THE CASE OF ZERO SURVIVORS IN PROBIT ASSAYS*

* Appendix to: Bliss, C.I. (1935) The calculation of the dosage-mortality curve. *Annals of Applied Biology*, 22: 134-167.

V. Appendix. The case of zero survivors, by R. A. Fisher.

The equations derived from the theory of large samples appropriate for plotting the points on the probit diagram, namely

$$q=\frac{s}{n}$$

and

$$\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{1}{2}t^2}\,dt=q,$$

give, for experiments with no survivors, $x=\infty$, with weight

$$\frac{z^2}{pq}\to zx\to 0.$$

It is evident that such values cannot, in this form, be used in fitting the regression line, and that the theory of large samples has broken down, as was to be expected, when the number in the class of survivors is small. A more exact treatment is necessary for such cases, and this is supplied by the Method of Maximum Likelihood.

If p is the probability of death, and q of survival, in any experiment, the probability that s survive out of n tested is

$$\frac{n!}{s!\,(n-s)!}\,p^{n-s}\,q^s. \qquad \text{......(I)}$$

In the method of maximum likelihood, we take the logarithm of the aggregate probability of all the experimental data, for any assigned series of probabilities of survival represented by the regression line, and estimate the position of the regression line by making this logarithm a maximum. This amounts to equating to zero the sum for the different experiments of the differential coefficients with respect to the value of x assigned. The exact form of the differential coefficient of (I) with respect to p is

$$\frac{n-s}{p}-\frac{s}{q}=\frac{qn-s}{pq}.$$

With respect to the probit value x, the differential coefficient involves also the factor dp/dx, and becomes

$$(qn-s)\frac{z}{pq}. \qquad \text{......(II)}$$

Now when both s and $n-s$ are so large that the distribution of s may be treated as normal, the factor $(qn-s)$, which is n times the difference between the proportion of survivors expected and observed, is taken to be proportional to the difference between the probit values expected and observed, according to the formula

$$(qn-s)=n\,(x-X)\,\frac{dp}{dx}=n\,(x-X)\,z, \qquad \ldots\ldots\text{(III)}$$

where X is the probit value expected, and x that observed. In such cases the equation for maximum likelihood is made up of such terms as

$$(x-X)\,\frac{nz^2}{pq},$$

and its solution consists merely in fitting the expected values, X, by least squares to observed values, x, obtained from each experiment, giving each observational point a weight nz^2/pq.

When, however, q is so small that s can frequently take values such as 0, 1, or 2, the equation (III) is not a satisfactory approximation, as is evident when $s=0$, for then x is infinite, while a finite value will be obtained from equation (III). If we write

$$n\,(x'-X)\,z=qn-s, \qquad \ldots\ldots\text{(IV)}$$

then x' is a fictitious deviate, which, if assigned to any experiment with no survivors, will allow that experiment to exert its proper influence on the regression line. It will be observed that x' is a function not only of an observed frequency s/n, but also of X, the corresponding point on the regression line. It is only fictitious in the sense that it is not calculated from the result of just a single experiment, but requires also a knowledge of the expected value X inferred by fitting the regression line to other experiments. When $s=0$, $(x'-X)$ is always positive, so that the fictitious frequency to which x' corresponds is always less than that expected, as is evidently proper when the observed frequency is zero. The fictitious value x', if used with its proper weight in recalculating a regression line of which an approximate estimate has already been made, will then allow experiments with few or no survivors to exert their proper influence in adjusting the line. It is of some importance to take this step, since the omission of experiments merely because they show no survivors must constantly bias our estimates in the sense of exaggerating the number of survivors to be expected.

When $s=0$, the value of x' depends only on X, though, of course, the weight assigned to the observation depends also on n, the whole number tested, equation (IV) becoming

$$x'-X=\frac{q}{z}.$$

These values are shown in Table II.

Table II.

Probit values when 100 per cent. mortality is observed experimentally. The provisional (graphic) dosage-mortality line, based on probits for dosages which were survived by one or more individuals, is extended to cover dosages from which no survivors were observed. The expected probit value indicated by the provisional line at each such dosage is then entered in column 1 and the correction in column 2 is added to it to give the value in probits for 0 survivors (column 3). When the provisional line has been read to 0·01 probits, the first differences in the last column are convenient for interpolation.

Curve value or probit for expected kill	Correction q/z	Probit for observed kill	First differences
5·5	0·8764	6·3764	466
5·6	0·8230	6·4230	519
5·7	0·7749	6·4749	564
5·8	0·7313	6·5313	604
5·9	0·6917	6·5917	640
6·0	0·6557	6·6557	670
6·1	0·6227	6·7227	699
6·2	0·5926	6·7926	723
6·3	0·5649	6·8649	745
6·4	0·5394	6·9394	764
6·5	0·5158	7·0158	782
6·6	0·4940	7·0940	799
6·7	0·4739	7·1739	812
6·8	0·4551	7·2551	825
6·9	0·4376	7·3376	838
7·0	0·4214	7·4214	848
7·1	0·4062	7·5062	857
7·2	0·3919	7·5919	867
7·3	0·3786	7·6786	874
7·4	0·3660	7·7660	883
7·5	0·3543	7·8543	889
7·6	0·3432	7·9432	895
7·7	0·3327	8·0327	901
7·8	0·3228	8·1228	906
7·9	0·3134	8·2134	912
8·0	0·3046	8·3046	916
8·1	0·2962	8·3962	920
8·2	0·2882	8·4882	924
8·3	0·2806	8·5806	928
8·4	0·2734	8·6734	931
8·5	0·2665	8·7665	935
8·6	0·2600	8·8600	938
8·7	0·2538	8·9538	940
8·8	0·2478	9·0478	943
8·9	0·2421	9·1421	