TFA 40 (1985-1987) 135-147

EXPERIMENTS IN GRADUATING THE DATA FOR THE ENGLISH LIFE TABLES (No. 14)

by

J. J. McCUTCHEON, M.A., Ph.D., F.F.A.

1. INTRODUCTION

By courtesy of the Government Actuary, the preliminary data relating to the proposed English Life Tables (No. 14) were made available to the author. In this paper we describe briefly the graduation of the underlying crude mortality rates by cubic splines and give an outline of the salient features of the graduations finally adopted. Complete details of the English Life Tables are given in reference 1, which also includes historical comparisons.

The investigation covered the calendar years 1980, 1981 and 1982. For each integer $x \geqslant 2$ the values of θ_x, the number of deaths during the investigation period, aged x last birthday at the time of death, and E_x^c, the corresponding *central* exposure to risk, were available for both sexes. The quotient θ_x/E_x^c gives the crude central death rate at exact age x.

For $x \geqslant 5$ the value of E_x^c was calculated by the simple quadratic-based formula

$$E_x^c = \frac{9}{8}\left\{P_x^{1980} + \frac{2}{3}P_x^{1981} + P_x^{1982}\right\},$$

where P_x^j denotes the number in the home population aged x last birthday at the mid-point of calendar year j. For $x = 2$, 3 and 4 the value of E_x^c was calculated by reference to both the number of deaths and the related births (cf. reference 5).

For both sexes numerous experiments were carried out, using the method of "variable-knot" spline graduation (cf. reference 6). These led to "best" graduations for each sex, which formed the bases for the resulting life tables.

2. MORTALITY AT YOUNG AGES

Special techniques were used for the measurement of mortality at ages 0 and 1. Because of the uneven incidence of deaths during the first year of life and of variations in the birth rate with time, for each sex the values of q_0 and q_1 were derived from the numbers of deaths at ages 0 and 1 during the investigation period and the associated births (i.e. births during the calendar years 1978 to 1982 inclusive).

The techniques used were based on those described in reference 5. Our results are summarised in the following table, in which E_x denotes the *initial* (as opposed to central) exposure to risk at age x last birthday and ϕ_0 is the average age at death of those dying in the first year of life.

TABLE 2.1

	England and Wales: 1980-82	
	Males	Females
θ_0	12,504	9,191
E_0	983,539	933,998
q_0	0·01271	0·00984
θ_1	828	662
E_1	970,789	922,367
q_1	0·00085	0·00072
ϕ_0	0·15	0·16

In comparison with the values used in the English Life Tables (No. 13), which relate to the period 1970-72, the reduction in q_0 is 36% for males and 35% for females. For q_1 the corresponding reductions are 29% and 32% respectively (cf. reference 2).

3. GRADUATION AT AGES 2 AND ABOVE BY CUBIC SPLINES

The ideas underlying the method of variable-knot spline graduation are described in some detail in reference 6. For completeness, however, we give below a brief summary of the method as applied to the present work.

At advanced ages the observed crude mortality rates for both sexes formed a very irregular series. (At these ages there is a paucity in the available data.) Initially, therefore, use was made of the data for individual ages only for ages 2 to 99 (inclusive). The calculation of the graduated rates of mortality at the highest ages is described in §4 below.

Suppose that n is a given positive integer and that $2 < x_1 < x_2 < \ldots < x_n < 99$. A cubic spline, $s(x)$, defined on the interval [2,99], with "knots" at the points $x_1, \ldots x_n$, is a function which is piecewise-cubic on each of the subintervals $[2,x_1]$, $[x_1,x_2], \ldots [x_{n-1},x_n]$, $[x_n,99]$. Moreover "adjacent" cubics fit together in such a way that $s(x)$ is *twice-differentiable* at each of the knots. (Thus, for example, in general greater smoothness is achieved than with King's classical method of osculatory interpolation, in which only *one* derivative exists at the knots. If, however, knots coincide, the number of derivatives which exist at a "multiple" knot is reduced (cf. reference 7).)

For *specified* knot positions $\{x_1, \ldots x_n\}$ a total of $(n + 4)$ parameters are needed to define a cubic spline (cf. reference 7). If, in addition, freedom of choice of the knot positions is allowed, then a total of $(2n + 4)$ parameters are required to specify precisely a cubic spline with n knots. For a given number of knots, n say, we define the "best" n-knot cubic spline to be that spline $s(x)$ which minimises

$$\chi^2 = \sum_{x=2}^{99} (z_x)^2,$$

where

$$z_x = \frac{\theta_x - \mathrm{E}_x^c\, s(x)}{\sqrt{\mathrm{E}_x^c\, s(x)}}$$

is the "relative deviation" at age x.

