Canadian Studies in Population, Vol. 20(1), 1993, pp. 43-57.

The Entropy of the Survival Curve: An Alternative Measure

Gerry Hill
Laboratory Centre for Disease Control
Health and Welfare Canada
Ottawa, Ontario, Canada

Abstract

The "entropy" of a life table, H_k, was introduced by Keyfitz as the elasticity of the expectation of life, E, with respect to uniform change in age-specific mortality rates. H_k is not a true measure of entropy in the probability sense, though $H_s = H_k + \ln(E)$ is the entropy of the distribution by age of the corresponding stationary population. A more direct measure is H_f, the entropy of the distribution of age at death. The mathematical relationships between these indexes are reviewed, and their behaviour compared using historical cohort life tables from various countries, with Gompertz functions being fitted to survival from age 30 to age 85 for each sex. It is shown that as E increases and H_k decreases, there is very little change in the "true" entropy measures H_s and H_f. Entropy and variance are conceptually similar and the variance of the distribution of the age at death, V, also changes very little. An interesting feature of the Gompertz distribution is that a simple mathematical relationship exists between H_f and the force of mortality at the expectation of life, m(E). It follows that the latter is also constant for most of the cohorts. A decrease in H_f and V and an increase in m(E) is seen for the most recent male cohorts in England and Wales. This may be due to the increasing prevalence of cigarette smoking in these cohorts.

Résumé

"L'entropie" d'une table de mortalité, H_k, a été introduite par Keyfitz pour mesurer l'élasticité de l'espérance de vie, E, par rapport à un changement uniforme des taux de mortalité par âge. H_k n'est pas une vraie mesure d'entropie au sens probabiliste, bien que $H_s = H_k + \ln(E)$ est l'entropie de la distribution par âge de la population stationnaire correspondante. Une mesure plus directe est H_f, l'entropie de la distribution de l'âge de décès. Les liens mathématiques entre ces indices sont passés en revue, et leur utilisation est comparée dans le contexte des tables de mortalité des cohortes historiques des pays différents, en utilisant la loi de Gompertz pour paramétriser la mortalité entre les âges de 30 et 85 ans pour chaque sexe. Tandis que E augmente et H_k diminue les "vraies" indices d'entropie, H_s et H_f, ne changent pas beaucoup. Il y a une liason conceptuelle entre l'entropie et la variance, et la variance de l'âge de décès, V, est aussi constante. Un attribut intéressant de la loi de Gompertz est l'existence d'un lien étroit entre H_f et la mortalité à l'espérance de vie, m(E). En conséquence celle-ci est aussi constante dans la plupart des cohortes. Une diminution de H_f et de V, et une augmentation de m(E), sont évidentes dans les cohortes masculines plus récentes en Angleterre et dans le Pays de Galles. Il est possible que ceci peut être expliqué par la prévalence croissante du tabagisme dans ces cohortes.

Key Words: life tables, entropy, cohort analysis

Introduction

A few years before his death, Dhruva Nagnur published a series of abridged life tables for Canada and the provinces (Nagnur, 1986a) as well as a paper in *Canadian Studies in Population* (Nagnur, 1986b), in which he used Keyfitz's index of entropy to demonstrate the rectangularization of the survival curve. Keyfitz' H (referred to in this paper as H_k, see below) was derived as a measure of the elasticity of the expectation of life to uniform change in mortality throughout life (Keyfitz, 1977). The use of H_k in this context, and further elucidation of its meaning, has been addressed by several authors (Mitra, 1978; Goldman and Lord, 1986; Vaupel, 1986; Hakkert, 1987).

The term "entropy" was applied to H_k following the work of Demetrius (1974, 1975, 1976). The form of the integral in its numerator resembles the definition of the (differential) entropy of a probability distribution used in information theory (e.g. Sveshnikov, 1978: 157). This use of the term is strictly incorrect, since the survival curve is not a probability density. However, as shown by Demetrius (1979), H_k plus the logarithm of the expectation of life (i.e., H_s) is the entropy of the stable age distribution generated by the survival curve. A probability distribution more directly related to the survival curve is that of the age at death, the entropy of which will be denoted by H_f. In this paper, some mathematical features of these various indexes are discussed and their application to several sets of cohort life tables described. An interesting feature of H_f is that under the Gompertz law, H_f and the logarithm of the hazard rate at the expectation of life sum to unity.

