Estimating the date of infection from individual response times

By G. G. MEYNELL

Guinness-Lister Research Unit, Lister Institute of Preventive Medicine,
Chelsea Bridge Road, London, S.W.1

AND TREVOR WILLIAMS

Department of Mathematics, Duke University, Durham,
North Carolina, 27706, U.S.A.

It is characteristic of both naturally occurring and experimental infections that the affected individuals do not fall ill or die at the same time. If one defines the 'response time' of an individual as the interval between the earliest date on which he could have been exposed to infection (as by eating contaminated food) and the date on which he fell ill, then the distribution of individual response times is always skewed with a long tail to the right. The true distribution has often been taken as log-normal, since probit proportion of responses plotted against logarithm of time since exposure approximates to a straight line (Sartwell, 1950, 1952; Meynell & Meynell, 1958; Meynell, 1963). Sartwell (1966) pointed out that, if the true distribution is indeed log-normal, an unknown date of exposure, a_L, can be estimated from the dates of the individual responses by the method of quantiles (Aitchison & Brown, 1963, §6.24). This is so but, owing to the actual distributions observed in practice, the earliest date of exposure can be equally well estimated by another method, and it is shown here that the two estimates necessarily disagree.

Assuming first that the true distribution is log-normal, then it is clear from Fig. 1 that, because the plot of the distribution gives a straight line, the log individual response times corresponding to, say, 10% and 90% responses are symmetrically placed about the median response time for 50% responses. However, this will only be so when the earliest date of exposure is correctly chosen for, if this is taken as either before or after the real date, the corresponding plots in Fig. 1 are convex or concave upwards (see Aitchison & Brown, 1963, Fig. 6.3). In general, therefore, if b_2 is the calendar date of the median response time and b_1, b_3 are the dates corresponding to per cent responses, q and $100-q$, then the earliest date of exposure, a_L, is the solution of the equation

$$\log (b_2 - a_L) - \log (b_1 - a_L) = \log (b_3 - a_L) - \log (b_2 - a_L),$$

which is readily seen to be

$$a_L = \frac{b_1 b_3 - b_2{}^2}{b_1 + b_3 - 2 b_2}. \tag{1}$$

The method can be illustrated by experiment 30 of Martin (1946) in which mice

9

were injected intravenously with *Mycobacterium tuberculosis*. The numbers of mice found dead on successive days were 1, 0, 2, 6, 6, 6, 4, 5, 4, 0, 0, 1, 0, 0, 1; the first death occurred 15 days after challenge; and all the inoculated mice died. Suppose all that is known is the number of mice dying on a given day of the month. If the first mouse is assumed to have died on the 20th, a plot of proportion of mice dying against calendar date shows the median date (b_2) to be 24·6 and the dates (b_1, b_3) corresponding to 10 % and 90 % deaths to be 21·8 and 28·2. Hence from (1)

$$a_L = 12.$$

That is, the estimated date of infection was the 12th compared to the assumed date, the 5th.*
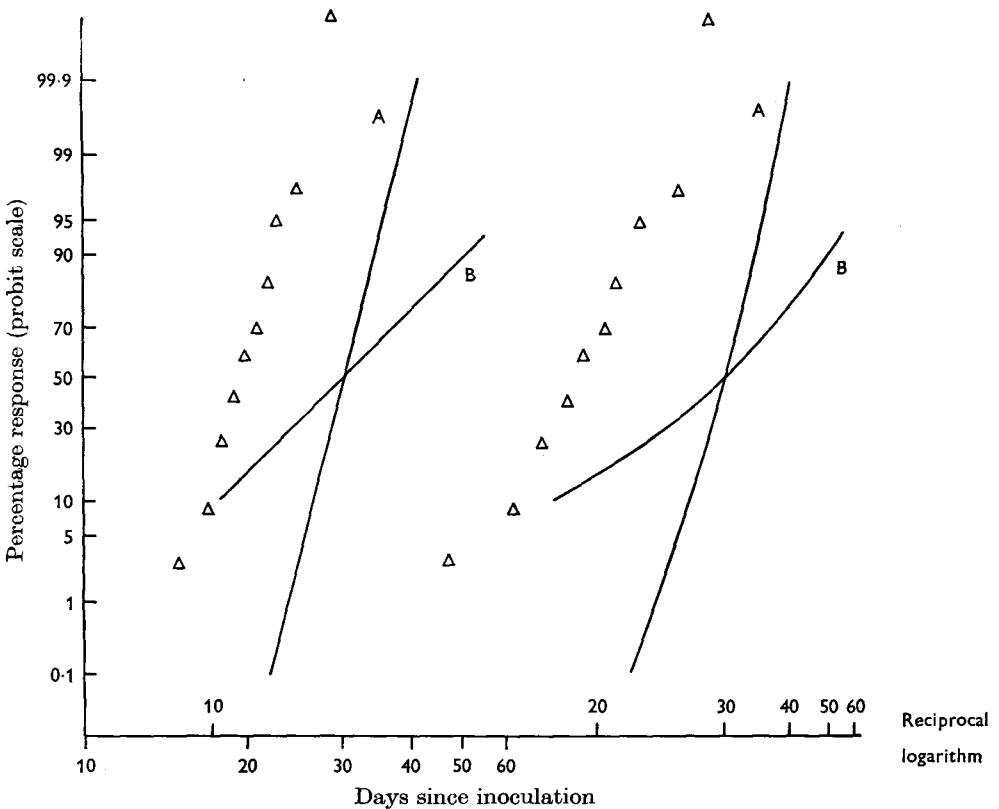

Fig. 1. Three distributions of response times plotted with either logarithm or recipro-cal of time since inoculation (curves on left and right of figure, respectively). Ordin-ate: percentage responses, on probit scale. Abscissa: logarithmic time below, reciprocal time above. $\triangle$, Experiment 20 of Martin (1946). Curves A and B, log-normal distributions with dispersion factors of 1·1 and 1·5 respectively.

