THE CALCULATION OF THE DOSAGE-MORTALITY CURVE

By C. I. BLISS.

(Galton Laboratory, University College, London.)

(With 3 Text-figures.)

CONTENTS.

		PAGE
I.	The interpretation of the dosage-mortality curve and its transformation to a straight line	135
II.	The provisional regression line	146
	(1) Probit values for 0 and 100 percentage kills	147
	(2) Weights for fitting the regression line	148
III.	The computation of the regression line	156
IV.	Accuracy of the regression line	158
	(1) The χ^2 test for comparing observations with the computed curve	159
	(2) The variances of position and slope	161
	(3) The zone of error of the regression line	162
V.	Appendix: The case of zero survivors, by R. A. Fisher	164
VI.	Summary	166
	References	167

Toxicological studies upon a large variety of organisms by many biologists have established the sigmoid character of the typical dosage-mortality curve, especially in the case of multicellular forms. Recently it has been shown in two different fields that such curves can easily be plotted as straight lines and their later analysis thereby facilitated (1, 5, 6). These methods, which are substantially the same, are developed more fully in the present paper. While the procedures have been selected on the basis of their statistical accuracy and efficiency, and accordingly follow the recent trends which are so closely associated with the name of R. A. Fisher, an attempt has been made to present them in sufficient detail to permit their use by biologists with a limited knowledge of statistics. The present paper is concerned with the calculation of the transformed dosage-mortality curve and its accuracy. Later papers in this series will deal with statistical methods for comparing dosage-mortality data, and with time-survival curves.

I. THE INTERPRETATION OF THE DOSAGE-MORTALITY CURVE AND ITS TRANSFORMATION TO A STRAIGHT LINE.

Action curves in pharmacology are those in which the amount of the response to any given degree of chemical or physical stimulation is expressed as a percentage of the maximum obtainable in that particular biological system. The action curve is frequently sigmoid, especially when it expresses the relationship of mortality to dosage, so that a graphic plot of the percentage of dead organisms on the ordinate against some function of dosage along the abscissa resembles the letter *S*, the change in percentage kill per unit of the abscissa being smallest near mortalities of 0 and 100 per cent., and largest near 50 per cent. Among multicellular organisms, it is practically universal for a diagram with these co-ordinates to show this characteristic shape, but the interpretation of such curves has varied widely. Since this controversy has been reviewed so fully by Clark (2), the ground need not be gone over again, and we may proceed at once to describe the viewpoint adopted here.

On this theory, the dosage-mortality curve is primarily descriptive of the variation in susceptibility between the individuals of a population. Let us suppose that, under uniform conditions, the susceptibility of each individual may be represented by the smallest dose which is just sufficient to kill it, the individual lethal dose. As in the case of any other biological characteristic, this susceptibility will vary from one individual to another in the population, and *a priori* we might expect the distribution curve of the number of individuals having each particular susceptibility to show the shape characteristic of the normal curve of error. If Fig. 1, which is the normal curve of error in its most usual form, is assumed, for the moment, to be an ideal representation of the variation in susceptibility, the ordinates will give the number of individual organisms corresponding to each particular individual lethal dose shown along the base in a graded series (assuming that the numbers along the base of the figure are equivalent to actual dosages in one form or another).

With intact animals, however, the experimental technique is usually not suitable for determining the exact minimum lethal dose for each individual, as would be required to secure the data for plotting this form of the normal frequency curve of error. As the experiment is actually conducted, the dosage applied to each separate lot of organisms kills not only those requiring at least this quantity of poison, but also all more susceptible individuals, *i.e.* those which could be killed with a smaller

dosage. Consequently, if Fig. 1 represents the hypothetical frequency distribution of susceptibility, as measured by the individual lethal dose, any given dose will split the sample of organisms into two categories of dead and alive, whose relative proportion will depend upon the relation of the dosage to the distribution of susceptibilities. If our dose had happened to come at the point marked x in Fig. 1, the ratio of the dead or more susceptible individuals to the total number in the sample treated—in other words, the percentage killed—would have been the ratio of the unshaded *area* to the total area under the curve. By varying our dosage

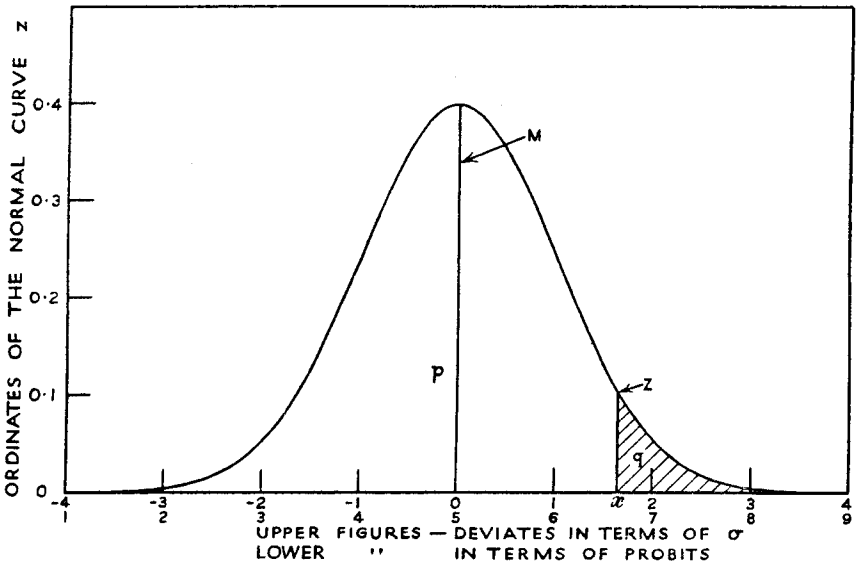

Fig. 1. The theoretical normal curve of error, in which p (0·95) and q (0·05) indicate areas under the curve to the left and right respectively of the ordinate z erected at the point on the abscissa indicated by x (1·645 σ). The position of the median (and also of the mean and the mode) is given by M which divides the area under the curve into halves.

along the base and using a succession of equivalent samples of organisms, it would be possible to determine a series of percentage kills (or proportionate areas, $\frac{p}{p+q}$, of the normal frequency curve) corresponding to the dosages applied experimentally. If these percentage kills were then plotted on the ordinate of another graph against the dosage on the abscissa as before, the result would be a cumulative normal frequency distribution such as Fig. 2. This type of curve, therefore, can be and frequently is obtained experimentally in the laboratory.

The assumption that the individual susceptibility to a poison is distributed normally may be tested by reversing our argument. From a given sample of 40 beetles, let us say, exposed to a known concentration of fumigant, 38, or 95 per cent., were killed. Temporarily neglecting the observed dosage, this percentage kill may be equated to a fraction of the total area under the theoretical normal curve of error, $\frac{p}{p+q}$, and the "expected" dosage, x, to which this mortality corresponds, read from the

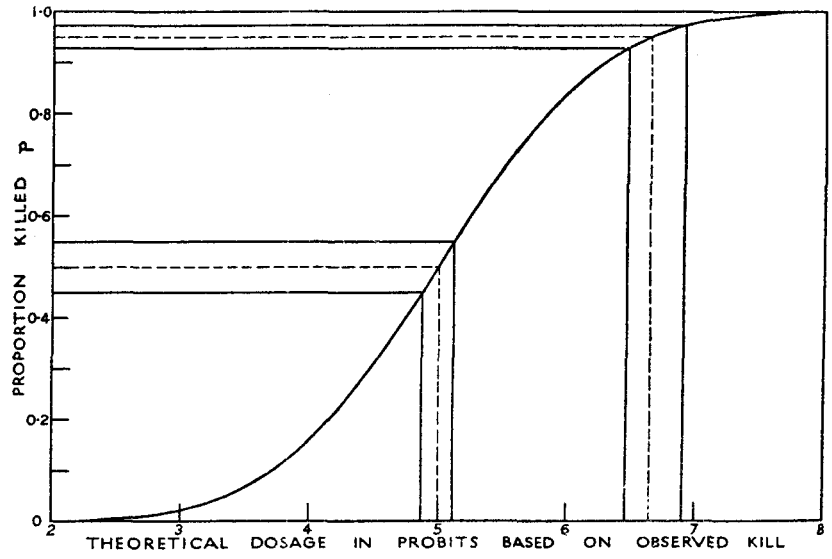

Fig. 2. The proportionate areas, $\frac{p}{p+q}$, of Fig. 1 plotted on the same abscissa as before (probit units). The "broken" lines are drawn at the same positions as the two ordinates, M and xz, of Fig. 1, while the solid parallel lines bounding the broken lines mark the corresponding limits of the standard error for a sample of 100 individuals.

base (Fig. 1). Because of the availability of statistical tables, this expected dosage is given most conveniently in units of standard deviations. The standard deviation, σ, corresponding to any observed mortality may be read directly from sources such as the Kelley-Wood Table (7) or the Shepard-Galton Table (9), and in this case would be 1·645 standard deviations. Similarly, another sample of 40 beetles at a lower dosage, may have shown a mortality of only 20 individuals or 50 per cent., and the expected dosage inferred from this mortality would be 0 standard deviations, since the standard deviation in the normal curve is measured from the median or mean as the origin.