Entropy Measures and Their Properties

Notation

x = age beyond some initial value (30 years in the examples)

m(x) = hazard rate (force of mortality) at age x

M(x) = cumulative hazard = $\int_0^x m(u)du$

S(x) = survivor function = Pr(survival to age x)

f(x) = probability density of age at death

$E_f(.)$ = expectation over f(x)

$E =$ $E_f(x)$ = expectation of life at initial age

$N_k =$ $- \int_0^\infty S(x) \ln\{S(x)\}dx)$ = numerator of Keyfitz' H

$H_k =$ Keyfitz' $H = N_k/E$

$H_f =$ $- \int_0^\infty f(x) \ln\{f(x)\}dx)$ = entropy of age at death

$G(u) =$ incomplete gamma function $= e^u \int_u^\infty y^{-1}e^{-y}\, dy)$

General properties

It is assumed throughout that the functions are defined for $x > 0$ and are well behaved, and that $x\ln(x)$ = zero when x = zero.

The following can be shown for any survivor function :

$$N_k = E_f\{M(x)/m(x)\} \qquad (1)$$

$$H_f = E_f\{M(x)\} - E_f[\ln\{m(x)\}]$$

$$= 1 - E_f[\ln\{m(x)\}] \qquad (2)$$

since $M(x)$ is always exponentially distributed with mean = 1.

Both results provide useful insights. Since $M(x)/m(x)$ remains constant under uniform proportional change in $m(x)$, it is clear from (1) that the resulting change in N_k is due to change in $f(x)$. Also, in the same situation, $\ln\{m(x)\}$ is unchanged except for the addition of a constant, so that from (2), H_f is similarly shifted except for changes in $f(x)$.

We see also from (2) that if $\ln\{m(x)\}$ is a polynomial in x, then H_f can be expressed in terms of the moments of $f(x)$. In particular for the Gompertz law where $\ln\{m(x)\} = a + bx$, $1 - H_f = a + bE(x) = \ln[m\{E(x)\}]$, as noted above.

Using the substitution $y = M(x)$ in the definitions, it is easy to show that the components of the entropy measures are of the form

$\int_0^\infty K(y)\, e^{-y} dy)$, with $K(y)$ defined as follows:

	$K(y)$
E	$1/m\{M^{-1}(y)\}$
N_k	$y/m\{M^{-1}(y)\}$
$1 - H_f$	$\ln[m\{M^{-1}(y)\}]$

Gompertz Survival

These expressions are useful when deriving formulae for the entropy measures with respect to specific survivor functions. For the Gompertz function, for example, with $m(x) = e^{a+bx} = Ae^{bx}$:

$$E = \{G(A/b)\}/b$$

$$N_k = b - A.G(A/b)$$

$$H_k = \{1/G(A/b)\} - A/b$$

$$H_f = 1 - a - G(A/b) = 1 - (a + b.E)$$

$$= 1 - \ln\{m(E)\} \text{ as seen from (2) above.}$$

Application to Cohort Life Tables

To illustrate the behaviour of the entropy measures to human survival from age 30 years onwards, the above formulae have been applied to various sets of abridged cohort life tables available from the literature. In each case, the published ${}_5q_x$ [x= 30(5)80] values have been used to estimate hazard rates using the formula, $m(x+2.5) = 0.4\, q/(2-q)$, and Gompertz coefficients estimated from the logarithms of the hazard rates using unweighted least squares.

The data came from the following sources:

Canada

The q values applicable to male and female cohorts born around 1891, 1896, and 1901 were extracted from the cross-sectional life tables published by Nagnur (1986a).

England and Wales

(a) Russell (1948: 404) estimated a cohort life table for males born in the period 1426-1450.

(b) Case (1962) used mortality data from 1841-1960 to calculate life tables for males and females born in five year intervals beginning in 1841. These have been updated using mortality data from 1961-1985.

