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 following notes are to serve as a postscript to the paper which appeared under the above title in the preceding number of the *Journal of Hygiene* (pages 1—42).

I. *On a modification of the scheme of interpolation proposed on page* 22.

The following alternative scheme may be suggested as entailing less labour, and while not giving, on minute analysis, a p_x curve so perfect as the other, leading to ultimate l_x and E_x values not materially differing.

Series 1 has five orders of differences, and the remaining series 2, 3, 4, and 5, have each but four orders of differences.

Series	
1	... ... $[p'_5 p'_{10} p'_{15} p'_{25} p'_{35}] p'_{45}$
2	... ... $p'_{15} [p'_{25} p'_{35} p'_{45}] p'_{55}$
3	... ... $p'_{25} [p'_{35} p'_{45} p'_{55}] p'_{65}$
4	... ... $p'_{35} [p'_{45} p'_{55} p'_{65}] p'_{75}$
5	... ... $p'_{45} [p'_{55} p'_{65} p'_{75} p'_{85} \cdots]$

The formulae on pages 22 and 24 are applicable to this scheme by simply eliminating respectively those relating to $\delta^6 u_0$ and δ^5.

The checking equations for u_{40} and u_{30} are given on page 25.

However, since the paper above referred to has appeared in print an idea has occurred to the writer that it may be possible to dispense altogether with any scheme of analytical interpolation after the foundation series of p'_x values have been obtained, and to arrive at the required

https://doi.org/10.1017/S0022172400001923 Published online by Cambridge University Press

series of p_x values for each separate year by a modification of the "graphic" method, thus combining in some degree the ease and simplicity of the latter with the accuracy of the former method. Accordingly what is next to be said may come under the following heading:

II. *On a suggested combination of the "analytical" and "graphic" methods in Life-Table construction.*

If it be assumed that the reader (1) has read and has at hand the paper already alluded to, and (2) is acquainted with the details of the graphic method, as explained in the last edition of Dr Newsholme's "Vital Statistics," the scheme now to be proposed may be made intelligible in very few words.

On reference to the preceding number of the *Journal* at pages 17—21 it will be evident that it is not a very difficult task to work out the foundation series of values of log p'_x given on page 21, especially as some simple rules will be given by which the actual calculation of log p'_5, log p'_{10}, and log p'_{15} may be dispensed with—at any rate they may or may not be calculated as preferred. The calculation of the values from log p'_{25} onwards is extremely simple and easy.

Having then obtained by an easy analytical method the series log p'_{25}...log p'_{85}, by differencing the series log p'_{45}, log p'_{55}, log p'_{65}, log p'_{75} and log p'_{85} and carrying down the differences the values of log p'_{95} and of log p'_{105} are readily fixed.

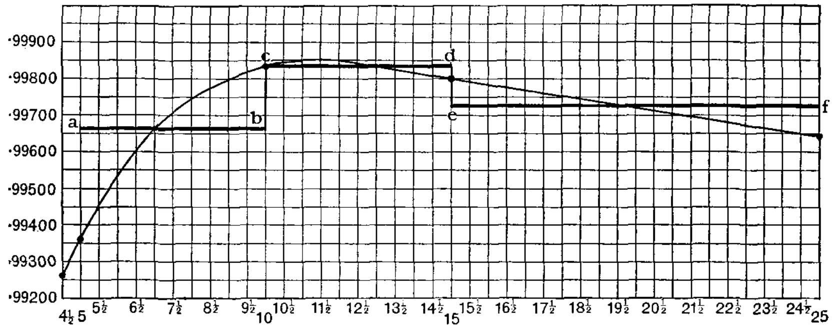
It is also necessary to have the value of log p_4 which will have been obtained at an earlier stage in the construction of the Life-Table, and as p_4 represents the chance of living from age 4 to age 5 it may be considered to be the chance of living a year at *exact age* 4½, therefore log p_4 may be called log $p'_{4\frac{1}{2}}$.