Note that, in determining the best-fitting n-knot cubic spline, we consider the knot positions as free parameters. The resulting minimum value of χ^2 is denoted by $\chi^2(n)$. Since in this case the χ^2 value has been obtained as a sum of squares over 98 ages, we regard it empirically as arising from a χ^2 distribution for which the number of degrees of freedom is $k = 98 - (2n + 4) = (94 - 2n)$. A test statistic for a χ^2 variable with k degrees of freedom is

$$t(\chi^2) = \sqrt{2\chi^2} - \sqrt{2k - 1}$$

(cf. reference 3). Accordingly, in order to determine the number of knots to be used, we consider not only the sequence $\{\chi^2(n)\}$ but also the sequence $\{t(\chi^2(n))\}$, where

$$\begin{aligned} t(\chi^2(n)) &= \sqrt{2\chi^2(n)} - \sqrt{2(94 - 2n) - 1} \\ &= \sqrt{2\chi^2(n)} - \sqrt{187 - 4n} \end{aligned}$$

For small values of n the values of both $\chi^2(n)$ and $t(\chi^2(n))$ decrease rapidly as n increases. However, although the values of $\chi^2(n)$ necessarily form a decreasing sequence, there is a "critical" point at which the value of $t(\chi^2(n))$ is first *greater* than the value of $t(\chi^2(n - 1))$. The existence of this critical point indicates that an increase in the number of knots has not produced a significantly lower value of χ^2 and determines both the appropriate number of knots to be used in the graduation and the best fitting spline.

Table 3.1 shows, for each sex, the values of $\chi^2(n)$ and $t(\chi^2(n))$ relevant to the above discussion.

TABLE 3.1

No. of knots	Statistics for best-fitting cubic splines			
	Males		Females	
n	$\chi^2(n)$	$t(\chi^2(n))$	$\chi^2(n)$	$t(\chi^2(n))$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
9	178·79	6·62	128·72	3·76
10	170·65	6·35	123·03	3·56
11	169·33	6·44	119·23	3·48
12	167·56	6·52	117·26	3·52
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

The values in the above table show that for males we obtain no significant improvement by increasing the number of knots from 10 to 11. This means that for males it is appropriate to use a 10-knot graduation. Likewise, for females we obtain no significant improvement by increasing the number of knots from 11 to 12, so that in this case it is appropriate to use an 11-knot graduation.

The positions of the knots in the best graduations for each sex are indicated in the table below.

TABLE 3.2

i	Knot position x_i	
	Males	Females
1	3·00	3·95
2	15·93	11·86
3	15·93	11·86
4	15·99	18·03
5	18·87	18·03
6	26·48	34·01
7	40·40	55·51
8	60·29	63·49
9	82·63	75·67
10	92·20	82·64
11	—	95·87

(Note: In Table 3.2 the knot-positions are given to only two decimal places. In the actual calculation of the graduated rates of mortality greater accuracy was retained. The difference is of no practical significance.)

For each sex the positions of the knots are of interest. For males there is a repeated knot at age 15·93 (at which point the spline $s(x)$ is thus only once differentiable) and two other knots relatively near this repeated knot. This is a reflection of the rapid increase in male mortality rates between ages 15 and 19. For females there are two repeated knots, at ages 11·86 and 18·03. Again this is a reflection of the shape of the curve for teenage mortality rates.

4. EXTRAPOLATION OF THE RATES OF MORTALITY TO THE HIGHEST AGES

At the highest ages (e.g. above age 95) the available data were suspect and, at these ages, the crude death rates formed a somewhat erratic set of values. Indeed for this reason it might have been better to ignore completely the data for ages 96 to 99, particularly as it is necessary in any event to extrapolate the graduated rates of mortality to even higher ages. For practical purposes, however, the inclusion in the original curve-fitting exercise of the data at the highest available ages is of little significance.

For males the highest knot occurs at age 92·20 (see Table 3.2). The extrapolation of the graduated central death rates was carried out at ages greater than 92 by a cubic polynomial, using the spline values at ages 90, 91 and 92 and the somewhat arbitrary value of 0·75 at age 105. This last value was chosen in the light of the recent investigation by the Registrar General into the mortality of centenarians (cf. reference 8). More precisely, the extrapolation cubic was determined by requiring it to have the same value and first and second derivatives at age 92 as the quadratic defined by the graduated values of m_{90}, m_{91}, m_{92} and by letting $m_{105} = 0{\cdot}75$.

