THE CALCULATION OF THE TIME-MORTALITY CURVE

By C. I. BLISS

Galton Laboratory, University College, London[1]

Appendix by W. L. Stevens

Galton Laboratory, University College, London

(With 5 Text-figures)

CONTENTS

An earlier paper in this series has described statistical methods for calculating the dosage-mortality curve (2) on the assumption that it owes its characteristic sigmoid shape to the variation in susceptibility between the individuals of a population. If living organisms are immersed in a toxic solution or gas and the time recorded at which each individual is paralysed, under conditions of equal exposure we may also assume that the differences in the survival period are an indirect measure of the individual susceptibility to the poison. When the percentage of organisms which has reacted up to and including each successive observation is plotted against time, with most multicellular organisms the resulting time-mortality curve is similar in shape to the dosage-mortality curve.

[1] Manuscript completed at the Institute of Plant Protection, Lenin Academy of Agricultural Sciences, Leningrad, with the approval of the Director.

It is reasonable to suppose, therefore, that the variation in this measure of susceptibility may also be distributed normally and to apply the normal curve to the survival period. Similar methods of analysis are available for many different measurements of reaction time, some of which are illustrated in the numerical examples. The basic statistical procedures differ considerably from those described for computing the dosage-mortality curve, and usually are much simpler. The time-mortality curve is of interest primarily in comparison with equivalent curves obtained at different concentrations or dosages, by the use of different poisons, or under varying experimental conditions. Then it may

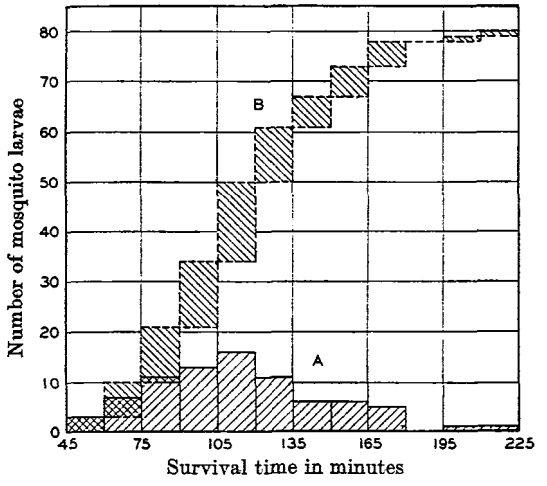

Fig. 1. Frequency distribution of the number of mosquito larvae surviving for different periods in a 5·1 millimolar solution of heptylic acid, showing the relation between the original or non-cumulative (*A*) and the cumulative (*B*) diagrams. Data in Table II.

become a valuable measure of toxicity and its correct computation essential for reaching reliable conclusions.

The methods of computation, for the most part, have been adapted from *Statistical Methods for Research Workers* by R. A. Fisher (11). For their application a calculating machine, Barlow's Tables (1), and a table of common logarithms are required.

I. The time-mortality curve

Time-mortality data may be defined as records of reaction time which can be reduced to a form showing the number of organisms which react to a toxicological stimulus in each of several successive periods of time. The record of a single experiment, therefore, can always be plotted in a form similar to curve *A* of Fig. 1. This figure, based on the data of

O'Kane *et al.* (14), shows the number of mosquito larvae in a total of 80 individuals which succumbed to a 5·1 millimolar concentration of heptylic acid in successive 15 min. periods. If the number of individuals in each successive interval is added to those which have already succumbed, the familiar "step diagram" (Fig. 1 *B*) is obtained, and when the points of intersection in each of these steps, transformed to the percentage of the total, is plotted against time, a sigmoid curve frequently is obtained quite similar in appearance to the dosage-mortality curve. Although time-mortality data can always be plotted as a cumulative curve which is typically sigmoid, dosage-mortality data cannot conversely be reduced to a non-cumulative frequency distribution, whether dosage is measured in time units or otherwise. Each observation in a dosage-mortality curve represents a different set of organisms, so that the individuals at the successive dosages are unrelated in susceptibility. It is possible for the mortality to be less at a longer than at a shorter period if by chance the first group of organisms should contain a large proportion of more resistant individuals. But this can never happen in a time-mortality curve since all of the points are merely different observations on a single set of animals and the percentage mortality at a given time can never be less than that recorded earlier in the experiment. The successive observations are strongly correlated with each other and in consequence the methods of computation that have been described for the dosage-mortality curve are not applicable here.

Although the dosage-mortality curve is the better standard of reference for most toxicological investigations, there are many instances where time-mortality curves may be used advantageously, either to give supplementary information or the same information more efficiently. The following examples will show how varied time-mortality data can be. Many investigators have timed the survival of organisms immersed in a toxic solution, or, more accurately, the period preceding some well-defined paralytic response. In this case the record of time included both the time for the fixation of an effective dose and the period required for it to produce its biological effect after fixation. Crustaceans (4), mosquito larvae (14) and goldfish (13) have been used recently in experiments of this type. The response of adult *Drosophila* has been timed in alcohol vapour (8) and in hydrocyanic acid gas (3), experiments which differ only in the use of a gaseous rather than a liquid medium. In other cases the poison has been applied as a contact spray to house-flies (16), injected into silkworm larvae (7), and fed in measured amounts to many kinds of mandibulate insects, and the preparalytic or survival period then

included the time for the penetration or translocation of given dosages from the point of application to the seat of action as well as the time to produce the effect that was recorded. Other investigators have timed the recovery of *Limnoria* after immersion in solutions of phenol and of other poisons (17) or of *Drosophila* after fumigation with HCN (6). Finally there have been many experiments in which the time was measured from the initial exposure to poisoned food rather than from the start or completion of feeding, the so-called "cage tests". In these and similar instances the basic measurements of time have involved the time-mortality curve, although the methods of computation have varied considerably and frequently insufficient data have been recorded to compute the curve. It is not suggested that it is always necessary to compute the time-mortality curve in full, but unless its characteristics are known and considered in the design of each experiment, there is no certainty that the results are unbiased.

Some limitations caused by not computing the time-mortality curve may be examined. Several investigators have based their conclusions upon the median reaction time, determined either during the experiment by recording the reaction time only of the median one or two individuals in each set or later by graphic interpolation from the middle section of an incomplete time-mortality curve in its sigmoid form. Although neither method is precise, especially when the total number in a distribution is small, both estimate without systematic error the time for 50% of the organisms to respond. However, they do not provide an estimate of the accuracy of each median nor is it possible to estimate other levels of effectiveness, such as the time for 95% to react.

Other investigators have computed the mean reaction time from their original records without determining whether these original units of time were, in fact, distributed normally. In such cases as have been tested, a normal distribution of the original units has been exceptional, although when the time units were transformed to logarithms or, less frequently, to reciprocals (rates), normal distributions were obtained in most cases. Consistent biological indices may then be computed from the transformed times by means of familiar statistical methods. When the distribution is not normal, the mean reaction time does not represent a constant level of effect, as is the case with the median, but a response displaced to a varying extent from the median. The extent of this displacement has been estimated in several cases in which the logarithm of the reaction time was distributed normally, by comparing the logarithm of the mean time with the mean of the logarithm time, which coincides presumably

with the median. From the ratio of the difference between the two means to the standard deviation of the logarithmic distribution, the percentage of the response represented by the arithmetic mean has been read from Pearson's Table II(15) for the cases listed in Table I. From the last column it is apparent that the mean did not represent a constant level of effect for different phenomena, but varied even for the same type of response at different concentrations of a single poison.

The standard deviation computed from a time-mortality distribution that is not normal loses most of its direct toxicological value. By means of probits or their equivalent a normal frequency distribution can be plotted as a straight line as will be described presently, but if the initial distribution is skewed to the left, as it frequently is, the observations at successive time intervals on the cumulative probit graph will not form a

Table I

Percentage response represented by the arithmetic mean of the observed reaction time when the logarithm of the time is found to be distributed normally. The data are from papers listed in the references under the number in the first column. The concentrations of the two heptylic acid solutions were (a) 5·1 and (b) 6·4 millimoles per l.

Source of data	Type of response	Insect	Poison	Mean log. min.	Log. mean min.	Standard deviation in log.	% response
(3)	Stupefaction	*Drosophila*	HCN	− 0·4050	− 0·3992	0·0701	53·3
(6)	Recovery	*Drosophila*	HCN	1·4776	1·4968	0·1349	55·7
(7)	Death	*Bombyx*	Sodium arsenate	2·4267	2·4293	0·0491	52·2
(14)	Paralysis	*Culex*	Heptylic acid (*a*)	2·0349	2·0555	0·1342	56·1
	(Death?)		(*b*)	1·6397	1·6496	0·0883	54·5

straight line but a curve that is convex upwards. The standard deviation for such a distribution, as computed from the original observations, gives the slope of the straight line that best fits this curving series of points. Since a straight line is not a satisfactory fit, it cannot be used to determine when any given percentage of individuals will succumb to the poison. However, if the survival times are first transformed to logarithms or to some other function of observed time and the points on the cumulative probit diagram then form a straight line, the standard deviation computed from these converted time units is itself of value. Since it is the slope of the best-fitting straight line, it can be used for computing the time at which any given percentage of individuals will succumb to the poison. In consequence, comparisons of different poisons or of different concentrations of the same poison need not be restricted to the time at which half of the organisms react but may be based with a calculable error upon a higher mortality.

It is not proposed that all records of the reaction time to toxicological stimuli can be reduced to normal distributions by relatively simple procedures. Thus, in a study of the resistance of *Drosophila* to alcohol by Crozier *et al.*(8), most of the records on time-to-death gave flat-topped distributions, quite different in appearance from normal curves. The authors suggest that this may be attributed to a product function of two independent types of variation, one a symmetrical distribution of the penetrability into the fly of alcohol vapour and the other a strongly skewed distribution of the resistance to absorbed alcohol. Although this is an interesting approach to the problem, the distributions given in their paper (Fig. 15) can be considered only as indicative, since each curve is based upon too few individuals for the differences from normality which they show to be significant. The five "flat-topped" distributions for younger flies, for example, could be fitted satisfactorily with the normal curve as tested both by graphic analysis and by the χ^2 test for the agreement of observed with expected frequencies. It would be necessary to use a larger number of individuals, ungrouped data, or both to establish these apparent departures from the normal distribution.

The reaction to a toxic stimulus probably always involves many components but this does not preclude either a relatively simple type of distribution or a change in its form at different levels of susceptibility. One of the first investigators to use methods such as are described here was W. P. Davey in a study of the prolongation of the life of *Tribolium confusum* by small doses of X-rays(9). Many of his distributions of longevity showed one or more sharp changes in slope when tested graphically with "probability paper", "breaks" which in Davey's opinion did not invalidate the method or his theory. If the time-to-death can be considered a function of the individual susceptibility to a poison and this susceptibility is distributed normally, then small dosages of any given poison must exist such that the distribution of time-to-death will be interrupted before all individuals have reacted due to the failure of the more resistant individuals to die. Since the response can change so markedly as a function of susceptibility, lesser qualitative changes in the nature of the response are to be expected, changes which may lead to discontinuities in the time-mortality curve. In order that the time-mortality curve may be an effective instrument in the analysis of toxicological phenomena, the methods for its study must be based upon the whole distribution of susceptibilities and expose their relation to the response. The most useful instrument for this purpose is graphic analysis.