* Even when the true distribution is log-normal, the estimation of a is recognized as being imprecise (Aitchison & Brown, 1963). In using quantiles, the values of b_1, b_2, and b_3 were obtained from a curve fitted to the observations. However, a can also be estimated directly from the data by Cohen's method (Aitchison & Brown, 1963, § 6.22). This was done here, but, although Martin's experiment was deliberately selected as an example because it appeared log-normal on probit paper, the estimate of a was wildly inaccurate, namely, more than 210 days before the first death.

It will be seen that the estimation of a_L depends, not upon the logarithm of the individual response time being normally distributed when measured from the correct time origin, but merely upon it being *symmetrically* distributed. The logarithm is far from being the only suitable function for, in many infections, the reciprocal also appears to be symmetrically distributed (Cavalli & Magni, 1943; Bryan, 1957). Host-pathogen systems are highly unlikely to fall into two fundamentally different classes, corresponding to the two transformations. It seems far more probable that many data can be fitted equally well in either way. On examination, this proved to be so, the reason being that, although response times are distributed, they always fall fairly close to the mean. Assuming a log-normal distribution, the degree of scatter can be expressed by the 'dispersion factor' (Sartwell, 1950) = time corresponding to 84% response/median response time. Its minimum value is 1·0, observed if all the hosts respond simultaneously, and the observed value rarely exceeds 1·5 in either epidemics (Sartwell, 1950, 1952, 1966; Williams, 1965) or experimental infections, and is often less. Given such small values, it is apparent from Fig. 1 that the corresponding log-normal distributions will appear virtually linear when plotted against reciprocal time, and, furthermore, that it will be almost impossible to determine which distribution truly describes the observations, remembering that these are always somewhat erratic and that the sample sizes tend to be small. Assuming the reciprocal of individual response time to be symmetrically distributed and the time of infection, a_R, correctly chosen, we have

$$\frac{1}{b_2 - a_R} - \frac{1}{b_1 - a_R} = \frac{1}{b_3 - a_R} - \frac{1}{b_2 - a_R}$$

whose solution is

$$a_R = \frac{2b_1 b_3 - b_2(b_1 + b_3)}{b_1 + b_3 - 2b_2}. \tag{2}$$

In Martin's experiment, $a_R = -0.6$. That is, the date of infection is given as 0·6 days before the 1st of the month.

Neither estimate is accurate to a useful degree in this example, but, quite apart from this, the whole procedure is open to a fundamental objection. We have seen that either the logarithm or the reciprocal of time can be used to estimate a. However, comparison of equations (1) and (2) shows

$$a_R = 2a_L - b_2.$$

In other words, a_R and a_L must always differ. The objection to these procedures is, therefore, that although either method is valid in itself, neither is of practical value because there is at present no reason *a priori* to accept one rather than the other.

SUMMARY

In principle, an unknown date of infection can be estimated from individual response times, provided some function of these with suitable origin is symmetrically distributed. Observed times are always skewly distributed, and either logarithm or reciprocal of time can be used to produce symmetry. Either is equally

justifiable but the resulting estimates are not only very imprecise but are also inconsistent, so that neither is of practical value.

REFERENCES

AITCHISON, J. & BROWN, J. A. C. (1963). *The Lognormal Distribution*. Cambridge University Press.

BRYAN, W. R. (1957). Interpretation of host response in quantitative studies on animal viruses. *Ann. N.Y. Acad. Sci.* **69**, 698.

CAVALLI, L. & MAGNI, G. (1943). Quantitative Untersuchungen über die Virulenz. III. Mitteilung: Analyse der Häufigkeitsverteilung der Absterbezeiten von infizierten Mäusen. *Zentbl. Bakt. ParasitKde, Abt. 1, Orig.* **150**, 353.

MARTIN, A. R. (1946). The use of mice in the examination of drugs for chemotherapeutic activity against *Mycobacterium tuberculosis*. *J. Path. Bact.* **58**, 580.

MEYNELL, G. G. (1963). Interpretation of distributions of individual response times in microbial infections. *Nature, Lond.* **198**, 970.

MEYNELL, G. G. & MEYNELL, E. W. (1958). The growth of micro-organisms *in vivo* with particular reference to the relation between dose and latent period. *J. Hyg., Camb.* **56**, 323.

SARTWELL, P. E. (1950). The distribution of incubation periods of infectious disease. *Am. J. Hyg.* **51**, 310.

SARTWELL, P. E. (1952). The incubation period of poliomyelitis. *Am. J. publ. Hlth* **42**, 1403.

SARTWELL, P. E. (1966). The incubation period and the dynamics of infectious disease. *Am. J. Epidem.* **83**, 204.

WILLIAMS, T. (1965). The basic birth-death model for microbial infections. *Jl R. statist. Soc.* **27**, 338.