In this fashion an expected dosage corresponding to every observed dosage measured experimentally may be determined from the observed mortality, and the inferred dosages, so derived, are called "normal equivalent deviations" or "N.E.D." by Gaddum (5) and by Hemmingsen (6). Many observations, however, will fall below 50 per cent. kill and by Gaddum's system would require negative expected dosages, which are inconvenient. In order to avoid this difficulty, a new table of statistical units called "probits" has been devised (1) in which the 0 of the usual statistical table of deviates has been equated to the digit 5, and the deviate of the normal curve, in terms of σ, added algebraically to secure the probit corresponding to each percentage kill (Table I). Because of their greater convenience, the expected dosages may be expressed in terms of probits and will not modify the proof or disproof of our basic assumption.

Table I.

Probits or probability units for transforming the sigmoid dosage-mortality curve to a straight line. In the body of the table is given the probit corresponding to each percentage mortality listed along the left edge and top.

	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0	—	1·9098	2·1218	2·2522	2·3479	2·4242	2·4879	2·5427	2·5911	2·6344
1	2·6737	2·7096	2·7429	2·7738	2·8027	2·8299	2·8556	2·8799	2·9031	2·9251
2	2·9463	2·9665	2·9859	3·0046	3·0226	3·0400	3·0569	3·0732	3·0890	3·1043
3	3·1192	3·1337	3·1478	3·1616	3·1750	3·1881	3·2009	3·2134	3·2256	3·2376
4	3·2493	3·2608	3·2721	3·2831	3·2940	3·3046	3·3151	3·3253	3·3354	3·3454
5	3·3551	3·3648	3·3742	3·3836	3·3928	3·4018	3·4107	3·4195	3·4282	3·4368
6	3·4452	3·4536	3·4618	3·4699	3·4780	3·4859	3·4937	3·5015	3·5091	3·5167
7	3·5242	3·5316	3·5389	3·5462	3·5534	3·5605	3·5675	3·5745	3·5813	3·5882
8	3·5949	3·6016	3·6083	3·6148	3·6213	3·6278	3·6342	3·6405	3·6468	3·6531
9	3·6592	3·6654	3·6715	3·6775	3·6835	3·6894	3·6953	3·7012	3·7070	3·7127
10	3·7184	3·7241	3·7298	3·7354	3·7409	3·7464	3·7519	3·7574	3·7628	3·7681
11	3·7735	3·7788	3·7840	3·7893	3·7945	3·7996	3·8048	3·8099	3·8150	3·8200
12	3·8250	3·8300	3·8350	3·8399	3·8448	3·8497	3·8545	3·8593	3·8641	3·8689
13	3·8736	3·8783	3·8830	3·8877	3·8923	3·8969	3·9015	3·9061	3·9107	3·9152
14	3·9197	3·9242	3·9286	3·9331	3·9375	3·9419	3·9463	3·9506	3·9550	3·9593
15	3·9636	3·9678	3·9721	3·9763	3·9806	3·9848	3·9890	3·9931	3·9973	4·0014
16	4·0055	4·0096	4·0137	4·0178	4·0218	4·0259	4·0299	4·0339	4·0379	4·0419
17	4·0458	4·0498	4·0537	4·0576	4·0615	4·0654	4·0693	4·0731	4·0770	4·0808
18	4·0846	4·0884	4·0922	4·0960	4·0998	4·1035	4·1073	4·1110	4·1147	4·1184
19	4·1221	4·1258	4·1295	4·1331	4·1367	4·1404	4·1440	4·1476	4·1512	4·1548
20	4·1584	4·1619	4·1655	4·1690	4·1726	4·1761	4·1796	4·1831	4·1866	4·1901
21	4·1936	4·1970	4·2005	4·2039	4·2074	4·2108	4·2142	4·2176	4·2210	4·2244
22	4·2278	4·2312	4·2345	4·2379	4·2412	4·2446	4·2479	4·2512	4·2546	4·2579
23	4·2612	4·2644	4·2677	4·2710	4·2743	4·2775	4·2808	4·2840	4·2872	4·2905
24	4·2937	4·2969	4·3001	4·3033	4·3065	4·3097	4·3129	4·3160	4·3192	4·3224

Table I (*cont.*).

	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
25	4·3255	4·3287	4·3318	4·3349	4·3380	4·3412	4·3443	4·3474	4·3505	4·3536
26	4·3567	4·3597	4·3628	4·3659	4·3689	4·3720	4·3750	4·3781	4·3811	4·3842
27	4·3872	4·3902	4·3932	4·3962	4·3992	4·4022	4·4052	4·4082	4·4112	4·4142
28	4·4172	4·4201	4·4231	4·4260	4·4290	4·4319	4·4349	4·4378	4·4408	4·4437
29	4·4466	4·4495	4·4524	4·4554	4·4583	4·4612	4·4641	4·4670	4·4698	4·4727
30	4·4756	4·4785	4·4813	4·4842	4·4871	4·4899	4·4928	4·4956	4·4985	4·5013
31	4·5041	4·5070	4·5098	4·5126	4·5155	4·5183	4·5211	4·5239	4·5267	4·5295
32	4·5323	4·5351	4·5379	4·5407	4·5435	4·5462	4·5490	4·5518	4·5546	4·5573
33	4·5601	4·5628	4·5656	4·5684	4·5711	4·5739	4·5766	4·5793	4·5821	4·5848
34	4·5875	4·5903	4·5930	4·5957	4·5984	4·6011	4·6039	4·6066	4·6093	4·6120
35	4·6147	4·6174	4·6201	4·6228	4·6255	4·6281	4·6308	4·6335	4·6362	4·6389
36	4·6415	4·6442	4·6469	4·6495	4·6522	4·6549	4·6575	4·6602	4·6628	4·6655
37	4·6681	4·6708	4·6734	4·6761	4·6787	4·6814	4·6840	4·6866	4·6893	4·6919
38	4·6945	4·6971	4·6998	4·7024	4·7050	4·7076	4·7102	4·7129	4·7155	4·7181
39	4·7207	4·7233	4·7259	4·7285	4·7311	4·7337	4·7363	4·7389	4·7415	4·7441
40	4·7467	4·7492	4·7518	4·7544	4·7570	4·7596	4·7622	4·7647	4·7673	4·7699
41	4·7725	4·7750	4·7776	4·7802	4·7827	4·7853	4·7879	4·7904	4·7930	4·7955
42	4·7981	4·8007	4·8032	4·8058	4·8083	4·8109	4·8134	4·8160	4·8185	4·8211
43	4·8236	4·8262	4·8287	4·8313	4·8338	4·8363	4·8389	4·8414	4·8440	4·8465
44	4·8490	4·8516	4·8541	4·8566	4·8592	4·8617	4·8642	4·8668	4·8693	4·8718
45	4·8743	4·8769	4·8794	4·8819	4·8844	4·8870	4·8895	4·8920	4·8945	4·8970
46	4·8996	4·9021	4·9046	4·9071	4·9096	4·9122	4·9147	4·9172	4·9197	4·9222
47	4·9247	4·9272	4·9298	4·9323	4·9348	4·9373	4·9398	4·9423	4·9448	4·9473
48	4·9498	4·9524	4·9549	4·9574	4·9599	4·9624	4·9649	4·9674	4·9699	4·9724
49	4·9749	4·9774	4·9799	4·9825	4·9850	4·9875	4·9900	4·9925	4·9950	4·9975
50	5·0000	5·0025	5·0050	5·0075	5·0100	5·0125	5·0150	5·0175	5·0201	5·0226
51	5·0251	5·0276	5·0301	5·0326	5·0351	5·0376	5·0401	5·0426	5·0451	5·0476
52	5·0502	5·0527	5·0552	5·0577	5·0602	5·0627	5·0652	5·0677	5·0702	5·0728
53	5·0753	5·0778	5·0803	5·0828	5·0853	5·0878	5·0904	5·0929	5·0954	5·0979
54	5·1004	5·1030	5·1055	5·1080	5·1105	5·1130	5·1156	5·1181	5·1206	5·1231
55	5·1257	5·1282	5·1307	5·1332	5·1358	5·1383	5·1408	5·1434	5·1459	5·1484
56	5·1510	5·1535	5·1560	5·1586	5·1611	5·1637	5·1662	5·1687	5·1713	5·1738
57	5·1764	5·1789	5·1815	5·1840	5·1866	5·1891	5·1917	5·1942	5·1968	5·1993
58	5·2019	5·2045	5·2070	5·2096	5·2121	5·2147	5·2173	5·2198	5·2224	5·2250
59	5·2275	5·2301	5·2327	5·2353	5·2378	5·2404	5·2430	5·2456	5·2482	5·2508
60	5·2533	5·2559	5·2585	5·2611	5·2637	5·2663	5·2689	5·2715	5·2741	5·2767
61	5·2793	5·2819	5·2845	5·2871	5·2898	5·2924	5·2950	5·2976	5·3002	5·3029
62	5·3055	5·3081	5·3107	5·3134	5·3160	5·3186	5·3213	5·3239	5·3266	5·3292
63	5·3319	5·3345	5·3372	5·3398	5·3425	5·3451	5·3478	5·3505	5·3531	5·3558
64	5·3585	5·3611	5·3638	5·3665	5·3692	5·3719	5·3745	5·3772	5·3799	5·3826
65	5·3853	5·3880	5·3907	5·3934	5·3961	5·3989	5·4016	5·4043	5·4070	5·4097
66	5·4125	5·4152	5·4179	5·4207	5·4234	5·4261	5·4289	5·4316	5·4344	5·4372
67	5·4399	5·4427	5·4454	5·4482	5·4510	5·4538	5·4565	5·4593	5·4621	5·4649
68	5·4677	5·4705	5·4733	5·4761	5·4789	5·4817	5·4845	5·4874	5·4902	5·4930
69	5·4959	5·4987	5·5015	5·5044	5·5072	5·5101	5·5129	5·5158	5·5187	5·5215
70	5·5244	5·5273	5·5302	5·5330	5·5359	5·5388	5·5417	5·5446	5·5476	5·5505
71	5·5534	5·5563	5·5592	5·5622	5·5651	5·5681	5·5710	5·5740	5·5769	5·5799
72	5·5828	5·5858	5·5888	5·5918	5·5948	5·5978	5·6008	5·6038	5·6068	5·6098
73	5·6128	5·6158	5·6189	5·6219	5·6250	5·6280	5·6311	5·6341	5·6372	5·6403
74	5·6433	5·6464	5·6495	5·6526	5·6557	5·6588	5·6620	5·6651	5·6682	5·6713