II. COMPLETE TIME-MORTALITY CURVES

In contrast with the dosage-mortality curve, the computation of the time-mortality curve (unless truncated) is not based upon the provisional or graphic time-mortality line but upon the original distribution from which it has been derived. The purpose of graphic analysis is to determine which function of time is to be used in computing the curve and to insure that none of the observations included in the computation depart systematically from the curve. If such deviations from normality do occur or if the curve is incomplete, methods suitable for the truncated distribution must be used rather than those to be described here.

Graphic analysis

Of the different ways in which the data can be plotted, the cumulative rectilinear graph is the most useful. The diagram then shows not only whether a given function of time is distributed normally, but if it is normal, it gives directly a graphic estimate of the mean (or median) reaction time, the standard deviation, and the time at which any given percentage of organisms between 0 and 100 has reacted. Moreover, the variance of each estimate can be derived with a minimum of calculation. If only part of the organisms react due to an insufficiently toxic dose of poison, if the record is incomplete for any other reason, or if in the course of a single experiment there is a change in the nature of the response which affects the time of its occurrence, graphic analysis is essential for the derivation of unbiased statistical constants. In some cases the data may be sufficiently uniform that graphic estimates of the essential statistical constants will answer all of the requirements of the investigator.

In preparing time-mortality data for plotting, the first step is to transform the initial frequencies into cumulative percentages and these in turn into probits. The procedure depends partly upon the manner of recording the experiment. In some cases the number of "dead" individuals is recorded periodically from the start of the exposure until all have reacted to the poison. The record then shows directly (or by subtraction) the numbers of animals which have reacted in the time intervals between successive observations. The transformation of this original, non-cumulative frequency distribution to units suitable for plotting are illustrated by the example in Table II (Figs. 1 and 2, curve *A*). The class limits in each case show the elapsed time up to and including each upper limit (Table II, col. 1); the original frequencies give the number of larvae paralysed within each interval (col. 2 and Fig. 1*A*); cumulative frequencies are placed

opposite the respective class limits rather than between them (col. 3); percentages are derived from the cumulative frequencies (col. 4) and then converted to probits (col. 5) by means of the table given in an earlier paper(2). Instead of the more convenient probits, it is possible to use directly the deviates of the normal curve, such as are given in Table I of Pearson's Tables(15), the mathematical constants ($prf^{-1}p$ in Wright's

Table II

Graphic analysis and computation of the time-mortality curve from data grouped during the experiment, as shown by the survival time of larvae of Culex pipiens *in a solution of* 5·1 *millimoles per l. of heptylic acid; data of O'Kane* et al.(14). *See Fig.* 1 *and Fig.* 2, *curve A*

		Data for graphic analysis				Data for computed curve			
Class limits in min.	Larvae dead in each interval f	Larvae dead at end of each interval: No.	%	Probit	Log. of class limits	Mid-point of log. intervals x	fx	Grouping interval i	fi
45		—	—	—	1·653				
	3					1·715	5·145	0·125	0·375
60*		3	3·7	3·21	1·778				
	7					1·826	12·782	0·097	0·679
75		10	12·5	3·85	1·875				
	11					1·915	21·065	0·079	0·869
90		21	26·3	4·37	1·954				
	13					1·988	25·844	0·067	0·871
105		34	42·5	4·81	2·021				
	16					2·050	32·800	0·058	0·928
120		50	62·5	5·32	2·079				
	11					2·105	23·155	0·051	0·561
135		61	76·2	5·71	2·130				
	6					2·153	12·918	0·046	0·276
150		67	83·8	5·99	2·176				
	6					2·197	13·182	0·041	0·246
165		73	91·3	6·36	2·217				
	5					2·236	11·180	0·038	0·190
180		78	97·5	6·96	2·255				
	0					—	—	—	—
195		78	97·5	6·96	2·290				
	1					2·306	2·306	0·032	0·032
210		79	98·75	7·24	2·322				
	1					2·337	2·337	0·030	0·030
225		80	100	—	2·352				

* An extra observation at 55 min. is omitted here but included in curve *A* of Fig. 2.

terminology(19) or N.E.D. in Gaddum's terminology(12)) in the body of the table being read as positive when above 50 % and as negative when below 50 % from the corresponding proportionate (percentage ÷ 100) frequencies along the margins. By the same reasoning as in earlier papers(19,12, 2), that function of time is distributed normally against which the probit can be plotted as a straight line and usually it can be considered a more direct measure of the biological factors determining susceptibility than units of

time which are not so distributed. In the paper from which the example in Table II has been taken the logarithm of the time is shown to be distributed normally, so that in col. 6 the class limits have been changed to logarithms and in curve *A* of Fig. 2 the probit has been plotted against the logarithm. In this case the plotted points fall so evenly along a straight line that the final computed curve practically coincided with the one fitted graphically.

In other cases, by means of continuous observation, the exact reaction time of each individual in the series may be recorded. If the number of individuals is large (say 50 or more), they should be grouped for analysis. When it is unknown in which units of time the distribution is normal and there is no clue either from previous experience or from theoretical

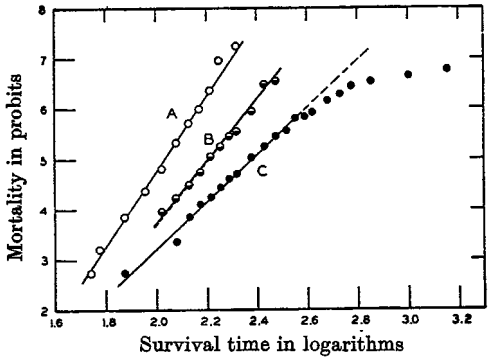

Fig. 2. Rectified time-mortality curves of the survival time of mosquito larvae in solutions of 5·1 (*A*), 4·1 (*B*), and 3·0 (*C*) millimoles per l. of heptylic acid, data of O'Kane *et al.* (14). The provisional curve for series *B* is indicated by a broken line, that for series *C* is graphically indistinguishable from the computed curve shown, as in *A* and *B*, by a solid line. See Tables II and IX.

considerations, several units may be tried empirically in succession, such as the reaction time directly, the rate of reaction, and the logarithm of the time (or of the rate). Class intervals that are equal in terms of any one of these are necessarily unequal in terms of the others, and markedly so if the distribution covers a wide range of time relative to the start of the experiment. Since the original records can be classified with almost equal facility into groups that are evenly spaced in terms of any of these units, the labour of testing several alternatives is not excessive. The procedure may be illustrated by the time in which 68 adult *Drosophila* recovered from a sublethal dose of hydrocyanic acid gas of 1·22 mg. per l. for 1 min. (6).[1]

[1] The series differs from that reported originally (6) in the omission of a single individual which recovered in 76 min. but was marked "very feeble" in the record, the only fly in the series that was so characterized and apparently an experimental error.

The original recovery times, listed in Table III, have been grouped in Table IV into equal units in terms of 5 min. intervals, of 0·5 rates or reciprocals (× 100), and of 0·05 logarithms of minutes. In each case the class interval includes its upper but not its lower limit. By writing opposite each class limit the recovery time in minutes to one more decimal than in the original record, here to 0·1 min., the frequencies in each grouping

Table III

Original data on the recovery period in minutes of individual adult Drosophila *after fumigation with* 1·22 *mg. of HCN per l. for* 1 *min.* (*Oct.* 19), *Broadbent & Bliss* (6)

45	33	29	26	34	55	21	30	30
77	32	41	18	29	25	19	26	17
63	34	21	25	32	19	27	33	47
39	30	28	33	24	29	22	34	24
33	32	31	38	19	27	33	20	—
32	28	13	32	37	29	29	33	—
36	30	26	25	19	33	40	24	—
39	25	27	28	32	32	28	32	—

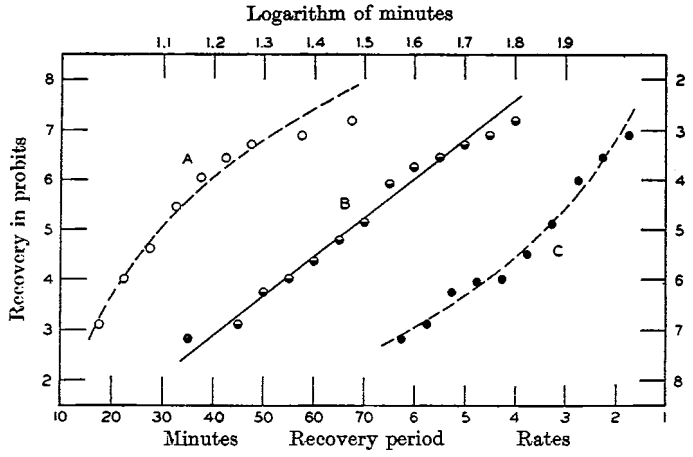

Fig. 3. Recovery of *Drosophila* from stupefaction with HCN, showing its relation to three different functions of recovery time. The rectilinear logarithmic curve (*B*) has been changed to minutes (*A*) and to rates (*C*) for comparison. Data in Table IV.

interval have been tallied directly. The grouped frequencies were then accumulated and transformed to percentages and probits as in the preceding example. Finally each probit has been plotted opposite its corresponding upper class limit in Fig. 3. In order that the different transformations may be compared more easily, the rate-mortality curve has been plotted with reversed co-ordinates. Of the three transformations, only the logarithmic can be fitted satisfactorily with a straight line.

Table IV

Method of handling data in Table III for graphic analysis and computation

Minutes								Reciprocals or rates								Logarithms									
		Cumulative curve						Class limits			Cumulative curve					Class limits			Cumulative curve						
Class limits	f	f	%	R	x	fx	fx^2	Rates	Min.	f	f	%	R	x	fx	fx^2	Logs.	Min.	f	f	%	R	x	fx	fx^2	fx^3
12·5								1·25	80·0								1·10	12·6								
	2				−3	− 6	18			2				−4	− 8	32			1				−7	− 7	49	−343
17·5		2	2·9	3·10				1·75	57·1		2	2·9	3·10				1·15	14·1		1	1·47	2·82				
	9				−2	−18	36			3				−3	− 9	27										
22·5		11	16·2	4·01				2·25	44·4		5	7·4	3·55				1·20	15·8								
	13				−1	−13	13			6				−2	−12	24			1				−5	− 5	25	−125
27·5		24	35·3	4·62				2·75	36·4		11	16·2	4·01				1·25	17·8		2	2·9	3·10				
	22				0					20				−1	−20	20			5				−4	−20	80	−320
32·5		46	67·6	5·46				3·25	30·8		31	45·6	4·89				1·30	19·9		7	10·3	3·74				
	12				1	12	12			16				0					4				−3	−12	36	−108
37·5		58	85·3	6·05				3·75	26·7		47	69·1	5·50				1·35	22·4		11	16·2	4·01				
	5				2	10	20			10				1	10	10			7				−2	−14	28	− 56
42·5		63	92·6	6·45				4·25	23·5		57	83·8	5·99				1·40	25·1		18	26·5	4·37				
	2				3	6	18			1				2	2	4			10				−1	−10	10	− 10
47·5		65	95·6	6·71				4·75	21·1		58	85·3	6·05				1·45	28·2		28	41·2	4·78				
										3				3	9	27			10				0			
52·5								5·25	19·1		61	89·7	6·26				1·50	31·6		38	55·9	5·15				
	1				5	5	25			5				4	20	80			18				1	18	18	18
57·5		66	97·06	6·89				5·75	17·4		66	97·1	6·89				1·55	35·5		56	82·4	5·93				
										1				5	5	25			5				2	10	20	40
62·5								6·25	16·0		67	98·53	7·18				1·60	39·8		61	89·7	6·26				
	1				7	7	49												2				3	6	18	54
67·5		67	98·53	7·18				6·75	14·8								1·65	44·7		63	92·6	6·45				
																			2				4	8	32	128
72·5								7·25	13·8								1·70	50·1		65	95·6	6·71				
	1				9	9	81			1				8	8	64			1				5	5	25	125
77·5		68	100	—				7·75	12·9		68	100	—				1·75	56·2		66	97·1	6·89				
																			1				6	6	36	216
																	1·80	63·1		67	98·53	7·18				
																	1·85	70·8								
																			1				8	8	64	512
																	1·90	79·4		68	100	—				

