VOL. VI MEETING NO. 14

TRANSACTIONS

APRIL, 1954

ON THE FORMULA FOR THE L-FUNCTION IN A SPECIAL MORTALITY TABLE ELIMINATING A GIVEN CAUSE OF DEATH

T. N. E. GREVILLE

AT THE end of my paper "Mortality Tables Analyzed by Cause of Death" (*RAIA* XXXVII, 283), I discussed the problem of obtaining a suitable approximate formula for ${}_nL_x^{(-i)}$, the number of years lived between ages x and $x+n$ by the $l_x^{(-i)}$ survivors at age x, where the superscript $(-i)$ indicates a special mortality table from which the ith cause of death has been eliminated. It is assumed that values of $l_x^{(-i)}$ have already been computed for the terminal ages of the age intervals employed. It is true that the function ${}_nL_x^{(-i)}$ is not often desired for its own sake, and is generally regarded as merely a step in the computation of $\mathring{e}_x^{(-i)}$, and that all reasonable formulas for ${}_nL_x^{(-i)}$ will probably give, in most cases, very nearly the same value of $\mathring{e}_x^{(-i)}$. However, the question of finding the most suitable formula for ${}_nL_x^{(-i)}$ is not entirely academic, as it may be desired to show these values, and this function has some applications in computing survival rates free from the influence of a given cause of death.[1]

The approximate formula I previously suggested is

$$ {}_nL_x^{(-i)} = \frac{{}_nd_x^{(-i)}}{{}_nd_x^{-i}}\, {}_nL_x\,, \tag{1} $$

where ${}_nd_x^{-i}$ denotes the number of deaths between ages x and $x+n$ in the main mortality table from all causes except the ith. I expressed some dissatisfaction with this formula on the ground that it consistently overstates the value, though apparently not by a significant amount. It is the purpose of this note to point out a further theoretical objection to formula (1), and to suggest as an alternative the approximate formula

$$ {}_nL_x^{(-i)} = nl_x^{(-i)} - \frac{{}_nd_x^{(-i)}}{{}_nd_x}\,(nl_x - {}_nL_x)\ . \tag{2} $$

[1] See, for example, "Effect of Cancer on Longevity," *Vital Statistics—Special Reports*, vol. 32, No. 7, July 28, 1950, Public Health Service, National Office of Vital Statistics, Washington.

Before giving formula (1), I pointed out that formulas for obtaining ${}_nL_x^{(-i)}$ by approximate integration based on neighboring values of $l_x^{(-i)}$ may sometimes produce a value of ${}_nL_x^{(-i)}$ less than ${}_nL_x$, especially at ages of low mortality from the ith cause. I have since observed that, in a few extreme cases, formula (1) and some other formulas can lead to anomalies of a different sort: values of ${}_nL_x^{(-i)}$ greater than $nl_x^{(-i)}$ or less than $nl_{x+n}^{(-i)}$.

To show that this is a theoretical possibility, we first point out that it is possible to imagine the deaths between ages x and $x+n$ so concentrated either at the end or at the beginning of the age interval that ${}_nL_x$ can be made as close as we please to nl_x or to nl_{x+n}. The right member of formula (1) can be expressed as

$$\frac{{}_nq_x^{(-i)}l_x^{(-i)}}{{}_nq_x^{-i}l_x}\,{}_nL_x \qquad \text{or} \qquad \frac{{}_nq_x^{(-i)}}{{}_nq_x^{-i}}\,nl_x^{(-i)}\cdot\frac{{}_nL_x}{nl_x}.$$

