J. Hyg., Camb. (1963), **61**, 401
Printed in Great Britain

Optimum moving averages for the estimation of median effective dose in bioassay

BY B. M. BENNETT
University of Washington, Seattle

(*Received* 27 *April* 1963)

1. INTRODUCTION AND STATEMENT OF PROBLEMS

Suppose that $x_1, x_2, \ldots, x_t$ represent the numbers of subjects in which an all-or-none response (e.g. death) is observed to occur amongst the $n_1, n_2, \ldots, n_t$ subjects independently tested at t different dose levels $d_1, d_2, \ldots, d_t$ of a certain substance (e.g. drug or hormone). The problems of quantal response in bioassay may be formulated as follows. If $F(d)$ represents an assumed distribution, or the proportion π expected to react at a dose d or less, then the x's each have independent binomical distributions

$$\frac{n_i!}{x_i!\,(n_i - x_i)!}\,\pi_i^{x_i}(1-\pi_i)^{n_i - x_i} \tag{1}$$

for $i = 1, 2, \ldots, t$ and $0 \leqslant x_i \leqslant n_i$.

Principal interest has been shown thus far in estimation of the parameters and the dose-response relation from the assumed distribution

$$F(d) = \frac{1}{\sqrt{2\pi^{\sigma}}}\int_{-\infty}^{d} \exp\left[-\frac{1}{2\sigma^2}(u-\mu)^2\right]du, \tag{2}$$

i.e. the cumulative normal distribution (e.g. Finney, 1952), and also from the 'logistic' distribution

$$F(d) = [1+\exp[-(\alpha+\beta d)]^{-1}] \tag{3}$$

(e.g. Berkson, 1953).

When nothing is assumed about the nature of the probabilities in the form of a specified response distribution $F(d)$ as above, Thompson (1947) proposed the use of 'moving averages' as a method of estimating the median effective or lethal dose (= LD 50). An (unweighted) 'moving average' of span k is defined as the successive sums

$$p'_i = \frac{1}{k}\sum_{r=0}^{k-1}\left(\frac{x_{i+r}}{n_{i+r}}\right) = \frac{1}{k}\sum_{r=0}^{k-1} p_{i+r} \tag{4}$$

in terms of the sample proportions p, and an associated dose

$$d'_i = \frac{1}{k}\left(\sum_{r=0}^{k-1} d_{i+r}\right). \tag{5}$$

In order to estimate LD 50 Thompson used ordinary linear interpolation between consecutive values of the moving average p' on either side of 0·50. Finney (1952, ch. 20) has discussed the efficiency of these unweighted moving averages with reference to the integrated normal distribution (2).

The present author (1952) has made one modification in moving averages by introducing weights proportional to the n's for odd-numbered spans of 3, 5, etc. Thus, for example, for three-term moving averages

$$p'_i = \left(p_{i-1}+\frac{n_i}{\bar{n}_i}p_i+p_{i+1}\right)\Big/\left(2+\frac{n_i}{\bar{n}_i}\right) \tag{6}$$

and a corresponding dose d'_i in terms of the harmonic mean $\bar{n}_i$ defined by

$$\frac{1}{\bar{n}_i} = \frac{1}{2}\left(\frac{1}{n_{i-1}}+\frac{1}{n_{i+1}}\right),$$

and the author has also discussed the efficiency of this form of moving averages in estimating LD 50. If the n's all happen to be equal, then the average (6) above coincides with the averages defined by Thompson.

It is the purpose of this paper to present some further theoretical aspects of determining new optimum forms of moving averages in estimation of LD 50 together with their approximate sampling errors.

2. GENERALITIES ON MOVING AVERAGES

For the $(2k+1)$ sample proportions $\{p_{i+r}\}(r = 0, \pm 1, \ldots, \pm k)$, consider the minimization of the weighted sums of squares

$$S_i = \sum_{r=-k}^{+k} \lambda_{i+r}(p_{i+r}-\pi_i)^2 \tag{7}$$

with respect to π_i for each $i = k+1, \ldots, (t-k)$, and subject to the further condition

$$\sum_{r=-k}^{+k} \lambda_{i+r} = 1.$$

This minimization results in estimates of the form $p'_i = \hat{\pi}_i = \sum_{r=-k}^{+k} \lambda_{i+r}\, p_{i+r}$, which will be called a moving average of span $(2k+1)$.

It should be mentioned that the condition $\sum_r \lambda_{i+r} = 1$ results from the natural requirement that if the observed proportions happen to coincide, i.e. if $p_{i+r} = p_i^0$, $r = 0, \pm 1, \ldots, \pm k$, then $p'_i = p_i^0$ as a reasonable condition of the moving average.

