VOLUME II JANUARY, 1902 No. 1

ON THE CONSTRUCTION AND USE OF LIFE-TABLES FROM A PUBLIC HEALTH POINT OF VIEW.

BY T. E. HAYWARD, M.B. (LOND.), F.R.C.S. ENG.

Medical Officer of Health of Haydock, Lancashire.

THE time may be considered opportune for bringing forward this subject in that the near approach of the publication of the completely classified results of the census of 1901 will cause the attention of many Medical Officers of Health to be directed to the possibility and desirability of using the census data for working out Life-Tables for their respective districts, and doubtless the construction of many more local Life-Tables will be contemplated than was the case after the census of 1891.

Anything, therefore, which will tend to remove or diminish the initial difficulties associated with such an undertaking will probably be acceptable to those concerned.

It is desirable at the outset to clearly limit and define the scope of what it is proposed to attempt in this paper.

To enter into a full explanation of the mathematical theory upon which the construction of Life-Tables is based, or to give an account of all the possible different methods which may be employed, with a discussion of the reasons for adopting one or the other, would necessarily occupy more space, if not time, than that available.

It appears, therefore, to the writer that the most serviceable position which he can take up is that of one who having made a special study of this subject, and having as the result of much laborious experimental work arrived at certain definite conclusions as to what is the best and most accurate method to adopt, is desirous of giving to those who may be willing to accept it, guidance as to the process of constructing a Life-Table, in such a way that it may be followed without any more mathematical knowledge being presupposed than an acquaintance with the ordinary rules of Arithmetic and with the use of common logarithms. With these considerations in view it is proposed

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

(1) To describe the construction of a complete or "extended" Life-Table (*i.e.* one for every separate year of life) by means of an "analytical" method.

(2) To give a short account of the practical uses to which such a Life-Table may be put when constructed.

(3) To show how by certain simple modifications of the "short" method first devised by the late Dr Wm. Farr, results can be arrived at which are practically identical with those worked out by the much more laborious extended method.

I. Construction of an extended life-table.

Data required.

The first and by no means the least important part of Life-Table construction is the compilation of the necessary foundations of numerical facts. These are the following:

(1) The total numbers of population as enumerated at two successive censuses, say 1891 and 1901.

These are required to be for each sex classified into the following age-groups:

0—5	25—35	65—75
5—10	35—45	75—85
10—15	45—55	85 and upwards
15—25	55—65	

It is better to take the age-group 15—25 together rather than in two groups as 15—20 and 20—25.

As the data after age 85 are unreliable it is best *not* to make a separate age-group of 85—95.

It may also be requisite in those districts within which are situated large public institutions, such as Hospitals or Lunatic Asylums, to correct the census numbers.

(2) The numbers of deaths registered in the district during the ten calendar years most nearly corresponding to the interval between the two censuses, say 1891—1900, classified into age-groups for each sex corresponding to those above given for the population-numbers.

The deaths of the age-group 0—5, however, require to be sub-classified as follows:

0—6 months	1—2 years
6—12 months	2—3 „
Total under 1 year	3—4 „
	4—5 „

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

It is necessary to take all possible care in getting the accurate numbers of deaths, by *excluding* all deaths of persons not properly belonging to the district, and also *including*, so far as can be ascertained, all deaths of persons properly belonging to the district which have occurred outside it, in Workhouses, Hospitals, &c. It may be noted in passing that it is greatly to be desired that the Registration arrangements in this country should be so amended as to facilitate these corrections.

It must be realized that an error of *one* in the death-number will have a very much greater effect in producing incorrect results than an error of *ten* in the population-number.

Thus, from the formulae to be shortly given it may be easily shown that with a population-number of 1000, and a death-number of 10, increasing the latter by one has the same effect as decreasing the former by 91, and decreasing the death-number by one is equivalent to *increasing* the population-number by 111.

(3) It is also requisite to have the following returns of deaths for some years preceding the decennial period.

Deaths at age 0—1 for each of the years 1887—90 inclusive
" " 1—2 " " 1888—90 "
" " 2—3 " " 1889—90 "
" " 3—4 for the year 1890

(4) The numbers of male and female births in each of the years 1886—1900 inclusive are also required.

How to calculate mean population-numbers from the census data.

After compiling and tabulating the data the next thing required is to deduce from the numbers enumerated at the two successive censuses such numbers as shall truly express the mean numbers living during the ten calendar years. In other words we have to calculate the "years of life" or the "lives at risk" during the decennium. This necessity has to be considered as applying (1) to the total population-numbers, and (2) to the numbers of the separate age and sex groups.

(*a*) The most simple and obvious method would be to take the arithmetical means of the two successive census enumerations. If this were done the sums of the parts, *i.e.* the separate age-groups, would necessarily equal the whole, *i.e.* the total population-number.

Although this method was used by the late Dr Wm. Farr, for the calculation of decennial death-rates, it cannot be considered accurate enough for Life-Table purposes for two reasons:

(1) On the assumption that population varies in the direction

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

of increase or decrease at a constant "rate," that is in Geometrical Progression, the true mean must necessarily be less than the arithmetical mean.

(2) Since the interval between two censuses does not exactly correspond with ten calendar years, but begins and ends a fourth part of a year *later*, the arithmetical mean will give a result *too great* in an *increasing* population and *too little* in a *decreasing* population. (See *Supplement to Fifty-fifth Annual Report of Registrar-General*, pp. xlii and xliii.)

(*b*) Admitting then the principle of Geometrical Progression, there are certain difficulties to be met with in its application. The chief difficulty is that the *sum of the parts cannot be made to coincide with the whole*, that is to say, the number arrived at by applying a process of calculation to the whole population-numbers differs from the sum of the numbers obtained by applying the same process to the numbers of the separate age and sex groups.

This difficulty has usually been overcome by making the *sum of the parts to be the true whole.*

Thus in working out the Brighton Life-Table (see Dr Newsholme's *Vital Statistics*, Third Edition, p. 263) each age-group was dealt with separately.

P_1 being the number in an age-group enumerated at the census of 1881, and P_2 being the number for the same age-group enumerated at the census of 1891, then $\frac{P_2}{P_1} = r$, and the years of life for that age-group are taken as the sum of the ten mid-year populations worked out by summing the series

$$P_1 . r^{\frac{1}{40}} + P_1 . r^{\frac{5}{40}} + \dots\dots\dots + P_1 . r^{\frac{37}{40}}$$

the sum of the series being $P_1 . \frac{r^{\frac{1}{40}}(r-1)}{r^{\frac{1}{10}}-1}$.

The total years of life were taken as the sum of the results of applying this method to each separate age-group. (This is not precisely the method given by Dr Newsholme, as he has given the calculation in two stages.) However, the sum of the ten mid-year populations is only an approximation to the true years of life which are to be obtained by the formula $\left(P_1 . \frac{r-1}{r^{\frac{1}{40}} . \log_e r}\right) \times 10.$

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

This practically was the method used by Dr Farr in his *English Life-Table*, p. xviii., applied to each age-group and the sum of the parts was taken as the true total.

Now, if there be any truth in the principle of geometrical progression at a constant rate applied to population-numbers, it would seem more rational to apply the principle *to the total population-numbers* and to devise some way of making the sum of the parts correspond to the whole.

The method which has been described by the writer in the *Journal of the Royal Statistical Society*, vol. LXII., Part iii., pp. 449—451, and also in Dr Newsholme's *Vital Statistics*, Third Edition, pp. 280—281, is based on the principle of first working out the true mean total population-number and then dividing this up by the method of mean proportions, which assumes that from one census to the next the proportion of the separate parts to the whole is changing uniformly in arithmetical progression, and in which the true mean total is finally divided up according to the proportions existing *on the above assumption* at $4\frac{3}{4}$ years after the earlier census, *i.e.* at the exact middle of the ten calendar years.

This is what until recently has appeared to the writer to be the best method, but it is defective in that it assumes that the proportion of the part to the whole taken at *one* point, is the true mean proportion for the *whole* period, and it gives identical results on reversing the proportions of the part to the whole at the two censuses.

The writer has been recently indebted, however, to Mr A. C. Waters for the knowledge of a more perfect method which he has worked out by an application of the Integral Calculus, and by which the true mean is arrived at by means of a full mathematical expression of the two assumptions, (1) that the *whole* population-number is changing in geometrical progression at a constant "rate," and (2) that the proportion of any selected part to the whole is uniformly changing in arithmetical progression.

As Mr Waters has since published a paper on his method in the *Journal of the Royal Statistical Society*, vol. LXIV., Part ii., June 1901, p. 293, those interested in the mathematical theory may be referred to this paper.

(For a general discussion of the various methods available for calculating mean population-numbers reference may also be made to the same *Journal* for September, 1901.)

The new method *might* be applied to the total population-numbers of any one district, but it is intended to be used on the assumption that the *whole* which is changing in Geometrical Progression is the population of the *entire country*, and that the *parts*, the proportions of which to the whole are changing in Arithmetical Progression, are the total populations of separate districts, as well as the sub-groups into which these latter are divided.

If Q_1 be the total population of the whole country (England and Wales) at the earlier census, and Q_2 the total population at the later census, then $\frac{Q_2}{Q_1} = r$, and the true mean total for the ten calendar years will be $Q_1 . \frac{r-1}{r^{\frac{1}{40}} . \log_e r}$.

(The "years of life" will be obtained by multiplying the last expression by 10 as the unit of calculation is a decennium.)

Now it can be shown that, on the two assumptions indicated above, two constant multiplying factors, m and n, can be worked out so that if P_1 be some selected part of Q_1, whether the total population of a district or some subdivision or age-group belonging to it, and if P_2 be the corresponding part of Q_2, then the true "years of life" for that part will be $(m . P_1 + n . P_2) \times 10$.

It is obvious that the *sum of all the separate parts* treated in this way must be equal to $(m . Q_1 + n . Q_2) \times 10$.

The only data needed for working out m and n are the total population-numbers for the whole country at two successive censuses; these being Q_1 and Q_2 respectively and r being $\frac{Q_2}{Q_1}$,

then
$$m = \frac{(r-1)\left(\frac{41}{40} + \frac{1}{\log_e r}\right) - r}{r^{\frac{1}{40}} . \log_e r},$$

and
$$n = \frac{r - (r-1)\left(\frac{1}{40} + \frac{1}{\log_e r}\right)}{r . r^{\frac{1}{40}} . \log_e r}.$$

If the above values of m and n be substituted in the expression $m . Q_1 + n . Q_2$ (i.e., $m . Q_1 + n . rQ_1$), and then simplified, the result will be found to work out to $Q_1 . \frac{r-1}{r^{\frac{1}{40}} . \log_e r}$.

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

(In order to work out these constants with the greatest possible degree of accuracy the values of r and of the hyperbolic logarithm of r have to be obtained to a large number of decimal places.)

When the final results of the census of 1901 are made known it will be possible for Medical Officers of Health to obtain the values of m and n without the trouble of working them out for themselves, as they will doubtless be published[1], and then the calculation of true mean numbers or years of life will be reduced to a very simple and easy matter.

To illustrate this the process may be shown as applied to an instance taken from the census enumerations of 1881 and 1891.

If reference be made to Dr Newsholme's *Vital Statistics* at page 262 the census data will be found on which his Life-Table for Brighton was based.