In small series the individual reaction times are arranged in an ascending order without grouping, as in the example of the time to stupefy the 18 *Drosophila* shown in Table V. However, if the frequencies were accumulated in the usual way, transformed to probits and plotted against the observed periods, the resulting curve would systematically under-estimate or over-estimate the reaction time depending upon whether each probit included or omitted the individual which reacted at that particular moment. This bias can be avoided by basing each percentage upon the total number which have responded at all shorter intervals plus one-half of the one or more individuals reacting at the time in question. Thus in Table V the observed reaction times are given in

Table V

Stupefaction time of adult Drosophila *exposed to* HCN *at a concentration of* 0·89 *mg. per l.; data of Bliss & Broadbent* (3)

Stupefaction time in sec.	No. of flies f	Cumulative curve				Logarithm sec.	Rate 100/ sec.
		No. at each observation	No. for plotting	%	Probit		
22	2	2	1·0	5·6	3·41	1·342	4·55
23	1	3	2·5	13·9	3·92	1·362	4·35
25	2	5	4·0	22·2	4·23	1·398	4·00
26	1	6	5·5	30·6	4·49	1·415	3·85
27	1	7	6·5	36·1	4·64	1·431	3·70
28	1	8	7·5	41·7	4·79	1·447	3·57
29	1	9	8·5	47·2	4·93	1·462	3·45
30	3	12	10·5	58·3	5·21	1·477	3·33
31	1	13	12·5	69·4	5·51	1·491	3·23
32	1	14	13·5	75·0	5·67	1·505	3·13
34	1	15	14·5	80·6	5·86	1·531	2·94
38	1	16	15·5	86·1	6·08	1·580	2·63
40	2	18	17·0	94·4	6·59	1·602	2·50

col. 1, the initial frequencies in col. 2, the frequencies accumulated by the end of each observed time in col. 3, and in col. 4 the number in col. 3 for the preceding time plus one-half the number in col. 2 for the time in question. Thus at 30 sec. the cumulative frequency for graphic analysis is $9+3/2=10{\cdot}5$. The cumulative frequencies are converted to percentages and thence to probits when they can be plotted directly against the observed reaction times or against a function of these times. By this convention it is possible to plot the last animal to die, since the corresponding percentage will always be less than 100. The results are shown in Fig. 4 with reversed co-ordinates for the rate of stupefaction to facilitate the comparison of the three curves, all of which have been fitted here by computation to indicate the full effect of the observations at the ends. On the time scale the lowest and the two highest points fall below the line but all intermediate values above it; this systematic divergence still

persists on the logarithmic scale although the fit is improved; in the rate diagram positive and negative departures occur equally in all parts of the curve. One cannot conclude from a single such diagram, however, that a given response is directly proportional to one rather than to another function of the reaction time, unless it can be verified by statistical computation.

From the straight line that is fitted by inspection to the plotted values of the transformed time-mortality curve, approximate values of its essential statistical constants can be estimated without much further

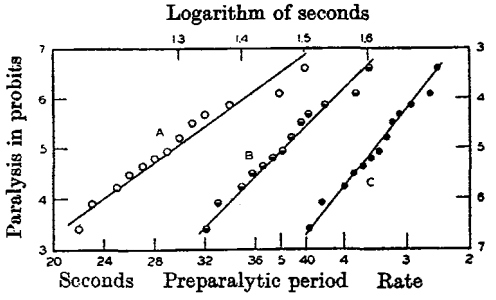

Fig. 4. Preparalytic or stupefaction period of adult *Drosophila* in HCN, comparing three different modes of plotting time. Data in Table V.

calculation. In transposed units the line will pass through the mean reaction time at 5 probits. This also represents the time for 50% of the organisms to respond, but within the range of the curve the time for other percentages of response (when changed to probits) can be read as easily. The interval on the abscissa that represents one probit on the ordinate, the slope of the line, is the standard deviation of the distribution. From the standard deviation and the total number of individuals in the series, the standard error of the mean, of the standard deviation, and of any given level of response can be estimated from equations (6), (7) and (8).

(2) *Grouping*

The reaction times observed in a given test are of interest to the extent that we can infer from these individuals what would be the response in the very much larger population of organisms from which they have been drawn. Obviously, large samples have a better chance of representing correctly the different levels of susceptibility in the population than small samples, and it will be easier to get a good sample if the original population is relatively uniform in its susceptibility. The size of the sample and its variability, therefore, control our estimates of the errors of random sampling, i.e. the standard errors of the mean and of

the standard deviation. They show how closely a given time-mortality curve approaches that which would be expected if the number of individuals upon which it is based could be increased without limit. Since the errors of random sampling are inherent, their effect can be reduced for a given population only by increasing the number of individuals in the test.

Estimates of the time-mortality curve may also be impaired by experimental errors, errors which are not necessarily inherent and which usually can be eliminated by a suitable experimental design. Many of these, such as inaccurate timing, unequal exposure to poison, use of a variable stock of organisms and biased selection within such a stock, are generally known and guarded against. One type of experimental error, however, is not recognized so widely and that is the grouping error introduced by the spacing of observations. Whether this error is introduced during or after the experiment, it leads to a measurable loss of information, just as would occur if fewer individuals were used. It is important to minimize this loss.

When each individual has been timed separately and there are few in the distribution, as in the example in Table V, the curve may be computed without grouping. With larger series, such as the example in Table IV, the labour of computation is reduced materially by grouping. In computing from a grouped distribution, the different individual reaction times within each class interval are replaced by a single value midway between the class limits, in contrast with the procedure required for graphic analysis where the upper limit of each class has been used. When the distribution is divided into 20–24 equal intervals, grouping introduces a negligible error, but as the number of intervals is reduced, it increases rapidly. The effect upon the standard deviation is numerically the same as the loss of $\frac{\text{S.A.}}{V(s)}$ individuals from the distribution, where S.A. is Sheppard's adjustment or the second term in equations (3) and (4) and $V(s)$ is the variance of the standard deviation defined in equations (7) or (7*a*). In Table IV the 68 records of recovery time have been grouped into 16 equal intervals of 0·05 log. units, so that in effect 1·7 flies or 2·5 % have been sacrificed to facilitate the computation, but if a coarser grouping had been used, such as 8 intervals of 0·1 log., the equivalent of 6·8 flies or 9·7 % of the potential information in the experiment would have been lost.

Unsatisfactory grouping may creep into the original record in two ways: (1) The use of periodic rather than continuous observations of

reaction time groups the data into unalterable classes at the time they are recorded. When these intervals are equally spaced on a time scale which later proves not to give a normal distribution of the response, conversion to another scale, such as to logarithms, necessarily leads to a series of unequal intervals and of unequally spaced mid-points between the logarithms of these class limits, as shown by the example in Table II. It is then impossible to simplify the later computation. (2) Many experiments have been made in which there was no observation after the beginning of the experiment until some or many individuals had reacted, and in such instances the lower limit of the first class interval would be 0 time. However, if graphic analysis should show either the logarithm or the reciprocal of time to be distributed normally, the mid-point of the first class would become indeterminate, since the logarithm of 0 is minus infinity and the reciprocal of 0 is plus infinity. This would have the effect of truncating the curve artificially at the time of the first observation and would necessitate the more complicated procedures appropriate for the truncated distribution. Curve *B* of Fig. 2 has been so computed as the lower limit of the first interval was not determinable from the published record.

These complications in the subsequent computation should be considered when recording the original data for time-mortality curves. When each lot of organisms is sufficiently small, the reaction of each individual often can be timed, but when this is impracticable, the observations are necessarily periodic. In such cases it is useful to use a small preliminary series for which a suitable transformation can be selected by graphic analysis and preliminary estimates obtained of the mean and standard deviation. With this guidance the main series of observations can be planned at equal intervals on the scale which probably will be used in later computations (the logarithm, reciprocal, etc.). As many periodic observations should be made when the number of individuals is small as when it is large, since a reduction in their number diminishes further and unnecessarily the reliability of the smaller experiment. These observations of the number which react should be started before the first individual responds and continue until the reactions have stopped. If the material and the phenomenon are so uniform that graphic analysis will be adequate, there may be advantages in not recording the beginning or the end of the experiment. But if computation will be necessary, it saves time to record the experiment in full so that statistical computations will be simple and direct, rather than to compute an incomplete record as a truncated distribution.

(3) *Statistics of the time-mortality curve*

The basic statistics of the time-mortality curve are the mean, $\bar{x}$, and the standard deviation, s. The square of the standard deviation, s^2, is also known as the variance of the distribution. They may be computed by the equations

$$\bar{x} = \frac{S(fx)}{N} \qquad \text{.....(1)}$$

and

$$s = \sqrt{\left(\frac{S(fx^2) - \bar{x}\,S(fx)}{N-1}\right)}, \qquad \text{.....(2)}$$

where x is the reaction time in units that give an apparently straight line in graphic analysis; f is the frequency of the number of individuals for each value of x; $(S(\ \ )$ indicates the sum of all quantities of the type enclosed in brackets; and $N = S(f)$ or the total number of individuals in the test. When the number of individuals is small, as in the example in Table V, x will be the transformed reaction time of each individual organism. When the final grouping intervals are necessarily unequal, as in Table II, x will be the mid-point of each class interval. In either case no short cut is available. When the final grouping intervals are equal as in Table IV, x may be set equal to 0 for a central class interval and to $+1$, $+2$, $+3$, ..., $+n$ and to -1, -2, -3, ..., $-n$ for successive intervals above and below 0. Equations (1) and (2) are solved as before but in terms of these arbitrary units; then multiplied by the grouping interval to convert them back to the units of the rectified time-mortality curve, and finally the mean is added to the mid-point of the interval which has been assigned the value of 0. These procedures are illustrated by the examples in Tables II, IV and V, $S(fx)$ in each case being the sum of the column under the heading "fx" and $S(fx^2)$ the sum of the individual products of the columns x by fx, which can be accumulated directly in the calculating machine.

Grouping has the effect of increasing our estimate of the standard deviation over its true value and this is corrected by subtracting Sheppard's adjustment for grouping. When the data have been grouped during the experiment and the transformation of the time scale has led to unequal intervals, the variance corrected for grouping, s_c^2, is

$$s_c^2 = s^2 - \frac{S(fi^2)}{12N}, \qquad \text{.....(3)}$$

where f is the frequency or number of individuals in each interval i, N is the total number of organisms in the distribution, and the summation $S(\ \ )$ extends over all intervals in the distribution. In the example of

Table II, the numerator of the correction term has been obtained by accumulating the products of the last two columns in the calculating machine. Then $s_c^2 = 0{\cdot}018546 - 0{\cdot}000369 = 0{\cdot}018177$, a correction in the standard deviation amounting to 13% of its standard error (equation (7)).