The mode of procedure is to take a sheet or sheets of Layton's actuarial paper ruled into exact $\frac{1}{8}$ inch squares and mark out a base line on the scale of one division for a half-year; commencing with 4½ (see diagram).

In order to construct the vertical scale the logs must be cut down to five figures and then each division may be made to represent ·00050. The inconvenience of having to work on more than one sheet of paper, which would necessarily attend the use of Layton's paper, even on the comparatively small scale mentioned, may be obviated by obtaining some excellent paper made by Ch. Fortin & Cie., "no. 2 ruled into

https://doi.org/10.1017/S0022172400001923 Published online by Cambridge University Press

$\frac{1}{8}$ inch squares." This may be obtained in lengths of several yards and it is wide enough to take in the whole curve from $4\frac{1}{2}$ to 95.



Assuming in the first instance that the values of $\log p'_5$, $\log p'_{10}$ and $\log p'_{15}$ have been calculated, it is now a simple matter to mark to scale the values of the logs in the ordinates corresponding to $4\frac{1}{2}$, 5, 10, 15, 25...105 and then a curved line may be drawn through the fixed points commencing with $4\frac{1}{2}$, the rule being simply to let the line be as little curved as possible consistently with running smoothly without angularities or breaks. It will be found that this curve is a much easier one to draw than those which have to be drawn to divide up population and deaths by the original graphic method.

The curve having been drawn it is simply necessary to measure the ordinates at $5\frac{1}{2}$, $6\frac{1}{2}$, $7\frac{1}{2}$, etc. etc., and these will give the required series of values of $\log p_5$, $\log p_6$, $\log p_7$, etc. etc. ready to be at once used for the next stage in the construction of the Life-Table.

However in order to avoid the necessity of calculating $\log p'_5$, $\log p'_{10}$ and $\log p'_{15}$, let the horizontal lines *ab*, *cd*, and *ef* be drawn as shown in the diagram, at heights above the base line corresponding to the values of the logs of the mean p_x values derived from the years of life and the total deaths in ten years for each of the age-periods respectively, 5—10, 10—15, and 15—25, by the fraction $\frac{2P-d}{2P+d}$. The formula being, (by logs), $\log p_{x \text{ to } x+n} = \log(2P-d) - \log(2P+d)$.

The position of $\log p'_{10}$ is fixed at the exact point *c* in the angle *bcd*, and the position of $\log p'_{15}$ at a point $\frac{3}{10}$ of the distance from *d* to *e* in the line *de*. It is not necessary to fix $\log p'_5$ at all, as its position will be indicated with sufficient accuracy by the point where the curve cuts the ordinate 5.

A curve of the shape shown in the diagram may now be drawn

https://doi.org/10.1017/S0022172400001923 Published online by Cambridge University Press

through the fixed points, starting with the fixed point in the ordinate $4\frac{1}{2}$.

This may seem a very rough and ready method, but it has been deduced from actual calculation and plotting out in a number of instances.

For example in the particular instance which has been used for the diagram,

	by calculation	by rule
$\log p'_{10}$	$= \bar{1}\cdot99836$	$\bar{1}\cdot99839$
$\log p'_{15}$	$= \bar{1}\cdot99807$	$\bar{1}\cdot99807$

As aids to accuracy in drawing this, the most difficult part of the curve, the following simple rules have been found to be applicable:

(1) The curve cuts the line *ab* on the ordinate 7, and the line *cd* on the ordinate 13.

(2) The highest point of the curve is on the ordinate $11\frac{1}{2}$.

(3) These rules apply to vertical and horizontal scales of similar relative proportions to those which have been recommended.

In the *Journal of the Royal Statistical Society*, Vol. LXII., Part iv. at pages 699—701, are to be found some remarks contributed by the writer by way of criticism of the graphic method as hitherto employed in Life-Table construction, the points alluded to being chiefly the extreme irregularity and want of symmetry of the p_x curves when plotted out from Life-Tables so constructed, and the impossible and absurd results obtained at the later ages of life.