Table I (*cont.*).

	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
75	5·6745	5·6776	5·6808	5·6840	5·6871	5·6903	5·6935	5·6967	5·6999	5·7031
76	5·7063	5·7095	5·7128	5·7160	5·7192	5·7225	5·7257	5·7290	5·7323	5·7356
77	5·7388	5·7421	5·7454	5·7488	5·7521	5·7554	5·7588	5·7621	5·7655	5·7688
78	5·7722	5·7756	5·7790	5·7824	5·7858	5·7892	5·7926	5·7961	5·7995	5·8030
79	5·8064	5·8099	5·8134	5·8169	5·8204	5·8239	5·8274	5·8310	5·8345	5·8381
80	5·8416	5·8452	5·8488	5·8524	5·8560	5·8596	5·8633	5·8669	5·8705	5·8742
81	5·8779	5·8816	5·8853	5·8890	5·8927	5·8965	5·9002	5·9040	5·9078	5·9116
82	5·9154	5·9192	5·9230	5·9269	5·9307	5·9346	5·9385	5·9424	5·9463	5·9502
83	5·9542	5·9581	5·9621	5·9661	5·9701	5·9741	5·9782	5·9822	5·9863	5·9904
84	5·9945	5·9986	6·0027	6·0069	6·0110	6·0152	6·0194	6·0237	6·0279	6·0322
85	6·0364	6·0407	6·0450	6·0494	6·0537	6·0581	6·0625	6·0669	6·0714	6·0758
86	6·0803	6·0848	6·0893	6·0939	6·0985	6·1031	6·1077	6·1123	6·1170	6·1217
87	6·1264	6·1311	6·1359	6·1407	6·1455	6·1503	6·1552	6·1601	6·1650	6·1700
88	6·1750	6·1800	6·1850	6·1901	6·1952	6·2004	6·2055	6·2107	6·2160	6·2212
89	6·2265	6·2319	6·2372	6·2426	6·2481	6·2536	6·2591	6·2646	6·2702	6·2759
90	6·2816	6·2873	6·2930	6·2988	6·3047	6·3106	6·3165	6·3225	6·3285	6·3346
91	6·3408	6·3469	6·3532	6·3595	6·3658	6·3722	6·3787	6·3852	6·3917	6·3984
92	6·4051	6·4118	6·4187	6·4255	6·4325	6·4395	6·4466	6·4538	6·4611	6·4684
93	6·4758	6·4833	6·4909	6·4985	6·5063	6·5141	6·5220	6·5301	6·5382	6·5464
94	6·5548	6·5632	6·5718	6·5805	6·5893	6·5982	6·6072	6·6164	6·6258	6·6352
95	6·6449	6·6546	6·6646	6·6747	6·6849	6·6954	6·7060	6·7169	6·7279	6·7392
96	6·7507	6·7624	6·7744	6·7866	6·7991	6·8119	6·8250	6·8384	6·85?2	6·8663
97	6·8808	6·8957	6·9110	6·9268	6·9431	6·9600	6·9774	6·9954	7·0141	7·0335

	0·00	0·01	0·02	0·03	0·04	0·05	0·06	0·07	0·08	0·09
98·0	7·0537	7·0558	7·0579	7·0600	7·0621	7·0642	7·0663	7·0684	7·0706	7·0727
98·1	7·0749	7·0770	7·0792	7·0814	7·0836	7·0858	7·0880	7·0902	7·0924	7·0947
98·2	7·0969	7·0992	7·1015	7·1038	7·1060	7·1084	7·1107	7·1130	7·1154	7·1177
98·3	7·1201	7·1224	7·1248	7·1272	7·1297	7·1321	7·1345	7·1370	7·1394	7·1419
98·4	7·1444	7·1469	7·1494	7·1520	7·1545	7·1571	7·1596	7·1622	7·1648	7·1675
98·5	7·1701	7·1727	7·1754	7·1781	7·1808	7·1835	7·1862	7·1890	7·1917	7·1945
98·6	7·1973	7·2001	7·2029	7·2058	7·2086	7·2115	7·2144	7·2173	7·2203	7·2232
98·7	7·2262	7·2292	7·2322	7·2353	7·2383	7·2414	7·2445	7·2476	7·2508	7·2539
98·8	7·2571	7·2603	7·2636	7·2668	7·2701	7·2734	7·2768	7·2801	7·2835	7·2869
98·9	7·2904	7·2938	7·2973	7·3009	7·3044	7·3080	7·3116	7·3152	7·3189	7·3226
99·0	7·3263	7·3301	7·3339	7·3378	7·3416	7·3455	7·3495	7·3535	7·3575	7·3615
99·1	7·3656	7·3698	7·3739	7·3781	7·3824	7·3867	7·3911	7·3954	7·3999	7·4044
99·2	7·4089	7·4135	7·4181	7·4228	7·4276	7·4324	7·4372	7·4422	7·4471	7·4522
99·3	7·4573	7·4624	7·4677	7·4730	7·4783	7·4838	7·4893	7·4949	7·5005	7·5063
99·4	7·5121	7·5181	7·5241	7·5302	7·5364	7·5427	7·5491	7·5556	7·5622	7·5690
99·5	7·5758	7·5828	7·5899	7·5972	7·6045	7·6121	7·6197	7·6276	7·6356	7·6437
99·6	7·6521	7·6606	7·6693	7·6783	7·6874	7·6968	7·7065	7·7164	7·7265	7·7370
99·7	7·7478	7·7589	7·7703	7·7821	7·7944	7·8070	7·8202	7·8338	7·8480	7·8627
99·8	7·8782	7·8943	7·9112	7·9291	7·9478	7·9677	8·9889	7·0114	8·1357	8·0618
99·9	8·0902	8·1214	8·1559	8·1947	8·2389	8·2905	8·3528	8·4316	8·5401	8·7190

The next step is to plot on the ordinate the probit of the expected dosage, inferred from the observed mortality, and on the abscissa some function of the amounts which were administered experimentally. These latter may be originally in terms of the concentrations of a toxic

substance in which the successive lots of organisms were immersed for a given time, a graded series of times of exposure to a fixed concentration of poison, doses administered individually at different units per gram of body weight, different concentrations of contact poison applied uniformly over the surface of the body, or in some other terms. When these units of measurement are plotted directly, the resulting curve is very seldom a straight line but is nearly always convex upwards, an effect which might have been anticipated from the markedly asymmetrical character of most sigmoid dosage-mortality curves.