This correction is very much easier to compute when the data have been grouped in equal intervals on the transformed time scale and these have been assigned a value of 1 for purposes of computation, for then i is equal to 1 for all intervals, $S(f)$ and N cancel out, and equation (3) reduces to the form

$$s_c^2 = s^2 - \tfrac{1}{12}. \qquad \ldots\ldots(4)$$

In Table IV, for example, $s^2 = 6{\cdot}57134$ in unit grouping intervals and $s_c^2 = 6{\cdot}57134 - 0{\cdot}08333 = 6{\cdot}48801$. Although the corrected variance, s_c^2, gives an improved estimate of the standard deviation of the population from which the data have been drawn, the uncorrected variance, s^2, is used in computing the standard errors of both the mean and the standard deviation.

With the computed mean, $\bar{x}$, and the standard deviation, s or s_c, the rectified graphic time-mortality curve can be corrected so that it will pass through $\bar{x}$ at 5 probits with a slope equal to s or s_c. The time at which any given percentage of organisms might be expected to react under the same conditions may either be read from this line or computed by the equation

$$X = \bar{x} + s\,(y - 5), \qquad \ldots\ldots(5)$$

where $\bar{x}$ is the mean, s the standard deviation after correction for grouping if computed from grouped data, and y the probit at the required percentage. In curve B of Fig. 2, for example, we might estimate the time at which the observations should have started to record the reaction of the first individual. Since there were 100 larvae in the test, this would be when 1% or 2·674 probits had responded. For the estimates of the mean and standard deviation we may anticipate their solution in Table IX and by substituting in equation (5) obtain

$$X = 2{\cdot}2076 + 0{\cdot}1634\ (2{\cdot}674 - 5) = 1{\cdot}8275$$

or 67·2 min. as the most probable time for the first larva in the series to have been paralysed by the heptylic acid. The first record was made only 105 min. after immersion in the poison and by that time 15% had reacted.

(4) *Errors of random sampling*

The reliability of a toxicological experiment on the reaction time depends upon how nearly the individuals in the test are representative of the larger population from which they have been drawn. Assuming that

the sample is truly a random one, how closely will the mean and the standard deviation agree with these values in similar future tests? This is measured by the variance of the mean, $V(\bar{x})$, and by the variance of the standard deviation, $V(s)$, as given by the equations

$$V(\bar{x}) = \frac{s^2}{N} \qquad \text{.....(6)}$$

and

$$V(s) = s^2 D, \qquad \text{.....(7)}$$

where s^2 is the square of the standard deviation as first computed (without Sheppard's correction for grouping), N is the number of individuals in the distribution, and D is a factor given in Table VI for values of N from 4 to 50. The values of D in Table VI have been newly calculated from the formula[1] for the variance of the standard deviation. They have not been computed for values of N above 50, but for these larger distributions the empirical formula

$$V(s) = \frac{s^2}{2N - 3/2} \qquad \text{.....(7a)}$$

will give sufficiently accurate estimates of $V(s)$. The variance of the standard deviation as given in equations (7) and (7*a*) is equally valid for determinations of the standard deviation before and after the application of Sheppard's correction. Both equations may be solved with graphic estimates of the standard deviation for use with constants obtained by graphic analysis.

The two most useful functions of the estimates of the variance are the square root and the reciprocal. The square roots of the variances of the mean and of the standard deviation are, of course, their standard errors, while the reciprocals show the relative amount of information contained in each of a series of time-mortality curves. The individual curves of a series are seldom of equal reliability and, in computing the relation of their means and standard deviations to any given third variable, each should be given a weight proportional to its reliability. This is measured by the reciprocal of the variances computed by equations (6) and (7) or (7*a*).

[1] The variance of the standard deviation of a normal distribution is given by the equation

$$V(s) = s^2 \left\{ 1 - \frac{2}{N-1} \left[\frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)} \right]^2 \right\},$$

for which I am indebted to Dr J. Neyman. The term in brackets for different values of N from 4 to 50 is given by D in Table VI. It should be noted that when N is less than 30, s is not distributed normally. A large part of this table has been computed for me by D. M. Kershner of the Institute of Plant Protection in Leningrad.

Table VI

Table for computing the variance of the standard deviation of a normal distribution, which is equal to the product of the square of the standard deviation and the constant D corresponding to the number in the distribution, N, when this varies from 4 to 50 individuals. For larger values of N use the approximation $V(s) = \frac{s^2}{2N-3/2}$

N	*D*	*N*	*D*
—	—	26	0·019794
—	—	27	0·019041
—	—	28	0·018343
4	0·151174	29	0·017694
5	0·116427	30	0·017090
6	0·094585	31	0·016526
7	0·079611	32	0·015997
8	0·068716	33	0·015501
9	0·060437	34	0·015035
10	0·053934	35	0·014597
11	0·048692	36	0·014183
12	0·044378	37	0·013792
13	0·040765	38	0·013421
14	0·037695	39	0·013070
15	0·035055	40	0·012738
16	0·032760	41	0·012422
17	0·030747	42	0·012121
18	0·028967	43	0·011834
19	0·027382	44	0·011561
20	0·025961	45	0·011299
21	0·024680	46	0·011048
22	0·023519	47	0·010809
23	0·022463	48	0·010580
24	0·021497	49	0·010361
25	0·020610	50	0·010152

Frequently the time in which 50% of the individuals react is of less interest than that at some earlier or later stage in the response, which may be computed by equation (5). The standard error of such a time, X, is affected by the error both in the mean and in the standard deviation, and is given by the square root of its variance, $\sqrt{(V(X))}$. The variance of the time, X, to reach any given level or percentage of response, y, measured in probits, is

$$V(X) = V(\bar{x}) + (y-5)^2 V(s). \qquad \text{.....(8)}$$

The reaction time of the most susceptible individual in the time-mortality curve B of Fig. 2 was estimated at 67 min. What is the accuracy of this estimate? From equation (8) and the final estimates of the mean and standard deviation,

$V(X) = 0{\cdot}0002748 + (2{\cdot}674-5)^2\, 0{\cdot}0001634 = 0{\cdot}001159$ and $s_X = 0{\cdot}0340$.

The first larva in the experiment in Fig. 2*B*, therefore, probably succumbed in the time represented by the antilogarithm of $1{\cdot}8275 \pm 0{\cdot}0340$. If we had wanted to be more certain of timing it, within odds of 19 out of 20, our observations should have started at the time given by the antilogarithm of $1{\cdot}8275 - 1{\cdot}645_x\, 0{\cdot}0340 = 1{\cdot}7716$ or 59·1 min. The constant 1·645 has been read from a table of deviates of the normal curve such as Table I in Fisher's text(11) at $P = 0{\cdot}10$. Since in this example we are interested in the chance only of a negative departure from the expected value, the deviate corresponding to $P = 0{\cdot}10$ must be used rather than that for $P = 0{\cdot}05$. The standard error itself shows the + and − limits for odds of a little better than 68 in 100 or about 2 in 3, but by means of a table of deviates the limits corresponding to any required odds (such as 9 in 10) can be computed. When the curve is based upon relatively few individuals, less than 30 for example, instead of "x" (Fisher's Table I), it is better to use in the same way the constant "t", given in Table IV of Fisher's text, with n equal to one less than the number of individuals used in computing the curve. The procedure is then quite equivalent to equation (12) in the first paper of this series(2).

(5) *Measures of the agreement between curve and hypothesis*

The validity of the foregoing procedures depends in large measure upon whether the logarithm, reciprocal, or other unit used for expressing time has in fact turned the distribution of observed reaction times into a normal curve. Although graphic analysis is probably the most efficient method for selecting a suitable function, it is sometimes desirable to confirm by computation whether a given transformation has or has not been effective, or, alternatively, whether the departures from another mode of plotting are significantly non-normal in character. The standard tests for this purpose depend upon carrying the computation two steps beyond the variance of the distribution, s^2, to the summation of the third and fourth powers. When the number of individuals is small, only big departures from a straight line will be significant, departures that will already have been recognized during graphic analysis, so that the computation is then seldom worth making. However, when the number of individuals is large enough to make grouping advisable (50 or more), a numerical measure of the agreement between the observations and the hypothesis that the transformed reaction times are distributed normally may give results that are not apparent from inspection. When the rectified data have been grouped equally, the entire computation is made in terms of arbitrary, unit class intervals and is not unreasonably onerous.

Two constants are computed, g_1 and g_2, each of which is equal to 0 in the theoretical normal distribution. If neither differs significantly from 0 when compared with its standard error, the data are in agreement with the hypothesis.

The first of these constants, g_1, measures the asymmetry of the curve. It may be computed from the equation

$$g_1 = \frac{[S(fx^3) - 3\bar{x}\ S(fx^2) + 2\bar{x}^2\ S(fx)]\ N}{s^3\ (N-1)\ (N-2)}, \qquad \text{.....(9)}$$

where the only new component, $S(fx^3)$, is the sum of the third powers of x, the other symbols having the same significance as before. The variance or square of the standard error of g_1 depends only upon the number of cases in the distribution:

$$V(g_1) = \frac{6N\ (N-1)}{(N-2)\ (N+1)\ (N+3)}. \qquad \text{.....(10)}$$

When the ratio of g_1 to its standard error ($\sqrt{(V(g_1))}$ exceeds 1·96 there is less than 1 chance in 20 that the function of time that has been used in the computation is distributed normally so far as can be judged from the experiment in question. The procedure may be illustrated by the logarithmic transformation of the data in Table IV. Substituting in equations (9) and (10), we have

$$g_1 = \frac{[131 - (-0{\cdot}3088)\ (441) + (0{\cdot}0212)\ (-7)]\ 68}{16{\cdot}845 \times 67 \times 66} = 0{\cdot}2438,$$

$$\sqrt{(V(g_1))} = \sqrt{\left(\frac{6 \times 68 \times 67}{66 \times 69 \times 71}\right)} = 0{\cdot}2908,$$

$$\frac{g_1}{\sqrt{(V(g_1))}} = \frac{0{\cdot}2438}{0{\cdot}2908} = 0{\cdot}838.$$

From the ratio of g_1 to its standard error, 0·838, it is apparent that the logarithmic transformation of recovery time has eliminated any asymmetry or skewness. The same calculation has been made for the original arithmetic and the reciprocal curves, for which $g_1 = 1{\cdot}9416 \pm 0{\cdot}2908$ and $1{\cdot}0832 \pm 0{\cdot}2908$ respectively. Since the ratio of each of these alternative units to its standard error exceeds 1·96 by a large margin, the computed constants confirm our graphic appraisal from Fig. 3 that neither function is distributed normally. We may expect, therefore, whenever the trend of the plotted points in graphic analysis is clearly curvilinear, that the function in question is distributed asymmetrically and therefore not normally.