Now, ${}_nq_x^{(-i)}$ is greater than ${}_nq_x^{-i}$, since ${}_nq_x^{(-i)}=\int_0^n {}_tp_x^{(-i)}\mu_{x+t}^{-i}\,dt$ and ${}_nq_x^{-i}=\int_0^n {}_tp_x\mu_{x+t}^{-i}\,dt$ and ${}_tp_x^{(-i)}$ is always greater than ${}_tp_x$. It is clear, therefore, that formula (1) could give a value of ${}_nL_x^{(-i)}$ greater than $nl_x^{(-i)}$ if ${}_nL_x$ were sufficiently close to nl_x. Similarly, the right member of formula (1) can be expressed as

$$\frac{{}_nq_x^{(-i)}\dfrac{l_{x+n}^{(-i)}}{{}_np_x^{(-i)}}}{{}_nq_x^{-i}\dfrac{l_{x+n}}{{}_np_x}}\,{}_nL_x \qquad \text{or} \qquad \frac{{}_np_x\,{}_nq_x^{(-i)}}{{}_np_x^{(-i)}\,{}_nq_x^{-i}}\,nl_{x+n}^{(-i)}\cdot\frac{{}_nL_x}{nl_{x+n}}.$$

This expression reduces to

$$\frac{{}_np_x^{(i)}-{}_np_x}{{}_nq_x-{}_nq_x^{i}}nl_{x+n}^{(-i)}\cdot\frac{{}_nL_x}{nl_{x+n}} \qquad \text{or} \qquad \frac{{}_nq_x-{}_nq_x^{(i)}}{{}_nq_x-{}_nq_x^{i}}\,nl_{x+n}^{(-i)}\cdot\frac{{}_nL_x}{nl_{x+n}},$$

since ${}_np_x={}_np_x^{(i)}\,{}_np_x^{(-i)}$ and ${}_nq_x={}_nq_x^{i}+{}_nq_x^{-i}$. Therefore, the formula could give a value less than $nl_{x+n}^{(-i)}$ if ${}_nL_x$ were close enough to nl_{x+n}.

An approximation which cannot produce anomalous results of either kind is given by

$$\frac{nl_x^{(-i)}-{}_nL_x^{(-i)}}{nl_x^{(-i)}-nl_{x+n}^{(-i)}}=\frac{nl_x-{}_nL_x}{nl_x-nl_{x+n}}. \tag{3}$$

This, of course, is essentially a form of straight-line interpolation between $nl_x^{(-i)}$ and $nl_{x+n}^{(-i)}$, and reduces at once to formula (2). Using the relation

$$nl_x - {}_nL_x = \int_0^n (n-t)\, l_{x+t}\mu_{x+t}dt\ ,$$

which is easily obtained through integration by parts, formula (3) can be written in the form

$$\frac{\int_0^n (n-t)\, l_{x+t}^{(-i)}\mu_{x+t}^{-i}dt}{\int_0^n l_{x+t}^{(-i)}\mu_{x+t}^{-i}dt} = \frac{\int_0^n (n-t)\, l_{x+t}\mu_{x+t}dt}{\int_0^n l_{x+t}\mu_{x+t}dt}.$$

Here it is apparent that each member is a weighted average of the durations from 0 to n, the weights being $l_{x+t}^{(-i)}\mu_{x+t}^{-i}$ and $l_{x+t}\mu_{x+t}$ in the respective cases. It is fairly clear that this assumption does not involve a consistent bias in either direction.

Another possible alternative would be to use "geometric" interpolation between $nl_x^{(-i)}$ and $nl_{x+n}^{(-i)}$, which gives

$$\frac{\log ({}_nL_x^{(-i)}/nl_x^{(-i)})}{\log (nl_{x+n}^{(-i)}/nl_x^{(-i)})} = \frac{\log ({}_nL_x/nl_x)}{\log (nl_{x+n}/nl_x)},$$

or

$$\frac{\log ({}_nL_x^{(-i)}/nl_x^{(-i)})}{\log ({}_nL_x/nl_x)} = \frac{\log {}_np_x^{(-i)}}{\log {}_np_x}. \tag{4}$$

Since I obtained ${}_np_x^{(-i)}$ by the relation $\log {}_np_x^{(-i)} = r_w^{-i} \log {}_np_x$, where r_w^{-i} denotes the proportion of deaths from causes other than the ith among the observed deaths in the age interval in question, the relation (4) reduces to

$${}_nL_x^{(-i)}/nl_x^{(-i)} = ({}_nL_x/nl_x)^{r_w^{-i}}. \tag{5}$$

This always gives a larger value than that based on formula (3) or its equivalent, formula (2). To show this, we note first that formula (2) can be put into the form

$${}_nL_x^{(-i)}/nl_x^{(-i)} = 1 - \frac{{}_nq_x^{(-i)}}{{}_nq_x}(1 - L_x/nl_x)\ . \tag{6}$$