Before discarding the normal curve as an adequate description of the variation between individuals in their susceptibility to a poison, let us question the assumption that the individual lethal dose is a satisfactory direct measure of susceptibility. The dosage units described above form an arithmetical scale of equal increments, and would not be a satisfactory index to the susceptibility if the structural or chemical constituents which determine the level of susceptibility of the individual in respect to a given drug were not to increase or decrease by equal additive increments. It was pointed out as long ago as 1879 by Galton that in biological material the variation often shows a geometrical rather than an arithmetical distribution, an observation which has been confirmed by several investigators in respect to toxicological characteristics. If, therefore, the changes in the substances or structures which determine susceptibility, whatever may be their nature, were ordinarily proportional in type, then they would be symmetrically distributed not on an arithmetical scale of individual lethal doses but only on a logarithmic scale. This possibility may be tested by converting the observed dosages to logarithms and again plotting the dosages inferred from mortality or probits against those secured experimentally. With this transformation, a straight line does result in a great majority of the cases which have been tested. Before the method of inferring "expected" doses from the percentage kills had been devised, Trevan (13) and others had shown that per cent. mortality plotted against the logarithm of the dose frequently results in symmetrical sigmoid curves, while in the descriptions (1, 5, 6) of the double transformation, many more cases were cited in which the logarithm of the individual dose was an adequate measure of susceptibility.

If the transformation of dosages to logarithms completes the transformation of the dosage-mortality curve to a straight line because it is an index to the inherent susceptibility of the individual animal to the poison, the poisoning process could be considered as an example of the Weber-Fechner law. This implies, however, a direct proportionality

between the concentration of the poison in the dose administered and the amount of poison fixed by the essential tissues of the animal, and there is no evidence in support of such a direct relationship. Moreover, if the poisoning of the individual multicellular animal can be attributed to the death of a certain proportion of its cells, then the susceptibility of the animal as a whole will be determined by the average susceptibility of its essential cells. Even though the susceptibility of these ultimate units, the cells, may vary geometrically rather than arithmetically, so that their distribution is highly asymmetrical, it is probable that the average susceptibilities of populations of these unit cells, the individual animals, are symmetrically and normally distributed, if we may judge from general statistical experience. *A priori*, therefore, the individual animals in a stock may be expected to vary normally in their susceptibility to a specific poison, since each animal is an "average" of its component cells. The justification of the logarithmic transformation may be sought in the relation between the dosage administered and the amount of poison fixed by the essential cells or tissues, rather than in the Weber-Fechner law.

The fixation of a drug or poison seems to be primarily a phenomenon of adsorption (2), and one of the two principal formulae for describing this process is that proposed by Freundlich. Freundlich's empirical formula is

$$KC^{\frac{1}{n}} = \frac{x}{m},$$

where, for our purposes, C may be equated to the concentration of the drug (or dosage), x = the amount fixed in the organism, m = the mass of adsorbing constituents within the organism, and K and n are constants. If the variation in susceptibility is attributed primarily to the reactions which follow the fixation of the poison, m will be constant from one individual test animal to the next. By combining constants, the Freundlich formula may be reduced to

$$\log C = n \log x + K',$$

from which it is apparent that there is a linear relation between the logarithm of the concentration (or dosage) and the logarithm of the amount fixed by the cells of the animal. The observed logarithmic conversion of the dosage-mortality curve is not due, therefore, to our using as the true individual lethal dose the amount fixed in the tissue, if this is related to the concentration by the Freundlich formula.

In many instances another adsorption equation, that proposed by Langmuir, has fitted the biological data on the fixation of drugs more satisfactorily than the Freundlich formula. Moreover, it is better

grounded theoretically. Langmuir's adsorption equation is given by Clark as

$$kx^n = \frac{y}{100-y},$$

where x = concentration of the drug, y = percentage of the maximum amount of drug which can be fixed by the cell, n is determined by the molecular state of the fixed drug as compared with its state before adsorption and is usually 1 or 2, and k is a constant. In order to compare the amount (percentage) fixed with the logarithm of dosage (y with log x), y was calculated for each of a series of hypothetical values of x when $k = 0{\cdot}0625$ and $n = 1$. A diagram of y against log x gave a sigmoid curve, symmetrical about 50 per cent. fixation, and very nearly a straight line between 20 and 80 per cent. fixation. If 100 per cent. kill on the dosage-mortality curve were to correspond to 100 per cent. fixation of the poison by the tissues of the experimental animals, all cases in which the logarithm-probit plot showed a straight line over a range of dosages that included kills of 90 per cent. and better—as very many of them do—would definitely rule out the Langmuir adsorption equation as an explanation. However, investigations have shown that live tissue is capable of adsorbing much more of the chemical than the amount which produces the maximum effect, in this case, the subsequent death of all individuals. If all experimental animals were to die before a dosage is reached which produces 80 per cent. or more adsorption, the logarithm-probit transformation would still be consistent with an interpretation based on the Langmuir adsorption equation, so far as the middle and higher kills—and dosages—are concerned.

The application of the Langmuir equation to the lower dosages presents a more involved problem. Usually the logarithm-probit plot of the dosage-mortality curve can be fitted by a single straight line over the entire range of mortalities, and it may then be reasonable to assume that the amount of poison fixed must exceed a threshold value of 20 per cent. of the maximum before even the most susceptible individuals will be killed. However, in many cases the transformed dosage-mortality line agrees with the higher kills very satisfactorily but indicates too small a mortality below 20 to 35 per cent. kill. At its lower end the otherwise straight line would need to bend up if it is to fit the entire range of observations. The similarity of this change in slope to the lower end of the theoretical curve secured by plotting the percentage of drug fixed against the logarithm of dosage suggests that in these cases the adsorption is less than 20 per cent. of the maximum at the threshold concentration of the

poison, and that if the observed dosage could be converted to the amount fixed by means of the Langmuir equation, a single straight line would be obtained by the use of probits.

Without measurements of the amount of poison adsorbed, the Langmuir equation cannot be tested critically, but an approximate graphic analysis has been applied successfully to several series of fumigation tests in which at the lower dosages there was a change of slope upon the logarithm-probit co-ordinates. For each series of points, the mortality in probits could be fitted satisfactorily (as in Fig. 3) with two intersecting straight lines when plotted against the logarithm of the concentration of the fumigant, the bend between the two lines being acute enough for there to be no hesitation in deciding which observations should be grouped. From a graphic comparison with the theoretical curve mentioned above (percentage fixed *v.* log. dosage) of the angle at which these two lines intersected, the observed concentrations were converted to terms of the percentages of maximum adsorption, and when the observed mortalities in probits were replotted against these theoretical dosage units, the data for each poison could be fitted adequately by a single straight line. This transformation of dosage to per cent. adsorbed introduces two additional constants, one attributable to the maximum adsorption which produces no lethal effect and the other to the minimum adsorption which is invariably fatal. On mathematical grounds alone, therefore, the agreement between observations and fitted curve should be as good as when two intersecting straight lines, also involving four constants, are fitted to the same data.

The use of the Langmuir equation need not necessarily eliminate the change in slope that is observed on occasion at the lower dosages upon the logarithm-probit plot. If a minimum of 15 to 20 per cent. adsorption were required to effect a kill, for example, the rectilinearity in the main portion of the curve and the change in slope at its lower end would be the same whether log. dosage or per cent. of maximum adsorption were plotted along the base. Since there is good experimental evidence, as in the case of protective stupefaction with hydrocyanic acid (10), that low concentrations frequently have an action qualitatively different from that of the higher dosages, the change in slope may very well have a biological reality and not be merely a mathematical artifact. Clark[1] thinks that "this break is a fairly common phenomenon. It suggests to me that the characteristic curve besides measuring individual variation also is affected by some relationship between concentration and amount of

[1] Personal communication.

action." Since without another kind of experimental data even an approximate conversion of dosage into percentage adsorption is possible only when there is a change in slope on the logarithm-probit co-ordinates, and may then be of doubtful theoretical significance, it is preferable at present to use the logarithm of the individual lethal dose as a measure of susceptibility with the understanding that its use can be interpreted in terms other than those of the Weber-Fechner law.

The above procedure should not be confused with another fundamentally different application of the Langmuir adsorption equation, which is hyperbolic, to similar data. If dosage is converted to logarithms, the percentage adsorption plotted against it is a sigmoid curve symmetrical about the 50 per cent. point, as has been described, and the percentage mortality plotted against it is a very similar sigmoid curve. In one case, Clark (p. 157) has considered these two measures as if they were identical, or the percentage mortality a direct measure of percentage adsorption. Yet elsewhere he has described experiments which show that adsorption frequently continues after the point is reached which produces maximum effect, and this possibility alone demonstrates that they are distinct[1]. Even if certain dosage-mortality data were fitted adequately by this use of the hyperbolic equation, they could still be considered from the "statistical" viewpoint adopted here. The abscissa, the logarithm of the dose, is the same in both methods of transformation, while the ordinate in both may be assumed to represent sigmoid frequency distributions which are experimentally inseparable between kills of 15 and 85 per cent.