If then the variance of each p'_i

$$V(p'_i) = \sum_{r=-k}^{+k} \lambda_{i+r}^2 \frac{\pi_{i+r}(1-\pi_{i+r})}{n_{i+r}} \tag{8}$$

is minimized for variation in the λ's, it is easy to verify that this occurs when

$$\hat{\lambda}_{i+r} = \frac{n_{i+r}\pi_i}{\pi_{i+r}(1-\pi_{i+r})}\Big/\sum_{r=-k}^{+k}\frac{n_{i+r}}{\pi_{i+r}(1-\pi_{i+r})}, \tag{9}$$

i.e. the weights are proportional to the reciprocals of the (binomial) weights for each sample proportion. The resulting minimum variance is

$$V(p'_i) = \left[\sum_{r=-k}^{+k}\frac{n_{i+r}}{\pi_{i+r}(1-\pi_{i+r})}\right]^{-1}.$$

https://doi.org/10.1017/S002217240002101X Published online by Cambridge University Press

If the true proportions π are known, then LD 50 is to be estimated by linear interpolation between the two consecutive p''s on either side of 0·5 in terms of the transformed doses

$$d'_i = \sum_{r=-k}^{+k} \hat{\lambda}_{i+r} d_{i+r}.$$

In particular, if the original doses are equally spaced, i.e. if $d_i = d_0 + (i-1)d$ in terms of an initial (log) dose level $(= d_0)$ and uniform spacing d, the resulting spacing for the d''s is

$$d'_i = d_0 + \left\{(i-1) + \left(\sum_{r=-k}^{+k} r\lambda_{i+r}\right)\right\} d \tag{10}$$

which will be in fact equal only in case of a completely symmetric set of the weights, i.e. $\lambda_{i-r} = \lambda_{i+r}$, for $r = 1, 2, \ldots, k$. This can occur when equal numbers are used with a symmetric dose-response or in terms of the population proportions, $\pi_{i-r} = \pi_{i+r}$.

3. UNBIASED MOVING AVERAGES

If now it is required only that each successive p'_i be an unbiased estimate of the corresponding π_i, i.e.

$$E(p'_i) = \sum_{r=-k}^{+k} \lambda_{i+r}\pi_{i+r} = \pi_i,$$

the weights resulting from minimizing the variance (8) subject to the condition of unbiasedness are

$$\hat{\lambda}_{i+r} = \frac{n_{i+r}\pi_i}{(1-\pi_{i+r})} \Big/ \left[\sum_{r=-k}^{+k} n_{i+r}\left(\frac{\pi_{i+r}}{1-\pi_{i+r}}\right)\right] \tag{11}$$

$$= \left(\frac{\pi_i}{\pi_{i+r}}\right) n_{i+r}\, \mathrm{e}^{Y_{i+r}} \Big/ \left[\sum_{r=-k}^{k} n_{i+r}\left(\frac{\pi_{i+r}}{1-\pi_{i+r}}\right)\right]$$

in terms of the usual 'logit' transformation $Y_i = \log_e[\pi_i(1-\pi_i)^{-1}]$. The minimum variance is then

$$V(p'_i) = (\pi_i^2)\left[\sum_{r=-k}^{k} n_{i+r}\left(\frac{\pi_{i+r}}{1-\pi_{i+r}}\right)\right]^{-1} = (\pi_i^2)\left[\sum_{-k}^{k} n_{i+r}\,\mathrm{e}^{Y_{i+r}}\right]^{-1}.$$

If the requirements on the λ's of being both unbiased and such that

$$\sum_r \lambda_{i+r} = 1$$

are superimposed, then minimum variance is attained whenever

$$\hat{\lambda}_{i+r} = \frac{n_{i+r}}{\pi_{i+r}(1-\pi_{i+r})}\left(\frac{p_{ir}}{\Delta_i}\right), \tag{12}$$

where

$$p_{ir} = \left[\sum_{j=-k}^{+k} \frac{n_{i+j}(\pi_{i+j}-\pi_i)(\pi_{i+j}-\pi_{i+r})}{\pi_{i+j}(1-\pi_{i+j})}\right],$$

$$\Delta_i = \left[\sum_r \frac{n_{i+r}}{\pi_{i+r}(1-\pi_{i+r})}\right]\left[\sum_r \frac{n_{i+r}\pi_{i+r}}{(1-\pi_{i+r})}\right] - \left[\sum_r \frac{n_{i+r}}{(1-\pi_{i+r})}\right]^2. \tag{13}$$

https://doi.org/10.1017/S002217240002101X Published online by Cambridge University Press

It is seen that the weights or coefficients $\hat{\lambda}$ obtained in equations (11) and (12) are parametric multiples of the binomial weights in (9).

In view then of the sampling complications of weights based on equations (11) or (12), it is felt that successive moving averages based on sample estimates of the binomial weights (9) are to be considered adequate in estimation of LD 50. These estimates together with their approximate standard errors will now be derived.