For females the highest knot occurs at age 95·87 (see Table 3.2). The extrapolation of the graduated rates at ages greater than 95 was again carried out using a cubic polynomial. This polynomial was defined by requiring it to have the same value and first and second derivatives at age 95 as the quadratic defined by the graduated values of m_{93}, m_{94}, m_{95} and by letting $m_{105} = 0{\cdot}65$. This last value was chosen in the light of the results of reference 8.

Summary statistics for the resulting graduations are given in Table 4.1.

TABLE 4.1

Graduation summary statistics

(over the age range 2 to 95 inclusive)

Statistic	Males	Females
$\Sigma\ E_x^c$	70,518,494	74,492,855
$\Sigma\ \theta_x$	854,385	844,116
$\chi^2 = \Sigma\ (z_x)^2$	163·51	118·92
No. of positive deviations	48	48
No. of sign changes in deviations	46	57
No. of relative deviations > 2	13	7
No. of relative deviations > 3	5	1
Total accumulated deviations	− 39·34	− 7.82

Note: The deviation at each age is defined to be $[\theta_x - E_x^c m_x]$, where m_x is the graduated central death rate at age x. The relative deviation is defined to be $z_x = [\theta_x - E_x^c m_x]/[E_x^c m_x]^{\frac{1}{2}}$.

5. CONSTRUCTION OF THE LIFE TABLES FROM THE GRADUATED RATES OF MORTALITY

In order to produce the life tables it was necessary to calculate values of $\{q_x\}$ from the graduated central death rates. The available rates of mortality were

(i) q_0 and q_1 (see §2)

(ii) $m_2, m_3, \ldots, m_{111}$ (the graduated central death rates)

(iii) ϕ_0 (the average age at death of those dying in the first year of life—see Table 2.1)

(iv) the limiting age (taken as 113 for both sexes).

The procedure for determining the remaining rates of mortality was as follows:

(a) Calculate

$$m_0 = \frac{q_0}{1 - (1 - \phi_0)q_0}$$

(b) Assume

$$m_1 = q_1\left[1 + \frac{5}{12}m_2\right] \Big/ \left[1 + \left(\frac{1}{2} - \frac{1}{3}q_1\right)m_2 - \frac{7}{12}q_1\right]$$

and

$$q_2 = m_2\left[1 - \frac{13}{12}q_1\right] \Big/ \left[\left(1 - q_1\right)\left(1 + \frac{5}{12}m_2\right)\right],$$

which equations are valid, if l_x is quadratic over the age interval [1, 3].

(c) For $x = 3, 4, \ldots, 111$ calculate q_x as

$$q_x = m_x\left(1 - \frac{1}{2}m_{x-1}\right)\Big/\left[1 + \frac{5}{12}\left(m_x - m_{x-1}\right) - \frac{1}{6}m_x\, m_{x-1}\right],$$

which is exact if ℓ_t is quadratic over the age interval $[x - 1, x + 1]$.

(d) Finally, assume

$$q_{112} = 1 \text{ and } m_{112} = 2\ ,$$

which holds if ℓ_t is linear over the final year of life.

These assumptions having been made, the arrays $\{m_x\}$ and $\{q_x\}$ are completely determined for $x = 0, 1, \ldots, 112$. The values of $\{\ell_x\}$ were found successively from the radix ℓ_0 as

$$\ell_{x+1} = \ell_x(1 - q_x)$$

and then the value of $L_x = \int_x^{x+1} \ell_t\, dt$ was calculated as

$$L_x = \frac{\ell_x - \ell_{x+1}}{m_x}\ (x = 0, 1, \ldots, 112).$$

Since

$$T_x = \sum_{y=x}^{112} L_y\ ,$$

and

$$\mathring{e}_x = T_x/\ell_x\ ,$$

the complete expectation of life at each age was readily calculated.

(A brief discussion of the above approximations may be found in reference 4.)

For males figures 5.1 and 5.2 show the logarithm (to the base 10) of the crude central death rates and the corresponding curve of the graduated mortality rates over the age-ranges 2 to 30 and 25 to 95 respectively. Figures 5.3 and 5.4 illustrate the corresponding rates for females.

Full details of the resulting life tables are given in reference 1. Table 5.1 shows abridged life tables extracted from the complete mortality tables.