[1] In a recent letter to *Nature* (CXXIV, 323), H. H. Shepard applies an equivalent method to original data that are similar to those quoted here in Table IV, except that he uses the dosage directly instead of the logarithm of the dose. When his data and fitted curve are plotted in a rectilinear form $\left(\text{logarithm of } \frac{\text{per cent. killed}}{\text{per cent. surviving}} \text{ against concentration}\right)$, it is apparent that the observed values are still distributed in a sigmoid manner about the straight line, despite his use of the hyperbola. However, when the probit values for percentage mortality are plotted against dosages which have been converted to hypothetical "percentages of poison adsorbed" $\left(\text{by means of the equation } kx^n = \frac{y}{100-y}\right)$, a very satisfactory fit can be obtained with $\log k = -18{\cdot}2$ and $n = 10{\cdot}2$. It should be noted that while Shepard used the same species of insect, the same poison, and apparently the same laboratory technique as in the data quoted here from Strand, his results agree in average susceptibility (the median lethal dose), but show a significantly larger range of variability within the population. Shepard apparently has totalled many individual experiments for each dosage, and if, over the period which this required, the average susceptibility in his stock of beetles had fluctuated as much as 10 to 15 per cent., the variability within his population at any one time might well have been consistent with Strand's earlier results which are quoted here.

They differ in mathematical treatment only in that the frequency distribution of susceptibilities in the interpretation followed here is assumed to be normal, while in the hyperbolic interpretation it is that of the z distribution (3).

On the basis of the above assumptions, we may proceed at once to a consideration of how to calculate the best-fitting dosage-mortality curve. The first step is to transform each percentage kill to its probit (Table I) and convert each dosage to its logarithm. The percentage kill will not, however, be the same as the percentage dead if there is an appreciable mortality among the untreated controls or checks. A convenient way of computing the percentage kill in such a case is to multiply the number of individuals used in a particular test by the proportion alive in the untreated controls, which gives the net total of organisms actually exposed to the action of the poison. When the number surviving the treatment is subtracted from this net total, the difference is the number killed, and the number killed (multiplied by 100), divided by the net number exposed is, of course, the percentage killed. The probit, or dosage inferred from mortality, is then plotted on co-ordinate paper against the logarithm of the dosage that was administered experimentally. Inspection of these points with the aid of a straight edge, such as the side of a celluloid triangle, will show very quickly whether they define a straight line over most or the whole of the range of dosages. In cases where the data for the lower dosages seems to be discordant with the straight line that is consistent with the rest of the observations, the straight line is fitted only to the higher dosages. A few cases may occur in which the points seem to be smoothly curvilinear throughout, and in such instances some other function of dosage should be tried which seems to have a toxicological significance. Having determined the range of dosage over which a rectilinear relation seems to hold good, a straight line is drawn through the points.

II. The provisional regression line.

The first estimate of the transformed dosage-mortality curve, which we will call the provisional regression line, is ordinarily not calculated, but represents the best judgment of the experimenter. When the data are consistent, the graphic provisional curve will often come surprisingly close to the corrected curve obtained after computation. Occasionally, however, the observations may be so scattered that the experimenter will prefer to calculate even the provisional regression line. The simplest procedure in this case is to give each experiment a weight of 1 and use

equations (3)–(6) of the next section. In other cases the data may be so uniform that the initial line will serve the needs of the experimenter. Usually, however, the graphic approximation will want correction, and to obtain this corrected curve we compute what is known in statistics as the regression line. The regression line in our case will show the probit which corresponds to any given logarithm of dosage as accurately as this relation can be determined from the experimental data used in its computation.

The provisional regression line serves two essential purposes: it determines what probit values are to be assigned to observed mortalities of 0 and 100 per cent., and it specifies what relative weights are to be given to the separate observations in a series.

(1) *Probit values for* 0 *and* 100 *percentage kills.* Although toxicological tests frequently include at one limit small dosages which kill no individuals or at the other limit large dosages which kill all individuals, these values cannot be listed in the standard table of probits (Table I). By means of the provisional regression line, the information in such observations may still be used in determining the corrected regression line. This possibility follows from our basic assumption that the distribution of susceptibility is normal and the fact that while the curve of the normal distribution (Fig. 2) approaches infinitely close to 100 per cent. kill—considering for convenience only the upper limit—it never quite reaches it mathematically at any finite dosage. Within the range of dosages and numbers of organisms ordinarily used in a laboratory test, this mathematical postulate agrees satisfactorily with the biological reality. Thus the smallest dosage giving 100 per cent. kill will be smaller in an experimental series with 30 organisms per dose than in a repetition of the same series using 300 specimens per dose, since in the larger numbers of the second case there is a greater chance of including the less susceptible individuals in each treatment. The mortality in probits that would be expected if we were dealing with very large numbers of organisms is given approximately by an extension of the provisional regression line over the range of these higher dosages. In a note on "The case of zero survivors," appended to the present paper, R. A. Fisher points out that when the number in the class of survivors is small, the theory of large samples breaks down if applied to the restricted numbers used in toxicological tests. He shows, however, that when zero survivors are observed the probit term for 100 per cent. kill may be derived by the method of maximum likelihood as a difference, which is added to the expected value in probits given by the provisional regression line.

An alternative method for plotting 100 per cent. kills in terms of probits or their equivalents has been proposed by Gaddum (5). His value is based upon the number of animals exposed to the treatment, but is not used whenever it indicates a smaller mortality than would be expected from the approximate regression line at this dosage. The method proposed here avoids this limitation and is mathematically the more exact.

The procedure to be followed in securing the probit value for 100 per cent. kills may be outlined briefly. The probit given by the extended provisional regression line is read from the graph at the logarithm for the dosage from which none survived. This probit is then entered in column 1 of Table II and the required probit for the observed kill is found in column 3. First differences are given in column 4 for convenience in interpolation if the provisional regression line has been read to 0·01 probit. These values will always fall above the provisional line as would be expected since no survivors were observed, and should be included in computing the corrected curve with a weight determined as described in the next section. The omission of such terms tends to bias the final regression line by exaggerating the number of survivors to be expected.

The same method is available, of course, at the opposite end of the curve, at dosages which fail to kill any individuals, except that the correction in column 2 of Table II is then subtracted from the probit value given by the provisional line. The correction to use in such a case will be that for the probit in column 1 which is as much greater than 5 as the one read from the provisional line is less than 5. These smaller dosages, however, are usually of little interest, and it frequently happens that, below 25 per cent. kill, the regression line which forms an adequate fit above that point is no longer applicable.

(2) *Weights for fitting the regression line.* The reliability of the probit for an observed percentage kill depends not only on how many individuals were counted to determine this percentage but also upon the corresponding probit value of the regression line, or, in actual practice, upon that of the provisional regression line. It is customary to consider the reliability of a percentage as proportional to the number of individuals tested, and the justification for thus weighting by the number of individuals rather than by the square root of the number of individuals is that the reliability of a measure is inversely proportional to the square of its standard error—the variance—and not to the standard error itself. The variance, in turn, is a function not only of the number of cases but also of several other factors, and it is these other factors which it is necessary to take into account. The principle of giving to individual

Table II.

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

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

observations weights that are proportional to their statistical reliability follows that described by Thompson (12) in his analysis of an experiment in sensory discrimination.

The required standard error is shown graphically on the cumulative form of the normal frequency distribution of Fig. 2, in which p, the pro-

portion killed, is plotted on the ordinate against x, the inferred dosage in probits, on the abscissa. The position of the paired horizontal lines cutting the ordinate on either side of 50 and 95 per cent. kill was calculated from the usual formula for the standard error of a proportion, $\sigma = \sqrt{\frac{pq}{N}}$, where p is the proportion killed, $q = 1 - p$, and $N = 100$ individuals exposed to treatment. However, in our transformed dosage-mortality curve, these percentages have been transformed to probits, which are given along the base of the figure, so that the standard error (and variance) which we need is not that for a proportion, p, but that for the corresponding inferred dosage or probit, x, a quantity equivalent to what statisticians call the percentile. From the points of intersection with the curve in Fig. 2 of the standard errors of the proportions (shown by the paired horizontal lines), we will draw paired vertical lines to cut the base at the standard errors of the probits (or percentiles) corresponding to these two proportions of 0·50 and 0·95. While the standard error of p is a maximum at 50 per cent. kill and diminishes toward either 0 or 100 per cent., that of the probit is smallest at 50 per cent. and increases toward either limit. Hence the accuracy of a given probit will increase as it approaches 50 per cent. kill.