By the use of the combined method herein suggested these difficulties would be obviated and results would be arrived at with less labour and probably greater accuracy than those obtained by the graphic division of population and deaths.

Although the present writer would personally prefer to use the previously described analytical method all throughout, for the sake of those who prefer to use a "graphic" method he can advance the following considerations to recommend the above-described scheme:

(1) The given *fixed points* are determined by a rigidly exact method.

(2) With a sufficient degree of technical skill in drawing and in measuring, a fairly close approximation can be obtained to the results which would be obtained by exact calculation—so close as to make but little difference in the ultimate E_x values.

(3) The differences from exact p_x values due to "personal equation" in drawing and in measuring, will certainly be less than the differences to be obtained by different systems of analytical interpolation.

By actual trial it has been found that the above described scheme gives very good results as far as age 65, but after this age there are difficulties in continuing the curve on the same scale. It would appear therefore preferable to construct the curve in two sections:—

(1) Let a curve be drawn as described as far as the ordinate 75. This is only to be used for measuring as far as 64½.

(2) Let another curve be drawn from 55 to 95 with the vertical scale reduced to half, *i.e.* so that each division corresponds to ·00100. This curve is to be used for measuring the ordinates from 65½ to 94½.

(3) By simply differencing the last five values obtained, viz. from $\log p'$ 90½ to $\log p'$ 94½, the series can be continued as far as may be required.

These or any other modifications of the proposed scheme must be left to the technical skill and to the discretion of those who may undertake to use it.

III. *Comparison of the results obtained by the simple rules which have been given for arriving at the values of* $\log p'_{10}$ *and* $\log p'_{15}$ *with the corresponding values worked out by exact calculation.*

Since the preceding pages have been in print the following Table has been worked out with the view of showing, by their application to an increased number of instances, to what extent the "rules" given, for arriving at the values of $\log p'_{10}$ and $\log p'_{15}$, may be relied upon for obtaining results approximating to those which would be got by calculation:

			(*a*) By calculation	(*b*) By rule	Differences of (*b*) from (*a*)
England and Wales 1881–90	Males	$\log p'_{10}$	$\bar{1}$·99870	$\bar{1}$·99872	+ ·00002
		$\log p'_{15}$	$\bar{1}$·99849	$\bar{1}$·99846	− ·00003
	Females	$\log p'_{10}$	$\bar{1}$·99868	$\bar{1}$·99865	− ·00003
		$\log p'_{15}$	$\bar{1}$·99840	$\bar{1}$·99841	+ ·00001
Selected Healthy Districts 1881–90	Males	$\log p'_{10}$	$\bar{1}$·99897	$\bar{1}$·99901	+ ·00004
		$\log p'_{15}$	$\bar{1}$·99886	$\bar{1}$·99877	− ·00009
	Females	$\log p'_{10}$	$\bar{1}$·99885	$\bar{1}$·99882	− ·00003
		$\log p'_{15}$	$\bar{1}$·99862	$\bar{1}$·99856	− ·00006
Manchester City 1881–90	Males	$\log p'_{10}$	$\bar{1}$·99836	$\bar{1}$·99839	+ ·00003
		$\log p'_{15}$	$\bar{1}$·99807	$\bar{1}$·99807	± ·00000
	Females	$\log p'_{10}$	$\bar{1}$·99838	$\bar{1}$·99840	+ ·00002
		$\log p'_{15}$	$\bar{1}$·99816	$\bar{1}$·99816	± ·00000
Brighton 1881–90	Males	$\log p'_{10}$	$\bar{1}$·99900	$\bar{1}$·99900	± ·00000
		$\log p'_{15}$	$\bar{1}$·99872	$\bar{1}$·99870	− ·00002
	Females	$\log p'_{10}$	$\bar{1}$·99898	$\bar{1}$·99890	− ·00008
		$\log p'_{15}$	$\bar{1}$·99877	$\bar{1}$·99882	+ ·00005
Glasgow 1881–90	Males	$\log p'_{10}$	$\bar{1}$·99730	$\bar{1}$·99760	+ ·00030
		$\log p'_{15}$	$\bar{1}$·99749	$\bar{1}$·99733	− ·00016
	Females	$\log p'_{10}$	$\bar{1}$·99753	$\bar{1}$·99768	+ ·00015
		$\log p'_{15}$	$\bar{1}$·99743	$\bar{1}$·99733	− ·00010