Figure 5.1: ELT14 (Males)

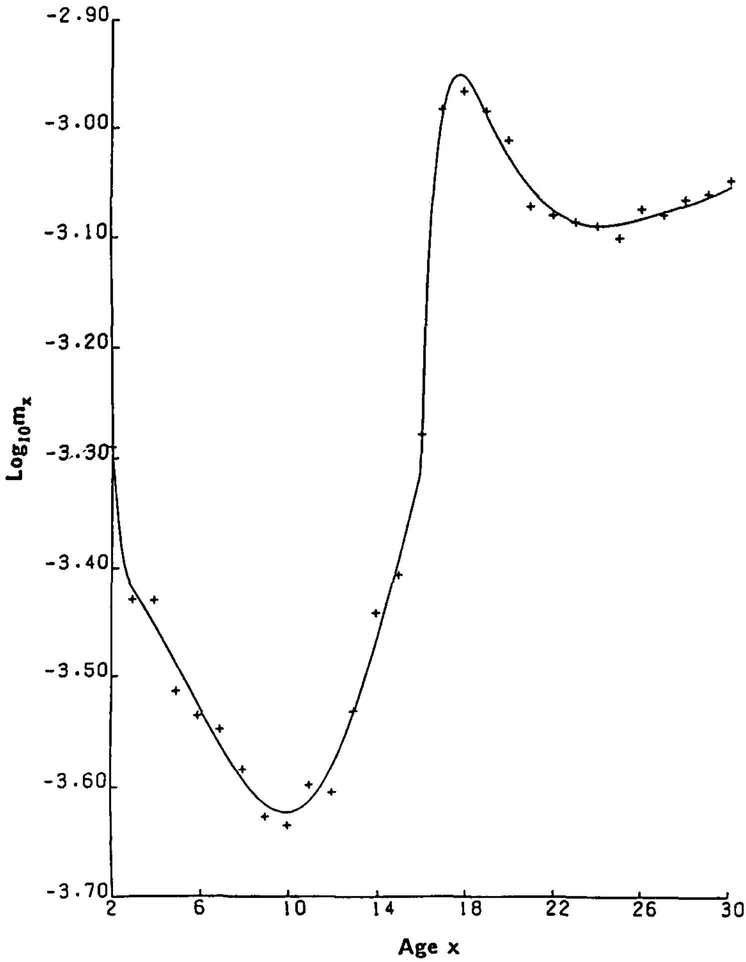

Figure 5.2: ELT14 (Males)