We find there,

Total population for males and females at census of 1881 = 128,350
" " " " " 1891 = 141,970

These two numbers have to be considered as forming parts of the entire census populations of the whole of England and Wales. Now the factors m and n for England and Wales, censuses 1881 and 1891 (the responsibility for their accuracy being the writer's) are

$$\log m = \bar{1}\cdot7354639 \qquad m = \cdot54383$$
$$\log n = \bar{1}\cdot6600871 \qquad n = \cdot45718$$

Therefore the years of life for Brighton during the ten years 1881—90

$$= \{(\cdot54383 \times 128{,}350) + (\cdot45718 \times 141{,}970)\} \times 10.$$

Since $\log m + \log 128{,}350 = \log (m \times 128{,}350)$

i.e. $\bar{1}\cdot7354639 + 5\cdot1083959 = 4\cdot8438598$

$\therefore$ $m \times 128{,}350 = 69{,}800\cdot7,$

and since $\log n + \log 141{,}970 = \log (n \times 141{,}970)$

i.e. $\bar{1}\cdot6600871 + 5\cdot1521966 = 4\cdot8122837$

$\therefore$ $n \times 141{,}970 = 64{,}905\cdot8$

and the total years of life $= (69{,}800\cdot7 + 64{,}905\cdot8) \times 10 = 1{,}347{,}065.$

[1] Based on the total population for England and Wales given in the Preliminary Report of the recent census, the factors "m" and "n" for the censuses of 1891 and 1901 have been worked out as follows:

$$m = \cdot5445944 \quad \log m = \bar{1}\cdot7360732$$
$$n = \cdot4564973 \quad \log n = \bar{1}\cdot6594383$$

(see Mr Waters' paper in the *Journal of the Royal Statistical Society* already referred to).

It is not likely that the finally corrected census number for 1901 will vary sufficiently from the number as yet given, to cause any material error through using the above values of m and n.

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

This number exceeds by 531 the number as worked out by Dr Newsholme.

By dealing similarly with the two census numbers for each of the age-groups the years of life belonging to each will be obtained, and the results will be checked by finding that the sum of the parts exactly corresponds with the total as already found.

It will now be evident that when once the factors m and n are obtained this method is to be preferred not only for its mathematical accuracy but for its simplicity and ease in application.

On the calculation of p_x values and on the relation between p_x and m_x.

Having now obtained and set down in tabular form the years of life and the deaths for each of the age and sex groups, the use which has to be made of these numbers is to calculate by their means the series of fractions which are set down in a Life-Table under the heading p_x, and which may be regarded as the most essential part of it. To work out these fractions for every single year of life constitutes by far the most laborious and difficult part of the task of constructing an extended Life-Table.

In Life-Table notation p_x simply means the chance (or probability) of surviving from the exact age x to the exact age $x+1$.

Thus if we have any number of persons l_x, of exact age x living at the beginning of a calendar year, and if a certain number d_x die during the year, then the chance of any individual of the l_x persons surviving to the end of the year is expressed thus

$$p_x = \frac{l_x - d_x}{l_x} = \frac{\text{survivors at end of year}}{\text{no. living at beginning of year}}.$$

We shall not be able, however, to deduce from our data the required p_x values in quite so simple a way, seeing that the population-numbers enumerated at the census and the deaths returned in the death-registers do not give us the numbers at exact age x, but the numbers at all ages between the fixed points x and $x+n$.

If then we have any number of persons P_x enumerated or estimated as living at the middle of a calendar year, this means that they may be of any age from x to $x+1$, and if a certain number d_x are returned as dying during the year who also may be of any age from x to $x+1$, the problem of calculating p_x is more complicated than the simple case previously considered.

In order to solve it we may assume two things—

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

(1) that at the middle of the calendar year the average age of the P_x persons was $x+\frac{1}{2}$, and

(2) that the number of deaths has been evenly distributed during the year, in other words that they have happened at equal intervals, half occurring in the first half of the year and half in the latter half of the year.

Therefore on these two assumptions, which when large numbers are being dealt with may be considered to be approximately true for every year of life *except the first*, the number living at the *beginning* of the year must be $P_x+\frac{1}{2}d_x$, and the number surviving at the *end* of the year must be $P_x-\frac{1}{2}d_x$ and therefore

$$p_x=\frac{P_x-\frac{1}{2}d_x}{P_x+\frac{1}{2}d_x}=\frac{2P_x-d_x}{2P_x+d_x}.$$

In actual practice the population and death-numbers are given in age-groups x to $x+n$, and if we are dealing with these groups as a whole it is assumed that the average age at the middle of the calendar year is $x+\frac{1}{2}n$. However, by certain processes of calculation the groups may be so divided up into numbers corresponding to each separate year of life that the problem is reduced to what has been just given.

It has been usual in Life-Table calculations to obtain the p_x values not *directly* from the population and death-numbers, but *indirectly* by first calculating m_x values. Now m_x is the "rate of mortality" per unit of the number living during the age x to $x+1$, and it is expressed by the fraction $\frac{d_x}{P_x}$. It is called in actuarial terminology the "central death-rate," meaning the rate at which people are dying at the central point of the age x to $x+1$.

We have already found that $p_x=\frac{P_x-\frac{1}{2}d_x}{P_x+\frac{1}{2}d_x}$, and dividing the numerator and denominator of this fraction by P_x, we have

$$p_x=\frac{\frac{P_x-\frac{1}{2}d_x}{P_x}}{\frac{P_x+\frac{1}{2}d_x}{P_x}}=\frac{1-\frac{1}{2}\frac{d_x}{P_x}}{1+\frac{1}{2}\frac{d_x}{P_x}}=\frac{1-\frac{1}{2}m_x}{1+\frac{1}{2}m_x}=\frac{2-m_x}{2+m_x},\text{ since, as we have seen, } m_x=\frac{d_x}{P_x}.$$

Conversely, $m_x=\frac{2(1-p_x)}{1+p_x}$.

If we are making calculations to deduce the "Law of mortality" from the observed facts with relation to the living and the dying at certain ages, the m_x values are necessary, but for the ordinary work of

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

Life-Table construction, such as is now being considered, and which is based on a *particular series of observed facts* without any necessary regard to other similar series of facts, there is no need to work out these values.

Calculation of the p_x values for the first five years of life.

Although at each census there are enumerated the numbers of those living at each of the first five years of life, *i.e.* at ages 0—1, 1—2, 2—3, 3—4, and 4—5, the results have hitherto been found, by reason of obvious misstatement of age, to be altogether impossible and unreliable, and they have to be discarded in Life-Table construction.

The only use which can be made of the census data for these five years is to give us the total number for each sex in the age-group 0—5, and from this, by means of the process of calculation already described, the mean number living, or the "years of life" can be deduced for this age-period.

Under these circumstances an approximate calculation has to be made based on the recorded numbers of births and deaths.

The principle of this will be more readily comprehended if the process be first of all considered with relation to one single calendar year.

The deaths under one year of age (at age 0—1) which were registered in the year 1891 occurred partly out of those born in 1890 and partly out of those born in 1891. It may therefore be a fair assumption to consider that these deaths occurred out of (1) either the births registered during the last half of 1890 *plus* those registered in the first half of 1891, or (2) more conveniently out of a number represented by the arithmetical mean of the births registered in the two years 1890—91.

For the purposes of Life-Table construction we have to consider that on 1st January, 1891, there existed *at birth* a number corresponding to half the sum of the births in 1890 and 1891, and that the deaths at age 0—1 which were registered during the year 1891 occurred out of these.

In order to get the mean number living at age 0—1 for the year 1891, or in other words the number living at the middle of the year 1891, we shall have to deduct from the number at birth on 1st January the deaths under 6 months of age which were registered in that year.

If we call this mean number living at age 0—1 P_0, then the chance of surviving from age 0 to age 1 will be represented by

$$\frac{P_0 - \text{deaths at 6 to 12 months}}{P_0 + \text{deaths at 0 to 6 months}} = \frac{\text{survivors at end of year}}{\text{number at birth}}.$$

It would of course come to the same thing if the p_0 fraction were represented thus

$$\frac{\text{number at birth} - \text{deaths at age 0 to 1}}{\text{number at birth}}.$$

Similarly half the sum of the births in 1889 and 1890 represents the number *at birth* on 1st January, 1890, and after deducting the deaths at age 0—1 during 1890 we shall have the number *at age* 1 surviving on January 1st, 1891 out of whom the deaths at age 1—2 will occur during 1891. The mean number living at age 1—2 at the middle of 1891 will be found by deducting half the deaths at age 1—2 occurring during 1891, and calling the resulting number P_1 then the chance of surviving from age 1 to age 2, *i.e.* p_1, will be represented by

$$\frac{P_1 - \frac{1}{2}\text{ deaths at age 1 to 2}}{P_1 + \frac{1}{2}\text{ deaths at age 1 to 2}} = \frac{\text{survivors at end of year}}{\text{survivors at beginning of year}}.$$

It would again come to the same thing to represent p_1 as being equal to

$$\frac{\text{number at age 1} - \text{deaths at age 1 to 2}}{\text{number at age 1}}.$$

Still calculating on the same principle, half sum of births in 1888 and 1889, less deaths at age 0—1 in 1889, and less deaths at age 1—2 in 1890, will give the number at age 2 living on 1st January, 1891, out of whom the deaths at age 2—3 will occur during 1891.

Commencing with $\frac{1}{2}$ sum of births in 1887 and 1888 by a similar method we arrive at the number of survivors at age 3 on 1st January, 1891, out of whom the deaths at age 3—4 will occur during 1891.

Finally by commencing with $\frac{1}{2}$ sum of births in 1886 and 1887 and deducting in succession deaths at age 0—1 during 1887, deaths at age 1—2 during 1888, deaths at age 2—3 during 1889, and deaths at age 3—4 during 1890, we obtain the number *at age* 4 surviving on 1st January 1891 out of whom the deaths at age 4—5 will occur during 1891.

The mean numbers living P_2, P_3, and P_4 and the expressions representing p_2, p_3, and p_4, are arrived at in the same way as P_1 and p

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

given above, the assumption being that for every year of life *except the first*, the deaths occur at equal intervals during the calendar year.

If the year 1892 be similarly dealt with and then every year in succession until and including 1900, and if finally the results are summed, the following scheme will be arrived at for the decennium 1891—1900.

(*a*) For the number *at Birth*
$\frac{1}{2}$ births in 1890 + all births in 1891 to 99 + $\frac{1}{2}$ births in 1900.

(*b*) For the number *at* 1 *year of age*
$\frac{1}{2}$ births in 1889 + all births in 1890 to 98 + $\frac{1}{2}$ births in 1899
less deaths at age 0—1 in the ten years 1890—99.

(*c*) For the number *at* 2 *years of age*
$\frac{1}{2}$ births in 1888 + all births in 1889 to 97 + $\frac{1}{2}$ births in 1898
less { deaths at age 0—1 in the ten years 1889—98
„ „ 1—2 „ „ 1890—99. }

(*d*) For the number *at* 3 *years of age*
$\frac{1}{2}$ births in 1887 + all births in 1888 to 96 + $\frac{1}{2}$ births in 1897
less { deaths at age 0—1 in the ten years 1888—97
„ „ 1—2 „ „ 1889—98
„ „ 2—3 „ „ 1890—99. }

(*e*) For the number *at* 4 *years of age*
$\frac{1}{2}$ births in 1886 + all births in 1887 to 95 + $\frac{1}{2}$ births in 1896
less { deaths at age 0—1 in the ten years 1887—96
„ „ 1—2 „ „ 1888—97
„ „ 2—3 „ „ 1889—98
„ „ 3—4 „ „ 1890—99 }

Let the five numbers deduced by the above scheme be denoted respectively as *a*, *b*, *c*, *d*, and *e*.

In order to obtain from them the sum of the ten mid-year or mean population-numbers,

from	*a*	subtract		deaths	at age	0—6 months	during	1891—1900
„	*b*	„	$\frac{1}{2}$	„	„	1—2 years	„	„
„	*c*	„	$\frac{1}{2}$	„	„	2—3 „	„	„
„	*d*	„	$\frac{1}{2}$	„	„	3—4 „	„	„
„	*e*	„	$\frac{1}{2}$	„	„	4—5 „	„	„

Let the five altered numbers thus obtained be denoted respectively by n_0, n_1, n_2, n_3, and n_4, and let their sum $= N$.