The formula for the variance of a percentile is given by Kelley (7) as

$$\frac{\sigma^2 pq}{z^2 N},$$

where σ is the standard deviation, z is the ordinate of the normal curve (see Fig. 1) and is given in tables of the probability integral, and the other terms have their previous significance. This will also be the variance for the probit of a single observed percentage mortality, but since the probit is already in terms of the standard deviation, σ^2 is always equal to 1 and the variance of a probit may be simplified to the form

$$\frac{pq}{Nz^2}.$$

In order, therefore, to give each observation a weight proportional to its true reliability, instead of multiplying it by N, we will multiply by the reciprocal of the variance as our weight, w. Hence

$$w = N\left(\frac{z^2}{pq}\right), \qquad \text{.....(1)}$$

where N is the number of organisms exposed to a given dosage of poison and z, p, and q have their previous significance as functions of the normal curve, which, in this case, are fixed by the probit value of the provisional

regression line at the same dosage. The term $\frac{z^2}{pq}$ we will call the weighting coefficient. It has been computed for each 0·1 probit within the useful range of probit values and is given in Table III (column 6). The procedure for determining the correct weights to be used in calculating the corrected regression line is thus made quite easy. After the provisional regression line has been drawn through the plotted points of the experimental series as described, the probit given by this line for the log. dosage used in each determination is read from the graph to the nearest 0·1 (or 0·01) probit and by reference to Table III is transformed directly to the weighting

Table III.

Weighting coefficients used in computing the dosage-mortality curve in terms of probits. The probit for the expected kill is read to the nearest 0·1 *or* 0·01 *from the provisional, graphic dosage-mortality line at the dosage used in a given test. Entering this in column* 1 *below, the weighting coefficient is read from column* 3 *(interpolating from the first differences in column* 4 *if the line has been read to* 0·01 *probit) and multiplied by the total number of organisms to secure the weight (w) of the test for use in computing the final curve. The weighting coefficients in column* 3 *have been abbreviated for ease of calculation from the five-place values of* z^2/pq *in column* 6. *Column* 5 *shows the relative number of individuals which must be used at different expected mortalities if all observations are to be weighted equally; while column* 2 *gives the percentage mortalities corresponding to the probits in column* 1.

Curve value or probit for expected kill	Expected percentage kill	Weighting coefficient	First differences	Relative no. of individuals for equal weights	$\frac{z^2}{pq}$
1·5	0·023	0·0033		1947	0·00327
			12		
1·6	0·034	0·0045		1412	0·00451
			16		
1·7	0·048	0·0061		1037	0·00614
			22		
1·8	0·069	0·0083		769	0·00828
			27		
1·9	0·097	0·0110		577	0·01104
			36		
2·0	0·135	0·0146		437	0·01457
			44		
2·1	0·187	0·0190		334	0·01903
			56		
2·2	0·256	0·0246		259	0·02459
			68		
2·3	0·347	0·0314		202	0·03143
			84		
2·4	0·466	0·0398		160	0·03977
			102		
2·5	0·621	0·050		128	0·04979
			12		
2·6	0·820	0·062		103	0·06169
			14		
2·7	1·072	0·076		84	0·07563
			16		
2·8	1·390	0·092		69	0·09179
			18		
2·9	1·786	0·110		58	0·11026
			21		
3·0	2·275	0·131		49	0·13112
			23		
3·1	2·872	0·154		41	0·15436
			26		
3·2	3·593	0·180		35	0·17994
			28		
3·3	4·457	0·208		31	0·20773
			30		
3·4	5·480	0·238		27	0·23753
			31		

Table III (*cont.*).

Curve value or probit for expected kill	Expected percentage kill	Weighting coefficient	First differences	Relative no. of individuals for equal weights	$\frac{z^2}{pq}$
3·5	6·681	0·269	31	24	0·26907
3·6	8·076	0·302	33	21	0·30199
3·7	9·680	0·336	34	19	0·33589
3·8	11·507	0·370	34	17	0·37031
3·9	13·567	0·405	35	16	0·40474
4·0	15·866	0·439	34	15	0·43863
4·1	18·406	0·471	32	14	0·47144
4·2	21·186	0·503	32	13	0·50260
4·3	24·196	0·532	29	12	0·53159
4·4	27·425	0·558	26	11	0·55788
4·5	30·854	0·581	23	11	0·58099
4·6	34·458	0·601	20	11	0·60052
4·7	38·209	0·616	15	10	0·61609
4·8	42·074	0·627	11	10	0·62742
4·9	46·017	0·634	7	10	0·63431
5·0	50·000	0·637	3	10	0·63662
5·1	53·983	0·634	3	10	0·63431
5·2	57·926	0·627	7	10	0·62741
5·3	61·791	0·616	11	10	0·61609
5·4	65·542	0·601	15	11	0·60052
5·5	69·146	0·581	20	11	0·58099
5·6	72·575	0·558	23	11	0·55788
5·7	75·804	0·532	26	12	0·53159
5·8	78·814	0·503	29	13	0·50260
5·9	81·594	0·471	32	14	0·47144
6·0	84·134	0·439	32	15	0·43863
6·1	86·433	0·405	34	16	0·40474
6·2	88·493	0·370	35	17	0·37031
6·3	90·320	0·336	34	19	0·33589
6·4	91·924	0·302	34	21	0·30199
6·5	93·319	0·269	33	24	0·26907
6·6	94·520	0·238	31	27	0·23753
6·7	95·543	0·208	30	31	0·20773
6·8	96·407	0·180	28	35	0·17994
6·9	97·128	0·154	26	41	0·15436
7·0	97·725	0·131	23	49	0·13112
7·1	98·214	0·110	21	58	0·11026
7·2	98·610	0·092	18	69	0·09179
7·3	98·928	0·076	16	84	0·07563
7·4	99·180	0·062	14	103	0·06169
7·5	99·379	0·050	12	128	0·04979
7·6	99·534	0·0398	102	160	0·03977
7·7	99·653	0·0314	84	202	0·03143
7·8	99·744	0·0246	68	259	0·02459
7·9	99·813	0·0190	56	334	0·01903
8·0	99·865	0·0146	44	437	0·01457
8·1	99·903	0·0110	36	577	0·01104
8·2	99·931	0·0083	27	769	0·00828
8·3	99·952	0·0061	22	1037	0·00614
8·4	99·966	0·0045	16	1412	0·00451
8·5	99·977	0·00327	123	1947	0·00327
8·6	99·984	0·00235	92	2709	0·00235
8·7	99·989	0·00167	68	3812	0·00167
8·8	99·993	0·00118	49	5395	0·00118
8·9	99·995	0·00082	36	7764	0·00082

coefficient. The weighting coefficient will be sufficiently accurate if read only to the first two or three significant figures as given in column 3 of Table III, interpolating from first differences (column 4) if the provisional curve justifies an estimate to the nearest 0·01 probit. Each weighting coefficient then is multiplied (most conveniently on the slide rule) by the number, N, in the test to secure its correct weight, w, for calculating the dosage-mortality curve.

It has been specified, without further explanation, that the weighting coefficient is determined from the provisional regression line rather than directly from each separate observation. With this important exception, the weighting coefficient described above is equivalent to that proposed by Gaddum (5) and by Hemmingsen (6) for the same purpose. Gaddum has based his coefficients directly upon the separate p's observed experimentally, so that above 50 per cent. kill the tests in which the mortality fell short of that expected would be weighted more heavily than those in which the mortality exceeded expectation. Conversely, below 50 per cent. kill. the excessive mortalities would carry greater weight than the deficient mortalities. Together these errors would bias the fitted regression line toward the horizontal. By using as a standard the probit (or mortality) determined from the experiment as a whole, instead of that shown by a single sample, the present weighting coefficients not only avoid this biasing error but give a suitable basis for comparing different dosage-mortality curves and for measuring their accuracy. Still another, though similar, weighting method has been used by McCallan and Wilcoxon (8) in the reciprocal of their "error in concentration."

In planning an experiment so as to secure equally reliable results at all dosages and thereby avoid the necessity of weighting—with a corresponding simplification in the computations—more individuals should be used at high and low dosages than at intermediate ones. Equalisation will result if the experimenter treats with the dosage at each expected kill some multiple of the number of individuals listed in the fifth column of Table III. This shows that it takes three times as many animals to get the same accuracy at 95 per cent. kill as at 50 per cent. kill and nearly ten times as many at 99 per cent. as at 50 per cent. It would not justify the procedure followed in the experiments reported by Hemmingsen ((6), p. 40), in which nearly twice as many mice were used for the two middle of four concentrations of insulin as for the largest and smallest.

In order that each step may be clearly understood, a numerical example has been selected from Strand's (11) experiments with *Tribolium confusum*. Two of his series, designated as I and II, give the mortality of the adult flour beetle after five hours'

exposure to gaseous carbon disulphide, and these will serve to illustrate the various procedures of the present paper. There was no appreciable mortality in the controls, so that this factor did not need correction. The original data are given in the first four columns of Table IV. The next column, x, is secured from column 3 by reference to a table of common logarithms. With the exception of the probit values corresponding to 100 per cent. kill, the sixth column, y, gives the percentages in terms of the probits of Table I. The observed values for x and y were then plotted on cross-section paper (Fig. 3), and it is apparent from inspection that the two series, I and II, did not differ

Table IV.