France

The cross-sectional life tables for French males and females from 1806-1936 published by Bourgeois-Pichat (1952: 326) have been used to obtain cohort q values which, in this case, are at ten-year age intervals with appropriate adjustment of the formula for the hazard rate.

United States

Manton and Stallard (1984: 260) published cohort tables for white males and females born in 1870, 1880, and 1890.

Results

Tables 1 and 2 give the estimates of the expectation of life at age 30 years (E), Keyfitz' entropy (H_k) and its numerator N_k, the entropy of the stable age distribution (H_s), and the entropy of the distribution of deaths (H_f) for each set of cohorts. Also shown is the estimated hazard rate at the expectation of life, m(E), expressed as a rate per 1,000 person-years. These results are also shown in Figures 1-6.

TABLE 1. ENTROPY MEASURES FOR SURIVAL FROM AGE 30 IN FEMALES

Cohort	E	N_k	H_k	H_s	H_f	m(E)*
France						
1776	32.8	14.0	.428	3.92	4.14	43.3
1781	33.3	14.0	.421	3.93	4.14	43.3
1786	33.4	13.8	.413	3.92	4.13	43.7
1791	35.6	14.1	.397	3.97	4.16	42.4
1796	33.7	13.8	.410	3.93	4.13	43.6
1801	33.9	14.2	.419	3.94	4.16	42.6
1806	33.7	14.1	.417	3.94	4.15	43.0
1811	34.2	14.0	.410	3.94	4.15	43.0
1816	34.1	14.1	.414	3.94	4.15	42.9
1821	34.5	13.6	.395	3.94	4.12	43.9
1826	34.2	13.8	.403	3.94	4.13	43.6
1831	34.6	13.4	.388	3.93	4.12	44.4
1836	34.5	13.6	.396	3.94	4.13	43.9
1841	34.2	13.6	.397	3.93	4.12	44.1
1846	35.0	13.6	.387	3.94	4.12	44.0
1851	36.1	13.4	.371	3.96	4.12	44.1
1856	35.7	13.7	.384	3.96	4.14	43.4
England & Wales						
1841	34.8	14.1	.406	3.95	4.16	42.6
1846	35.6	14.3	.402	3.97	4.17	41.9
1851	36.3	14.2	.392	3.98	4.17	42.0
1856	37.2	14.2	.382	4.00	4.17	41.9
1861	39.0	13.9	.356	4.02	4.16	42.3
1866	39.3	14.1	.358	4.03	4.17	41.8
1871	40.4	13.8	.341	4.04	4.16	42.3
1876	41.4	13.8	.332	4.05	4.16	42.3
1881	42.5	13.7	.323	4.07	4.16	42.3
1886	43.4	14.0	.323	4.09	4.18	41.4
1891	44.5	13.5	.302	4.10	4.15	42.8
1896	45.5	13.4	.294	4.11	4.15	43.0
1901	46.6	13.2	.284	4.13	4.14	43.3
United States (White)						
1870	40.6	14.6	.360	4.06	4.21	40.3
1880	42.5	14.9	.349	4.10	4.23	39.4
1890	45.2	15.2	.335	4.15	4.26	38.4
Canada						
1891	44.7	14.8	.331	4.13	4.24	39.3
1896	46.1	15.1	.327	4.16	4.26	38.4
1901	47.7	15.1	.317	4.18	4.26	38.3

