THE STANDARD ERROR OF AN ESTIMATE OF EXPECTATION OF LIFE, WITH SPECIAL REFERENCE TO EXPECTATION OF TUMOURLESS LIFE IN EXPERIMENTS WITH MICE

BY J. O. IRWIN, SC.D., D.SC., *Statistical Research Unit, Medical Research Council*

1. DERIVATION OF THE RESULT

An expectation of life limited to n years is properly defined by

$$(1/l_0)\int_0^n l_x dx, \tag{1}$$

to which the approximation usually used is

$$E=(1/l_0)\{\tfrac{1}{2}l_0+l_1+l_2+\ldots+l_{n-1}+\tfrac{1}{2}l_n\}. \tag{2}$$

The complete expectation is obtained by substituting $(\omega+1)$ for n, where ω is the greatest age at which there are survivors.

It is proposed to find the standard error of the expression (2). We may write

$$l_r=l_0p_0p_1p_2\cdots p_{r-1}, \tag{3}$$

where l_0 is the fixed base number supposed born at any moment of time and $p_0, p_1, \ldots, p_{r-1}$ are the probabilities of surviving one year at ages 0, 1, 2, ..., $(r-1)$. Let us suppose that the estimates $p_0, p_1, \ldots, p_{r-1}$ have been obtained from respectively $N_0, N_1, N_2, \ldots, N_{r-1}$ exposed to risk. It is advisable to distinguish between the true values of l_r, p_r and the estimates of them made from the data. Accordingly, for the true values we shall write

$$l_r=l_0P_0P_1P_2\ldots P_{r-1}.$$

The sampling errors of the p's or q's may be taken as uncorrelated (in contradistinction to the actual numbers of deaths at any year of age). The values of N are of course also subject to sampling errors, which will in general be correlated and whose exact values depend on how the numbers exposed are determined; for instance, whether there are entrances to the age group by immigration and exits by emigration as well as death. However, the effect of variation in the N's on the standard error of E will be small. For the variance of any N (whose expected value is $\bar{N}$) will in general be of order $\bar{N}$, say $k\bar{N}$, and if $q=d/N$, $p=1-q$

$$V(p)=(V(d)/\bar{N}^2)+(d^2V(N)/\bar{N}^4)$$
$$=(pq+kq^2)/\bar{N}=0(q/\bar{N}).$$

At most ages q^2 will be relatively small compared with pq, and in any case, if the formula obtained is to be applied, N must be substituted for $\bar{N}$. It therefore seems best to determine the standard error of E for a fixed set of N's, that is, to regard the values of the N's as ancillary information.

The p's being uncorrelated it follows that

$$\mu'_k(l_r)=l_0\prod_{t=0}^{n-1}\mu'_k(p_t). \tag{4}$$

Taking $k=1$, we find

$$\left.\begin{aligned}(l_r/l_0)&=\prod_{t=0}^{r-1}P_t,\\ (1/l_0^2)\,V(l_r)&=\prod_{t-0}^{r-1}\left(P_t^2+\frac{P_tQ_t}{\bar{N}_t}\right)-\prod_{t=0}^{r-1}P_t^2,\end{aligned}\right\} \tag{5}$$

and $$\operatorname{cov}(l_rl_s)=(l_s/l_r)\,V(l_r)\quad(s>r). \tag{6}$$

(5) and (6) are exact expressions. As a rule it will be adequate to take the approximate formula

$$V(l_r)=l_r^2\sum_{t=0}^{r-1}(Q_t/N_tP_t). \tag{5 bis}$$

Now

$$\begin{aligned}V(E)=(1/l_0^2)\Big[&V(l_1)+V(l_2)+\ldots+V(l_{n-1})+\tfrac{1}{4}V(l_n)\\&+(l_n/l_1)\,V(l_1)+(l_n/l_2)\,V(l_1)+\ldots+(l_n/l_{n-1})\,V(l_n)\\&+2\sum_{r=0}^{n-1}\{(l_{r+1}/l_r)+(l_{r+2}/l_r)+\ldots+(l_{n-1}/l_r)\}\,V(l_r)\Big]\\=(1/l_0^2)\Big[&2\sum_{r=1}^{n-1}\{V(l_r)/l_r\}\{\tfrac{1}{2}l_r+l_{r+1}+\ldots+l_{n-1}+\tfrac{1}{2}l_n\}\\&+\tfrac{1}{4}V(l_n)\Big]\\=(1/l_0^2)\Big[&2\sum_{r=1}^{n-1}V(l_r)\,E_{n/r}+\tfrac{1}{4}V(l_n)\Big],\end{aligned} \tag{7}$$

where $E_{n/r}$ is the expectation of life at age r limited to n years.

