California State Department of Public Health, reported at the American Public Health Association's Annual Meeting in 1961 [4]. The other study was made on a 10 percent sample of United States deaths in 1963 compiled by the National Center for Health Statistics of the Department of Health, Education, and Welfare.

In both studies the exact number of days lived by every person who died was recorded and the mean value and variance of days lived were computed for each age. With the exception of the first year of life, both studies showed that the fraction a_i is subject to little variation with respect to race, sex, cause of death, geographical location and other demographic variables. Therefore, once a_i is determined for each age interval it can be used for many populations, with revision being made every ten years.

The application of the present method is illustrated with the construction of the life table for the total United States population, 1967. The values of a_i are given in Column 5 of Tables 2 and 3. In Table 2 the estimate of probability of death, $\hat{q}_i$, is computed from (1.6). The quantities in Table 3 are computed from the respective formulas (1.7) through (1.11).

3. ABRIDGED LIFE TABLE FOR TOTAL UNITED STATES POPULATION, 1967

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	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}	$\hat{q}_i$	ℓ_i	d_i	a_i	L_i	T_i	$\hat{e}_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0 - 1	.02191	100000	2191	.09	98006	7058889	70.59
1 - 5	.00345	97809	337	.39	390413	6960883	71.17
5 - 10	.00210	97472	205	.46	486806	6570470	67.41
10 - 15	.00203	97267	197	.54	485881	6083664	62.55
15 - 20	.00507	97070	492	.57	484292	5597783	57.67
20 - 25	.00641	96578	619	.49	481311	5113491	52.95
25 - 30	.00673	95959	645	.50	478182	4632180	48.27
30 - 35	.00836	95314	796	.52	474659	4153998	43.58
35 - 40	.01215	94518	1148	.54	469949	3679339	38.93
40 - 45	.01828	93370	1706	.54	462926	3209390	34.37
45 - 50	.02842	91664	2605	.54	452328	2746464	29.96
50 - 55	.04392	89059	3911	.53	436104	2294136	25.76
55 - 60	.06651	85148	5663	.52	412148	1858032	21.82
60 - 65	.09660	79485	7678	.52	378997	1445884	18.19
65 - 70	.14286	71807	10258	.52	334415	1066887	14.86
70 - 75	.20679	61549	12728	.51	276561	732472	11.90
75 - 80	.28466	48821	13897	.51	210057	455911	9.34
80 - 85	.40135	34924	14016	.48	138178	245854	7.04
85+	1.00000	20908	20908		107676	107676	5.15

3. THEORETICAL JUSTIFICATION

For an individual alive at (exact) age x_i, the probability that he will die in interval (x_i, x_i+n_i) is given by

$$q_i = 1 - \exp\left\{-\int_0^{n_i} \mu(x_i + \tau)d\tau\right\} \tag{3.1}$$

where $\mu(x_i+\tau)$ is the mortality intensity function (force of mortality, failure rate) at age $x_i+\tau$. For derivation of q_i, see [3, pp. 60–61].

The theoretical age specific death rate for the interval (x_i, x_i+n_i), denoted by m_i, is defined as the ratio of the probability q_i to the expected duration of the interval lived by an individual. To derive the formula for m_i, define an indicator function $I_i(y)$ such that

$$I_i(y) = 1 \quad \text{if the individual alive at } x_i \text{ survives to } x_i + y$$
$$= 0 \quad \text{if he dies before } x_i + y.$$

The individual will live a time element $(x_i+y,\ x_i+y+dy)$ once he survives to x_i+y. The probability of surviving from x_i to x_i+y is

$$\exp\left\{-\int_0^y \mu(x_i+\tau)d\tau\right\}.$$

Therefore, the expectation of $I_i(y)$ is given by

$$E[I_i(y)] = \exp\left\{-\int_0^y \mu(x_i+\tau)d\tau\right\}.$$

It is clear that the expected duration of the interval $(x_i,\ x_i+n_i)$ lived by the individual is

$$\int_0^{n_i} E[I_i(y)]dy = \int_0^{n_i} \exp\left\{-\int_0^y \mu(x_i+\tau)d\tau\right\}dy.$$

It follows from the definition that the theoretical age specific death rate is given by

$$m_i = \frac{q_i}{\int_0^{n_i} \exp\left\{-\int_0^y \mu(x_i+\tau)d\tau\right\}dy}\,. \tag{3.2}$$

Graphically, the denominator in (3.2) is the area under the survival curve between the ordinates at x_i and x_i+n_i.