It is possible for a given function of the reaction time to be distributed symmetrically but yet not normally. The original time-mortality curve, in which the percentage of individuals which have reacted is plotted against the elapsed time, is sigmoid in shape. Presumably the transformation to probits eliminates the S-shaped component just as the change of minutes to logarithms in the last example eliminated any convexity or concavity. But sometimes the points may still tend to curve like an S or a reversed S around the straight line that fits the general trend rather well, much as if the probit transformation had undercorrected or overcorrected the original sigmoid shape. The upper end of the logarithmic curve in Fig. 3, for example, looks a little as if the original S had not been corrected sufficiently by the use of probits, but due to the correlation between successive points secondary twists of this kind are frequently artifacts. The statistic, g_2, is used to determine whether a symmetrical frequency distribution, such as in this example, is also a normal distribution.

The statistic, g_2, is computed from the sum of the fourth powers, $S(fx^4)$ and the constants that have already been determined. In convenient form, the equations may be written as

$$S_4 = S(fx^4) - 4\bar{x}\,S(fx^3) + 6\bar{x}^2\,S(fx^2) - 3\bar{x}^3\,S(fx) \qquad \ldots\ldots(11)$$

and

$$g_2 = \frac{N(N+1)\,S_4 - 3(N-1)^3\,s^4}{(N-1)(N-2)(N-3)\,s^4}. \qquad \ldots\ldots(12)$$

The variance of g_2 is larger than that for g_1 although it, too, depends only upon the number of individuals in the distribution.

$$V(g_2) = \frac{24\,N(N-1)^2}{(N-3)(N-2)(N+3)(N+5)}. \qquad \ldots\ldots(13)$$

When the numerical values for the logarithmic curve in Fig. 3 are substituted in equations (11) to (13), we have

$$S_4 = 11541 - (-0{\cdot}4118)(131) + (0{\cdot}0636)(441) - (-0{\cdot}0033)(-7) = 11623{\cdot}0,$$

$$g_2 = \frac{68 \times 69 \times 11623{\cdot}0 - 902289 \times 43{\cdot}1824}{67 \times 66 \times 65 \times 43{\cdot}1824} = 1{\cdot}2546$$

and

$$\sqrt{(V(g_2))} = \sqrt{\left(\frac{24 \times 68 \times 4489}{65 \times 66 \times 71 \times 73}\right)} = \sqrt{(0{\cdot}329482)} = 0{\cdot}5740.$$

This value of $g_2 = 1{\cdot}255 \pm 0{\cdot}574$ is significantly greater than 0, since the ratio

$$\frac{1{\cdot}255}{0{\cdot}574} = 2{\cdot}186$$

is greater than the customary limit of 1·96 and is equivalent to a probability of $P = 0{\cdot}029$ by interpolation from Table I in Fisher's text. Since g_2 is positive, the use of probits has "undercorrected" the sigmoid character of the original distribution, although the discrepancy could not be appraised from the graphic evidence alone. Presumably a distribution similar to the normal curve but converging less rapidly to 0 frequency at the upper and lower ends would fit satisfactorily. It may be suggested, however, that since each fly in this series was fumigated independently of the others, a small number of treatments might have differed accidentally from the required level, an experimental error which could easily have given the observed results.

III. Incomplete time-mortality curves

In the curves discussed so far, each was complete in the sense that the reaction times were recorded for all individuals and could be reduced to a single normal frequency distribution. Curves for which this is not true are called truncated distributions, when one or both ends are missing or depart more or less abruptly from the straight line which fits the remainder of the distribution. The graphic analysis of such truncated distributions follows the same procedure as that described for complete distributions, although the methods for computing the constants of the two types differ materially. In each case it is assumed that the total number of individuals has been recorded and the percentages are based upon the total number as before.

Two different types of truncation may occur, artificial and biological. In the first type, truncation is not inherent in the process under investigation but is due to the method of experimentation and presumably could be eliminated by a change in technique. In some cases, an experiment may not be completed on the assumption that if it were continued long enough, the remaining individuals would react in essentially the same manner as those which have already been recorded. In his investigations on the action of pyrethrum upon the house-fly, H. H. Richardson (16) timed the reactions of all individuals in his first experiments and obtained complete sigmoid curves, so that in his later tests he cut short his record when three-fourths of the flies were stupefied. Cage tests are sometimes discontinued before all individuals have died, but in many instances there is no certainty that 100% mortality would have been obtained in any case. Some time-mortality curves are truncated by delaying the first observation after the beginning of the experiment until a considerable proportion of the individuals have reacted, as we have seen. The reaction time of the more susceptible individuals is then known quite as indefinitely as that of the most resistant individuals in the preceding case.

Quite distinct from artificial truncation is the second type, biological truncation. In the case of biological truncation, no change in the technique of recording the experiment will give a complete distribution. Truncation is inherent in the biological process. It occurs in the record of survival time when the dosage of poison is small enough to be survived by one or more of the individuals in the experiment, or in that of recovery time when the dosage is so large that some insects fail to recover from an initial stupefaction. In relating the mean and standard deviation of each of a series of time-mortality curves to some third variable in which we are interested, our measurements will have a consistent biological significance only if they represent the reaction times of those individuals that conform to the particular response under investigation. These individuals are related to the basic population as a whole by including the other animals in the sample, i.e. those which did not react typically if at all, as if their reaction had been normal but delayed beyond the limits of the experiment. To omit the non-reactors from the total has the effect of assigning them a mean reaction time equal to that of the more susceptible individuals which actually did respond. For this reason, the percentages usually should be based upon all of the individuals in the series.

In adopting the above procedure we assume that the time of response and the presence or absence of response are two different indices of the same toxicological process. It is possible, of course, that two relatively

independent processes may be involved, one determining when the individual will react to a particular dose and the other the minimum dose which will provoke the reaction. In this case animal *a* might die much more quickly than its brother *b*, if both are given a large dose of poison, and yet be able to survive a somewhat smaller dose which would be fatal to *b*. If the same mean reaction time (when computed only from those which died) were observed after poisoning with the median lethal dose, with the minimum lethal dose, and with a dosage again as large, reaction time could be considered as independent of susceptibility. Reaction time would then be of minor interest and the computation of incomplete time-mortality curves as truncated distributions not justified. Some experiments on stomach insecticides have been reported which give indirect support to this alternative viewpoint(10), but the weight of evidence is against the condition being a common one and in favour of basing most time-mortality curves upon all of the organisms, including those which fail to respond.

The transition in the time-mortality curve between the individuals which give the characteristic response, as judged by their reaction time, and those which fail to respond at all is not always an abrupt one. The reaction time of the transitional individuals may be limited by qualitatively different processes, a condition which would be indicated graphically by a marked change in the slope of an otherwise straight line. In such cases, the time-mortality curve descriptive of the main response may be drawn as if these individuals had not reacted at all, or the curve truncated artificially at the time of the last normal observation. So far as the main process is concerned, either a change of slope or a cessation of response or both may be considered as a biological truncation.

When the change in slope is not pronounced, it may also indicate an incomplete correlation between the two criteria, reaction time and percentage response, rather than a qualitative change in the limiting processes. This possibility can be tested graphically by computing the successive percentages for the time-mortality curve not only from the total number of organisms exposed but also from trial totals intermediate between the total number which reacted and the number exposed. Should one of these give a satisfactory straight line when plotted in terms of probits and a suitable function of time, it would show that the "break" in the curve indicated not a qualitative change in the nature of the response which is measured by the reaction time, but partial independence between the two criteria of susceptibility: reaction time and percentage response.

Two curves for the recovery of *Drosophila* from stupefaction with HCN are shown in Table VII and in Fig. 5. The logarithm of the rate has been plotted rather than the logarithm of the time, to facilitate the later computation of the curves; in all other respects the procedure is the same as that described for plotting Fig. 4. In Fig. 5, curve *A*, at the higher dosage, is truncated abruptly, but the three slowest individuals in curve *B* have been disregarded in fitting a curve intended to represent the principal response to this particular dosage.

Table VII

Data for the graphic analysis of small truncated distributions, showing the individual recovery times of adult Drosophila *following fumigation, in series A with* 0·98 *mg. of* HCN *per l. for* 5 *min.* (16 *May*), *in series B with* 1·01 *mg. of* HCN *per l. for* 2 *min.* (23 *Aug.*); *data of Broadbent & Bliss* (6)

Series *A*						Series *B*					
Recovery time in min.	No. of flies *f*	Rate of recovery: Cum. no.	%	probit	Log. of rate	Recovery time in min.	No. of flies *f*	Rate of recovery: Cum. no.	%	Probit	Log. of rate
82	1	19·5	97·5	6·96	1·086	32	2	18·0	94·7	6·62	1·495
112	1	18·5	92·5	6·44	0·951	41	1	16·5	86·8	6·12	1·387
119	1	17·5	87·5	6·15	0·924	49	1	15·5	81·6	5·90	1·310
138	1	16·5	82·5	5·93	0·860	58	1	14·5	76·3	5·72	1·237
158	1	15·5	77·5	5·76	0·801	59	2	13·0	68·4	5·48	1·229
159	1	14·5	72·5	5·60	0·799	61	1	11·5	60·5	5·27	1·215
161	1	13·5	67·5	5·45	0·793	69	1	10·5	55·3	5·13	1·161
177	1	12·5	62·5	5·32	0·752	79	1	9·5	50·0	5·00	1·102
178	1	11·5	57·5	5·19	0·750	82	1	8·5	44·7	4·87	1·086
205	1	10·5	52·5	5·06	0·688	86	1	7·5	39·5	4·73	1·065
206	1	9·5	47·5	4·94	0·686	219	1	6·5	34·2	4·59	0·660
214	1	8·5	42·5	4·81	0·670	244	1	5·5	28·9	4·44	0·613
216	1	7·5	37·5	4·68	0·665	311	1	4·5	23·7	4·28	0·507
267	1	6·5	32·5	4·55	0·573	NR	4	—	—	—	—
281	1	5·5	27·5	4·40	0·551	—	—	—	—	—	—
NR*	5	—	—	—	—	—	—	—	—	—	—

* Flies which did not recover designated as "NR"

Another example of biological truncation is that in curve *C* of Fig. 2, which shows the survival time of mosquito larvae in a 3·0 millimolar solution of heptylic acid. A comparison of this curve with others in the same series (not shown here) suggests that the first change in slope may be an experimental error and that biologically only the last three or four observations indicate a qualitative difference in the response. The method of analysis of the normal section of a curve, however, is independent of the nature of the response in the individuals outside of this part of the distribution.

Both Pearson (15) and Fisher (5) have described methods for computing the normal distribution when one end is missing and is of unknown size. In the time-mortality curve more information is available since in each experimental sample the number of individuals in the blank or aberrant end of the curve is known. A new solution which takes advantage of this additional information has been developed by W. L. Stevens in an appendix to the present paper, to which the reader is referred. By Stevens's method, the mean and standard deviation are determined by computing corrections for the preliminary graphic estimates with the aid of the constants in Table IX of Pearson's Tables (15). The curve can be computed when truncated at either end or at both ends.

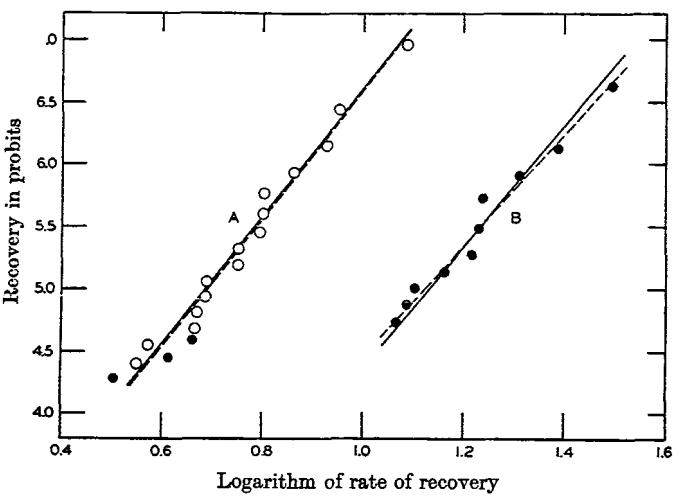

Fig. 5. Truncated time-mortality curves of the recovery of *Drosophila* from stupefaction with HCN. The provisional graphic curves are shown by broken lines, the final computed curves by solid lines. Data in Tables VII and IX.

