ON THE ANALYTICAL CONSTRUCTION OF POPULATION AND DEATH CURVES TO FORM THE BASIS OF A LIFE-TABLE.

By T. E. HAYWARD, M.B. (Lond.), F.R.C.S. (Eng.),
Medical Officer of Health of Haydock, Lancashire.

The "graphic" method of constructing a Life-Table which is ably and lucidly expounded in the paper by Drs Newsholme and Stevenson (published in the *Journal of Hygiene*, vol. III. No. 3, pp. 297—324), is one way of dividing up foundation figures of population and deaths, given in groups of ages, into numbers belonging to each separate year of age, so that, either *directly* by the fraction $\frac{2P_x - d_x}{2P_x + d_x}$, or *indirectly* by first obtaining $m_x = \frac{d_x}{P_x}$, and then finding the value of $\frac{2 - m_x}{2 + m_x}$, the series of p_x values is obtained forming the connecting link between the data and the results of the Life-Table.

In this method the curve is drawn, guided in its course by previously constructed parallelograms, and the required ordinates are measured from points in the base-line corresponding to mid-yearly intervals.

In the alternative "analytical" method the mid-yearly ordinates are measured by processes of calculation, and when so obtained they may be plotted out, one by one, to scale; then, by drawing a curved line through the extremities of these ordinates, and by the further addition of parallelograms representing to scale the foundation numbers as given in groups of ages, a diagram of a population or of a death curve is obtained corresponding to a diagram belonging to the "graphic" method, but arrived at by a reverse process.

There was described in a paper contributed by the present writer to the *Journal of Hygiene*, vol. II. No. 1, a method of constructing analytically an extended Life-Table, which consisted essentially (1) in calculating at certain intervals ordinates of population and death curves,

(2) in obtaining from these the corresponding p_x values, and (3) in completing the yearly series of p_x values by interpolations in series where $u_x = \log p_x$.

However the writer is now of opinion that if it be worth while to set out to construct an extended Life-Table by analytical methods it is best to begin by interpolating population and death numbers for each year of age after the first five.

The object of this paper is first to give a series of formulae (based on the method of finite differences) reduced to their lowest terms of simplicity, by means of which population and death curves may be calculated before being graphically represented, and then to illustrate the results by one pair of diagrams chosen out of a considerable number which have been prepared in the evolution of the scheme.

It is of some importance, however, to comprehend the difference in the geometrical aspect of the "graphic" and "analytical" methods.

Whereas in the graphic method the foundation numerical data are first represented as parallelograms and then the curve is drawn through, and by means of these, in the analytical method only "linear quantities" can be dealt with, therefore by the successive addition to the number belonging to the age-group 85 and upwards of the number belonging to the age-group 75—85, and then of the number belonging to the age-group 65—75, and so on, a series of linear quantities will be obtained representing, as the case may be, the numbers of population or deaths *at age x and upwards*. These may be regarded as ordinates of a curve erected at points in a base-line corresponding to exact ages (or values of x), 4, 5, 10, 15, 20, 25...85.

It is convenient to use the symbol u_x to represent the number of population or of deaths at age x and upwards.

If now by processes of calculation we can interpolate the intermediate ordinates u_6, u_7, u_8, etc., so as to complete the series for each exact year of age, then it is obvious that $u_5 - u_6$ will give the number of population or deaths, as the case may be, belonging to the year of age from 5 to 6, and so on; and if we are plotting out by analytical means a population or a death curve we start with a base-line and measure out at the point $5\frac{1}{2}$ in this base-line an ordinate equal to $u_5 - u_6$, and so on.

In actual practice, however, it is best to use *not* the numbers but the *logarithms* of the numbers expressing population or deaths at age x and upwards, because it is only thus possible to obtain a rational continuation

https://doi.org/10.1017/S0022172400002837 Published online by Cambridge University Press

of the series beyond the point where the actual data cease to be available.

Further, instead of effecting interpolations in two series in one of which u_x = the log of the number representing population at age x and upwards, and in the other u_x = the log of the corresponding number of deaths at age x and upwards, the two series may be so modified that u_x = the log of the number representing twice population *minus* deaths at age x and upwards in one series, and in the other u_x = the log of the number representing twice population *plus* deaths at age x and upwards.

If *numbers* were being dealt with the results would be the same whether population and deaths were dealt with separately or in combination, as $2P-d$ and $2P+d$, but if the *logarithms* of the numbers are used the results as regards the values of P and d for each year of age must necessarily differ.