Now N represents the sum of the ten mid-year population-numbers of those at all ages from birth to age 5.

We have previously obtained from the census data a mean population-number of those at all ages from birth to age 5, and the

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

multiplication of this by 10 gave us the "years of life" or the total number of those living at ages 0—5 during the decennium. Let this latter number be denoted by C.

Were it not for disturbing influences, N and C ought to at least very nearly correspond. However as a matter of fact N is usually found to be the greater[1]. The principal causes of this difference are (1) excess of emigration over immigration and (2) over-statement of age in the death-registers.

However, we must take the total C as a fixed quantity, and divide it up in the proportions which n_0, n_1, n_2, n_3 and n_4 respectively bear to N, the resulting numbers being P_0, P_1, P_2, P_3 and P_4.

Thus $$n_0 : N :: P_0 : C,$$

or $$\log P_0 = \log n_0 + (\log C - \log N),$$

similarly $$\log P_1 = \log n_1 + (\log C - \log N),$$

&c. &c.

Then $$p_0 = \frac{P_0 - \text{deaths at 6 to 12 months during 1891 to 1900}}{P_0 + \text{deaths at 0 to 6 months during 1891 to 1900}},$$

$$p_1 = \frac{P_1 - \frac{1}{2}\text{ deaths at age 1 to 2 during 1891 to 1900}}{P_1 + \frac{1}{2}\text{ deaths at age 1 to 2 during 1891 to 1900}},$$

&c. &c.

The distribution of the difference between N and C in proportion to n_0, n_1, n_2, n_3 and n_4, is open to the objection that migration or other disturbing influences may have not existed actually in quite the same proportions. However, after a good deal of labour expended in trying to find a better way the writer is of opinion that the uncertainty of the various disturbing factors is too great to justify any departure from the method which has usually been adopted.

In case any reader should note a difference between the above description and what has previously been given by the writer (see *Journal of the Royal Statistical Society*, September, 1899, and Dr Newsholme's *Vital Statistics*, Third Edition, pp. 271—273), it simply amounts to this—formerly N and C were made *comparable* by *adding* deaths to C—now they have been made comparable by *subtracting* deaths from N.

[1] The data for England and Wales for 1841—50 and 1851—60 produce a value of N less than C, the difference being most marked in 1841—50. This is probably explained by the births having not all been registered. For each succeeding decennium up to 1881—90 there has been a progressively increasing excess of N over C.

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

For the sake of future convenience in calculating, the p_0 value may be taken in two stages, thus:

$$(1)\quad p_{0\text{ to }\frac{1}{2}} = \frac{P_0}{P_0 + \text{deaths at 0 to 6 months}},$$

$$(2)\quad p_{\frac{1}{2}\text{ to }1} = \frac{P_0 - \text{deaths at 6 to 12 months}}{P_0}.$$

How to calculate the p_x values from p_5 onwards.

Up to the point which has now been reached the work is the same for a short or extended Life-Table, or whether the "graphic" or an "analytical" method is to be afterwards followed.

(For a description of the "graphic" method reference may be made to Dr Newsholme's *Vital Statistics*, as the present writer has only undertaken to try to explain an analytical method.)

The data available in either case are the "years of life," *i.e.* ten times the mean annual population-number, and the deaths for the ten calendar years, as classified in the given age-groups.

It is obvious that a mean value of $p_{x\text{ to }x+n}$ might be easily calculated for each age-period by the formula

$$\frac{2P_{x\text{ to }x+n} - d_{x\text{ to }x+n}}{2P_{x\text{ to }x+n} + d_{x\text{ to }x+n}}.$$

As will be shown afterwards this is the plan upon which the short method of Life-Table construction is based, but for an extended Life-Table it is necessary to adopt some process of calculation so that the p_x values when plotted out to scale shall show not a series of step-like abrupt ascents or descents, as would be the case by the simple method just referred to, but an even curve without any sudden transitions or breaks in its symmetry.

Such a curve when obtained may not be exactly such as the true facts for each separate year of life would show (if we could get them), for some age-periods are more "critical" to life than others. However, it is the nearest approach which we can make to the hypothetical true curve and probably does not diverge very greatly from it.

The "analytical" method consists in deducing from the given numerical facts the required p_x values by means of some adaptation of the mathematical process known as "Interpolation" in a series of quantities by the method of "Finite Differences."

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

On "Finite Differences" and Interpolation.

As it would scarcely be possible for anyone to calculate by this process without having at least some general idea, clear and precise in so far as it goes, of its essential principles, some preliminary attempt must be made to explain these, in spite of the limitations laid down at the commencement of this paper.

Suppose that we take a series of numerical quantities, for example's sake the fourth powers of the numbers 3, 4, 5, 6, and 7, and set them down in inverse order, as in the left-hand column below, marking them in succession by the symbols u_0, u_1, u_2, u_3, and u_4.

		Δ	Δ^2	Δ^3	Δ^4
u_0	2401	-1105	$+434$	-132	$+24$
u_1	1296	-671	$+302$	-108	
u_2	625	-369	$+194$		
u_3	256	-175			
u_4	81				

Now let this column of figures be "differenced," that is, change the sign of the upper quantity, and take the algebraical sum of it and the one below, that is their *sum if the signs are like,* and their *difference if the signs are unlike.* Let this be done in succession all down the column. In the next column to the right we have now the series of "first differences," marked by the symbol Δ. The first difference opposite u_0 is called Δu_0, and the first difference opposite u_1 is called Δu_1, etc., etc. Let a similar operation be carried out with the column of first differences, then we get the "second differences," marked as Δ^2, and the one in a line with u_0 is called $\Delta^2 u_0$. Let the process be repeated until we come to the last difference Δ^4, beyond which the process cannot be carried.

We have thus a series of five quantities giving a "constant fourth difference," or "with four orders of differences," and generally for n orders of differences we must have $n+1$ terms.

Now several considerations will at once be obvious.

(1) Having had given the five quantities u_0, u_1, u_2, u_3, and u_4 we have obtained by successive differencing the values of Δu_0, $\Delta^2 u_0$, $\Delta^3 u_0$, and $\Delta^4 u_0$, but if we had had given u_0 and the line of differences *opposite to* u_0 we could just as easily have carried these differences down so as to have obtained the values of u_1, u_2, u_3, and u_4.

Thus adding Δu_0, *i.e.* -1105, to u_0, *i.e.* 2401, we obtain u_1, *i.e.* 1296.

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

Again, adding $\Delta^2 u_0$, *i.e.* $+434$, to Δu_0, *i.e.* -1105, we obtain Δu_1, *i.e.* -671, and adding this to u_1 we obtain u_2, etc., etc.

The most convenient way to proceed, having given $u_0 = 2401$, $\Delta u_0 = -1105$, $\Delta^2 u_0 = +434$, $\Delta^3 u_0 = -132$, and $\Delta^4 u_0 = +24$, is shown as follows, since it is easier to work from left to right.

			$+2401 = u_0$
			-1105
		-1105	$+1296 = u_1$
		$+434$	-671
	$+434$	-671	$+625 = u_2$
	-132	$+302$	-369
-132	$+302$	-369	$+256 = u_3$
$+24$	-108	$+194$	-175
-108	$+194$	-175	$+81 = u_4$

(2) By carrying down the differences in the above table one stage further, *i.e.* by continuous *addition* of the successive differences, the value of another term, u_5, could be obtained, which will be found to be 16, the fourth power of the number 2, and so on for as many terms as we please, all the terms of the series having the constant fourth difference $+24$.

On the other hand by successive *subtraction* of the differences, as set down in the table first given opposite u_0, *i.e.* by changing their signs and adding, we could carry the series one stage *upwards* and obtain the term u_{-1}, which will be found to be 4096, *i.e.* the fourth power of the number 8, and so on ad infinitum.

The general equation which expresses the relation of the terms of a series of quantities differing from each other by a constant *n*th difference is the following, which follows the law of the "Binomial Theorem."

$$u_x = u_0 + x\,\Delta u_0 + \frac{x(x-1)}{2}\Delta^2 u_0 + \frac{x(x-1)(x-2)}{2\,.\,3}\Delta^3 u_0 + \frac{x(x-1)(x-2)(x-3)}{2\,.\,3\,.\,4}\Delta^4 u_0 + \ldots$$

By means of this equation any term u_x of the series can be calculated without the trouble of working up or down to it by the method above indicated. This may be verified by being applied to the numerical instances already used.

If working upwards, *i.e.* if the value of x is negative, care is needed with regard to the signs.

Again, if the above equation be expanded and reduced to its simplest

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

form, it will be found that each power of x has a *constant coefficient*—that is the equation can be expressed in the following form

$$u_x = A + Bx + Cx^2 + Dx^3 + Ex^4 + \ldots$$

Thus B, or the coefficient of $x = \Delta u_0 - \frac{1}{2}\Delta^2 u_0 + \frac{1}{3}\Delta^3 u_0 - \frac{1}{4}\Delta^4 u_0 + \ldots$

Certain formulae will afterwards be given the use of which is to obtain the line of differences opposite to u_0. If the object of the above brief remarks had been to explain *how* these formulae had been *worked out* they would have had to be expanded much beyond the limits laid down. It may be possible, however, to *work from* them with such explanation as has been given.

The process of interpolation may be applied to the population and death-numbers in such a way as to obtain a value of $2P - d$ and $2P + d$ for each year of age and therefore the value of p_x by the fraction

$$\frac{2P - d}{2P + d}.$$

The chief objection to this method is that it entails the labour of a *double* series of interpolations all throughout.

Another method is to first work out p_x values at certain fixed ages, $p_5, p_{10}, p_{15}, p_{25}$, etc. and then by applying the formulae of interpolation to these to obtain the required complete series of p_x values.

Such a method is the one proposed to be now explained. It is based on an application of the Differential Calculus devised originally by Mr A. C. Waters. A full description of this, with some proposed slight modifications, is given in the *Journal of the Royal Statistical Society*, for December 1900, to which any reader is referred who may be desirous of going into the mathematical theory.

In this communication the terminology of the Differential Calculus will be eliminated and only such explanations will be attempted as may suffice for the arithmetical work required.

In order to demonstrate the method of procedure it will be necessary to take some assumed set of foundation figures.

In the two left-hand columns given below we are supposed to have the "years of life" and the deaths in ten calendar years at the age-periods indicated for males.

From these are constructed the two right-hand columns representing twice population (*i.e.* years of life) minus deaths, and twice population plus deaths, *at age x and upwards*.

The reason for constructing the two right-hand columns which represent population and deaths at age x and upwards will be com-

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

At age	Years of life	Deaths	At age and upwards	$2P-d$	$2P+d$
4—5	330,090	5,645	4	23,608,335	24,015,655
5—10	1,571,715	11,980	5	22,953,805	23,349,835
10—15	1,450,170	5,375	10	19,822,355	20,194,425
15—25	2,614,970	16,135	15	16,927,390	17,288,710
25—35	2,222,620	24,505	25	11,723,585	12,052,635
35—45	1,669,670	32,640	35	7,302,850	7,582,890
45—55	1,102,130	34,325	45	3,996,150	4,210,910
55—65	621,470	33,810	55	1,826,215	1,972,325
65—75	264,820	27,185	65	617,085	695,575
75—85	59,210	10,790	75	114,630	138,750
85 and upwards	4,135	1,270	85	7,000	9,540

prehended if they are regarded from a geometrical point of view as perpendicular lines or "ordinates" erected upon a base line or "abscissa." The process of interpolation by finite differences consists practically in drawing a curve of the nth degree through the upper extremities of $n+1$ of these ordinates. The ordinate erected at any intermediate point, *i.e.* at any intermediate age, measured from the abscissa to the curve, is the measure of the corresponding value of $2P-d$ or $2P+d$.