The computation can be simplified for truncation at only one end of the distribution, and lower end truncations have been selected for this purpose. The rate or speed of toxic action is more likely to be normally distributed and biologically interesting than the reaction time directly. Biological truncation would then be expected at diminishing rates of reaction which would be preceded at the lower end by a complete failure of response or by truncation. If the reaction times are distributed normally when changed to logarithms, the same is true of the rates, so that an upper end truncation in a logarithmic distribution can easily be converted to a lower-end truncation for purposes of computation. Since

the constants from the graphic solution of a truncated curve are required for determining the computed values, they may be described first.

(1) *Statistics from the graphic solution*

It is assumed that the data have been plotted on co-ordinate paper and a provisional time-mortality curve fitted by inspection to those points which are distributed normally, as illustrated by the provisional curves in Figs. 2 and 5. Three terms are read from these lines, the mean, the standard deviation, and the point of truncation, all in terms of the rectified time units given on the abscissa. The provisional mean, which will here be designated by the symbol m_1, since it is not based exclusively upon the sum of the x's, is read as before from the intersection of the line with the ordinate for 5 probits, even though it may be necessary to prolong the line beyond the observations until it intersects the ordinate for 5 probits. The standard deviation, s_1, as for the complete curve, is the abscissal distance separated by 1 probit. The point of truncation, T, in the case of grouped data is the lower limit of the lowest class interval to agree with the distribution that is being fitted; in the case of ungrouped data it is the rectified reaction time of the individual or individuals represented by the lowest point that is included in the curve. From these three graphic estimates the point of truncation is converted into terms of the standard deviation by the equation:

$$x' = \frac{T - m_1}{s_1}. \qquad \text{.....(14)}$$

In Fig. 2, curve B was truncated artificially at the beginning of the experiment and $x' = \frac{2{\cdot}021 - 2{\cdot}210}{0{\cdot}163} = -1{\cdot}1595$ has been based on the time of the first observation following exposure to the poison. Curve C in Fig. 2 departed systematically from normality after the observation at 360 min. Since this is an "upper end" truncation in the present form, for purposes of computation the distribution has been inverted by subtracting the logarithm of each survival time from the digit 3. The point of truncation is $\log \frac{1000}{360} = 0{\cdot}444$ and $x' = \frac{0{\cdot}444 - 0{\cdot}618}{0{\cdot}214} = -0{\cdot}8131$. In Fig. 5 the data for the recovery time have been plotted in terms of the logarithm of the rate. Both were truncated biologically. In curve A the point of truncation has been based upon the slowest individual

$$(x' = -0{\cdot}7165),$$

and in curve B upon the slowest normal individual ($x' = -0{\cdot}2478$), disregarding the last three flies to recover.

The statistic x' is used both in determining the variances of the graphic estimates of the curve and later in computing corrections for these provisional estimates. The variances of the provisional mean, $V(m_1)$, and of the provisional standard deviation, $V(s_1)$, are given by the equations:

$$V(m_1) = \frac{{s_1}^2}{N} E \qquad \text{.....(15)}$$

and

$$V(s_1) = \frac{{s_1}^2}{N} G, \qquad \text{.....(16)}$$

where E and G are constants given in Table VIII[1] for different values of x'. The successive steps in computing these values for each of the distributions in Figs. 2 and 5 are shown in full in the first 12 rows of Table IX. When the data are moderately uniform and there is biological truncation, it is frequently a more profitable expenditure of time to increase the number in each curve or the number of curves and to rely upon graphic analysis than to carry the calculations beyond this stage.

(2) *The calculation of improved estimates*

The calculation of the truncated distribution consists of computing corrections to the graphic estimates of the parameters of the curve. These corrections are not definitive in the sense that a single solution leads at once to precise estimates of the truncated curve. On the contrary the process is one of successive approximations, each approaching more closely to the best available estimate. They converge so rapidly to the correct value, however, especially when the truncated area is less than half of the total distribution, that a single computed estimate is usually sufficient.

[1] The constants of Table VIII have been computed from the values of the incomplete normal moment function in Pearson's Table IX and from the constants of the normal curve in the British Association Tables by the formulae:

$$\alpha = 1{\cdot}5 + 2m_2(x) - 3m_4(x) - q + \frac{x^2 z^2}{q}. \qquad E = \frac{\alpha}{\| i \|}.$$

$$\beta = 0{\cdot}3989423 + m_1(x) - 2m_3(x) + \frac{xz^2}{q}. \qquad F = \frac{\beta}{\| i \|}.$$

$$\gamma = 0{\cdot}5 - m_2(x) + \frac{z^2}{q}. \qquad G = \frac{\gamma}{\| i \|}.$$

$$\| i \| = \alpha\gamma - \beta^2.$$

In this case x is equivalent to x' in the terminology used elsewhere in the present paper.

An improved estimate depends upon the calculation of two correction factors, A and B, by the equations:

$$A=\frac{1}{s_1}\left\{\frac{S(fx)-m_1 n_p}{s_1}-n_q\frac{z}{q}\right\} \qquad \text{.....(17)}$$

and

$$B=\frac{1}{s_1}\left\{\frac{S(fx^2)-m_1\,[2S(fx)-m_1 n_p]}{{s_1}^2}-n_p-n_q\,x'\frac{z}{q}\right\}, \qquad \text{.....(18)}$$

where $S(fx)$ is the sum of the known, normal values of x, each multiplied by its frequency (f) and $S(fx^2)$ is the corresponding sum of the squares of x, using the same definitions of x as in equations (1) and (2). The graphic estimates, m_1 and s_1, have already been described; n_p is the number of individuals in the normal part of the distribution and is equal to $S(f)$, n_q the number in the truncated portion for which the reaction time is unknown, indeterminate or aberrant, and $n_p+n_q=N$; z/q is interpolated from Table VIII for the value of x' that has been computed by equation (14). The nimimum number of components that need be written down in solving equations (17) and (18) with a standard calculating machine has been given in a convenient order in rows 14–18 and 20–23 of Table IX. The corrections, A and B, should be small numbers and will approach 0 in successive approximations. Before these corrections can be used one more term is needed, the co-variance of the mean and standard deviation,

$$W\,(m_1 s_1)=\frac{{s_1}^2}{N}F, \qquad \text{.....(19)}$$

where F is interpolated from Table VIII for x' as in the case of E and G. With these terms we determine improved estimates of the mean, m_2, and of the standard deviation, s_2, by the equations:

$$m_2=m_1+AV(m_1)+BW(m_1 s_1) \qquad \text{.....(20)}$$

and

$$s_2=s_1+AW(m_1 s_1)+BV(s_1). \qquad \text{.....(21)}$$

The validity of m_2 and s_2 depends in part upon the accuracy of the initial graphic estimates, and this is indicated by the magnitude of the differences between m_1 and m_2 and between s_1 and s_2 in relation to the standard errors of m_1 and s_1 respectively. The estimates of m_2 and s_2 for the numerical examples are given in rows 25 and 26 of Table IX. From the next two rows we see that the graphic estimates of the mean have been changed by 2·3 to 17·2% of their standard errors and of the standard deviation by 0·8 to 41·6% of their standard errors.

The two curves in Table IX in which the computed differed from the graphic estimates by more than 10% of the standard errors may be carried to the next approximation. This is obtained by solving equations

Table VIII

Constants for computing the mean, the standard deviation, and their standard errors when the normal frequency distribution is truncated at the lower end but the number of missing records is known. The statistic x′ is computed by equation (14) *from provisional estimates of the curve as explained in the text.**

x'	z/q	Δ_2	E	Δ_2	Δ_4	F	Δ_2	Δ_4	G	Δ_2	Δ_4
−2·5	2·82274	43	1·00056	6	—	− 0·00171	− 15	− 1	0·50528	36	1
−2·4	2·73186	46	1·00078	8	—	− 0·00231	− 19	− 1	0·50693	44	1
−2·3	2·64143	49	1·00107	10	1	− 0·00310	− 24	− 1	0·50903	53	1
−2·2	2·55150	52	1·00147	13	1	− 0·00412	− 30	− 1	0·51166	64	2
−2·1	2·46208	56	1·00200	17	1	− 0·00544	− 37	− 2	0·51493	76	2
−2·0	2·37322	59	1·00270	23	1	− 0·00712	− 46	− 2	0·51896	90	2
−1·9	2·28495	63	1·00363	29	2	− 0·00927	− 56	− 2	0·52390	106	2
−1·8	2·19731	68	1·00485	37	2	− 0·01197	− 69	− 3	0·52990	124	2
−1·7	2·11036	73	1·00645	48	3	− 0·01537	− 85	− 3	0·53714	144	3
−1·6	2·02413	78	1·00852	61	4	− 0·01961	− 103	− 4	0·54583	167	3
−1·5	1·93868	83	·01120	78	5	− 0·02488	− 125	− 5	0·55618	193	4
−1·4	1·85406	89	1·01467	100	6	− 0·03141	− 152	− 6	0·56847	222	4
−1·3	1·77033	95	1·01914	127	8	− 0·03946	− 185	− 7	0·58298	256	5
−1·2	1·68755	102	1·02488	162	10	− 0·04936	− 224	− 8	0·60004	294	6
−1·1	1·60580	109	1·03224	207	13	− 0·06149	− 272	− 11	0·62005	338	7
−1·0	1·52514	117	1·04168	265	16	− 0·07634	− 330	− 13	0·64344	390	9
−0·9	1·44564	125	1·05376	339	21	− 0·09450	− 402	− 16	0·67072	450	11
−0·8	1·36740	134	1·06923	434	28	− 0·11667	− 490	− 21	0·70251	521	13
−0·7	1·29050	143	1·08904	557	37	− 0·14374	− 598	− 26	0·73952	606	16
−0·6	1·21503	153	1·11442	717	48	− 0·17680	− 733	− 33	0·78257	706	20
−0·5	1·14108	163	1·14696	925	63	− 0·21718	− 901	− 42	0·83269	827	25
−0·4	1·06876	173	1·18876	1196	84	− 0·26658	− 1111	− 54	0·89108	972	31
−0·3	0·99817	184	1·24252	1552	112	− 0·32708	− 1374	− 69	0·95918	1148	38
−0·2	0·92942	195	1·31180	2019	149	− 0·40133	− 1707	− 89	1·03877	1363	48
−0·1	0·86262	207	1·40127	2635	199	− 0·49264	− 2128	− 115	1·13198	1626	61