When population and deaths are taken separately, the series of u_x values (logs) having been completed by interpolation, the numerical values of the series are taken out and then by subtracting the value corresponding to u_{x+1} from the value corresponding to u_x the value of P or d is obtained for the year of age x to $x+1$.

On the other hand when the *combined* method is used similar procedures result in the obtaining of values representing $2P-d$ and $2P+d$ for each year of age.

Then $$P = \tfrac{1}{4}\left[(2P-d)+(2P+d)\right]$$

and $$d = \tfrac{1}{2}\left[(2P+d)-(2P-d)\right],$$

that is to say, from the two combined values the population number is obtained by taking a fourth part of their sum, and the death number by taking half their difference.

It is found that the values of P obtained by the two methods very nearly coincide, but that the values of d may differ considerably. However, when the two sets of values are plotted out, it is seen that the two curves only diverge from each other at intervals, and that they alternately intersect each other.

Seeing that a third diagram originally prepared to illustrate these points (as regards the two death curves) could not be satisfactorily reproduced on the diminished scale necessary for publication, it has had to be left out. The practical point brought out in the large diagram is that the death curve obtained by the *combined* method of interpolation (as $2P-d$ and $2P+d$) is on the whole *decidedly a smoother and better* *curve* than the other obtained by interpolations in deaths taken separately. In the actual construction of a Life-Table the smoother curve of deaths would give the smoother m_x and p_x curves, and therefore, from this point of view, the combined method is to be preferred. It may also be noted that a comparison having been made of the results obtained by the application of the "Improved method of constructing shortened Life-Tables" (described in this *Journal*, vol. V. No. 1), to the data of the Life-Table for Brighton (males) for 1891—1900, first with P and d taken separately, and then as $2P-d$ and $2P+d$, *the latter gave the nearer approach to the results of the original Life-Table*[1].

The two diagrams of population and death curves given respectively in Figs. 1 and 2 (pp. 223, 225) represent numbers obtained by interpolations according to the *combined* method.

After these preliminary observations the proposed scheme of interpolation and the formulae belonging to it may now be given. In order that the scheme may be comprehended, some degree of knowledge of the theory of "finite difference" and "interpolation" must be presupposed. Reference may be made by those without such knowledge to some attempted brief explanations given in the paper already alluded to as having been published in this *Journal*, vol. II. No. 1, January 1902, pp. 15—18 and 23—25.

Scheme of Interpolation.

Series 1...$u_4\ [u_5\ u_{10}\ u_{15}]\ u_{20}$

,, 2...... $u_5\ [u_{10}\ u_{15}\ u_{20}]\ u_{25}$

,, 3......... $u_{10}\ [u_{15}\ u_{20}\ u_{25}]\ u_{35}$

,, 4............ $u_{15}\ [u_{20}\ u_{25}\ u_{35}]\ u_{45}$

,, 5.................. $u_{20}\ [u_{25}\ u_{35}\ u_{45}]\ u_{55}$

,, 6..................... $u_{25}\ [u_{35}\ u_{45}\ u_{55}]\ u_{65}$

,, 7........................... $u_{35}\ [u_{45}\ u_{55}\ u_{65}]\ u_{75}$

,, 8.............................. $u_{45}\ [u_{55}\ u_{65}\ u_{75}\ u_{85}\ldots]$

The general principles of the scheme are:

(1) To take the given series of u_x values in eight series consisting of five terms in each, symmetrically overlapping each other, the orders of differences used being thus limited to four.

(2) To use only the central part of each series.

[1] For a previous discussion of the same point see *Journal of the Royal Statistical Society*, vol. LXII. Part IV. pp. 696—699, where the same conclusion is arrived at as to the preferability of the *combined* method of interpolation.

(3) To weld or combine the overlapping series by pairs of factors derived from the "curve of cosines" so that one series shall gradually pass into the other.

In the above scheme the part of each series actually used is indicated by brackets, and the *under*lined part of the series above is to be welded with the corresponding *over*lined part of the series below.

With 5-yearly intervals there are of course four intermediate terms to be combined, and with 10-yearly intervals nine.

In order that the simplest formulae of interpolation may be used it is necessary to have five consecutive *equidistant* terms in each series.

Having given five such terms it is a simple matter to set them down and "difference" them until the fourth difference is obtained.

As in each series of the given scheme the interpolation is required to begin at the *second* line of differences, this line must be completed by filling up the blank space by the constant fourth difference.