If the expression (5) be substituted for $V(l_r)$ in (7), we have the exact sampling variance of E; if the expression (5 *bis*) be substituted, we reach

$$(1/l_0^2)\left[2\sum_{r=1}^{n-1}\left(\sum_{t=0}^{r-1}(Q_t/N_tP_t)\,l_r^2E_{n/r}\right)+\frac{1}{4}\sum_{t=0}^{n-1}(Q_t/N_tP_t)\,l_n^2\right]. \tag{8}$$

https://doi.org/10.1017/S0022172400014443 Published online by Cambridge University Press

The expression (8) can be transformed into a simpler form. Writing

$$\sum_{t=0}^{r-1}(Q_t/N_tP_t)=U_r,$$

$$l_0^2V(E)=2\sum_{r=1}^{n-1}U_rl_r(\tfrac{1}{2}l_r+l_{r+1}+\dots+l_{n-1}+\tfrac{1}{2}l_n)+\tfrac{1}{4}U_nl_n^2$$

$$=\sum_{r=0}^{n-1}(S_{r+1}^2Q_r/N_rP_r), \qquad (9)$$

where

$$S_r=l_r+l_{r+1}+\dots+l_{n-1}+\tfrac{1}{2}l_n.$$

We conclude finally that the standard error of E is given by

$$\sigma_E=(1/l_0)\left[2\sum_{r=1}^{n-1}V(l_r)\,E_{n,r}+\tfrac{1}{4}V(l_n)\right]^{\frac{1}{2}}, \qquad (10)$$

where $V(l_r)$ is given by (5), while an adequate and usually very close approximation is given by

$$\sigma_E=(1/l_0)\left[\sum_{r=0}^{n-1}S_{r+1}^2Q_r/N_rP_r\right]^{\frac{1}{2}}, \qquad (11)$$

where $S_r=\sum_{r=1}^{n-1}l_r+\tfrac{1}{2}l_n$. This is because for most age groups P_t/Q_tN_t is very small compared with P_t^2, even if N_t is very small. In the extreme case when $N_t=2$, $P_t=\tfrac{1}{2}$, the ratio Q_t/N_tP_t is only $\tfrac{1}{2}$.

showing that the approximate formula (11) is quite good enough in practice. Accordingly, only the details of the calculation using formula (11) are shown in the table. The expectation of tumourless life limited to 25 weeks was required; in fact, there were no tumourless survivors after the end of the 23rd week.

In applying the formula the observed values l_r, p_t must of course be substituted for the expected values L_r, P_t.

2. A NUMERICAL EXAMPLE

In the previous treatment, the unit of age was taken as a year, but it might equally be any other period. In animal experiments it will usually be convenient to take a much shorter period, for example, a week or a month. In the analysis published in this *Journal* (Irwin, 1946) of the results of C. C. and J. M. Twort's experiments in which mice were painted with oils and tars and a record kept of the tumours which developed, much use was made of the *expectation of tumourless life*. This is the average time an animal would remain tumourless in the course of an experiment if there were no deaths. An example of the calculation of this expectation is given in Table 1 of the paper cited. There the first week of the experiment was designated week 1. It is better to call it week 0 in accordance with standard actuarial

Table 1. *Expectation of tumourless life and standard error for bi-weekly tar experiment C 51—based on all tumours*

Week ual	Week Working	Animals exposed to risk (N)	Animals getting tumours	q_x	p_x	l_x	S_x	N_xp_x	q_x/N_xp_x	$(S_{x+1}/l_0^2)(q_x/N_xp_x)$
9	0	41	1	0·0244	0·9756	10,000	—	40	0·000610	0·02147
0	1	37·5	6	0·1600	0·8400	9756	59,320	31·5	0·005079	0·12477
1	2	30	2	0·0667	0·9333	8195	49,564	28	0·002382	0·04077
2	3	28	1	0·0357	0·9643	7648	41,369	27	0·001322	0·01503
3	4	27	4	0·1481	0·8519	7375	33,721	23	0·006439	0·04469
4	5	23	2	0·0870	0·9130	6283	26,346	21	0·004143	0·01668
5	6	20·5	5	0·2439	0·7561	5736	20,063	15·5	0·015735	0·03230
6	7	14	6	0·4286	0·5714	4337	14,327	8	0·053575	0·05347
7	8	7	0	0	1·0	2478	9990	7	0	0
8	9	6·5	1	0·1538	0·8462	2478	7512	5·5	0·027964	0·00709
9	10	5	1	0·2000	0·8000	2097	5034	4	0·050000	0·00431
0	11	4	2	0·5000	0·5000	1678	2937	2	0·250000	0·00396
1	12	2	1	0·5000	0·5000	839	1259	1	0·50000	0·00088
2	13	1	1	1·0000	0·0000	420	420	0	—	0

$E=9+6{\cdot}43=15{\cdot}43$ $\qquad V(E)=0{\cdot}36542$ $\qquad \sigma_E=0{\cdot}605$

The exposed to risk were obtained by deducting from the animals surviving tumourless at the beginning of each week half the deaths among tumourless animals in that week. The standard error has been calculated using both formulae (10) and (11). The results differ by only 5 in the third decimal place,

notation. The last week at the beginning of which no tumours had occurred was the 10th or week 9; this is called week 0 for purposes of calculation. The expectation from this point of time onwards can then be calculated and 9 added to the results. The relevant data and necessary calculations are shown in Table 1.

REFERENCE

IRWIN, J. O. (1946). *J. Hyg., Camb.*, **44**, 366.

(*MS. received for publication* 23. II. 49.—Ed.)