When the upper limit of the outer integral is infinity, the denominator in (3.2) becomes the expectation of life at age x_i, denoted by e_i. That is,

$$e_i = \int_0^\infty \exp\left\{-\int_0^y \mu(x_i+\tau)d\tau\right\}dy. \tag{3.3}$$

Formula (3.3) may be derived directly as follows. Let a random variable Y_i be the future life time of an individual now aged x_i, so that x_i+Y_i is the entire life span of the individual. It is easy to verify that the density function, $f(y)$, of Y_i is given by

$$f(y)dy = \exp\left\{-\int_0^y \mu(x_i+\tau)d\tau\right\}\mu(x_i+y)dy \quad \text{for } 0 \le y < \infty, \tag{3.4}$$

with

$$\int_0^\infty f(y)dy = 1. \tag{3.5}$$

The mathematical expectation of Y_i, denoted by e_i, is the expectation of life at x_i; or

$$e_i = E(Y_i) = \int_0^\infty y\exp\left\{-\int_0^y \mu(x_i+\tau)d\tau\right\}\mu(x_i+y)dy. \tag{3.6}$$

Now let $u=y$, $du=dy$,

$$v = -\exp\left\{-\int_0^y \mu(x_i+\tau)d\tau\right\}$$

and

$$dv = \exp\left\{-\int_0^y \mu(x_i+\tau)d\tau\right\}\mu(x_i+y)dy. \tag{3.7}$$

Integrating (3.6) by parts gives

$$e_i = -y\exp\left\{-\int_0^y \mu(x_i+\tau)d\tau\right\}\Bigg|_0^\infty + \int_0^\infty \exp\left\{-\int_0^y \mu(x_i+\tau)d\tau\right\}dy. \tag{3.8}$$

The first term on the right-hand side of (3.8) vanishes and the second term is the same as (3.3), as required to be shown.

To derive a relation between q_i and m_i we introduce a random variable α_i denoting the fraction of the interval $(x_i,\ x_i+n_i)$ lived by an individual who dies in the interval. Clearly the density function of α_i is

$$g_{\alpha_i}(t)dt = \left[\frac{\exp\left\{-\int_0^{n_it} \mu(x_i+\tau)d\tau\right\}}{q_i}\right]\cdot\mu(x_i+n_it)dn_it, \quad 0 \le t \le 1. \tag{3.9}$$

The quantity inside the brackets in (3.9) is the conditional probability that an individual alive at x_i will survive to x_i+n_it if he dies in $(x_i,\ x_i+n_i)$, and $\mu(x_i+n_it)dn_it$ is the probability that he will die in $(x_i+n_it,\ x_i+n_it+dn_it)$. In other words, the product on the right-hand side of (3.9) is the probability that α_i will assume a value between t and $t+dt$, that is the density function $g_{\alpha_i}(t)dt$. When the integral in the numerator of (3.9) is evaluated, we have

$$\int_0^1 \exp\left\{-\int_0^{n_it} \mu(x_i+\tau)d\tau\right\}\mu(x_i+n_it)n_idt = 1 - \exp\left\{-\int_0^{n_i} \mu(x_i+\tau)d\tau\right\} = q_i \tag{3.10}$$

so that

$$\int_0^1 g_{\alpha_i}(t)dt = 1, \tag{3.11}$$

and α_i is a proper random variable.

The expectation of α_i, which is called *the fraction of the last age interval of life*, denoted by a_i, may be computed directly:

$$\begin{aligned} a_i &= E(\alpha_i) \\ &= \int_0^1 t\,\frac{\exp\left\{-\int_0^{n_it} \mu(x_i+\tau)d\tau\right\}}{q_i}\,\mu(x_i+n_it)n_idt \\ &= \frac{1}{q_i}\left[-\exp\left\{-\int_0^{n_i} \mu(x_i+\tau)d\tau\right\} + \frac{1}{n_i}\int_0^{n_i} \exp\left\{-\int_0^y \mu(x_i+\tau)d\tau\right\}dy\right]. \end{aligned} \tag{3.12}$$

Substituting (3.1) and (3.2) in (3.12) gives

$$a_i = 1 - \frac{1}{q_i} + \frac{1}{n_i m_i} \cdot \qquad (3.13)$$

Equation (3.13) is solved for q_i to give the basic relationship between q_i and m_i,

$$q_i = \frac{n_i m_i}{1 + (1 - a_i) n_i m_i} \cdot \qquad (3.14)$$

The relationship between the theoretical values q_i and m_i in (3.14) is exactly the same as the relationship between the observed values $\hat{q}_i$ and M_i in (1.6).