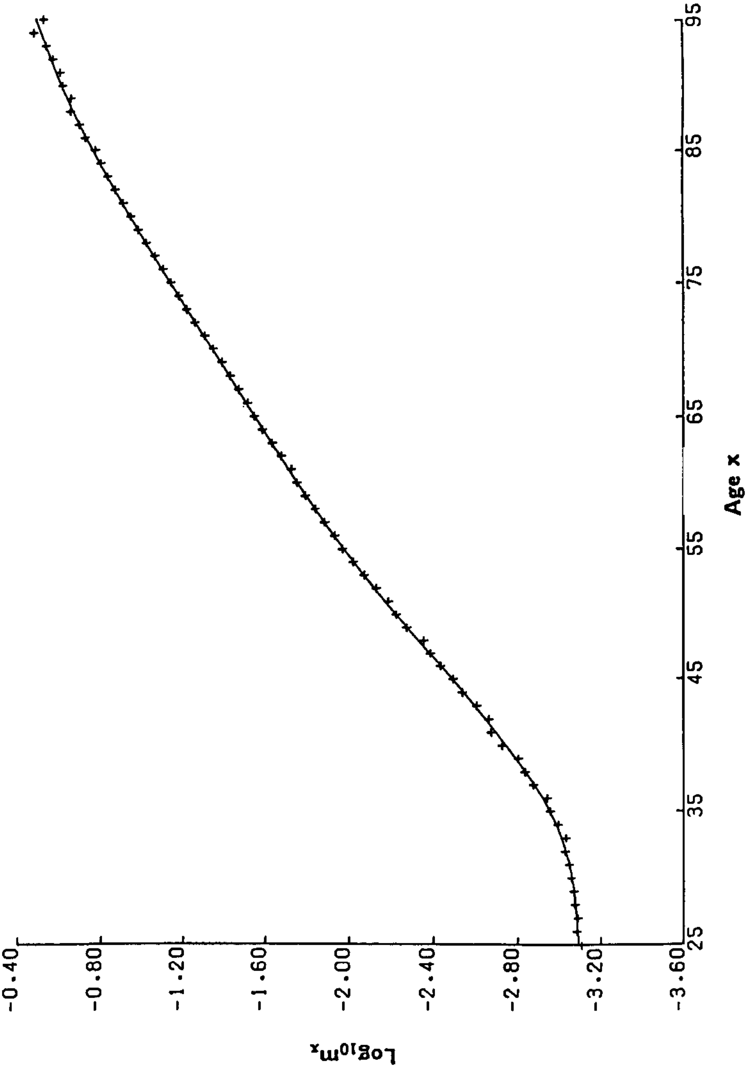

Figure 5.3: ELT14 (Females)

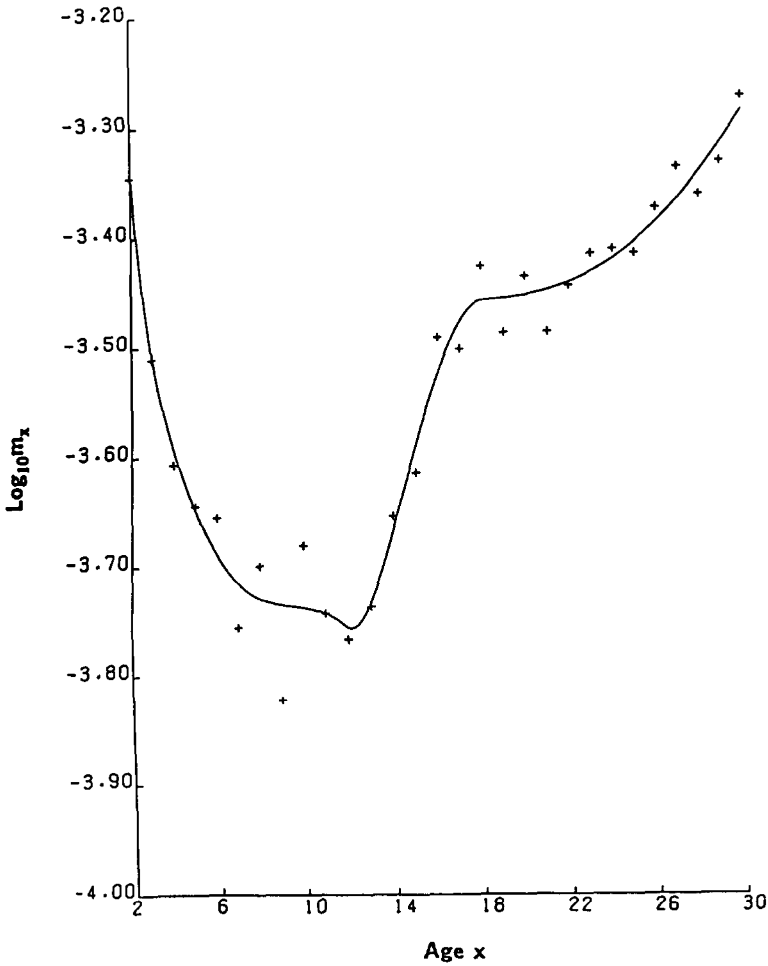

Figure 5.4: ELT14 (Females)