It is thus evident that, as regards the first four Life Tables, the results obtained by "rule" are very close to those arrived at by calculation, but that in the case of Glasgow, the results of the rules are rather more divergent from the true values than is desirable.

It would be, of course, more satisfactory *in every case* to take the trouble of calculating the values of $\log p'_5$, $\log p'_{10}$ and $\log p'_{15}$.

Some idea could be formed beforehand as to how closely the results of the rules would approximate to those to be calculated, by an inspection of the *death-rates* at the age-periods 5—10, 10—15, and 15—25. If these latter should be, as in the case of Glasgow, *very high*, the rules will probably produce less satisfactory results.

See the following *comparative Table of death-rates per thousand.*

Males.

Age-periods	England and Wales	Healthy Districts	Manchester City	Brighton	Glasgow
5—10	5·35	3·88	7·62	4·83	10·65
10—15	2·95	2·28	3·71	2·30	5·52
15—25	4·97	4·13	6·18	4·57	7·59

Females.

Age-periods	England and Wales	Healthy Districts	Manchester City	Brighton	Glasgow
5—10	5·26	3·86	7·40	4·45	10·14
10—15	3·11	2·71	3·69	2·53	5·33
15—25	4·97	4·73	5·51	3·19	8·03

There is, however, no *direct* proportion between the magnitude of the death-rates and the accuracy of the results obtained by rule, for it will be noted that in the case of Manchester City, in which the death-rates come next below in order of magnitude, the results of the rules come out with extreme accuracy.

In the case of Glasgow the death-rates are not only very high but differ from those of Manchester in their relative proportions to each other.

IV. *Further suggestions for increasing the accuracy of the previously described "graphic" method.*

It is of course obvious that in drawing a curve through a number of fixed points, the shorter the intervals between these points the less is the possibility of variation in the curve when drawn by different individuals.

Thus, assuming that we have calculated the series of values $\log p'_5$, $\log p'_{10}$, $\log p'_{15} \ldots \log p'_{85}$, while it is possible, by means of these fixed points alone, to draw a fairly satisfactory curve, it would be of great advantage, if it can be done without too much trouble, to fix intermediate points in each of the 10-yearly intervals.

The following simple formulae have been worked out so as to enable this to be done with comparatively little labour.

(1) *To obtain the value of* $\log p'_{20}$, using the symbol u_x to represent $\log p'_x$,

$$u_{20} = \frac{105u_5 - 576u_{10} + 1260u_{15} + 630u_{25} - 84u_{35} + 9u_{45}}{1344}.$$

https://doi.org/10.1017/S0022172400001923 Published online by Cambridge University Press

(2) *To obtain the values of* $\log p'_{30}$, $\log p'_{40}$, &c.

If there be six *equidistant* terms at an interval from each other of $2a$, then a central term, u_o, can be interpolated by the formula:

$$u_o = \frac{300(u_{-a}+u_a) - 50(u_{-3a}+u_{3a}) + 6(u_{-5a}+u_{5a})}{512}.$$

Thus to obtain $\log p'_{30}$, or u_{30}, since $a=5$,

$$u_{30} = \frac{300(u_{25}+u_{35}) - 50(u_{15}+u_{45}) + 6(u_5+u_{55})}{512}.$$

In order to be able to obtain the values of u_{70}, u_{80}, u_{90}, and (if required) u_{100}, the series u_{45}, u_{55}, u_{65}, u_{75}, and u_{85} must be carried on by differencing so as to obtain the terms u_{95}, u_{105}, u_{115}, and u_{125}.