Procedure for fitting the transformed dosage-mortality curve to kills of Tribolium confusum *following 5-hour exposures to known concentrations of carbon disulphide. The computations in columns 7 to 10 and at the end of the table show the steps for fitting the regression line to the upper range of dosages from the data of both series (Fig. 3). Data from Strand (11).*

	Original data								
Series no.	Total no. of insects	CS_2 mg. per litre	% kill	x log. of dosage	y Probit kill	Weighting coefficient	w Weight	wx Column 8 × Column 5	wy Column 8 × Column 6
I	29	49·06	6·9	1·6907	3·517	—	—	—	—
	30	52·99	23·3	1·7242	4·271	—	—	—	—
	28	56·91	32·9	1·7552	4·557	0·555	15·5	27·20560	70·6335
	27	60·84	51·9	1·7842	5·048	0·633	17·1	30·50982	86·3208
	30	64·76	76·7	1·8113	5·729	0·500	15·0	27·16950	85·9350
	31	68·69	93·6	1·8369	6·522	0·292	9·1	16·71579	59·3502
	30	72·61	96·7	1·8610	6·838	0·125	3·8	7·07180	25·9844
	29	76·54	100·0	1·8839	7·952	0·0398	1·2	2·26068	9·5424
II	30	49·06	13·3	1·6907	3·888	—	—	—	—
	30	52·99	20·0	1·7242	4·158	—	—	—	—
	34	56·91	26·5	1·7552	4·372	0·555	18·9	33·17328	82·6308
	29	60·84	48·3	1·7842	4·957	0·633	18·4	32·82928	91·2088
	33	64·76	87·9	1·8113	6·170	0·500	16·5	29·88645	101·8050
	28	68·69	85·7	1·8369	6·067	0·292	8·2	15·06258	49·7494
	32	72·61	100·0	1·8610	7·447	0·125	4·0	7·44400	29·7880
	31	76·54	100·0	1·8839	7·952	0·0398	1·2	2·26068	9·5424

$S(w) = 128{\cdot}9$

$S(wx) = 231{\cdot}58946$

$\bar{x} = \dfrac{S(wx)}{S(w)} = 1{\cdot}796660$

$S(wy) = 702{\cdot}4907$

$\bar{y} = a = \dfrac{S(wy)}{S(w)} = 5{\cdot}449889$

$S(wx^2) = 416{\cdot}237456$

$A = S(wx^2) - \bar{x}S(wx) = 0{\cdot}1499627$

$S(wxy) = 1265{\cdot}962620$

$S(wxy) - \bar{x}S(wy) = 3{\cdot}825756$

$b = \dfrac{S(wxy) - \bar{x}S(wy)}{A} = 25{\cdot}511392$

$S(wy^2) = 3931{\cdot}653109$

$S(wy^2) - \bar{y}S(wy) = 103{\cdot}156728$

$\chi^2 = [S(wy^2) - \bar{y}S(wy)] - b[S(wxy) - \bar{x}S(wy)] = 5{\cdot}5564$

n' (see text) $= 9$

$n = n' - 2 = 7$

$V(a) = \dfrac{1}{S(w)} = 0{\cdot}007758$

$V(b) = \dfrac{1}{A} = 6{\cdot}668327$

t (at $P = 0{\cdot}05$) $= 2{\cdot}365$

consistently. In comparison with the remaining observations, the two lowest concentrations gave an exceptionally high kill. Over the remaining concentrations, the plotted values seemed to form a moderately straight line, so that the data were handled as two separate sets, only the results at 56·91 mg. of CS_2 per litre being included in both sets. The provisional regression lines were drawn in with the aid of a straight edge, but these provisional curves, indicated by the broken lines, agreed quite

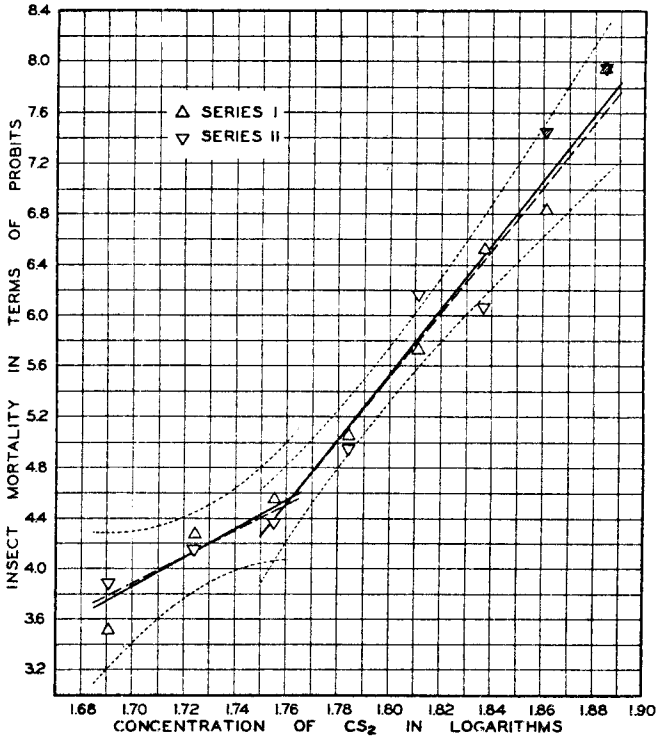

Fig. 3. A transformed dosage-mortality curve, showing the effect upon adult flour beetles of 5-hour exposures to different concentrations of gaseous carbon disulphide. The broken straight regression lines were placed graphically by inspection, the solid ones by computation, while the dotted curved lines show the limits within which the solid lines have been determined by the data. The shaded triangles represent treatments from which no beetles survived. Data from Strand(11).

closely with those arrived at by computation, the solid lines, in both the upper and the lower range of dosages.

Restricting our attention for the moment to the more important, upper range of dosages, the approximate curve was used first to secure probit values for 100 per cent. kills. We find that at a concentration of 72·61 mg. of CS_2 per litre, there was 1 survivor in Series I but 0 survivors in Series II, while no survivors were found in either series at 76·54 mg. per litre. The provisional curve showed that at a log. concentration of 1·8610 (72·61 mg.), 7·03 probits was expected and at a 1·8839 log. concentration

(76·54 mg.), 7·61 probits. Entering these values in column 1 of Table II, the two required probit values of 7·447 and 7·952 were obtained from column 3, using the first differences of column 4 for interpolation.

From the original plot of the provisional curve (on millimetre cross-section paper), the probit for each observed dosage could be read without difficulty to the nearest 0·01 probit. These were then entered directly in column 1 of Table III to secure the weighting coefficients from column 3 of the same table, interpolating with the aid of the adjoining column of first differences. The weighting coefficients so obtained were written down in column 7 of Table IV and multiplied on the slide rule by the corresponding number of insects (column 2) to determine the true weights in column 8. The last two columns of Table IV contain the products x multiplied by w, and y multiplied by w.

III. The computation of the regression line.

In toxicological experiments of the type which we have been considering, the mortality among a limited number of organisms is measured after treatment with known amounts of a toxic agent. These results have significance primarily because they form a sample from an infinitely larger group of organisms for which we are interested in determining the toxicological relationships. The fitting of a dosage-mortality curve is an attempt to infer from a given experiment the conditions obtaining in a class or species of organisms, and the calculated regression line of the dosage-probit diagram is the most accurate estimate which can be drawn from the data, granted that our basic assumptions are correct. In some cases it will be very near the first graphic approximation which has already been described, but oftentimes it will represent a rather important correction to this initial estimate, especially when the material is variable and fitting by eye less reliable. Moreover, in a calculated regression line, each separate observation can be weighted accurately, as has been shown, and the limits determined within which will lie the true curve for an infinitely larger population.

In describing the arithmetical procedure of fitting, the methods and symbols employed by Fisher(4) have been adapted to the present purposes. Short-cut methods, suitable for use with a calculating machine, are described. With a machine, these should enable one to fit the regression line without previous experience.

The formula for the regression line may be expressed as

$$Y = a + b\,(X - \bar{x}), \qquad \text{.....(2)}$$

where, in this case, Y is the mortality in probits on the regression line (or transformed dosage-mortality curve) which corresponds to any given dosage X, usually expressed in logarithms; $a = \bar{y}$ = numerically the average probit for all determinations in that part of the experiment which is being

fitted by a straight line; $\bar{x}$ is the average of the dosages administered (in logarithms) for the same section of data; and b is the regression coefficient or the slope of the line, the amount by which the probit of mortality is increased for every unit increase in log. dosage. It is necessary, therefore, to calculate from the experimental data the quantities $\bar{x}$, $\bar{y}$, and b. The formulae are as follows:

$$\bar{x} = \frac{S(wx)}{S(w)}, \qquad \text{.....(3)}$$

$$\bar{y} = \frac{S(wy)}{S(w)}, \qquad \text{.....(4)}$$

$$b = \frac{S(wxy) - \bar{x}S(wy)}{A}, \qquad \text{.....(5)}$$

$$A = S(wx^2) - \bar{x}S(wx), \qquad \text{.....(6)}$$

where the symbols are defined as:

S = "the sum of" and indicates that all quantities of the type in the brackets after the S are to be added,

w = weight of a given observation, the product of the weighting coefficient multiplied by the number of killed plus survived,

x = a function of the dosage administered experimentally, usually its logarithm, and

y = the probit corresponding to the observed percentage mortality.