0	0·79788	218	1·51709	3451	268	− 0·60523	− 2663	− 149	1·24145	1949	77
0·1	0·73533	229	1·66743	4535	361	− 0·74446	− 3348	− 194	1·37042	2350	97
0·2	0·67507	241	1·86310	5979	490	− 0·91717	− 4227	− 255	1·52288	2847	124
0·3	0·61722	251	2·11857	7914	663	− 1·13214	− 5360	− 336	1·70381	3468	159
0·4	0·56188	262	2·45318	10514	909	− 1·40071	− 6829	− 444	1·91942	4248	204
0·5	0·50916	271	2·89293	14024	1248	− 1·73757	− 8742	− 590	2·17751	5233	264
0·6	0·45915	279	3·47293	18782	1723	− 2·16185	− 11244	− 792	2·48793	6482	347
0·7	0·41192	286	4·24075	25263	2380	− 2·69858	− 14539	− 1058	2·86318	8078	449
0·8	0·36756	291	5·26121	34124	3320	− 3·38069	− 18892	− 1436	3·31921	10123	596
0·9	0·32611	294	6·62291	46305	4639	− 4·25173	− 24681	− 1949	3·87647	12764	787
1·0	0·28760	296	8·44766	63124	6512	− 5·36957	− 32418	− 2656	4·56136	16191	1043
1·1	0·25205	294	10·90365	0·86456	9231	− 6·81160	− 0·42812	− 3669	5·40817	20662	1411
1·2	0·21944	291	14·22420	1·19019	13041	− 8·68174	− 0·56875	− 5024	6·46159	26544	1872
1·3	0·18974	285	18·73495	1·64623	18745	− 11·12064	− 0·75962	− 7061	7·78045	34298	2586
1·4	0·16288	276	24·89192	2·28971	26738	− 14·31916	− 1·02111	− 9763	9·44228	44639	3464
1·5	0·13879	265	33·33860	3·20058	38748	− 18·53879	− 1·38022	− 13819	11·55050	58443	4807
1·6	0·11735	252	44·98586	4·49892	0·56278	− 24·13864	− 1·87752	− 19556	14·24316	0·77054	6641
1·7	0·09844	237	61·13204	6·36004	0·82104	− 31·61600	− 2·57039	− 27791	17·70635	1·02306	9197
1·8	0·08189	221	83·63826	9·04220	1·20389	− 41·66376	− 3·54117	− 39724	22·19260	1·36755	12841
1·9	0·06756	203	115·18668	12·92825	1·78260	− 55·25269	− 4·90919	− 57438	28·04640	1·84045	18159
2·0	0·05525	184	159·66335	18·59690	2·65147	− 73·75081	− 6·85159	− 83390	35·74065	2·49494	25771

* *Note on interpolation.* Seven or more decimal places have been used in computing most of Table VIII, although in the last few entries the final decimal is open to question. To retain this accuracy in interpolation, central differences are needed, so that the second (Δ_2) and fourth (Δ_4) differences are given for each constant except for z/q for which second differences suffice. With a table of coefficients, such as that provided by Thompson(18), interpolation is no more tedious than the other steps in the calculation. If the truncated portion comprises less than 30% of the entire distribution, Δ_4 can be disregarded, but beyond this point, especially when x' is positive, it should be used. Even with fourth difference interpolation, the last decimal may not always be accurate from $x'=0{\cdot}7$ to 1·3 or the last two decimals from 1·3 to the end of the table.

When determining the variance of graphic estimates of the mean and standard deviation by equations (15) and (16), the constants need not be so precise as in the calculation of improved estimates. Simple interpolation by proportional parts may then suffice in many cases. Where less than one-third of the organisms fall in the truncated part, this is usually accurate to the third decimal and up to two-thirds truncation to the second decimal.

Table IX

Procedure for fitting the truncated distribution to the survival times of mosquito larvae (Culex) *plotted in curves B and C of Fig.* 2, *and to the recovery times of* Drosophila *shown in Table VI and plotted in Fig.* 5. *Roman numerals indicate the first and second computed approximations*

Row no.	Statistic	Reference	*Culex* B I	*Culex* B II	*Culex* C I	*Drosophila* A I	*Drosophila* B I	*Drosophila* B II
1	m_1	Figs. 2 and 5	2·210	2·2077	0·618	0·690	1·121	1·1311
2	s_1	Figs. 2 and 5	0·163	0·1639	0·214	0·194	0·226	0·2045
3	T	—	2·021	2·021	0·444	0·551	1·065	1·065
4	x_1'	Eq. (14)	−1·1595	−1·1391	−0·8131	−0·7165	−0·2478	−0·3232
5	s_1^2	Eq. (15) and (16)	0·026569	0·026863	0·045796	0·037636	0·051076	0·041820
6	N	Eqs. (15) and (16)	100	100	80	20	19	19
7	s_1^2/N	Eqs. (15) and (16)	0·00026569	0·00026863	0·00057245	0·0018818	0·0026882	0·0022011
8	E	Table VIII	1·02764	1·02914	1·06698	1·08542	1·27644	1·22881
9	G	Table VIII	0·60776	0·61185	0·69806	0·73302	0·99917	0·94243
10	$V(m_1)$	Eq. (15)	0·00027303	0·00027646	0·00061079	0·0020425	0·0034313	0·0027047
11	$V(s_1)$	Eq. (16)	0·00016148	0·00016436	0·00039960	0·0013794	0·0026860	0·0020744
12	$\sqrt{V(m_1)}$	—	0·01652	0·01663	0·02471	0·04519	0·05858	0·05201
13	$\sqrt{V(s_1)}$	—	0·01271	0·01282	0·01999	0·03714	0·05183	0·04555
14	$S(fx)$	—	191·675	191·675	44·001	11·549	15·011	15·011
15	$S(fx^2)$	—	433·569019	433·569019	32·361955	9·181023	19·013141	19·013141
16	n_p	—	85	85	63	15	12	12
17	n_q	—	15	15	17	5	7	7
18	z/g	Table VIII	1·65432	1·63764	1·37756	1·30309	0·96204	1·01439
19	F	Table VIII	−0·05398	−0·05645	−0·11351	−0·13889	−0·36393	−0·31193
20	A	Eq. (17)	−8·27354	−0·20988	1·21020	−1·72710	0·72544	−0·34192
21	$2S(fx) - m_1 n_p$	Eq. (18)	195·50	195·6955	49·068	12·748	16·57	16·4488
22	$x^1 z/q$	Eq. (18)	−1·91820	−1·86544	−1·12007	−0·93367	−0·23839	−0·32785
23	B	Eq. (18)	4·64710	0·04082	2·52985	−0·53956	−7·75429	0·04874
24	$W(m_1 s_1)$	Eq. (19)	−0·00001434	−0·00001516	−0·00006498	−0·0002614	−0·0009783	−0·0006866
25	m_2	Eq. (20)	2·20767	2·20764	0·61857	0·68661	1·13108	1·13014
26	s_2	Eq. (21)	0·16387	0·16391	0·21493	0·19371	0·20446	0·20484
27	% change in m	(In terms of $\sqrt{V(m_1)}$)	−14·1	−0·2	2·3	−7·5	17·2	−1·8
28	% change in s	(In terms of $\sqrt{V(s_1)}$)	6·8	0·3	4·7	−0·8	−41·6	0·8
29	Final m	—	—	2·2076	2·3814	0·6866	—	1·1301
30	Final s	—	—	0·1634	0·2143	0·1937	—	0·2048

(14 to (21) a second time, except that m_2 and s_2 are substituted for m_1 and s_1 throughout to get m_3 and s_3. Some of the terms from the original distribution, such as $S(fx)$ and $S(fx^2)$, are unchanged, so that the calculation is shorter than before. In terms of their standard errors, the mean and standard deviation of the experiment "*Culex* B" have been altered by only 0·2 and 0·3% respectively, and in the experiment "*Drosophila* B" by 1·8 and 0·8%. This demonstrates the rapid convergence obtainable by the method. In both cases the first computed approximations would have been sufficiently reliable. Unless there is a marked difference between graphic and computed estimates and more than half of the distribution is lost by truncation, a second calculated estimate is probably not worth the labour of computation.

When the estimates of the standard deviation have been computed from grouped data, it is necessary to subtract Sheppard's correction for grouping from the square of the best estimate of the standard deviation, just as was done in the case of the complete distribution. It is computed by equation (3), except that the number of individuals represented in the numerator, $S(f)$, will be equal to n_p instead of N. The examples "*Culex* B" and "*Culex* C" in Table IX have been computed from grouped data similar to that in Table II. The corrections for grouping to be subtracted from the variances 0·026866 (B) and 0·046195 (C) are 0·000159 and 0·000271 respectively, giving corrected estimates of the standard deviation of 0·1634 and 0·2143. It may be noted that the correction for grouping is here of the same order of magnitude as the corrections in the graphic estimate of the standard deviation.

The last stage in solving the truncated distribution is to compute the variances and standard deviation from the final estimates. These are used first to recompute x' by equation (14) and then with the corresponding values of E and G, the required statistics are calculated from equations (15) and (16).

IV. APPENDIX: THE TRUNCATED NORMAL DISTRIBUTION

BY W. L. STEVENS

Galton Laboratory, University College, London

The following three cases of truncated normal distribution should be distinguished:

(A 1 and 2). The point of truncation is predetermined by the method of the experiment or the selection of the data, and the number of individuals in the truncated portion either (1) is or (2) is not known.

(B) The truncation arises from the nature of the experimental material, and the point of truncation can be determined only from the data themselves.

The solution of (A 2) has been given by R. A. Fisher in the introduction to Vol. I of the *British Association Mathematical Tables.*

The solution of (A 1) is given herewith. The case would arise in practice if a time-mortality experiment were discontinued at a pre-arranged instant, and the number noted of organisms which had not reacted.

Case (B) is the one arising from the present paper. A solution has been found, but as four parameters need to be estimated, it is too cumbersome for practical use. As a compromise, it is proposed to determine the biological point of truncation by inspection of the probit graph, and then to treat the data as though case (A 1) were appropriate. The truncated portion is here taken to include not merely the individuals that never react, but also those that are abnormally slow, as indicated by the falling away of their probit plots from the straight line. This method cannot be regarded as anything more than a compromise, and the calculated standard errors therefore overrate the precision.

The normal curve truncated at both ends by predetermined limits

The theoretical distribution of the variate u is:

$$q_1 = \int_{-\infty}^{x_1} \frac{1}{\sqrt{(2\pi)}} e^{-x^2/2}.dx$$

= probability of observation lying below the lower point of truncation $\mu + \sigma x_1$.

$$zdx = \frac{1}{\sqrt{(2\pi)}} e^{-x^2/2}.dx = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(u-\mu)^2/2\sigma^2}.du$$

= probability of observation lying in range $(u,\ u+du)$ between the points of truncation.

$$q_2 = \int_{x_2}^{\infty} \frac{1}{\sqrt{(2\pi)}} e^{-x^2/2}.dx$$

= probability of observation lying above the upper point of truncation $\mu + \sigma x_2$.

The data consists of n_0 individual observations, and n_1 and n_2 respectively in the lower and upper truncated portions.