When this is done there will be a series of absolutely accurate fixed points, at intervals of 5 years, to start the curve with.

If desired by making $a=2\frac{1}{2}$, the same formula might be used with advantage, especially in the section of the curve after age 65, to fix the intermediate points $67\frac{1}{2}$, $72\frac{1}{2}$, $77\frac{1}{2}$, &c., &c.

The formula would be applicable to obtaining $u_{17\frac{1}{2}}$ to $u_{92\frac{1}{2}}$ inclusive, but in order to complete the series the following special formulae are needed:

$$u_{7\frac{1}{2}} = u_{17\frac{1}{2}} - \left[\frac{(u_{20}-u_5) - 5(u_{10}-u_{15})}{4}\right],$$

$$u_{12\frac{1}{2}} = \frac{5(u_{20}+9u_{15}+3u_{10}) - (u_5+24u_{17\frac{1}{2}})}{40},$$

$$u_{97\frac{1}{2}} = \frac{5(u_{90}+9u_{95}+3u_{100}) - (u_{105}+24u_{92\frac{1}{2}})}{40},$$

$$u_{102\frac{1}{2}} = u_{92\frac{1}{2}} - \left[\frac{5(u_{95}-u_{100}) + (u_{90}-u_{105})}{4}\right].$$

As a point of practical convenience, in drawing the second section of the curve, from age 55 to age 95, it may be noted that it is desirable not only to reduce the vertical scale, but to increase the horizontal scale, so that $\frac{1}{8}$ inch will correspond to ·00100 and $\frac{1}{4}$ year respectively.

With fixed points at shorter intervals than 10 years it is possible for each draughtsman according to his own convenience or preference to draw the curve in different sections and according to different scales.

To sum up it may be said that there are *three grades* in the application of the proposed graphic method:

(1) $\log p'_5$, $\log p'_{10}$, and $\log p'_{15}$ not calculated but determined by rule; $\log p'_{25} \ldots \log p'_{85}$ calculated; $\log p'_{95}$ found by differencing; curve drawn through these fixed points at 10-yearly intervals after age 15.

(2) $\log p'_5$, $\log p'_{10}$ and $\log p'_{15}$ calculated, and intermediate values $\log p'_{20}$ to $\log p'_{100}$ also fixed by calculation, so that the curve is drawn through fixed points at 5-yearly intervals.

(3) Intermediate values also fixed by calculation at intervals of $2\frac{1}{2}$ years. This, however, may be considered to involve too much labour for the purpose in view. Grade (2) will be found the most satisfactory.

https://doi.org/10.1017/S0022172400001923 Published online by Cambridge University Press

The *practical use* of working out the series $\log p'_{7\frac{1}{2}}$, $\log p'_{12\frac{1}{2}}$, &c., &c., would be to compare the results obtained by grade (2) with those to be obtained by a complete analytical interpolation, without the trouble of going through the whole scheme as set down in Note I., and obtaining $\log p_x$ values for each separate year.

V. *Note on the calculation of* $\log p'_{95}$ *and* $\log p'_{105}$.

Instead of adopting the method already given, viz. by differencing the series $\log p'_{45} \ldots \log p'_{85}$, an alternative procedure, theoretically better, would be to calculate $\log p'_{95}$ and $\log p'_{105}$ *directly* from population and deaths.

Thus, as the series of logs of $2P-d$ and $2P+d$ at age x and upwards, represented by the symbols $u_{45} \ldots u_{85}$ (and $U_{45} \ldots U_{85}$), will already have been carried down by differencing, to u_{95} and u_{105}, it will be a simple matter to carry on the series two stages further and thus obtain the terms u_{115} and u_{125}. When this has been done $\log p'_{95}$ and $\log p'_{105}$ can be worked out in exactly the same way as the series $\log p'_{25}$, &c.