* rate per 1,000

TABLE 2. ENTROPY MEASURES FOR SURVIVAL FROM AGE 30 IN MALES

Cohort	E	N_k	H_k	H_s	H_f	m(E)*
France						
1776	32.3	13.9	.429	3.90	4.12	44.0
1781	32.6	14.0	.429	3.91	4.13	43.7
1786	33.5	13.8	.414	3.92	4.13	43.6
1791	33.2	14.0	.420	3.92	4.14	43.4
1796	33.1	13.6	.411	3.91	4.12	44.3
1801	33.4	14.0	.420	3.93	4.14	43.3
1806	33.3	13.7	.411	3.92	4.12	44.1
1811	33.4	13.7	.411	3.92	4.12	44.0
1816	33.4	13.8	.415	3.92	4.13	43.7
1821	33.6	13.2	.393	3.91	4.10	45.2
1826	33.0	13.5	.408	3.90	4.11	44.7
1831	33.4	13.0	.388	3.90	4.08	45.9
1836	33.2	13.2	.398	3.90	4.09	45.4
1841	33.1	13.3	.403	3.90	4.10	45.1
1846	33.2	13.2	.399	3.90	4.09	45.3
1851	33.0	13.4	.406	3.90	4.10	44.9
1856	33.2	13.5	.407	3.91	4.11	44.5
England & Wales						
1426-50	24.4	15.8	.647	3.84	4.10	45.2
1841	32.1	13.5	.421	3.89	4.10	44.8
1846	32.8	13.7	.417	3.91	4.12	44.2
1851	33.3	13.4	.404	3.91	4.11	44.7
1856	34.0	13.3	.393	3.92	4.11	44.8
1861	34.8	13.3	.382	3.93	4.11	44.7
1866	35.8	13.3	.371	3.95	4.11	44.6
1871	36.7	13.1	.357	3.96	4.10	44.9
1876	37.1	12.9	.347	3.96	4.09	45.5
1881	37.5	12.7	.340	3.96	4.08	45.8
1886	37.7	12.8	.338	3.97	4.09	45.7
1891	38.7	12.1	.312	3.97	4.04	47.9
1896	39.3	11.8	.300	3.97	4.02	48.8
1901	39.8	11.6	.292	3.98	4.01	49.4
United States (White)						
1870	36.7	13.7	.372	3.98	4.14	43.3
1880	37.8	13.4	.355	3.99	4.13	43.7
1890	39.0	13.1	.337	4.00	4.12	44.4
Canada						
1891	40.9	12.4	.303	4.01	4.07	46.5
1896	41.4	12.3	.300	4.02	4.06	46.6
1901	41.4	12.4	.300	4.02	4.07	46.4

* rate per 1,000

The data from France, England and Wales provide a consistent series for cohorts born between 1776 and 1901. The data from the United States and Canada complement the latter part of the series from England and Wales. Among females, a gradual increase in E, with the expectation of life at age 30, is evident for successive cohorts born in the first part of the nineteenth century, accelerating for those born in the second half of the century. For males, there is no evidence of the earlier change seen in females, but for those born after the middle of the century, E increases steadily as for their female contemporaries, but less steeply and with some evidence of a plateau in the latter cohorts.

In both sexes, as E increases, H_k decreases due partly to the fact that E forms the denominator of H_k, and also due to a decrease in the numerator N_k. H_s increases slightly with E (more so in females than males). H_f remains virtually unchanged in females throughout the series. In the male cohorts, there is also very little change in H_f for those born before 1876, after which H_f declines slightly. The hazard rate at the expectation of life necessarily shows the same constancy as H_f for most cohorts, decreasing slightly for the later female cohorts and increasing slightly for the later male cohorts.

TABLE 3. **AVERAGE ANNUAL PERCENTAGE CHANGE IN ENTROPY INDEXES FOR COHORTS BORN IN ENGLAND AND WALES BETWEEN 1841 AND 1901**

	Males	Females
E	0.36	0.48
N_k	0.26	-0.12
H_k	-0.62	-0.61
H_s	0.04	0.07
H_f	-0.04	-0.01
m(E)	0.16	0.03

Discussion

The general pattern of increase in E, decrease in H_k, and little change in H_s, H_f or m(E), is emphasized by the comparison of the estimates for males in England and Wales between the fifteenth and nineteenth centuries. The constancy of H_f and m(E) over this long period is very striking. In contrast, the changes in these parameters for recent male (not female) cohorts is interesting. The male cohorts in England and Wales were the first to become addicted to cigarette smoking and this might account for the sex differential. The internal combustion engine, the other main agent of death for males in this century, weighed most heavily at ages below 30 years.

It is clear that the fall in H_k as E increases has nothing to do with a decrease in "entropy". Table 3 shows the mean annual percentage change in the various measures for both sexes born in England and Wales between 1841 and 1901. Three-fifths of the change in H_k among males and four-fifths of the change in females, is due directly to the fact that E is in the denominator of H_k.