Now consider the expression

$$f(z) = \frac{1-(1-z)^{r_w^{-i}}}{z}.$$

Upon expanding the binomial, this becomes an infinite power series in z with only positive coefficients (since $r_{\bar{w}}^{-i} < 1$). Therefore, if $z_1 > z_2$, we have $f(z_1) > f(z_2)$. Since ${}_np_x < {}_nL_x/nl_x$, clearly ${}_nq_x > 1 - {}_nL_x/nl_x$. Taking $z_1 = {}_nq_x$ and $z_2 = 1 - {}_nL_x/nl_x$ and rearranging, $f(z_1) > f(z_2)$ becomes

$$1 - \frac{{}_nq_x^{(-i)}}{{}_nq_x}(1 - {}_nL_x/nl_x) < [1 - (1 - {}_nL_x/nl_x)]^{r_{\bar{w}}^{-i}} = ({}_nL_x/nl_x)^{r_{\bar{w}}^{-i}},$$

since ${}_nq_x^{(-i)} = 1 - (1 - {}_nq_x)^{r_{\bar{w}}^{-i}}$. In view of the relation (6), this shows that formula (5) always gives a larger value than formula (2).

I have suggested elsewhere[2] still another approximation in connection with the method of construction of an abridged mortality table by reference to a standard table which is used in the preparation of the annual abridged life tables for the United States. This might be adapted for the present purpose in the form

$${}_nL_x^{(-i)} = \frac{l_x^{(-i)} + l_{x+n}^{(-i)}}{l_x + l_{x+n}}\,{}_nL_x.$$

Clearly, the value of ${}_nL_x^{(-i)}$ by this formula is always greater than ${}_nL_x$. However, the right member can be expressed as

$$\frac{1 + {}_np_x^{(-i)}}{1 + {}_np_x}\, nl_x^{(-i)} \cdot \frac{{}_nL_x}{nl_x},$$

showing that this approximation could exceed $nl_x^{(-i)}$ if ${}_nL_x$ were sufficiently close to nl_x. Similarly, it can be shown that it could be made less than $nl_{x+n}^{(-i)}$ by taking ${}_nL_x$ close enough to nl_{x+n}.

In theory, this sort of anomaly could happen also in the construction of an abridged mortality table by reference to a standard table, though this would appear to be unlikely if the table under construction closely resembles the standard table, and it is only in such a situation that the method referred to was recommended. However, there would seem to be nothing lost, and perhaps something gained, by using instead the following analogue of formula (2):

$${}_nL_x = nl_x - \frac{{}_nd_x}{{}_nd_x^s}(nl_x^s - {}_nL_x^s),$$

where the superscript s indicates functions based on the standard table.

[2] "Method of Constructing the Abridged Life Tables for the United States, 1949," *Vital Statistics—Special Reports*, vol. 33, No. 15, June 30, 1953, p. 258, Public Health Service, National Office of Vital Statistics, Washington.

Returning to the mortality table analyzed by cause of death, a brief reference to the final age group of the oldest ages is in order. Of the formulas mentioned in this note, formula (1) is the only one which is applicable to this case. However, the type of anomaly sometimes produced by this formula in other age intervals cannot occur here, since the value of n is not defined. In this case, formula (1) reduces to

$$\mathring{e}_y^{(-i)} = \frac{\mathring{e}_y}{r_w^{-i}},$$

where the final age group consists of all ages beyond y. There is a certain logic in this formula, since, for $r_w^{-i} = 1$, which would mean that there are no deaths beyond age y from the ith cause, we have $\mathring{e}_y^{(-i)} = \mathring{e}_y$; while, for $r_w^{-i} = 0$, which would mean that all deaths beyond age y are from this cause, $\mathring{e}_y^{(-i)}$ becomes infinite!