Therefore as we have the height of the ordinate given for age x and upwards, denoted by the symbol u_x, if we can get the measure of the corresponding ordinate for age $x+1$ and upwards, u_{x+1}, then $u_x - u_{x+1}$ will give the measure of $2P-d$, or $2P+d$ for age x to age $x+1$ and so from the fraction $\dfrac{2P-d}{2P+d}$ the value of p_x is easily deduced.

It is found best however to work *not* directly from the numbers but from the corresponding Logarithms, one chief reason for this being that there is thus possible to be obtained a rational continuation of the series below the point at which the actual data cease to be available, that is after the age 85.

The two right-hand columns in the above table are therefore translated into their corresponding logarithms as below.

$u_4 = 7{\cdot}3730654$	$U_4 = 7{\cdot}3804945$
$u_5 = 7{\cdot}3608547$	$U_5 = 7{\cdot}3682838$
$u_{10} = 7{\cdot}2971553$	$U_{10} = 7{\cdot}3052315$
$u_{15} = 7{\cdot}2285900$	$U_{15} = 7{\cdot}2377626$
$u_{25} = 7{\cdot}0690604$	$U_{25} = 7{\cdot}0810820$
$u_{35} = 6{\cdot}8634924$	$U_{35} = 6{\cdot}8798347$
$u_{45} = 6{\cdot}6016418$	$U_{45} = 6{\cdot}6243759$
$u_{55} = 6{\cdot}2615519$	$U_{55} = 6{\cdot}2949785$
$u_{65} = 5{\cdot}7903451$	$U_{65} = 5{\cdot}8423439$
$u_{75} = 5{\cdot}0592983$	$U_{75} = 5{\cdot}1422330$
$u_{85} = 3{\cdot}8450980$	$U_{85} = 3{\cdot}9795484$

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

Log $2P-d$ at age x and upwards is denoted by the symbol u_x, and log $2P+d$ at age x and upwards by the symbol U_x.

At this point it must be noted that whereas we have used the symbol p_x to denote the chance of surviving from age x to age $x+1$, the values to be obtained by the method to be immediately described give the chance of survival as existing at the *exact age x*, or what may be taken practically as the chance of surviving from age $x-\frac{1}{2}$ to age $x+\frac{1}{2}$, and these values will hereafter be denoted by the symbol p'_x.

Now the general formula by which the *logarithms* of the p'_x values are to be obtained from the tables of u_x and U_x values given above is this:

$$\log p'_x = (u_x + \log b) - (U_x + \log B).$$

Thus for p'_5 the values of u_5 and U_5 can be at once written down in the formula.

From what has been previously said it will be understood that B is the coefficient of x in the expansion of the equation $u_x = A + Bx + Cx^2 + \ldots$. For the sake of distinction the small letter b is used to denote the coefficient in the series $2P-d$, and the capital letter B the corresponding coefficient in the series $2P+d$. For the purposes of the equation for p'_x above given the values of b and B require a *separate calculation for each value of x*.

Two points may be here noted with regard to the above formula for p'_x. (1) The coefficient b (or B) is a negative quantity, but it may be left positive since $\frac{-b}{-B} = \frac{b}{B}$.

(2) Any *identical* multiples of b and B may be used in the equation since $\frac{xb}{xB} = \frac{b}{B}$.

The series from which p'_5, p'_{10}, and p'_{15} have to be calculated is for b', u_4, u_5, u_{10}, u_{15}, u_{25}, u_{35} and for B, the corresponding series commencing with U_4.

The values of b and B can be thus expressed in terms of the series u_4, u_5, &c.

(1) In the equation for p'_5

$$b = \left[\begin{pmatrix} 883{,}190u_5 \\ +152{,}768u_{10} \\ +3{,}410u_{25} \end{pmatrix} - \begin{pmatrix} 1{,}000{,}000u_4 \\ +39{,}060u_{15} \\ +308u_{35} \end{pmatrix}\right] \div 1{,}432{,}200.$$

(2) In the equation for p'_{10}

$$b = \left[\begin{pmatrix} 1{,}250{,}000u_4 \\ +343{,}728u_{10} \\ +585{,}900u_{15} \\ +\ 2{,}772u_{35} \end{pmatrix} - \begin{pmatrix} 2{,}148{,}300u_5 \\ +34{,}100u_{25} \end{pmatrix}\right] \div 5{,}728{,}800.$$

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

(3) In the equation for p'_{15}

$$b=\left[\begin{pmatrix}1{,}575{,}420u_5\\+1{,}035{,}090u_{15}\\+75{,}020u_{25}\end{pmatrix}-\begin{pmatrix}1{,}000{,}000u_4\\+1{,}680{,}448u_{10}\\+5{,}082u_{35}\end{pmatrix}\right]\div 4{,}296{,}600.$$

From what has been said above it will be obvious that only the *numerators* of the above fractions need be worked out.

In order to save labour the formulae given above for the numerators have been re-arranged as below so as to give the multiples of b indicated by the denominators.

(1) for p'_5

$$\left\{(220-3)\times\begin{bmatrix}+4{,}070u_5\\+704u_{10}\end{bmatrix}\right\}+3{,}410u_{25}-\begin{bmatrix}1{,}000{,}000u_4\\+39{,}060u_{15}\\+308u_{35}\end{bmatrix},$$

(2) for p'_{10}

$$\left\{(220-3)\times\begin{bmatrix}+(11\times 144)u_{10}\\+(3{,}000-300)u_{15}\\-(10{,}000-100)u_5\end{bmatrix}\right\}+\begin{bmatrix}1{,}250{,}000u_4\\+(3{,}080-308)u_{35}\end{bmatrix}-34{,}100u_{25},$$

(3) for p'_{15}

$$\left\{(220-3)\times\begin{bmatrix}\begin{pmatrix}+(700-40)u_5\\-704u_{10}\end{pmatrix}\times 11\\+(5{,}000-230)u_{15}\end{bmatrix}\right\}+(3{,}410\times 22)u_{25}-\begin{bmatrix}1{,}000{,}000u_4\\+(3{,}080+2{,}002)u_{35}\end{bmatrix}.$$

The labour may be still further diminished by subtracting *any one* of the given terms u_4, u_5, &c. *from all the terms* of the series, thus reducing the one subtracted from itself to zero, and then applying the formulae to the *new* series of values.

If the calculations have been first made with u_4 eliminated it is well to check the results by repeating the work with another term (say u_{15} or u_{35}) eliminated.

The working out of the above formula for p'_5 gives the following result:

$$\text{multiple of } b=-17617{\cdot}5017318$$
$$\text{,,}\quad\text{,,}\quad B=-17573{\cdot}8315096.$$

$$\therefore\quad \log p'_5=(7{\cdot}3608547+\log 17617{\cdot}502)-(7{\cdot}3682838+\log 17573{\cdot}832)$$
$$=(7{\cdot}3608547+4{\cdot}2459444)-(7{\cdot}3682838+4{\cdot}2448665)$$
$$=\bar{1}{\cdot}9936488.$$

$$\therefore\quad p'_5={\cdot}98548.$$

After having calculated the logs of p'_5, p'_{10}, and p'_{15} the remaining values from log p'_{25} onwards are much more easily arrived at.

The formula given below in each case produces the value of $b \times -120$, and as before the similar multiples of b and B are used.

Formulae for multiples of b and B

for p'_{25} $8(u_{15}-u_{35})-(u_5-u_{45})$
for p'_{35} $8(u_{25}-u_{45})-(u_{15}-u_{55})$
for p'_{45} $8(u_{35}-u_{55})-(u_{25}-u_{65})$
for p'_{55} $8(u_{45}-u_{65})-(u_{35}-u_{75})$
for p'_{65} $8(u_{55}-u_{75})-(u_{45}-u_{85})$
for p'_{75} $8(u_{65}-u_{85})-(u_{55}-u_{95})$
for p'_{85} $8(u_{75}-u_{95})-(u_{65}-u_{105})$.

(For B of course the corresponding U_x values are to be used.)

In order to obtain the values of u_{95} and u_{105} it is only necessary to difference the series u_{45}, u_{55}, u_{65}, u_{75}, u_{85} and to carry the differences down for two stages.

After having completed this stage of the work we shall have the following series of values as the foundation for the subsequent process of interpolation.

$\log p'_5 = \bar{1}\cdot9936488$
$\log p'_{10} = \bar{1}\cdot9983627$
$\log p'_{15} = \bar{1}\cdot9980733$
$\log p'_{25} = \bar{1}\cdot9965108$
$\log p'_{35} = \bar{1}\cdot9933917$
$\log p'_{45} = \bar{1}\cdot9893048$
$\log p'_{55} = \bar{1}\cdot9823833$
$\log p'_{65} = \bar{1}\cdot9663823$
$\log p'_{75} = \bar{1}\cdot9359880$
$\log p'_{85} = \bar{1}\cdot8842469$.

Now it would be quite possible to take these ten values and by one scheme of interpolation with nine orders of differences to obtain the required continuous series of intermediate values (in fact they have been worked through by the writer as far as p'_{35}). However the labour is too great, and the results so obtained are to be practically arrived at by an easier method, which consists in effecting interpolations in *several overlapping series* and then *joining* or "*welding*" these series with each other at certain parts, the final result being very nearly what would have been arrived at by the more laborious method.

The scheme which is recommended for adoption after the trial of others more elaborate as well as more simple, is to be represented as follows: the portion of each series used being included in brackets and the parts of series to be combined (by a method to be afterwards

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

described) being indicated by over-lines and under-lines, *i.e.* the part *under*-lined is to be welded with the part of the series below it which is *over*-lined.

Series 1. 6 orders of differences

$$[p'_5,\ p'_{10},\ p'_{15},\ \underline{p'_{25},\ p'_{35}}]\ p'_{45},\ p'_{55}$$

Series 2. 5 orders of differences

$$p'_{15},\ [\overline{p'_{25},\ p'_{35}},\ \underline{p'_{45},\ p'_{55}}]\ p'_{65}$$

Series 3. 5 orders of differences

$$p'_{25},\ p'_{35},\ [\overline{p'_{45},\ p'_{55}},\ \underline{p'_{65}}]\ p'_{75}$$

Series 4. 5 orders of differences

$$p'_{35},\ p'_{45},\ [\overline{p'_{55},\ p'_{65}},\ p'_{75},\ p'_{85}\ldots\ldots]$$

It will thus be evident that the values of p'_6 to p'_{24} are derived from the first series alone, p'_{26} to p'_{34} by combining series 1 with series 2, p'_{36} to p'_{44} from series 2 alone, p'_{46} to p'_{54} by combining series 2 with series 3, p'_{56} to p'_{64} by combining series 3 with series 4, and from p'_{66} onwards from series 4 alone.

Series 1 is the one which entails most labour as the intervals between the terms are unequal.