Then, indicating the second term of the series (relatively to the other terms) as u_0, and denoting the differences corresponding to 5-yearly or 10-yearly intervals by the symbol Δ, the subdivided line of differences corresponding to yearly intervals, represented by the symbol δ, may be obtained by the following formulae.

5-yearly intervals.

$$\delta^4 u_0 = \cdot 0016\Delta^4 u_0$$
$$\delta^3 u_0 = \cdot 008\Delta^3 u_0 - 6\delta^4 u_0$$
$$\delta^2 u_0 = \cdot 04\Delta^2 u_0 - 4\delta^3 u_0 - 8\delta^4 u_0$$
$$\delta u_0 = \cdot 2\Delta u_0 - 2\delta^2 u_0 - 2\delta^3 u_0 - \delta^4 u_0$$

10-yearly intervals.

$$\delta^4 u_0 = \cdot 0001\Delta^4 u_0$$
$$\delta^3 u_0 = \cdot 001\Delta^3 u_0 - 13\cdot 5\delta^4 u_0$$
$$\delta^2 u_0 = \cdot 01\Delta^2 u_0 - 9\delta^3 u_0 - 44\cdot 25\delta^4 u_0$$
$$\delta u_0 = \cdot 1\Delta u_0 - 4\cdot 5\delta^2 u_0 - 12\delta^3 u_0 - 21\delta^4 u_0$$

The pairs of "welding" factors required are as follows:

5-yearly intervals.		*10-yearly intervals.*	
		·976	·024
·904	·096	·904	·096
		·794	·206
·654	·346	·654	·346
		·500	·500
·346	·654	·346	·654
		·206	·794
·096	·904	·096	·904
		·024	·976

It will be noted that the given terms in series 1, 3, 4 and 5 are not equidistant. Therefore, before the simple formulae applicable to an equidistant series of terms can be made use of, the following formulae, which have been worked out by means of the general formula of

Lagrange, must be employed to obtain the required *relative* terms in each series.

For series 1,

$$u_0 = \frac{625u_4 + 11u_{20} - 64u_{15}}{44} - 4(4u_5 - u_{10}).$$

The bringing in of the term u_4 avoids the break in the regularity of Life-Table curves between the ages of 4 and 5 which may otherwise occur.

The required five consecutive equidistant terms to be differenced are u_0, u_5, u_{10}, u_{15}, u_{20}.

For series 3,

$$u_{30} = u_{15} - 2\left[(u_{20} - u_{25}) + \tfrac{1}{10}(u_{10} - u_{35})\right].$$

The required five terms are then u_{10}, u_{15}, u_{20}, u_{25}, u_{30}.

For series 4,

$$u_{30} = \frac{5(u_{15} + 9u_{25} + 3u_{35}) - (24u_{20} + u_{45})}{40}.$$

The terms to be differenced are, then, u_{15}, u_{20}, u_{25}, u_{30}, u_{35}.

For series 5,

$$u_{15} = \frac{4(32u_{20} - 7u_{45}) + 5u_{55}}{35} - 2(2u_{25} - u_{35}).$$

(This u_{15} is of course only a *relative* term in series 5; it must not be confounded with the *true* u_{15} in the preceding series.)

The terms to be differenced are u_{15}, u_{25}, u_{35}, u_{45}, u_{55}.

The foundation figures of population and deaths to which the preceding formulae have been applied to obtain the results graphically represented in Figs. 1 and 2 (pp. 223, 225), are those relating to England and Wales (males) for the decennium 1881—90.

The "years of life" for the decennium have been calculated from the enumerated population at the censuses of 1881 and 1891 (as given finally corrected in Table 13, page 206 of Appendix A in *The General Report of the Census of* 1901), by the method of Mr A. C. Waters (described in *The Journal of the Royal Statistical Society*, vol. LXIV. Part II. June 1901).

It must be noted, however, that the *enumerated* figures for ages from 0 to 15 inclusive have been dealt with as *one* age-group, and that the "years of life" therein included have been divided up between the age-groups 0—5, 5—10, and 10—15 in proportion to the *calculated* numbers of survivors in these age-groups given in Table 19, page 211 of the *Census Report* for 1901 above alluded to.

The "years of life" for the age-group 0—5 have been further subdivided into numbers belonging to each of the first five years of age in proportion to numbers obtained by deaths and of births.

The attempt thus made to eliminate errors due to misstatements of age during the first 15 years of life would appear to have been successful from the regularity of the population curve in Fig. 1.

For representation on the scale employed the proportionate distribution of half a million of the total of 133,020,965 "years of life" has been shown.