The entropy of a distribution is a measure of uncertainty and is conceptually close to that of variance or dispersion. The fact that the entropy of the distribution of age at death, H_f, remains constant suggests that the variance of that distribution does likewise. Table 4 shows that this is indeed the case for the England and Wales cohorts. The improvements in diet and public health experienced by these cohorts has shifted the mean survival to the right with little change in dispersion, except for the recent male cohorts.

TABLE 4. MEAN (E) AND VARIANCE (V) OF DISTRIBUTION OF THE YEARS SURVIVED FROM AGE 30 FOR COHORTS BORN IN ENGLAND AND WALES BETWEEN 1841 AND 1901

Cohort	Males		Females	
	E	V	E	V
1841	32.1	233.7	34.8	257.7
1846	32.8	240.3	35.6	265.8
1851	33.3	233.8	36.3	264.6
1856	34.0	232.5	37.2	266.4
1861	34.8	233.9	39.0	260.7
1866	35.8	234.9	39.3	266.7
1871	36.7	231.7	40.4	260.6
1876	37.1	225.3	41.4	261.2
1881	37.5	222.9	42.5	262.3
1886	37.7	223.5	43.4	273.2
1891	38.7	205.2	44.5	257.8
1896	39.3	198.0	45.5	256.0
1901	39.8	193.9	46.6	253.4