Using the small letter δ instead of the capital Δ as the symbol representing the successive differences, and using the symbol u_0 to represent p'_5, u_5 to represent p'_{10}, u_{10} to represent p'_{15}, u_{20} to represent p'_{25}, &c., &c., the following formulae will give the line of differences opposite to u_0, viz. δu_0, $\delta^2 u_0$, $\delta^3 u_0$, $\delta^4 u_0$, $\delta^5 u_0$ and $\delta^6 u_0$.

$$\delta^6 u_0 = \begin{bmatrix} -0{\cdot}001{,}024u_5 \\ -0{\cdot}000{,}18u_{40} \\ +0{\cdot}000{,}028u_{50} \end{bmatrix} \div 21 + \begin{bmatrix} 0{\cdot}000{,}012u_0 \\ +0{\cdot}000{,}06u_{10} \\ +0{\cdot}000{,}024u_{30} \end{bmatrix} - 0{\cdot}000{,}04u_{20}$$

$$\delta^5 u_0 = \begin{bmatrix} +0{\cdot}002{,}56u_5 \\ +0{\cdot}000{,}1u_{40} \end{bmatrix} \div 7 - \begin{bmatrix} 0{\cdot}000{,}1u_0 \\ +0{\cdot}000{,}4u_{10} \\ +0{\cdot}000{,}08u_{30} \end{bmatrix} + 0{\cdot}000{,}2u_{20} - 15\delta^6 u_0$$

$$\delta^4 u_0 = \begin{bmatrix} +0{\cdot}008u_0 \\ +0{\cdot}002{,}4u_{10} \\ +0{\cdot}000{,}16u_{30} \end{bmatrix} - \begin{bmatrix} 0{\cdot}002{,}56u_5 \\ +0{\cdot}000{,}8u_{20} \end{bmatrix} - 11\delta^5 u_0 - 64{\cdot}5\delta^6 u_0$$

$$\delta^3 u_0 = \begin{bmatrix} -0{\cdot}006u_0 \\ -0{\cdot}012u_{10} \end{bmatrix} + \begin{bmatrix} 0{\cdot}016u_5 \\ +0{\cdot}002u_{20} \end{bmatrix} - 7{\cdot}25\delta^4 u_0 - 28\delta^5 u_0 - 75\delta^6 u_0$$

$$\delta^2 u_0 = \begin{bmatrix} +0{\cdot}04u_0 \\ +0{\cdot}04u_{10} \end{bmatrix} - 0{\cdot}08u_5 - 4\delta^3 u_0 - 8\delta^4 u_0 - 10\delta^5 u_0 - 8{\cdot}4\delta^6 u_0$$

$$\delta u_0 = -0{\cdot}2u_0 + 0{\cdot}2u_5 - 2\delta^2 u_0 - 2\delta^3 u_0 - \delta^4 u_0 - 0{\cdot}2\delta^5 u_0.$$

In order to verify the correctness of the values obtained the follow-

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

ing checking equation may be used which has been worked out from the general equation previously given.

$$u_{50} = u_0 + 50\,\delta u_0 + 1{,}225\,\delta^2 u_0 + 19{,}600\,\delta^3 u_0 + 230{,}300\,\delta^4 u_0$$
$$+ 2{,}118{,}760\,\delta^5 u_0 + 15{,}890{,}700\,\delta^6 u_0.$$

In applying the above formulae the labour may be diminished by subtracting u_0 from all the terms, thus reducing u_0 itself to zero, and then working from the *new* terms. Thus:

Original terms			New terms after subtracting u_0
$u_0 = p'_5 = \bar{1}$·9936488 :	...	...	0 :
$u_5 = p'_{10} = \bar{1}$·9983627 :	...	...	+47139 :
$u_{10} = p'_{15} = \bar{1}$·9980733 :	...	...	+44245 :
$u_{20} = p'_{25} = \bar{1}$·9965108 :	...	...	+28620 :
$u_{30} = p'_{35} = \bar{1}$·9933917 :	...	...	−2751 :
$u_{40} = p'_{45} = \bar{1}$·9893048 :	...	...	−43440 :
$u_{50} = p'_{55} = \bar{1}$·9823833 :	...	...	−112655 :

It will require extreme care with regard to the signs, a negative coefficient multiplied by a negative quantity giving a + result, &c.

The working out of the formulae gives the following results:

$\delta^6 u_0 =$	− :6282552
$\delta^5 u_0 =$	+14:2743429
$\delta^4 u_0 =$	−154:2905086
$\delta^3 u_0 =$	+1046:5677300
$\delta^2 u_0 =$	−5090:7329360
$\delta u_0 =$	+17667:5660519

The sign : is conveniently used to denote the end of seven places of decimals.

If reference be made to the simple illustrative instance previously given no difficulty should be experienced in regard to the mode of procedure.

The differences are successively carried down by algebraical addition starting with $\bar{1}$·9936488 as u_0, thus:

		$\bar{1}$·9936488	$=u_0=p'_5$
		+ 17667:566052	
	+17667:566052	$\bar{1}$·9954155:566052	$=u_1=p'_6$
	− 5090:732935	+ 12576:833116	
−5090:732936	+12576:833116	$\bar{1}$·9966732:399168	$=u_2=p'_7$
+1046:567730	− 3944:165206	+ 8632:667910	
−3944:165206	+ 8632:667910	$\bar{1}$·9975365:067078	$=u_3=p'_8$

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

It is obvious of course that when each of the given values of p'_{10}, p'_{15} &c. is reached it will at once be seen whether any errors have been made, as the hypothetical curve being drawn should pass through the fixed points.

The formulae for the series (2), (3), and (4) are much simpler than those for series (1), and are the same for each series as the intervals between the given terms are equal.

In proceeding to deal with series (2), the first step is to set down and difference the given values of p'_{15} to p'_{65}, thus:

		Δ	Δ^2	Δ^3	Δ^4	Δ^5
p'_{15}	$\bar{1}$·9980733:	−15625:	−15566:	+5888:	−24556:	−19225:
p'_{25}	$\bar{1}$·9965108:	−31191:	−9678:	−18668:	−43781:	
p'_{35}	$\bar{1}$·9933917:	−40869:	−28346:	−62449:		
p'_{45}	$\bar{1}$·9893048:	−69215:	−90795:			
p'_{55}	$\bar{1}$·9823833:	−160010:				
p'_{65}	$\bar{1}$·9663823:					

The problem is to subdivide these differences represented by the symbol Δ^n into smaller differences δ^n corresponding to the tenth part of the interval.

The key to the solution of the problem is the equation

$$\Delta = \{(1+\delta)^{10} - 1\}.$$

The working out of this equation to five orders of differences produces the following simple formulae:

$$\begin{aligned}
\delta^5 &= {\cdot}00001\,\Delta^5 \\
\delta^4 &= {\cdot}0001\,\Delta^4 - 18\delta^5 \\
\delta^3 &= {\cdot}001\,\Delta^3 - 13{\cdot}5\delta^4 - 96{\cdot}75\delta^5 \\
\delta^2 &= {\cdot}01\,\Delta^2 - 9\delta^3 - 44{\cdot}25\delta^4 - 150\delta^5 \\
\delta &= {\cdot}1\,\Delta - 4{\cdot}5\delta^2 - 12\delta^3 - 21\delta^4 - 25{\cdot}2\delta^5.
\end{aligned}$$

It is essential to remember that these formulae must be applied to *a complete line of* Δ^n *values.*

In the instance now being dealt with the interpolation is required to begin at p'_{25} and it is only necessary to fill in the constant Δ^5 value −19225: at the blank space opposite to p'_{25}.

If as in series (3) and (4) the interpolation had had to begin at the *third* line the first blank space in the line would have had to be filled by the sum of −43781: and −19225:, and the last blank space as before by −19225:.

The following are the needful checking equations so as to carry the checking process *to the end of the series,* which is necessary to avoid error.

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

(*a*) For series (2)

u_{40} (*i.e.* p'_{65}) $= u_0$ (*i.e.* p'_{25}) $+ 40\delta + 780\delta^2 + 9{,}880\delta^3 + 91{,}390\delta^4 + 658{,}008\delta^5$.

(*b*) For series (3) and (4)

u_{30} (*i.e.* p'_{75}) $= u_0$ (*i.e.* p'_{45}) $+ 30\delta + 435\delta^2 + 4{,}060\delta^3 + 27{,}405\delta^4 + 142{,}506\delta^5$.

(In series (4) u_{30} is p'_{85} and u_0 is p'_{55}.)

When the values of δ^5 to δ are obtained the interpolation is proceeded with precisely as shown before.

On "Welding" or combining two series.

The next matter demanding explanation is the process to which allusion has been made, of combining overlapping portions of two adjacent series so that one curve shall pass into the other without any abrupt transition.

This method is one of the many improvements in Life-Table construction devised by Mr. A. C. Waters.

It is an adaptation of what is known as the "curve of cosines."

The figures in illustration given below show the process as applied to welding series (1) with series (2) at the interval between p'_{25} and p'_{35}, *i.e.* p'_{26} to p'_{34}.

It will be noted that the sum of each pair of multipliers = 1.

From series (1)		From series (2)		Combined values
p'_{26} = ·9963336:32	× 0·976	+ ·9962487:15	× 0·024	= $\bar{1}$·9963316
p'_{27} = ·9961318:92	× 0·904	+ ·9959728:09	× 0·096	= $\bar{1}$·9961166
p'_{28} = ·9959012:47	× 0·794	+ ·9956843:13	× 0·206	= $\bar{1}$·9958565
p'_{29} = ·9956385:47	× 0·654	+ ·9953843:67	× 0·346	= $\bar{1}$·9955506
p'_{30} = ·9953419:96	× 0·500	+ ·9950740:01	× 0·500	= $\bar{1}$·9952080
p'_{31} = ·9950112:59	× 0·346	+ ·9947541:14	× 0·654	= $\bar{1}$·9948431
p'_{32} = ·9946475:09	× 0·206	+ ·9944254:55	× 0·794	= $\bar{1}$·9944712
p'_{33} = ·9942534:08	× 0·096	+ ·9940886:06	× 0·904	= $\bar{1}$·9941044
p'_{34} = ·9938330:27	× 0·024	+ ·9937439:59	× 0·976	= $\bar{1}$·9937461

This process is facilitated by deducting the largest possible *common value* from each pair of p'_x values before multiplying, and then adding the two products to the common value. Thus

$$p'_{26} = \cdot 9960000 + (3336{:}32 \times 0\cdot 976) + (2487{:}15 \times 0\cdot 024) = \bar{1}\cdot 9963316.$$

(*See* Diagram *A* for a graphic illustration of the above figures, the height of the ordinates representing the numerical values of the logarithms.)

It is hoped that sufficient explanation and illustration have now

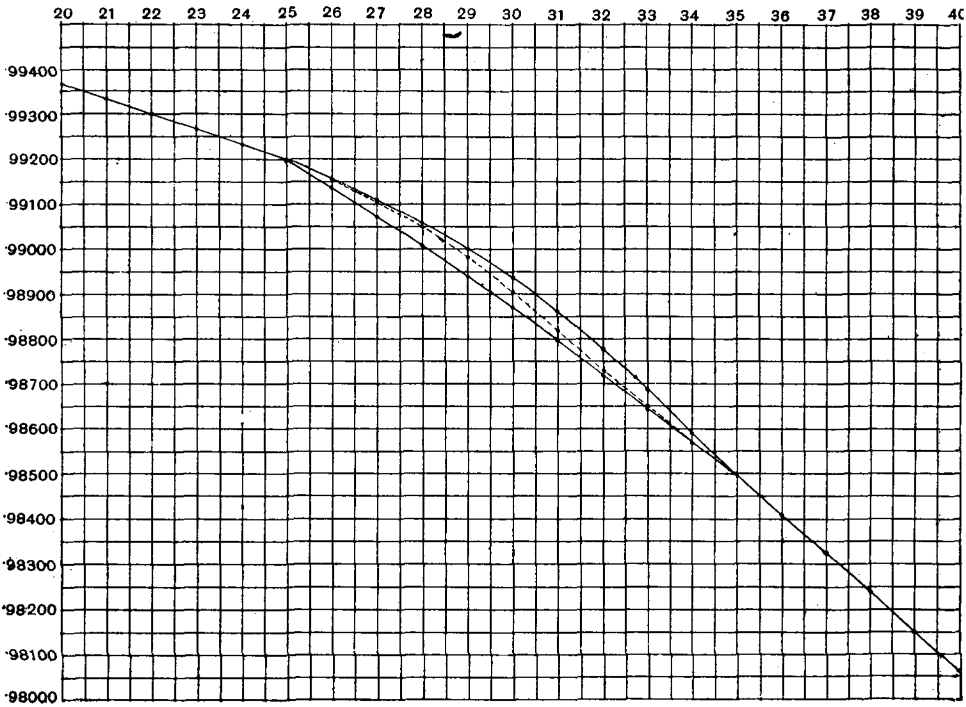
https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

been given to render it possible to calculate the required series of p'_x values from p'_5 right on to the end.