4. SAMPLING THEORY WHEN BINOMIAL WEIGHTS ARE USED

In applying the results of § 3 for the binomial weights (9), we replace the π's by their corresponding sample estimates p. The moving averages p' are then computed from the sample weights

$$\hat{\lambda}_{i+r} = \frac{n_{i+r}}{p_{i+r}(1-p_{i+r})} \Bigg/ \left[\sum_{r=-k}^{k} \frac{n_{i+r}}{p_{i+r}(1-p_{i+r})}\right] \tag{14}$$

excluding any observed zero or one values for the sample p's. In these latter cases the value of the observed numbers x should be replaced by 1 or $(n-1)$, respectively in the calculations.

If then two successive moving average proportions p'_i, p'_{i+1} are such that $p'_i < 0{\cdot}5 < p'_{i+1}$, the estimate of LD 50 obtained by linear interpolation is

$$m' = d'_i + \frac{(0{\cdot}5 - p'_i)}{(p'_{i+1} - p'_i)}(d'_{i+1} - d'_i) \tag{15}$$

with variance

$$V(m') = (d'_{i+1} - d'_i)^2 V\left(\frac{0{\cdot}5 - p'_i}{p'_{i+1} - p'_i}\right). \tag{16}$$

Using the formula for the approximate variance of a ratio of two random variables, we obtain

$$V(m) \simeq \frac{(d'_{i+1} - d'_i)^2}{(\pi'_{i+1} - \pi'_i)^2}[(1-\tau)^2 V(p'_i) + 2\tau(1-\tau)C(p'_i, p'_{i+1}) + \tau^2 V(p'_{i+1})], \tag{17}$$

where π'_i are the corresponding moving averages of the true proportions π again estimated from the sample and

$$V(p'_i) = \left[\sum_{r=-k}^{+k} \frac{n_{i+r}}{\pi_{i+r}(1-\pi_{i+r})}\right]^{-1}, \tag{18}$$

$$C(p'_i, p'_{i+1}) = \left[\sum_{r=-(k-1)}^{k} \frac{n_{i+r}}{\pi_{i+r}(1-\pi_{i+r})}\right] \Bigg/ \left[\sum_{r=-k}^{k} \frac{n_{i+r}}{\pi_{i+r}(1-\pi_{i+r})}\right]\left[\sum_{-k}^{k} \frac{n_{i+r+1}}{\pi_{i+r+1}(1-\pi_{i+r+1})}\right]$$

if we denote the fraction $\tau = (0{\cdot}5 - p'_i)/(p'_{i+1} - p'_i)$.

5. OPTIMUM MOVING AVERAGES BASED ON ANGULAR TRANSFORMATION

Finally, it is of interest to point out an optimum form of moving average based on the arc-sine transformation of proportions (e.g. Snedecor, 1956, p. 318), which is well known for the property of approximately stabilizing the variance. In terms of the sample proportions $p_i = x_i/n_i \, (i = 1, \ldots, c)$, the arc-sine transformation

$$y_i = \sin^{-1}\sqrt{p_i}$$

is such that the resulting y's have approximately constant variance $V(y_i) = c^2/n_i$, where

$$c^2 = 0{\cdot}25 \quad \text{if } y \text{ in radians}$$
$$= 821 \quad \text{if } y \text{ in degrees.}$$

Moving averages on the y's of the form $y'_i = \sum_{r=-k}^{+k} \lambda_{i+r} y_{i+r}$ attain minimum variance when

$$\hat{\lambda}_{i+r} = n_{i+r} \Big/ \left[\sum_{r=-k}^{k} n_{i+r}\right],$$

and in this case

$$V(y'_i) = c^2 \Big/ \left[\sum_{r=-k}^{k} n_{i+r}\right]. \tag{19}$$

The corresponding doses are

$$d_i^* = \left(\sum_{r=-k}^{k} n_{i+r} d_{i+r}\right) \Big/ \left[\sum_{r=-k}^{k} n_{i+r}\right] \tag{20}$$

which, in case the d's are equally spaced $d_i = d_0 + (i-1)\,d$ with interval d, reduce to

$$d_i^* = d_0 + \left\{(i-1) + \frac{(\Sigma r n_{i+r})}{(\Sigma n_{i+r})}\right\} d. \tag{21}$$