Figure 1A. Expectation of life at age 30 Females

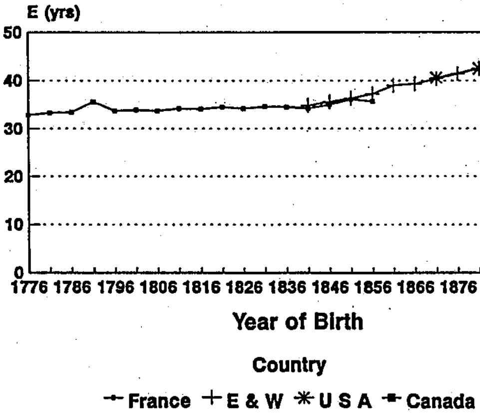


Figure 1B. Expectation of life at age 30 Males

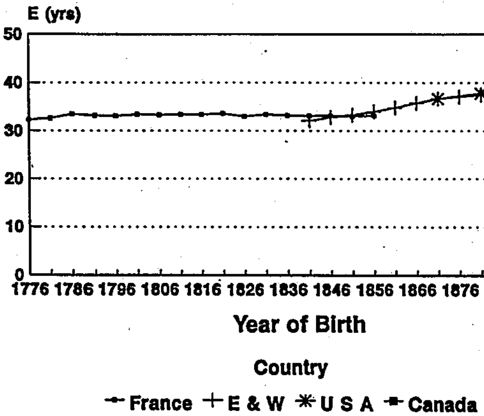


Figure 2A. Keyfitz' H
Females aged 30 and over

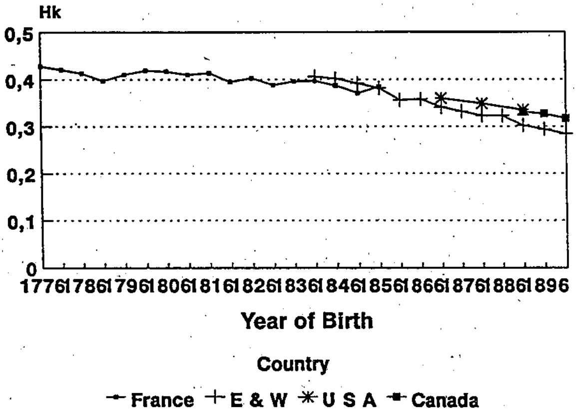


Figure 2B. Keyfitz' H
Males aged 30 and over

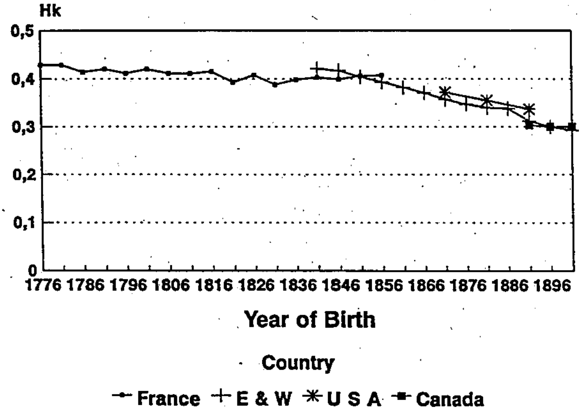


Figure 3A. Entropy of Stationary Age Distribution. Females aged 30 and over

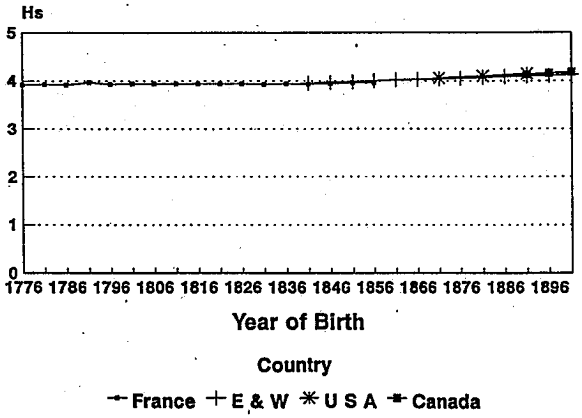


Figure 3B. Entropy of Stationary Age Distribution. Males aged 30 and over

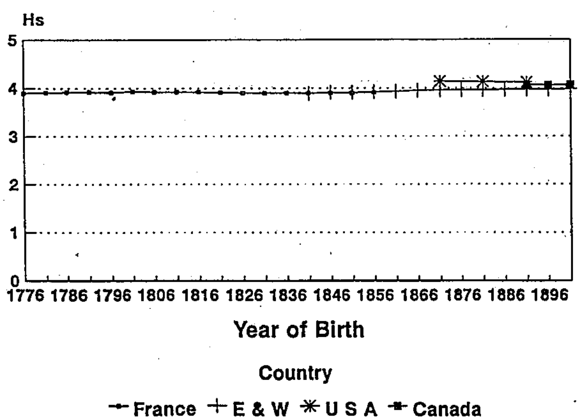


Figure 4A. Entropy of Distribution of Age at Death. Females aged 30 and over

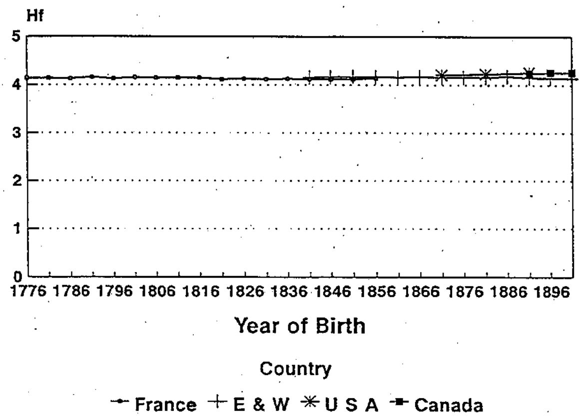


Figure 4B. Entropy of Distribution of Age at Death. Males aged 30 and over

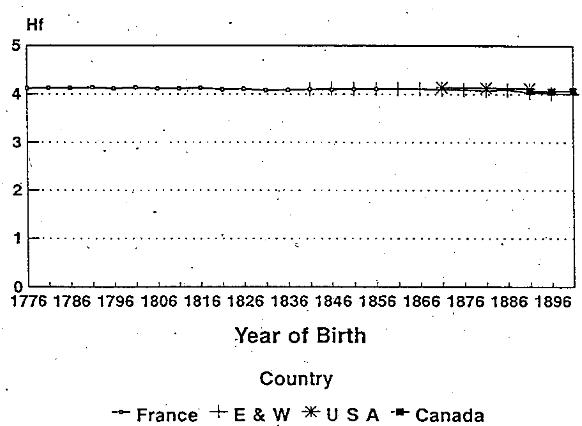


Figure 5A. Hazard at Expectation of Life Females aged 30 and over

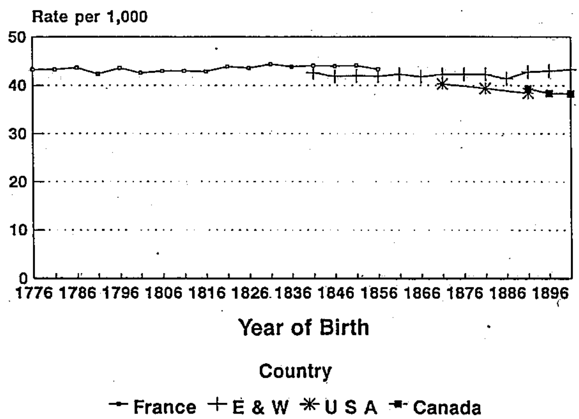
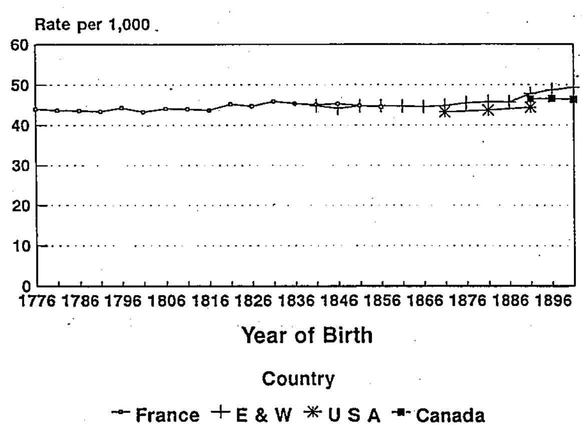


Figure 5B. Hazard at Expectation of Life Males aged 30 and over



Acknowledgements

The assistance of Jeffrey Bertram, who helped with the calculations, is gratefully acknowledged. This paper is dedicated to the memory of Dhruva Nagnur—a fine demographer and a charming colleague.

References

Bourgeois-Pichat, J. 1952. Note sur l'evolution generale de la population francaise depuis le XVIIIe siecle. Population 7:319-329.

Case, R.A., Coghill, C., Harley, J.L. and Pearson, J.T. 1962. The Chester Beatty Research Institute Serial Abridged Life Tables, England and Wales 1841-1960. London, England : Royal Cancer Hospital.