Precisely how far the last series will have to be carried can only be determined when the l_x column comes to be dealt with.

DIAGRAM A.

In this diagram the p'_x curve from p'_{20} to p'_{35}, derived from series 1 (see page 22), and the p'_x curve from p'_{25} to p'_{40}, derived from series 2, are shown separately by full lines, and the curve obtained by "welding" these two series, from p'_{26} to p'_{34}, is shown by an intermediate dotted line.



On the reduction of p'_x to p_x values.

However when these p'_x values are obtained and tabulated it must be remembered that they give the chance of living from age $x - \frac{1}{2}$ to $x + \frac{1}{2}$, and that before they can be made use of in the next stage of Life-Table construction, viz. the calculation of the number of survivors at each age out of a given number at birth, *i.e.* the l_x column, they

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

must be reduced to such values as shall represent the chance of living from age x to age $x+1$.

The simplest mode of effecting this would be to take the geometrical mean of two consecutive p'_x values, or in other words the arithmetical mean of their logarithms, as the required intermediate value.

However, it is found that a more accurate and more evenly graduated curve is to be obtained by making each of the new p_x values to occupy the central point in the interval between the extremes of a series of four consecutive p'_x values.

The working out of the first value, viz. p_5 requires a special formula, so that the previously calculated p_4 value may be brought in. (It must be understood that p_x and p'_x are written down for the sake of brevity in the following formulae instead of $\log p_x$ and $\log p'_x$, and that the same remark applies to the preceding pages from page 22 onwards to this point, for in all these processes of interpolation and welding the logarithms are dealt with as if they were common numbers).

$$p_5 = \frac{-4p_4 + 15p'_5 + 10p'_6 - p'_7}{20}.$$

For the remaining p_x values only one formula is required which may be adequately represented by that for p_6,

$$p_6 = \frac{10(p'_6 + p'_7) - (p'_5 + p'_6 + p'_7 + p'_8)}{16}.$$

The general rule simply is in order to deduce a p_x value from four p'_x values.

From ten times the sum of the two middle terms subtract the sum of all four terms and divide the remainder by 16, *the result is the central term required.*

These last interpolations can be very easily and quickly effected.

In order to avoid errors an important practical point is to difference the numerical values of the logs of the p_x values as one proceeds, as well as to plot them out to scale on paper ruled into squares.

(See diagram B for graphic illustration of the p_x curve from p_5 to p_{24}. This diagram also shows what extreme variations in this part of the p_x curve occur from different methods of interpolation, and that the method herein described at least gives a curve having some rational relation to the mean values of p_x deduced from the total population and death-numbers for the age-periods 5—10, 10—15, and 15—25.)

Calculation of the l_x column.

After having obtained the complete series of logs of p_x values the remaining work is of a comparatively simple nature.

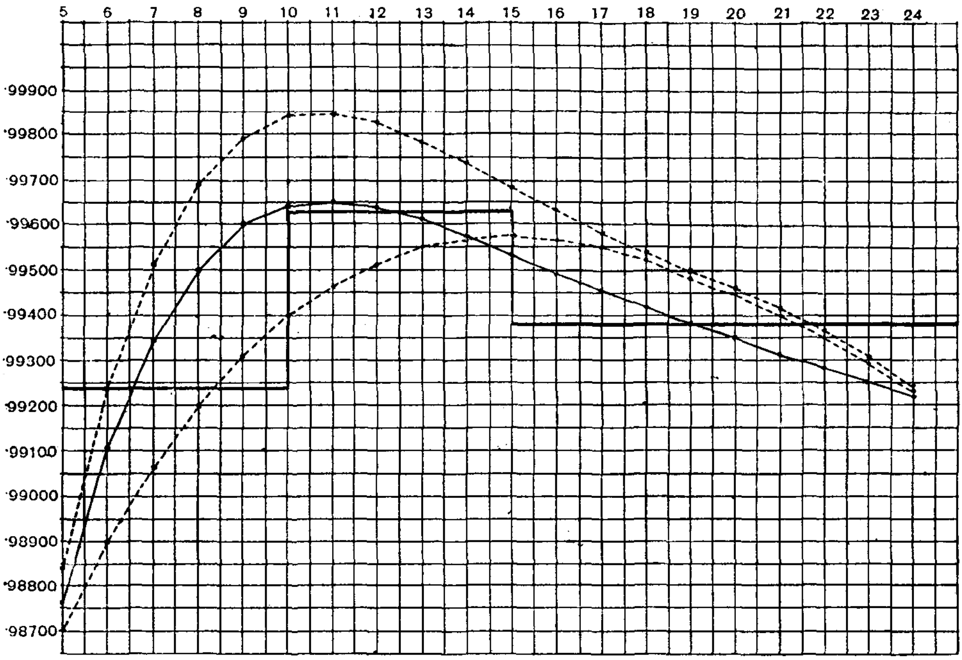
Diagram B.

In this diagram are shown :

(1) by means of the full straight horizontal lines 5 to 10, 10 to 15, and 15 to 25, the mean p_x values derived from the years of life and the total deaths in ten years for each of the age-periods 5—10, 10—15, and 15—25, by the fraction $\frac{2P-d}{2P+d}$;

(2) by means of the full curved line, the p_x curve from p_5 to p_{24}, as derived by the final interpolation in p'_x values from series 1 (see page 27);

(3) by means of dotted curved lines, the p_x curves derived from the same data by two other methods of interpolation which have been employed in Life-Table construction, dealing with the data for the age-period 5—15 in *one* group instead of in the *two* groups 5—10 and 10—15.



The next step is to take some arbitrary number of individuals, assumed to set out together on the journey of life, as the l_0 number, *i.e.* the number "at birth," in actuarial terminology the "radix" of the Life-Table.

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

It does not matter what number is so taken, however, for the sake of being able to afterwards easily combine the results of the male and female sections of the Life-Table into one relating to *persons*, it is usual to divide up a hundred thousand or a million in proportion to the numbers of male and female births which have been registered during the decennium being dealt with.

The number l_0 which has been taken is next translated into the corresponding logarithm and then the logs of the p_x values are in succession added. Thus:

$$\log l_0 + \log p_{0 \text{ to } \frac{1}{2}} = \log l_{\frac{1}{2}} = \text{log survivors at age 6 months.}$$
$$\log l_{\frac{1}{2}} + \log p_{\frac{1}{2} \text{ to } 1} = \log l_1 = \text{log survivors at age 1 year.}$$
$$\log l_1 + \log p_1 = \log l_2 = \text{log survivors at age 2 years.}$$
&c. &c.

The process is continued until a negative "characteristic" is obtained, *i.e.* until the number of survivors falls below unity.

The logs of the l_x values are then to be translated into their corresponding numerical values.

In order to have a uniform standard of comparison for males and females as to the number of survivors at each age x it is usual to construct two other l_x columns, each starting with a hundred thousand or a million at birth.

This can be done

(*a*) in precisely the same way as above described, or

(*b*) assuming that we are taking a million as the l_0 value, by adding to each log l_x value already obtained (log 1,000,000 − log l_0 value previously taken). This will give the log of the new l_x value.

The d_x column, i.e. the successive numbers of those dying from age x to age $x+1$.

This is simply obtained by the formula $d_x = l_x - l_{x+1}$. It is obvious that $\Sigma d_x = l_x$, that is the sum of those dying from age x to the end of the Life-Table must equal those living at age x.

The P_x column.

This column, which represents the mean number living from age x to age $x+1$, for every year of life *except the first*, is taken as the *arithmetical mean* of l_x and l_{x+1}, *i.e.* $P_x = \frac{1}{2}(l_x + l_{x+1})$. (The arithmetical

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

mean is not only simpler but it can be shown to be *more accurate* than the geometrical mean, which has been used in some Life-Tables.)

The same number which expresses the *mean population* will also represent the *years of life lived* during the year x to $x+1$ by the number of survivors l_x entering upon that year of age; for, the years of life lived $= l_{x+1} + \frac{1}{2} d_x$,

$$= l_{x+1} + \tfrac{1}{2}(l_x - l_{x+1}) = \tfrac{1}{2}(l_x + l_{x+1}) = P_x.$$

The value of P_0, that is the mean population for, or the years of life lived in, the *first* year of age is best calculated by the following simple formula which by an application of the Integral Calculus has been deduced by Mr A. C. Waters from the data, deaths at 0—3 mos., deaths at 3—6 mos., and deaths at 6—12 mos.

$$P_0 = l_1 + \left(\frac{d_{0 \text{ to } \frac{1}{2}} + 5d_{\frac{1}{2} \text{ to } 1}}{6}\right),$$

i.e. $= \frac{1}{6}$ deaths 0 to 6 mos. $+ \frac{5}{6}$ deaths 6 to 12 mos. + survivors at age 1. As $l_{\frac{1}{2}}$ and l_1 will have been obtained, the values of $d_{0 \text{ to } \frac{1}{2}}$ and $d_{\frac{1}{2} \text{ to } 1}$ are to be deduced thus:

$$d_{0 \text{ to } \frac{1}{2}} = l_0 - l_{\frac{1}{2}}, \text{ and } d_{\frac{1}{2} \text{ to } 1} = l_{\frac{1}{2}} - l_1.$$

The Q_x column.

This is now to be constructed from the P_x column by successive additions beginning from below.

Reference to the previously explained double significance of the P_x numbers will make it apparent that each Q_x value will represent

(*a*) the years of life lived by l persons of exact age x during the year of age x to $x+1$ and *during all the years of age afterwards to the end of the Life-Table*, and

(*b*) the complete Life-Table population at age x and upwards.

The E_x column.

From the definition (*a*) of the Q_x column given above it is obvious that the expectation of life at age x, or in other words the mean after-lifetime of l individuals of exact age x will be equal to $\frac{Q_x}{l_x}$,

or $$\log E_x = \log Q_x - \log l_x.$$

Distribution of the total expectation of life at birth over the different periods of life.

It is of some importance in comparing different Life-Tables with each other to show how the total average expectation of life at birth (or E_0) is distributed over the different age-periods of life.

The periods of life from a practical point of view may be taken thus:

0—5, infancy
5—15, school age
15—65, working period of life
65— decline.

By dividing $Q_0 - Q_5$, $Q_5 - Q_{15}$, $Q_{15} - Q_{65}$, and Q_{65} respectively by l_0 the required values will be obtained, and, as will be obvious, the sum of the parts will equal the whole.

How to deduce from the Q_x column the average expectation of life of the individuals at all ages from x to $x + n$ comprised within the age-groups x to $x + n$.

By the methods of calculation which have been described it has been shown how from the data of the numbers living and dying at all intermediate ages within certain age-groups it is possible to calculate the mean after-lifetime of individuals at the *exact* age x.

However, for some of the most striking and important applications of a Life-Table it is necessary to re-translate these E_x values into those which represent the mean expectation of life of all the individuals comprised within the age-groups x to $x + n$, or, in other words, to make them applicable to a census population.

How this can be effected will be apparent from the following considerations.

(1) The future lifetime of P_x persons, *i.e.* of P persons living at all ages from x to $x + 1$ must be equal to $Q_x - \frac{1}{2}P_x$; for, on the assumption which is made in Life-Table construction, the average age of the P persons is $x + \frac{1}{2}$ at the middle of the year, therefore at the middle of the year they will on the average have each expended half of that year of life, and $\frac{1}{2}P_x$ must be deducted from the Q_x value.