If now two successive values y'_i, y'_{i+1} of the sequence of moving averages are such that $y'_i < \tfrac{1}{4}\pi < y'_{i+1}$, then the estimate m^* of LD 50 will be

$$m^* = d_i^* + \frac{(\tfrac{1}{4}\pi - y'_i)}{(y'_{i+1} - y'_i)}(d_{i+1}^* - d_i^*) \tag{22}$$

with approximate variance

$$V(m^*) = \frac{(d_{i+1}^* - d_i^*)^2}{(\theta'_{i+1} - \theta'_i)^2}\left[(1-\nu)^2\, V(y'_i) + 2\nu(1-\nu)\, C(y'_i, y'_{i+1}) + \nu^2 V(y'_{i+1})\right], \tag{23}$$

where

$$\theta_i = \sin^{-1}\sqrt{\pi_i},$$

$$V(y'_i) = (c)^2 \Big/ \left(\sum_{r=-k}^{k} n_{i+r}\right), \tag{24}$$

$$C(y'_i, y'_{i+1}) = (c^2)\left[\sum_{r=-(k-1)}^{k} n_{i+r}\right] \Big/ \left[\sum_{r=-k}^{k} n_{i+r}\right]\left[\sum_{r=-k}^{k} n_{i+r+1}\right]$$

and θ'_i is the weighted average $= (\Sigma n_{i+r}\theta_{i+r})/(\Sigma n_{i+r})$, and ν is the fraction $\nu = (\tfrac{1}{4}\pi - \theta'_i)/(\theta'_{i+1} - \theta'_i)$.

Table 1. *Toxicity of rotenone* (*Finney*, 1947)

(D) concentration mg./l.	$\log\left(\frac{D}{D_0}\right) = d_i$	No. of insects n_i	No. affected x_i	$100\, p_i$	Degrees y_i	y'_i	d_i^*
$2{\cdot}6 = D_0$	0·0000	50	6	12·0	20·27	—	—
3·8	0·1644	48	16	33·3	35·24	33·56	0·1482
5·1	0·2923	46	24	52·2	46·26	49·35	0·3107
7·7	0·4713	49	42	85·7	67·78	61·63	0·4566
10·2	0·5933	50	44	88·0	69·73	—	—

https://doi.org/10.1017/S002217240002101X Published online by Cambridge University Press

6. EXAMPLE

The following example of an optimum three-term moving average using the arc-sine transformation is based on the data (Table 1) on toxicity of rotenone when sprayed on *Macrosiphoniella sanborni* in batches of approximately fifty insects each (Finney, 1947, p. 26).

Using equations (22) and (23), we obtain the estimates

$$m^* = \log_{10}(\text{LD } 50) = \log 2{\cdot}6 + 0{\cdot}1482 + \frac{(45 - 33{\cdot}56)}{(49{\cdot}35 - 33{\cdot}56)}(0{\cdot}1625)$$

$$= 0{\cdot}4150 + 0{\cdot}1482 + (0{\cdot}7245)(0{\cdot}1625) = 0{\cdot}6809,$$

$$V(m^*) = \frac{(0{\cdot}1625)^2}{(15{\cdot}79)^2}[(1 - 0{\cdot}7245)^2 V(y_i') + 2(0{\cdot}7245)(0{\cdot}2755)\, C(y_i', y_{i+1}) + (0{\cdot}7245)^2 V(y_{i+1})]$$

$$= \frac{(0{\cdot}02641)}{249{\cdot}32}(4{\cdot}9424) = 5{\cdot}235 \times 10^{-4}.$$

Table 2 compares this estimate and its variance with the corresponding maximum likelihood estimate (Finney, 1947) and the minimum logit χ^2 value (Berkson, 1953) with their respective variances.

Table 2. *Comparison of estimates of log (LD 50)*

Method	Estimate	Variance ($\times 10^4$)
Maximum likelihood (probits)	0·6862	4·849
Minimum (logit) χ^2	0·6848	5·215
Optimum moving average (arc-sine)	0·6809	5·235

Table 2 demonstrates the relatively high efficiency of a three-term moving average based on the angular transformation when compared with an assumed cumulative normal (92·6 %) and the logit (99·6 %) in this example.

SUMMARY

Optimum forms of moving averages are derived for the estimation of LD 50 in the situation where no assumptions are made about the form of the dose-response distribution. The theory is also applied to uses of moving averages based on the angular transformation of the percentage response. A numerical example illustrates this application, and its results are compared with the corresponding probit and logit estimates.

REFERENCES

BENNETT, B. M. (1952). Estimation of LD 50 by moving averages. *J. Hyg., Camb.*, **50**, 157–64.

BERKSON, J. (1953). A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function. *J. Amer. statist. Ass.* **48**, 565–99.

FINNEY, D. J. (1947). *Probit Analysis*. Cambridge University Press.

FINNEY, D. J. (1952). *Statistical Method in Biological Assay*. Griffin and Co., Ltd.

SNEDECOR, G. W. (1956). *Statistical Methods*, 5th edition. Iowa State Press, Ames, Iowa.

THOMPSON, W. R. (1947). Use of moving averages and interpolation to estimate median-effective dose. *Bact. Rev.* **11**, 115–45.