The following results may be verified:

$$\left.\begin{aligned} \frac{\partial}{\partial\mu}\left(\frac{z}{\sigma}\right) &= \frac{xz}{\sigma^2} \\ \frac{\partial}{\partial\sigma}\left(\frac{z}{\sigma}\right) &= \frac{z}{\sigma^2}(x^2-1) \end{aligned}\right\} \quad \left.\begin{aligned} \frac{\partial q_1}{\partial\mu} &= \frac{-z_1}{\sigma} \\ \frac{\partial q_1}{\partial\sigma} &= \frac{-x_1 z_1}{\sigma} \end{aligned}\right\} \quad \left.\begin{aligned} \frac{\partial q_2}{\partial\mu} &= \frac{z_2}{\sigma} \\ \frac{\partial q_2}{\partial\sigma} &= \frac{x_2 z_2}{\sigma} \end{aligned}\right\}. \qquad \text{.....(1)}$$

The likelihood function is

$$L = -S\left\{\frac{(u-\mu)^2}{2\sigma^2}\right\} - n_0 \log \sigma + n_1 \log q_1 + n_2 \log q_2, \quad \ldots\ldots(2)$$

where summation proceeds over the n individual observations. The maximum likelihood estimates of μ and σ are therefore roots of the equations:

$$\left.\begin{aligned} \frac{\partial L}{\partial \mu} &= S\left(\frac{u-\mu}{\sigma^2}\right) - \frac{n_1 z_1}{\sigma q_1} + \frac{n_2 z_2}{\sigma q_2} = 0 \\ \frac{\partial L}{\partial \sigma} &= S\left\{\frac{(u-\mu)^2}{\sigma^3}\right\} - \frac{n_0}{\sigma} - \frac{n_1 x_1 z_1}{\sigma q_1} + \frac{n_2 x_2 z_2}{\sigma q_2} = 0 \end{aligned}\right\}. \quad \ldots\ldots(3)$$

The components of the information matrix are:

$$\left.\begin{aligned} i_{\mu\mu} &= \frac{1}{q_1}\left(\frac{z_1}{\sigma}\right)^2 + \int_{x_1}^{x_2} \frac{\sigma}{z}\left(\frac{xz}{\sigma^2}\right)^2 \sigma dx + \frac{1}{q_2}\left(\frac{z_2}{\sigma}\right)^2 \\ i_{\mu\sigma} &= \frac{1}{q_1}\left(\frac{z_1}{\sigma}\right)\left(\frac{x_1 z_1}{\sigma}\right) + \int_{x_1}^{x_2} \frac{\sigma}{z}\left(\frac{xz}{\sigma^2}\right)\frac{z}{\sigma^2}(x^2-1)\,\sigma dx + \frac{1}{q_2}\left(\frac{z_2}{\sigma}\right)\left(\frac{x_2 z_2}{\sigma}\right) \\ i_{\sigma\sigma} &= \frac{1}{q_1}\left(\frac{x_1 z_1}{\sigma}\right)^3 + \int_{x_1}^{x_2} \frac{\sigma}{z}\left(\frac{z}{\sigma^2}\right)^2 (x^2-1)^2\,\sigma dx + \frac{1}{q_2}\left(\frac{x_2 z_2}{\sigma}\right)^2 \end{aligned}\right\}. \quad \ldots(4)$$

If $\int_{x_1}^{x_2} x^n \frac{1}{\sqrt{(2\pi)}} e^{-\frac{x^2}{2}} . dx$ is written $f_n(x_1 x_2)$, then

$$\left.\begin{aligned} i_{\mu\mu} &= \frac{1}{\sigma^2}\left[\frac{{z_1}^2}{q_1} + f_2(x_1 x_2) + \frac{{z_2}^2}{q_2}\right] \\ i_{\mu\sigma} &= \frac{1}{\sigma^2}\left[\frac{x_1 {z_1}^2}{q_1} + f_3(x_1 x_2) - f_1(x_1 x_2) + \frac{x_2 {z_2}^2}{q_2}\right] \\ i_{\sigma\sigma} &= \frac{1}{\sigma^2}\left[\frac{{x_1}^2 {z_1}^2}{q_1} + f_4(x_1 x_2) - 2f_2(x_1 x_2) + f_0(x_1 x_2) + \frac{{x_1}^2 {z_1}^2}{q_2}\right] \end{aligned}\right\}. \quad \ldots(5)$$

The functions $f_n(x_1 x_2)$ may be evaluated from Pearson's *Tables for Statisticians and Biometricians*. Table IX gives the functions:

$$m_n(x) = \frac{\int_0^x x^n \frac{1}{\sqrt{(2\pi)}} e^{-x^2/2} . dx}{(n-1)(n-3)\ldots},$$

$$m_n(-x) = (-1)^{n+1} m_n(x), \quad \ldots\ldots(6)$$

$$f_n(x_1 x_2) = (n-1)(n-3)\ldots\{m_n(x_2) - m_n(x_1)\}; \quad \ldots\ldots(7)$$

$f_0(x_1 x_2)$ is the area of the normal curve between x_1 and x_2.

When truncation is at the top end only $-\infty$ is substituted for x_1; when at the bottom end only $+\infty$ is substituted for x_2.

From the information matrix are found the variances and co-variances of the maximum likelihood estimates:

$$\left.\begin{aligned} v_{\mu\mu} &= \frac{1}{n}\frac{i_{\sigma\sigma}}{i_{\mu\mu}i_{\sigma\sigma}-i_{\mu\sigma}^2} \\ v_{\mu\sigma} &= -\frac{1}{n}\frac{i_{\mu\sigma}}{i_{\mu\mu}i_{\sigma\sigma}-i_{\mu\sigma}^2} \\ v_{\sigma\sigma} &= \frac{1}{n}\frac{i_{\mu\mu}}{i_{\mu\mu}i_{\sigma\sigma}-i_{\mu\sigma}^2} \end{aligned}\right\} \qquad \text{.....(8)}$$

The solution of the Maximum Likelihood Equations

The straight line probit graph which has been fitted by eye gives the first approximations m_1, s_1 to the maximum likelihood estimates of μ and σ. These estimates give approximate values for $v_{\mu\mu}$, $v_{\mu\sigma}$, $v_{\sigma\sigma}$ and make the left-hand sides of equations (3) equal to small quantities A and B. The second approximations are now given by

$$\left.\begin{aligned} m_2 &= m_1 + \delta m \\ s_2 &= s_1 + \delta s \end{aligned}\right\}, \qquad \text{.....(9)}$$

where

$$\left.\begin{aligned} \delta m &= v_{\mu\mu}A + v_{\mu\sigma}B \\ \delta s &= v_{\mu\sigma}A + v_{\sigma\sigma}B \end{aligned}\right\}. \qquad \text{.....(10)}$$

A single improvement is usually sufficient, but if necessary the process may be repeated.

V. Summary

When the result of a toxicity test is measured in terms of the reaction time, the data can be plotted so as to show the percentage of animals which has reacted at different times from the beginning to the end of the experiment. This may be called a "time-mortality" curve and with most animals is sigmoid in its original form. On the hypothesis that it measures the individual variation in susceptibility, it frequently can be plotted as a straight line by converting the percentages to probits and the observed time to logarithms or to rates. When not based upon this type of graphic analysis, time-mortality measurements are often of indeterminate reliability, represent different degrees of effectiveness, conceal changes in the nature of the response necessary to its understanding, or cannot be reduced to a consistent biological formulation covering both partially and fully effective levels of dosage.

The preparation of original data for plotting depends in part upon whether they have been grouped during the experiment or afterwards in preparing the frequency distribution. In the first case equally spaced

observations usually lead to unequal grouping intervals when converted to the function of time that is distributed normally, but for either case methods are given by which the loss of information due to grouping can be measured and minimized. Small distributions of individual reaction times can be plotted without grouping. The same procedures are available when the distribution is truncated, either artificially because of the experimental technique or biologically at a given level of susceptibility due to a change in the nature of the response or to its complete cessation. In either case graphic analysis leads directly to consistent and comparable approximate estimates of the mean and standard deviation and of their variances.

The time-mortality curve is computed directly from the non-cumulative frequency distribution of the rectified reaction times rather than from the cumulative curve that is plotted. The effect of grouping upon these calculations is discussed with particular reference to developing an efficient experimental design. Sheppard's correction of the variance for grouping is given in a form applicable to both unequal and equal grouping intervals. From the parameters of the time-mortality curve, the mean and the standard deviation, the reaction time for any given proportion of the population between 0 and 100% can be computed. The accuracy of the time-mortality curve is measured by the errors of random sampling of the mean and of the standard deviation. For the determination of the latter a newly computed table and a corrected formula are provided. Errors in both parameters reduce to a measurable degree the accuracy of an estimated reaction time earlier or later than that for 50% of the population. The agreement of a time-mortality curve with the hypothesis upon which it has been computed may be tested by means of the statistics g_1 and g_2. The first measures the asymmetry or skewness of the supposedly normal distribution and determines whether the main trend of the points in the transformed cumulative curve is really rectilinear; the second shows whether or not the secondary trends and twists about the rectilinear curve are statistically significant.

By means of the truncated time-mortality curve the toxicological value of studies on the reaction time can be extended considerably. Sometimes graphic analysis will supply sufficiently accurate estimates from an incomplete curve of its mean and standard deviation and of their standard errors, but when the data are less regular it is desirable to compute corrections for these graphic solutions. A new method for computing the truncated normal distribution by successive approximations has been developed by W. L. Stevens and is described by him in an

appendix. On the basis of Stevens's method tables have been prepared for computing time-mortality curves that are truncated at their lower ends, a form which covers most cases. Usually the first approximation is sufficiently precise.

The different procedures are illustrated by numerical examples.

REFERENCES

(1) *Barlow's Tables of Squares, Cubes, etc.* (1935). 3rd ed. E. and F. N. Spon.
(2) BLISS, C. I. (1935). *Ann. appl. Biol.* XXII, 134, 307.
(3) BLISS C. I. & BROADBENT, B. M. (1935). *J. econ. Ent.* XXVIII, 989.
(4) BREUKELMAN, JOHN (1932). *Physiol. Zool.* V, 207.
(5) *British Association for Advancement of Science: Mathematical Tables,* I. London.
(6) BROADBENT, B. M. & BLISS, C. I. (1936). *J. econ. Ent.* XXIX, 143.
(7) CAMPBELL, F. L. (1926). *J. gen. Physiol.* IX, 433.
(8) CROZIER, W. J., PINCUS, G. & ZAHL, P. A. (1936). *J. gen. Physiol.* XIX, 523.
(9) DAVEY, W. P. (1919). *J. exp. Zool.* XXVIII, 449.
(10) ELMORE, J. C. & RICHARDSON, C. H. (1936). *J. econ. Ent.* XXIX, 426.
(11) FISHER, R. A. (1936). *Statistical Methods for Research Workers,* 6th ed. Oliver and Boyd.
(12) GADDUM, J. H. (1933). *Med. Res. Council Spec. Rep. No.* 183. His Majesty's Stationery Office.
(13) GERSDORFF, W. A. (1935). *J. agric. Res.* L, 881.
(14) O'KANE, W. C., WESTGATE, W. A. & GLOVER, L. C. (1934). *Tech. Bull. N.H. agric. Exp. Sta.* No. 58.
(15) PEARSON, K. (1924). *Tables for Statisticians and Biometricians,* Part I, 2nd ed. Cambridge.
(16) RICHARDSON, H. H. (1931). *J. econ. Ent.* XXIV, 97.
(17) SHACKELL, L. F. (1923). *J. gen. Physiol.* V, 783.
(18) THOMPSON, A. J. (1921). *Tracts for Computers,* No. V. Dept. of Applied Statistics, University College, London.
(19) WRIGHT, SEWAL (1926). *J. Amer. statist. Ass.* XXI, 162.

(*Received* 31 *August* 1936)