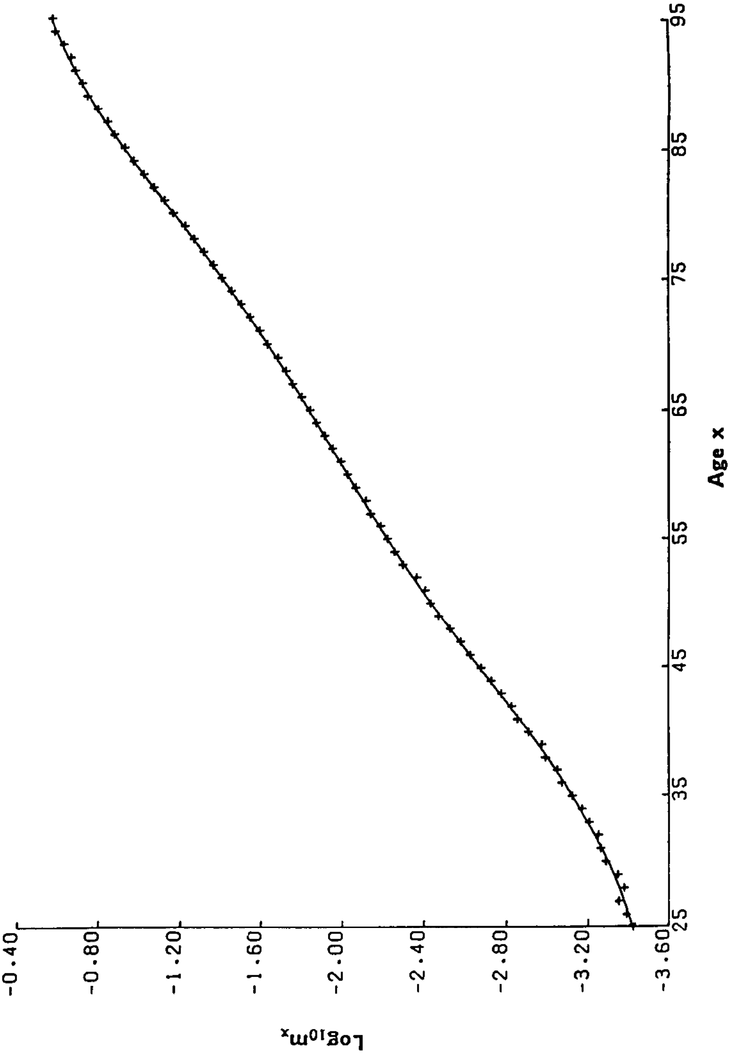

TABLE 5.1

Abridged form of the English Life Tables (No. 14)

	Males			Females			
x	l_x	$10^5 q_x$	$\mathring{e}_x$	l_x	$10^5 q_x$	$\mathring{e}_x$	x
0	100,000	1,271	71·04	100,000	984	77·00	0
1	98,729	85	70·96	99,016	72	76·77	1
5	98,522	32	67·10	98,844	22	72·90	5
10	98,385	24	62·19	98,746	18	67·97	10
15	98,250	41	57·27	98,653	26	63·03	15
20	97,849	93	52·50	98,497	35	58·12	20
25	97,432	81	47·71	98,318	39	53·22	25
30	97,027	88	42·90	98,105	52	48·33	30
35	96,564	113	38·09	97,807	78	43·47	35
40	95,907	184	33·33	97,436	127	38·67	40
45	94,787	332	28·70	96,573	219	33·95	45
50	92,758	615	24·26	95,244	378	29·39	50
55	89,152	1,098	20·13	93,034	624	25·02	55
60	83,199	1,843	16·38	89,564	986	20·89	60
65	74,261	2,949	13·04	84,384	1,528	17·01	65
70	61,864	4,703	10·12	76,864	2,443	13·41	70
75	46,123	7,416	7·70	65,886	4,110	10·21	75
80	28,965	11,334	5·78	50,623	6,982	7·49	80
85	14,153	16,591	4·35	31,978	11,922	5·38	85
90	4,928	22,693	3·33	14,608	18,468	3·95	90
95	1,156	28,971	2·57	4,457	24,914	3·02	95
100	165	38,087	1·86	893	32,252	2·21	100

ACKNOWLEDGEMENT

I wish to thank the Government Actuary's Department for making available before publication the crude data on which my experiments were based. I am indebted to the Numerical Algorithms Group for computer subroutines used in this work.

REFERENCES

1. English Life Tables No. 14 (1980-82): Series DS No. 7, H.M.S.O., 1987.
2. English Life Tables No. 13 (1970-72): decennial supplement 1970-72 England and Wales (Series DS No. 2), H.M.S.O., 1979.
3. "Considerations affecting the preparation of standard tables of mortality", (Continuous Mortality Investigation Committee), Journal Inst. Actuaries, 101(2), 1974, pp. 135-216 and Trans. Fac. Actuaries, 34(2), 1975, pp. 135-235.
4. McCutcheon, J. J., "Some elementary life table approximations", Trans. Fac. Actuaries, 35(3), 1977, pp. 297-301.
5. McCutcheon, J. J., "On the measurement of infant mortality", Trans. Fac. Actuaries, 34(4), 1976, pp. 433-442.
6. McCutcheon, J. J., "Spline graduation with variable knots", Proc. 22nd Int. Congress of Actuaries, 1984, Vol. 4, pp. 47-53.
7. McCutcheon, J. J., "Some remarks on splines", Trans. Fac. Actuaries, 37(4), 1981, pp. 421-438.
8. Thatcher, A. R., "Centenarians", Population Trends, (25), 1981, pp. 11-14.