In the life table where there are ℓ_i individuals alive at x_i, let $E(L_i)$ be the expected number of years to be lived by ℓ_i in the interval $(x_i, x_i + n_i)$. An individual alive at x_i will live n_i years of the interval if he survives to the end of the interval (for which the probability is $1 - q_i$), or he will live an expected duration $a_i n_i$ if he dies in the interval (for which the probability is q_i). It follows that

$$E(L_i) = n_i(1 - q_i)\ell_i + a_i n_i q_i \ell_i. \qquad (3.15)$$

For the final interval (x_w and over) in the life table, the expectation e_w is directly related to the death rate, m_w. We see from (3.1) that $q_w = 1$, and the death rate in (3.2) becomes

$$m_w = \frac{1}{\int_0^\infty \exp\left\{-\int_0^y \mu(x_w + \tau)d\tau\right\} dy} = \frac{1}{e_w} \qquad (3.16)$$

so that

$$e_w = \frac{1}{m_w}, \qquad (3.17)$$

which is a well-known formula. Formulas (3.14), (3.15) and (3.17) provide the theoretical basis for the method of constructing current life tables described in Section 1.1 and Equations (1.6), (1.7), and (1.11).

Remark. The similarity between the definition of the expectation of life in (3.6) and the definition of the fraction of the last age interval of life in (3.12) suggests a relation between the two quantities. The relation is as follows [cf. 3, p. 237]:

$$e_i = a_i n_i + \sum_{j=i}^{w-1} [(1 - a_j)n_j + a_{j+1} n_{j+1}] p_{i,j+1} \qquad (3.18)$$

where $a_w n_w = e_w$ and p_{ij}, the probability of surviving the interval (x_i, x_j), is given by

$$p_{ij} = \exp\left\{-\int_{x_i}^{x_j} \mu(\tau)d\tau\right\}. \qquad (3.19)$$

To prove (3.18), we start with (3.6) and write

$$\begin{aligned} e_i &= \int_0^\infty y \exp\left\{-\int_0^y \mu(x_i + \tau)d\tau\right\} \mu(x_i + y)dy \\ &= \sum_{j=i}^{w} \int_{x_j - x_i}^{x_{j+1} - x_i} y \\ &\quad \cdot \exp\left\{-\int_0^y \mu(x_i + \tau)d\tau\right\} \mu(x_i + y)dy \end{aligned} \qquad (3.20)$$

with $x_{w+1} = \infty$. Each of the integrals in (3.20), when evaluated and simplified, becomes

$$\begin{aligned} \int_{x_j - x_i}^{x_{j+1} - x_i} y \exp\left\{-\int_0^y \mu(x_i + \tau)d\tau\right\} \mu(x_i + y)dy \\ = (x_j - x_i + a_j n_j)(p_{ij} - p_{i,j+1}) \end{aligned} \qquad (3.21)$$

for $j = i, \cdots, w-1$; and

$$\begin{aligned} \int_{x_w - x_i}^{\infty} y \exp\left\{-\int_0^y \mu(x_i + \tau)d\tau\right\} \mu(x_i + y)dy \\ = (x_w - x_i + e_w)p_{iw}, \qquad \text{for } j = w. \end{aligned} \qquad (3.22)$$

Substituting (3.21) and (3.22) in (3.20) and simplifying the resulting expression, we recover (3.18).

[*Received August 1971. Revised February 1972.*]

REFERENCES

[1] Chiang, C.L., "A Stochastic Study of the Life Table and Its Applications: I. Probability Distributions of the Biometric Functions," *Biometrics*, 16 (December 1960), 618–5.

[2] ———, "A Stochastic Study of the Life Table and Its Applications: II. Sample Variance of the Observed Expectation of Life and Other Biometric Functions," *Human Biology*, 32 (September 1960), 221–38.

[3] ———, *Introduction to Stochastic Processes in Biostatistics*, New York: John Wiley and Sons, Inc., 1968.

[4] ———, Norris, F.D., Olsen, F., Shipley, P.W., and Supplee, H.E., "Determination of the Fraction of Last Year of Life and Its Variation by Age, Race, and Sex," Unpublished manuscript presented at the Annual Meeting of the American Public Health Association, Detroit, Michigan, November 1961.

[5] Coale, A.J. and Demeny, P., *Regional Life Tables and Stable Populations*, Princeton, N.J.: Princeton University Press, 1966.

[6] Grenander, U., "On the Theory of Mortality Measurements," *Skandinavisk Aktuarietidskrift*, 39, No. 1 (1956), 1–55.

[7] Greville, T.N.E., "Short Methods of Constructing Abridged Life Tables," *Record of American Institute of Actuaries*, 32 (June 1943), 29–43.

[8] Keyfitz, N., "A Life Table That Agrees With the Data," *Journal of the American Statistical Association*, 61 (June 1966), 305–11.

[9] King, G., "On a Short Method of Constructing an Abridged Mortality Table," *Journal of the Institute of Actuaries*, 48 (1914), 294.

[10] Reed, J.L., and Merrell, M., "A Short Method of Constructing an Abridged Life Table," *American Journal of Hygiene*, 30 (September 1939), 33–62.

[11] Sirken, M., "Comparison of Two Methods of Constructing Abridged Life Tables," *Vital and Health Statistics*, Ser. 2, No. 4 (1964), 1–11, National Center for Health Statistics.