For the curve of deaths, as out of 133,020,968 "years of life" there occurred 2,698,316 deaths, the number of deaths out of half a million "years of life" was 10,142·45, and the proportionate distribution of this number has been shown in the diagram.

Concluding remarks as to the preferability of "analytical" or "graphic" methods.

In so far as the present writer is concerned he must confess that the result of having laboured to devise what might appear the simplest and best analytical method has been to increase his admiration for the beauty and simplicity of Milne's method of graphically constructing population and death curves as it has been explained by Mr George King, F.I.A., etc.

However, to satisfactorily apply this graphic method there would appear to be required a high degree of technical skill in drawing and measuring the curves, especially in *drawing* them.

If it should appear that the method of analytically calculating population and death curves described in this paper, when the results are graphically represented, gives curves such as might have been drawn by a master of the art of graphic construction, it may be that some who are doubtful of their graphic powers will care to avail themselves of the perhaps more common ability to calculate according to a given scheme.

It is further to be considered that the application of such a scheme to several sets of data will eliminate variations due to "personal equation" in drawing curves and measuring ordinates, and will ensure that the results obtained will be strictly comparable.

https://doi.org/10.1017/S0022172400002837 Published online by Cambridge University Press

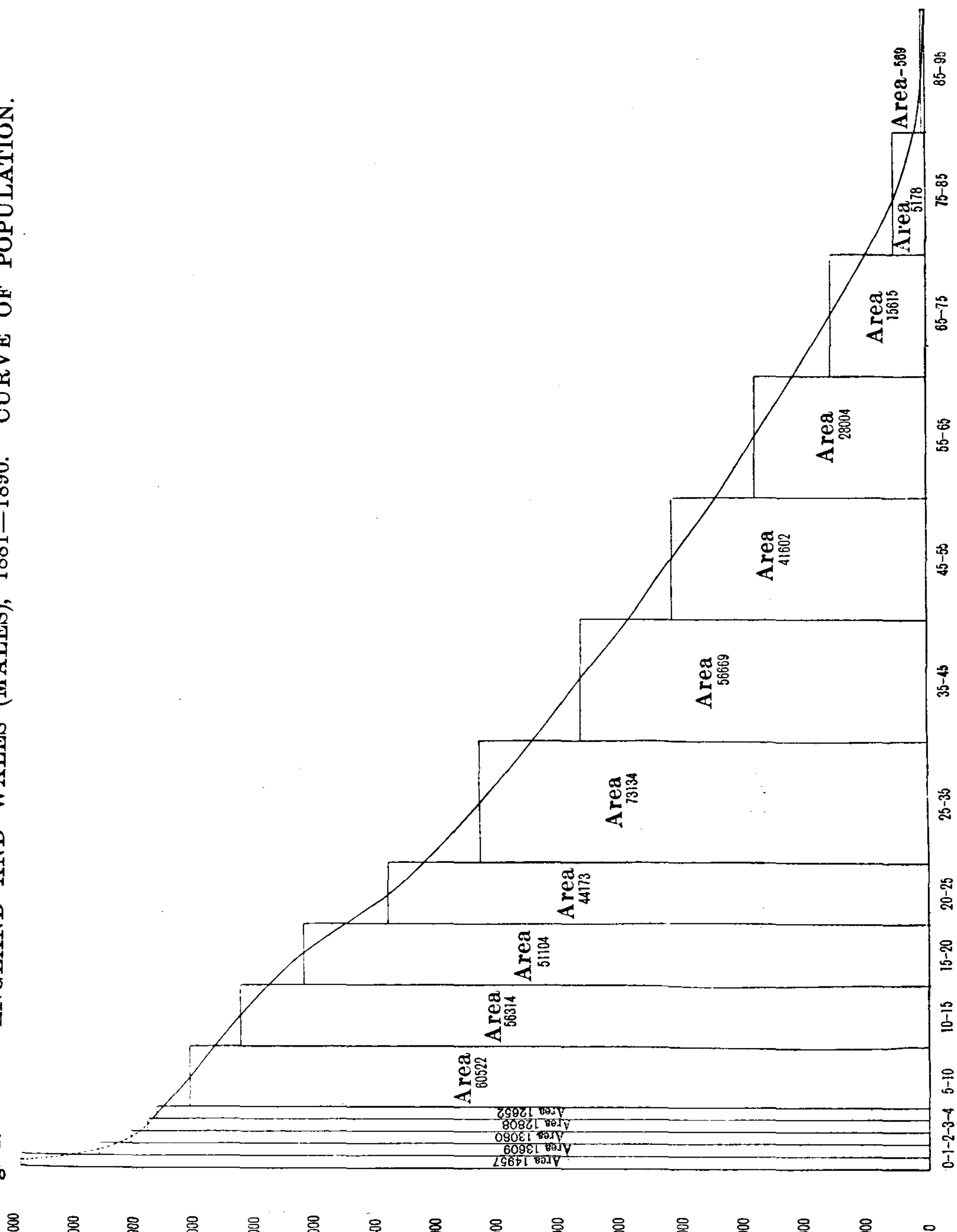
In the following tables are given the numbers of population and of deaths, calculated for each year of age, which have been used in the construction of the two curves represented respectively in the diagrams.