(2) It is therefore obvious that their mean expectation of life individually must be equal to

$$\frac{Q_x - \frac{1}{2}P_x}{P_x} = \frac{Q_x}{P_x} - \frac{1}{2}.$$

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

(3) Similarly the future lifetime of $P_x + P_{x+1} + \ldots + P_{x+n-1}$ persons, living at all ages between x and $x + n$, is represented by

$$(Q_x + Q_{x+1} + \ldots + Q_{x+n-1}) - \tfrac{1}{2}(P_x + P_{x+1} + \ldots + P_{x+n-1})$$

and their mean expectation of life =

$$\frac{(Q_x + Q_{x+1} + \ldots + Q_{x+n-1})}{(P_x + P_{x+1} + \ldots + P_{x+n-1})} - \tfrac{1}{2} = \frac{(Q_x + Q_{x+1} + \ldots + Q_{x+n-1})}{(Q_x - Q_{x+n})} - \tfrac{1}{2}.$$

In actual practice of course the value of n is either 5 or 10, and from the Q_x column the average expectation of life can be readily calculated for the age-groups 0—5, 5—10, 10—15, 15—25 &c. &c.

For the purpose of this calculation it is sufficiently accurate to take as the mean expectation of life for the age-group 0—5

$$\frac{Q_0 + Q_1 + Q_2 + Q_3 + Q_4}{Q_0 - Q_5} - \tfrac{1}{2}.$$

II. ON THE USES OF A LIFE-TABLE FROM A PUBLIC HEALTH POINT OF VIEW.

Assuming that the laborious task which it has been the object of the preceding pages to explain has been completed, it remains to indicate in very brief outline the practical uses of the Life-Table.

That it is a most valuable statistical instrument for a community to possess will be apparent from the following considerations. By means of it the possibility exists of making exact comparisons

(1) With the Life-Table for the whole country for the same decennial period.

(2) With all other Local Life-Tables which may have been worked out for the same period.

(3) And, what is of very special importance, with any Life-Tables *for the same district* which may have been already calculated for previous decennial periods, thus giving the most exact measure possible for marking the effects of advance or retrogression in the conditions affecting health and life.

The special lines along which such comparisons may be with advantage made are the following.

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

(1) *The p_x values.*

These afford the means of testing the vitality of a community at each age or age-period.

They depend neither upon antecedent nor upon consequent conditions, but simply upon the "force of mortality" which has prevailed at each special age or age-period.

(2) *The l_x values.*

These depend upon the rates of mortality at *preceding* age-periods. Thus high death-rates during the earlier years of life diminish the number of survivors at later ages.

(3) *The E_x values.*

These are affected by the death-rates at all *following* age-periods.

Therefore p_x, l_x, and E_x are measures respectively of *present, past,* and *future.* (The death-rates of the decennial period on which the Life-Table is based are of course *simultaneous,* but they are assumed to exist in succession.) As has been already pointed out the distribution of the total expectation of Life at Birth over the successive age-periods of life may be readily arrived at and is an important point for comparison.

The expectation of Life at Birth having been obtained for males and females and the values being supposed to hold good for succeeding years until the next Life-Table is calculated, a balance of gain and loss can readily be struck for each year. Each Birth represents so many years of prospective lifetime, and the total prospective gain is readily calculated. Since the same number which expresses the estimated mean population for a year expresses also the years of life lived or expended in that year we have thus the loss to set against the gain.

Seeing that the mean expectation of Life has been obtained for the individuals of all ages within the usual age-groups, it is a simple matter, having given the estimated population living at the middle of a year, classified into the same age-groups[1], to calculate the total "Life-capital"

[1] It is of course necessary to assume that the estimated population for each year is composed of age and sex groups in the same proportions to the total as those ascertained at the preceding census.

of the community, and the division of this total by the whole population-number obviously gives the average Life-capital, or future life-time, of each individual of the population.

Finally if a calculation be made of the number of deaths which should have occurred in each age-group if the mean death-rates for the 10-yearly period of the Life-Table had continued unchanged, and if then a comparison be made between these numbers and the numbers of deaths which *actually have occurred* in the year being dealt with, it is a simple matter to strike the balance of gain or loss of Life-capital. It is obvious that lives lost or gained in the earlier age-groups have greater weight in the balance sheet than those at later ages.

III. On certain modifications of Dr Farr's "short" method of life-table construction by means of which, as regards Expectation of Life at quinquennial age-intervals, results can be obtained practically identical with those arrived at by the previously described "extended" method.

Going back to the point at which it had been described how to obtain the p_x values for the first five years of age, and from which proceeded the laborious path by which the remaining p_x values are to be one by one arrived at, it will be noted that attention has been drawn to the ease with which a mean value of $p_{x \text{ to } x+n}$ can be worked out from the "years of life" and the total deaths in the ten years for each age-period by the fraction $\dfrac{2P-d}{2P+d}$.

The first step to be taken in the construction of a short Life-Table is to work out such a series of mean p_x values for the usual age-groups, viz. 5—10, 10—15, 15—25 and so on, ending with 85—.

To obtain p_{95-} it is best to put down the logs of p_{55-}, p_{65-}, p_{75-} and p_{85-} in a column, and then by differencing them and carrying down the differences for one stage, log p_{95-} is arrived at.

For the first few age-groups there is but little difference between the p_x values so found and the means of the separate yearly values obtained by the extended method, but afterwards the former become more and more in excess of the latter. See Table below.

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

Mean p_x values.

Age-periods	By extended method (*a*)	By short method (*b*)	Differences of (*b*) from (*a*)
5—10	·99247	·99241	− ·00006
10—15	·99630	·99630	± ·00000
15—25	·99380	·99384	+ ·00004
25—35	·98890	·98903	+ ·00013
35—45	·98037	·98064	+ ·00027
45—55	·96872	·96933	+ ·00051
55—65	·94514	·94704	+ ·00190
65—75	·89665	·90236	+ ·00571
75—85	·81742	·83299	+ ·01557
85—95	·70007	·72985	+ ·02978
95—	·58632	·59463	+ ·00831

Calculation of l_x values.

In using the above given mean p_x values it is simply necessary to take 5 times the mean value for a stage of 5 years, or 10 times the mean value for a stage of 10 years, etc.

Thus, commencing with

$$l_5 = 34{,}467$$
$$\log l_5 + (\log p_{5-10} \times 5) = \log l_{10}$$
$$\log l_{10} + (\log p_{10-15} \times 5) = \log l_{15}$$
$$\log l_{15} + (\log p_{15-25} \times 10) = \log l_{25}$$
&c. &c.

It is evident that by this method the l_x numbers will tend to differ more and more in the direction of excess as compared with those obtained by the extended method.

See Table given below.

Comparison of l_x values obtained by extended and short methods.

	By extended method (*a*)	By short method (*b*)	Differences of (*b*) from (*a*)
$l_{10}=$	33,190	33,178	−12
$l_{15}=$	32,580	32,569	−11
$l_{25}=$	30,615	30,617	+2
$l_{35}=$	27,381	27,420	+39
$l_{45}=$	22,458	22,551	+93
$l_{55}=$	16,344	16,516	+172
$l_{65}=$	9,296	9,585	+289
$l_{75}=$	3,123	3,431	+308
$l_{85}=$	416	552	+136
$l_{95}=$	12	24	+12

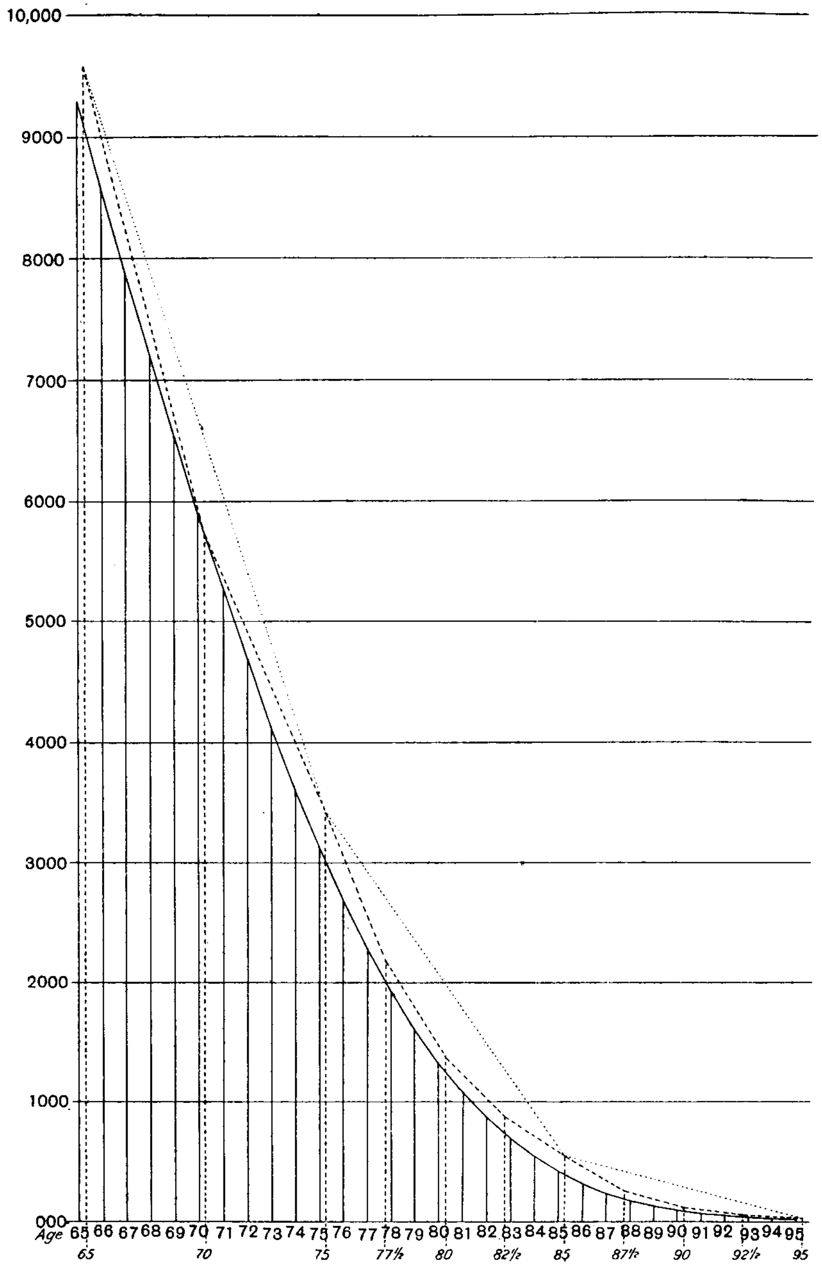
https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

Diagram C.

In diagram C the numbers shown at the base-line represent ages, and the vertical lines, or ordinates, represent the respective l_x values.

On the scale employed it has been impossible to show the construction further than age 95.

For other details of description see page 38.



https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

Calculation of P_x values.

It will be remembered that in the extended method P_x was taken as $\frac{1}{2}(l_x + l_{x+1})$. This in geometrical construction is equivalent to joining the extremities of the ordinates l_x and l_{x+1} by a straight line. If measured with strict mathematical accuracy account would have to be taken of the *curve* passing through the extremities of the ordinates, but as the interval of one year is proportionally small the series of straight lines approximate very closely to the true curve. (See Diagram C.)

P_x was also found to be equal to $l_{x+1} + \frac{1}{2} d_x$ which is equivalent to considering that those dying between age x and age $x+1$ on the average live half through the interval.

Now if we come to apply these same assumptions to an interval of ten years,

$$P_{x \text{ to } x+10} = \tfrac{1}{2}(l_x + l_{x+10}) = l_{x+10} + \tfrac{1}{2} d_{x \text{ to } x+10},$$

and it is evident by reference to the diagram "C" that this geometrical construction must diverge more and more from the truth.

It will thus be evident why the E_x values obtained by Dr Farr's original short method diverged more and more from the values of an extended method in the direction of excess.