Thus $$\log p'_{95} = (u_{95} + \log b) - (U_{95} + \log B),$$

and $$b = 8\,(u_{85} - u_{105}) - (u_{75} - u_{115}),$$

or, to use the most convenient form in working,

$$b = 10\,(u_{85} - u_{105}) - [2\,(u_{85} - u_{105}) + (u_{75} - u_{115})].$$

In order to calculate the hypothetical terms $\log p'_{115}$ and $\log p'_{125}$, the series $\log p'_{65} \ldots \log p'_{105}$ can be differenced. It will be noted that all these values are derived from the series of population and deaths, $u_{45} \ldots u_{85}$, and $U_{45} \ldots U_{85}$.

In order to make the new analytical scheme of interpolation set down in Note I. at the beginning, on page 206, to correspond, two additional series, 7 and 8, commencing respectively with $\log p'_{55}$ and $\log p'_{65}$, would have to be appended, with "weldings" symmetrically arranged.

This method has been mentioned to show that it has been taken into consideration, and after having tried it in several instances the writer would still recommend the use of the simpler method already described.

VI. *An easy first lesson in constructing a Life-Table from the series of p_x values when obtained.*

For the sake of those who may find any difficulty in clearly comprehending the processes of calculation by which a Life-Table is built up on the foundation of the series of p_x values, the following simple Table may be of service.

Age	p_x	l_x	P_x	Q_x	$\frac{Q_x}{l_x}=E_x$
x	$\frac{8}{10}=0{\cdot}8$	10	9	25	$\frac{25}{10}=2{\cdot}5$
$x+1$	$\frac{6}{8}=0{\cdot}75$	8	7	16	$\frac{16}{8}=2{\cdot}0$
$x+2$	$\frac{4}{6}=0{\cdot}\dot{6}$	6	5	9	$\frac{9}{6}=1{\cdot}5$
$x+3$	$\frac{2}{4}=0{\cdot}5$	4	3	4	$\frac{4}{4}=1{\cdot}0$
$x+4$	$\frac{0}{2}=0$	2	1	1	$\frac{1}{2}=0{\cdot}5$
$x+5$		0			

The processes of calculation are as follows :

(1) $l_x \times p_x = l_{x+1}$; $l_{x+1} \times p_{x+1} = l_{x+2}$; &c.

(2) $P_x = \frac{1}{2}(l_x + l_{x+1})$; $P_{x+1} = \frac{1}{2}(l_{x+1} + l_{x+2})$; &c.

(3) $Q_{x+4} = P_{x+4}$; $Q_{x+3} = Q_{x+4} + P_{x+3}$; &c.

(4) $\frac{Q_x}{l_x} = E_x$; $\frac{Q_{x+1}}{l_{x+1}} = E_{x+1}$: &c.

It is of importance to comprehend :

(1) that the p_x values have relation to the *present* tense ; i.e. each value depends upon the rate of mortality at the middle of the age x to $x+1$;

(2) that the l_x values have relation to the *past* tense, in that they are only affected by the *preceding* p_x values, from p_0 to p_{x-1} ;

(3) that the E_x values have relation to the *future* tense, in that they are only affected by p_x and the *following* values to $p_{x+\omega-1}$ (where $x+\omega$ is the age at which there are no more survivors).

It is obvious that the value $l_x=10$ can only depend upon the arbitrary number l_0 with which the Life-Table will have been commenced, and upon the values of p_0 to p_{x-1}. Assuming these to have been so different that l_x has become 20 instead of 10, this could not affect the values of E_x to E_{x+4}, for the values of p_x to p_{x+4} remain the same, and therefore all the results of the calculations in the above Table will be doubled, and $E_x=\frac{50}{20}=2{\cdot}5$ as before.

Therefore Q_x is always in direct proportion to l_x.

https://doi.org/10.1017/S0022172400001923 Published online by Cambridge University Press