England and Wales (Males), 1881—1890. *Population.*

Age x to $x+1$		
0	14957	
1	13609	
2	13080	
3	12808	
4	12652	
5	12451	
6	12268	
7	12097	60522
8	11933	
9	11773	
10	11611	
11	11445	
12	11272	56314
13	11090	
14	10896	
15	10696	
16	10497	
17	10262	51104
18	9977	
19	9672	
20	9374	
21	9074	
22	8798	44173
23	8566	
24	8361	
25	8154	
26	7952	
27	7757	
28	7570	
29	7387	73134
30	7208	
31	7032	
32	6859	
33	6690	
34	6525	
35	6365	
36	6212	
37	6063	
38	5913	
39	5759	56669
40	5599	
41	5434	
42	5269	
43	5106	
44	4949	

Age x to $x+1$		
45	4800	
46	4661	
47	4526	
48	4389	
49	4247	41602
50	4098	
51	3945	
52	3792	
53	3643	
54	3501	
55	3368	
56	3242	
57	3121	
58	2998	
59	2872	28004
60	2742	
61	2610	
62	2479	
63	2349	
64	2223	
65	2100	
66	1979	
67	1859	
68	1738	
69	1618	15615
70	1499	
71	1380	
72	1262	
73	1147	
74	1033	
75	923	
76	816	
77	715	
78	619	
79	530	5178
80	448	
81	374	
82	307	
83	248	
84	198	
85	155	
86	119	
87	90	
88	66	
89	48	56
90	34	
91	24	
92	16	
93	10	
94	7	

Fig. 1. ENGLAND AND WALES (MALES), 1881—1890. CURVE OF POPULATION.



https://doi.org/10.1017/S0022172400002837 Published online by Cambridge University Press

England and Wales (Males), 1881—1890. *Deaths.*

Interpolations in $2P-d$ *and* $2P+d$.

Age x to $x+1$		
4	143·28	
5	106·65	
6	79·11	
7	59·16	327·09
8	45·44	
9	36·73	
10	32·67	
11	31·97	
12	32·74	166·44
13	33·82	
14	35·24	
15	38·16	
16	42·20	
17	45·48	220·74
18	47·09	
19	47·81	
20	48·84	
21	49·86	
22	50·78	253·34
23	51·54	
24	52·32	
25	53·10	
26	53·48	
27	53·80	
28	54·22	
29	55·06	568·37
30	56·21	
31	57·86	
32	59·68	
33	61·60	
34	63·36	
35	64·82	
36	66·19	
37	67·59	
38	68·88	
39	70·11	702·78
40	71·25	
41	72·21	
42	73·08	
43	73·86	
44	74·79	

Age x to $x+1$		
45	75·76	
46	76·28	
47	77·16	
48	78·14	
49	79·36	806·45
50	80·81	
51	82·40	
52	84·00	
53	85·58	
54	87·16	
55	88·67	
56	90·00	
57	91·40	
58	93·15	
59	95·29	972·45
60	97·78	
61	100·45	
62	103·02	
63	105·37	
64	107·32	
65	108·91	
66	110·23	
67	111·24	
68	111·93	
69	112·18	1100·87
70	111·98	
71	111·25	
72	109·93	
73	107·96	
74	105·26	
75	101·84	
76	97·67	
77	92·76	
78	87·17	
79	80·97	759·90
80	74·27	
81	67·22	
82	59·93	
83	52·63	
84	45·44	
85	38·55	
86	32·10	
87	26·21	
88	20·96	
89	16·39	171·08
90	12·54	
91	9·36	
92	6·81	
93	4·83	
94	3·33	

https://doi.org/10.1017/S0022172400002837 Published online by Cambridge University Press

Fig. 2. ENGLAND AND WALES (MALES), 1881—1890. CURVE OF DEATHS.