The general principle of the modification of Dr Farr's method which has been proposed by the writer is simply to take the 10-yearly intervals at subdivided stages. Thus each 10-yearly period from 15—25 to 65—75 inclusive is to be taken in *two* stages, the periods 75—85, and 85—95 are to be each taken in *four* stages, and from age 95 onwards *yearly* stages are to be used.

The series of l_x values required, therefore, will be as follows, having commenced the calculation with l_5:

	l_{10}	l_{75}	$l_{87\frac{1}{2}}$	l_{96}
	l_{15}	$l_{77\frac{1}{2}}$	l_{90}	l_{97}
	l_{20}	l_{80}	$l_{92\frac{1}{2}}$	l_{98}
	l_{25}	$l_{82\frac{1}{2}}$	l_{95}	l_{99}
and so on to	l_{70}	l_{85}		and so on.

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

The process of calculation is indicated as follows:

$$\begin{aligned}
\log l_5 &+ (\log p_{5-10} \times 5) = \log l_{10} \\
\log l_{10} &+ (\log p_{10-15} \times 5) = \log l_{15} \\
\log l_{15} &+ (\log p_{15-25} \times 5) = \log l_{20} \\
\log l_{20} &+ (\log p_{15-25} \times 5) = \log l_{25} \\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\log l_{75} &+ (\log p_{75-85} \times 2\tfrac{1}{2}) = \log l_{77\frac{1}{2}} \\
\log l_{77\frac{1}{2}} &+ (\log p_{75-85} \times 2\tfrac{1}{2}) = \log l_{80}
\end{aligned}$$

and so on.

The P_x values may be then easily arrived at, but it must be borne in mind that when the intervals are more than one year the *mean number living from age x to age $x+n$* will not express the *years of life lived from age x to $x+n$,* which must be obtained by multiplying P_x by n, and therefore before we can calculate the Q_x column there must be an intermediate column which may be called Y_x. Thus:

$$\begin{aligned}
P_{5-10} &= \tfrac{1}{2}(l_5 + l_{10}) & Y_{5-10} &= P_{5-10} \times 5 \\
P_{10-15} &= \tfrac{1}{2}(l_{10} + l_{15}) & Y_{10-15} &= P_{10-15} \times 5 \\
P_{15-20} &= \tfrac{1}{2}(l_{25} + l_{20}) & Y_{15-20} &= P_{15-20} \times 5
\end{aligned}$$

and so on to P_{70-75}.

$$\begin{aligned}
P_{75-77\frac{1}{2}} &= \tfrac{1}{2}(l_{75} + l_{77\frac{1}{2}}) & Y_{75-77\frac{1}{2}} &= P_{75-77\frac{1}{2}} \times 2\tfrac{1}{2} \\
P_{77\frac{1}{2}-80} &= \tfrac{1}{2}(l_{77\frac{1}{2}} + l_{80}) & Y_{77\frac{1}{2}-80} &= P_{77\frac{1}{2}-80} \times 2\tfrac{1}{2}
\end{aligned}$$

and so on to P_{90-95}.

$$\begin{aligned}
P_{95} &= \tfrac{1}{2}(l_{95} + l_{96}) = Y_{95} \\
P_{96} &= \tfrac{1}{2}(l_{96} + l_{97}) = Y_{96}
\end{aligned}$$

and so on.

Let the P_x and Y_x values be tabulated opposite to the corresponding l_x values. The Q_x column may now be constructed from the Y_x column by successive additions beginning from below, and then the E_x values may be readily calculated as before by the formula $E_x = \frac{Q_x}{l_x}$. These values, however, are only to be worked out for ages 5, 10, 15, 25, 35, 45, 55, 65, 75, 85 and 95.

In the diagram "C" the geometrical construction of the l_x, P_x, Y_x and Q_x columns is shown from age 65 onwards.

(*a*) By the extended method in full black lines.

(*b*) By the modified short method in interrupted black lines. (It must be supposed that this diagram has been superimposed upon the preceding and then moved a little to the right so as to show the construction.)

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

(*c*) The dotted black lines joining the ordinates at 10-yearly intervals indicate the construction of Dr Farr's original method.

Interpolation of intermediate quinquennial E_x values.

The method above described is only intended and adapted for obtaining E_x values at *decennial* intervals from E_{15} onwards.

However, by using very simple formulae like those which have already been described with relation to the interpolation of p_x values in a series of p'_x values, it is readily possible to obtain the values of E_{20}, E_{30}, &c., &c.

From E_{30} to E_{80} the following formula is applicable (with obvious successive changes in the suffixes):

$$E_{30} = \frac{10(E_{25}+E_{35}) - (E_{15}+E_{25}+E_{35}+E_{45})}{16},$$

E_{20} and E_{90} require special formulae, thus:

$$E_{20} = \frac{E_{15}+E_{35}}{4} + 1\tfrac{1}{2}E_{25} - E_{30},$$

$$E_{90} = \frac{E_{75}+E_{95}}{4} + 1\tfrac{1}{2}E_{85} - E_{80}.$$

Comparison of results obtained by the above described short method with those to be arrived at by an extended method.

In order to show the value of the modified short method its results as applied to the data for England and Wales 1881—90 are given below, and these are contrasted with those of an extended method. The values of E_0, E_5, E_{10} and E_{15} shown for the extended method have been recalculated by a method to some extent similar to what has been already described. The other values from E_{20} onwards have been taken from the official Life-Table.

Comparative Table. Section A.

Comparison of E_x values, *i.e.* mean expectation of life, or mean after-lifetime at exact age x obtained by (*a*) extended method, and (*b*) a modified short method.

N.B. The comparison is based on a new set of p_x values from p_5 to p_{24} which have been worked out for (*a*).

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

England and Wales, 1881—90.

Age x	Males (a)	Males (b)	Differences of (b) from (a)	Females (a)	Females (b)	Differences of (b) from (a)
0	43·28	43·32	+0·04	46·66	46·67	+0·01
5	52·24	52·30	+0·06	54·26	54·27	+0·01
10	48·59	48·65	+0·06	50·64	50·65	+0·01
15	44·28	44·33	+0·05	46·40	46·40	±0·00
20	40·27	40·28	+0·01	42·42	42·40	−0·02
25	36·28	36·34	+0·06	38·50	38·51	+0·01
30	32·52	32·53	+0·01	34·76	34·74	−0·02
35	28·91	28·87	−0·04	31·16	31·08	−0·08
40	25·42	25·38	−0·04	27·60	27·51	−0·09
45	22·06	22·04	−0·02	24·05	24·01	−0·04
50	18·82	18·79	−0·03	20·56	20·50	−0·06
55	15·74	15·71	−0·03	17·23	17·12	−0·11
60	12·88	12·84	−0·04	14·10	14·00	−0·10
65	10·31	10·24	−0·07	11·26	11·17	−0·09
70	8·04	7·98	−0·06	8·77	8·71	−0·06
75	6·10	6·06	−0·04	6·68	6·62	−0·06
80	4·52	4·53	+0·01	5·00	4·98	−0·02
85	3·29	3·32	+0·03	3·71	3·69	−0·02
90	2·37	2·40	+0·03	2·75	2·75	±0·00
95	1·72	1·72	±0·00	2·05	1·97	−0·08

Comparative Table. Section B.

The next table shows a similar comparison with regard to the distribution of the total expectation of life *at Birth* over the age-periods indicated.

Age-period	Males (a)	Males (b)	Differences of (b) from (a)	Females (a)	Females (b)	Differences of (b) from (a)
0—5	4·02	4·02	±0·00	4·16	4·16	±0·00
5—15	7·33	7·34	+0·01	7·63	7·65	+0·02
15—25	7·05	7·03	−0·02	7·35	7·33	−0·02
25—35	6·62	6·61	−0·01	6·90	6·89	−0·01
35—45	5·97	5·97	±0·00	6·30	6·30	±0·00
45—55	5·10	5·10	±0·00	5·55	5·54	−0·01
55—65	3·91	3·91	±0·00	4·49	4·48	−0·01
65 and upwards	3·28	3·34	+0·06	4·28	4·32	+0·04
Totals	43·28	43·32	+0·04	46·66	46·67	+0·01

It is thus evident that as regards the two applications of a Life-Table indicated in the preceding two tables the short method gives results so close to those of the extended method as to be practically identical with them in most cases.

Until recently the writer had thought that it was not possible without the aid of an extended Life-Table to arrive at those striking results which are connected with the term "Life-capital." However, it has been found that by certain simple methods to be afterwards described the results are to be obtained set down in the following Table.

Comparative Table. Section C.

Mean expectation of Life of the individuals at all ages from x to $x+n$ comprised within the age-groups indicated.

	Males			Females		
Age-group x to $x+n$	(*a*)	(*b*)	Differences of (*b*) from (*a*)	(*a*)	(*b*)	Differences of (*b*) from (*a*)
0—5	51·76	51·76	±0·00	54·07	54·07	±0·00
5—10	50·57	50·58	+0·01	52·60	52·56	−0·04
10—15	46·43	46·49	+0·06	48·52	48·53	+0·01
15—25	40·27	40·33	+0·06	42·44	42·46	+0·02
25—35	32·60	32·61	+0·01	34·83	34·82	−0·01
35—45	25·51	25·48	−0·03	27·66	27·65	−0·01
45—55	18·95	18·93	−0·02	20·67	20·63	−0·04
55—65	13·09	13·06	−0·03	14·30	14·24	−0·06
65—75	8·35	8·34	−0·01	9·08	9·06	−0·02
75—85	4·94	4·94	±0·00	5·41	5·41	±0·00
85—95	2·80	2·78	−0·02	3·16	3·11	−0·05
95—	1·57	1·62	+0·05	1·86	1·87	+0·01

Methods of arriving at above results, i.e. those of section C, column (b).

(1) It is presumed that a short Life-Table has been constructed by the method previously described, and that therefore the values of E_x have been obtained which are set down in "A" the first of the three sections of the comparative table of which "C" is the last.

(2) In the construction of this short Life-Table P_x and Q_x values have been obtained and set down in columns opposite to the l_x values as already described.

(3) The value of E_{0-5} is of course the same for (*b*) as for (*a*), as it is calculated alike for both methods (see description already given).

For E_{5-10} take the arithmetical mean of E_5 and E_{10} in Section A, column (*b*), and *add* 0·1.

For E_{10-15} and E_{15-25} take the arithmetical means respectively of E_{10} and E_{15} and E_{15} and E_{25} in Section A, column (*b*).

https://doi.org/10.1017/S0022172400001662 Published online by Cambridge University Press

For E_{25-35}, find the value of $\frac{Q_{25}+Q_{30}}{P_{25}+P_{30}} - 2\frac{1}{2}$ and take the arithmetical mean of this and E_{30} in Section A, column (*b*).

Similarly for E_{35-45} to E_{65-75}.

For E_{75-85}, find the value of $\frac{Q_{75}+Q_{77\frac{1}{2}}+Q_{80}+Q_{82\frac{1}{2}}}{P_{75}+P_{77\frac{1}{2}}+P_{80}+P_{82\frac{1}{2}}} - 1\frac{1}{4}$ and take the arithmetical mean of this and of E_{80} in Section A, column (*b*).

Similarly for E_{85-95}.

For E_{95-} take the E_{95} of Section A, column (*b*), and *subtract* 0·1.

Conclusion.

The question may now well arise in the minds of those who are contemplating the construction of merely local Life-Tables as to whether it is worth while to embark on the undertaking of working by the extended method.

The choice must be left to "personal equation." If anyone should try both methods he will at least appreciate how much labour is saved by the modified short method.

But for necessary limitation of space the difficult subject which it has been endeavoured to elucidate might have been dealt with more fully. The writer can only express the hope in conclusion that what has been given may be sufficient to render this paper a practical guide to Life-Table construction.