Demetrius, L. 1974. Demographic parameters and natural selection.Proceedings of the National Academy of Sciences 71: 4645-4647.

________. 1975. Reproductive strategies and natural selection. American Naturalist 109: 243-249.

________. 1976. Measures of variability in age-structured populations. Journal of Theoretical Biology 63: 397-404.

________. 1979. Relations between demographic parameters. Demography 16: 329-338.

Goldman, N. and Lord, G. 1986. A new look at entropy and the life table. Demography 23: 275-282.

Hakkert, R. 1987. Life table transformations and inequality measures: some noteworthy formal relationships. Demography 24: 615-622.

Keyfitz, N. 1977. Applied Mathematical Demography. New York, NY: John Wiley and Sons.

Manton, K.G. and Stallard, E. 1984. Recent Trends in Mortality Analysis. Orlando, FL: Academic Press.

Mitra, S. 1978. A short note on the Taeuber Paradox. Demography 15: 621-623.

Nagnur, D. 1986a. Longevity and Historical Life Tables. 1921-1981(Abridged). Canada and the Provinces. Statistics Canada Catalogue 89-506. Ottawa,ON : Minister of Supply and Services.

________. 1986b. Rectangularization of the survival curve and entropy: the Canadian experience, 1921-1981. Canadian Studies in Population 13: 83-102.

Russell, J.C. 1948. Demographic pattern in history. Population Studies 1: 388-404.

Sveshnikov, A.A. (ed.) 1978. Problems in Probability Theory, Mathematical Statistics and Theory of Random Functions. New York, NY: Dover.

Vaupel, J.W. 1986. How change in age-specific mortality affects life expectancy. Population Studies 40: 147-157.

Received, April 1991; revised, September 1992.