The position of the regression line, in the sense in which we will use the term, is determined by $\bar{x}$ and $\bar{y}$, since it must pass through the point on the diagram given by these two means. They fix the degree of susceptibility to a toxic agent shown by the population as a whole. From a statistical viewpoint, b is the slope or the tangent of the angle with which the regression line will pass through the point established by $\bar{x}$ and $\bar{y}$; from a biological viewpoint, b measures how closely the individual organisms in the experiment agree with one another in their sensitivity to the toxic agent. It is convenient to express this toxicological characteristic as the percentage increase in dosage that is required to increase kill by one probit. This is the ratio of $100 \log_e 10$ to b, $\frac{230{\cdot}26}{b}$.

Returning to our numerical example, the solution of equations (3)–(6) has been given at the bottom of Table IV in the order which has been found the most convenient. The first, second, and fourth quantities are the totals of the last three columns of the table, while the two means were determined in order, without clearing the lower dials of the calculator, when the totals first appeared (in machines such as the Monroe and the Marchant). $S(wx^2)$ was obtained by placing wx on the keyboard of the calculator and multiplying by the corresponding x, then clearing the keyboard and upper dials

and repeating the process with the next pair of values until the total of the products, $S(wx^2)$, had been accumulated in the lower dials. Leaving this sum in the lower dials, $S(wx)$ was placed on the keyboard and subtracted $\bar{x}$ times to secure A. Repeating the process with wx on the keyboard and multiplying this time by y, the sum, $S(wxy)$, was obtained directly. From $S(wxy)$ in the lower dials, $S(wy)$ on the keyboard was subtracted $\bar{x}$ times to secure the next term, which, in turn, was divided by A to obtain the regression coefficient, b. In checking the arithmetic of these various operations, other short-cuts will soon suggest themselves for facilitating the work and reducing the possibility of error. It is important in this method that computations be carried out to six or more significant figures in the means and regression coefficient in order to insure sufficient accuracy throughout. From $\bar{x}$, $\bar{y}$, and b the equation of the corrected regression line was solved as $Y = 5{\cdot}450 + 25{\cdot}51\ (X - 1{\cdot}7967)$, holding for concentrations of carbon disulphide above approximately 57·8 mg. per litre of air. In this range, an increase in dosage of 9·03 per cent. (230·26/25·5114) increased kill by 1 probit.

The change in slope at a kill of about 33 per cent. (Fig. 3) is a frequent phenomenon for which no explanation will be attempted here. A separate curve has been calculated for the lower concentrations, including the smallest dosage of the main curve. The regression coefficient, b, was less than one-half that for the higher dosages. Usually this lower section of the toxicity curve will be of too little practical or theoretical importance to warrant calculating its equation, and it may be questioned whether a straight line is the correct relationship when the mortality below 25 to 35 per cent. kill differs from the rectilinearity of the higher dosages. Assuming a straight line in the present case, the regression equation was $Y = 4{\cdot}186 + 11{\cdot}35\ (X - 1{\cdot}7286)$.

The two experimental series have been listed separately, although the same dosages were used in Series I and in Series II. If the number of living and dead for each dosage had been combined before calculating the percentage kill and transforming to probits, the regression equation would have been determined from half as many separate observations. The result should be practically the same. Tested arithmetically, the new equation, $Y = 5{\cdot}436 + 25{\cdot}33\ (X - 1{\cdot}7967)$, differed so slightly that both regression lines could not be shown in Fig. 3. When it is evident from the similarity of different experimental series that the stocks of test animals are the same, the results at each separate dosage may be combined into a single percentage and probit for placing the first regression line by eye and for reducing the labour of computing the curve, although for estimating the errors of this curve the longer form is preferred.

IV. Accuracy of the regression line.

The fitting of a dosage-mortality curve to a series of experimental observations, however crude or refined the technique, is an attempt to infer, from a limited number of individuals, the "true" empirical relationship of dosage and mortality for a given toxic agent in an infinitely larger population from which they represent only a sample. The regression equation and line is the closest we can approximate this "true" relationship, but all determinations of this type are not equally reliable. If the experimental points are quite close to the line and the number of individuals is large, we have greater confidence that a second or third

determination will agree with our first estimate than if the points are scattered and based on fewer animals. We will want to compute from our experimental data not only the most likely position (the regression line) of the "true" dosage-mortality curve, but also how accurately this most likely position has been determined.

(1) *The χ^2 test for comparing observations with the computed curve.* The first step is to determine whether the observed mortalities agree with our original assumption of a rectilinear relationship on the logarithmic-probability scale within the limits of sampling error; in other words, do the experimental observations vary significantly from our fitted straight line? Since each observation has been weighted by the reciprocal of its variance (Nz^2/pq), which, in turn, is based upon a regression line at the observed dosage, the most satisfactory criterion is the chi-square (χ^2) test. At each dosage the observed mortality is compared with that expected from the regression equation, but instead of calculating separately each expected probit (mortality) from equation (2), and then subtracting it from the observed probit (mortality), a short-cut method for securing the sum of the squares of these differences may be adapted from the one given by Fisher (4). When this is combined with the weighting procedure above, which gives the part of the equation corresponding to the "expectation," χ^2 may be calculated quite easily as follows:

$$\chi^2 = [S(wy^2) - \bar{y}S(wy)] - b\,[S(wxy) - \bar{x}S(wy)]. \qquad \ldots\ldots(7)$$

Nearly all of the components of equation (7) have already been computed in determining the regression equation. The first parenthesis contains $S(wy^2)$, which is the sum of the products of columns 6 and 10 in our example of Table IV. The second part is the numerator of the equation for the regression coefficient (equation (5)) multiplied by the regression coefficient, b. Although in this equation for χ^2 the weights, and therefore the expected probit values, are based upon the initial, graphic regression line, while the differences between expectation and observation depend upon the later, calculated regression line, the discrepancy thus introduced is not a serious one.

The computation of χ^2 is a relatively straightforward operation without statistical complications, but its significance depends upon a term known as the number of "degrees of freedom," n, which may be more difficult to evaluate. If the regression line were calculated from one set of data and then drawn on the same graph with the individually plotted points of a second, entirely independent series of determinations of toxicity, the second series could differ from the line in as many ways—

or in as many degrees of freedom (n)—as there are plotted points or observations (n'). Under these circumstances n would equal n'. If, however, the average log. dose and the average probit were calculated from the second series, and the regression line drawn through the point established by these two averages with a slope which had previously been computed from other data, the separate tests in the second series could not differ as freely from the line as before, because the position of the line has been determined from the observations with which it is being compared. The number of degrees of freedom would then be one less than the number of tests in the second series or $n = n' - 1$, for one degree of freedom has been used up in locating the position of the line. Finally, when not only the position of the regression line but also its slope have been computed from a given series of observations, the extent to which these latter may differ from the transformed dosage-mortality line is still more restricted. In this case, the one with which we have been dealing, the number of degrees of freedom would be equal to the number of separate tests less one which was sacrificed in using these same observations to determine the position of the regression line and less a second degree of freedom lost in establishing the slope of the line. The number of degrees of freedom in the regression line of our computations will be equal, therefore, to the number of separate tests in the series establishing the curve less 2, or $n = n' - 2$.

This rule is simple and easy to apply, but is complicated by another requirement, *i.e.* that the calculated distributions of χ^2, upon which the tests of significance depend, are not very closely realised when very small numbers are expected. In fact, such tests are not rigidly exact when the number expected is less than 5. In toxicological experiments, the expected number of survivors at the higher dosages will regularly fall below this ideal limit, especially when zero survivors are obtained. If each of these particular tests is assigned a value of 1 in determining the number of degrees of freedom, the apparent goodness of fit will be exaggerated by the inclusion of observations which, because of their small weight, contribute little to the observed χ^2. The exact procedure is to exclude from the computation both of χ^2 and of n the results of those dosages at which the number of *expected* survivors, based on the number of organisms counted and the regression line, is less than 3 to 5 individuals. An alternative, which is more convenient though possibly less precise, is to include these small contributions to χ^2 with their standard weights as before, but for the purpose of determining n' and n to group those in which the survival expectancy is small, so that there will be no contribu-

tions to n' or n which are based upon a survival expectancy of less than one individual. The limit of expectancy is lowered here because the separate observations will contribute somewhat more to χ^2, despite their small weights, than they would if the variation between them could be smoothed out by combining them into as few terms as their contributions to n'. The same considerations would hold at the opposite end of the curve when the expectancy of death is very small.

Having secured χ^2 and n, it is a simple matter by reference to a table of χ^2, such as Table III in Fisher's text, to determine if the observations depart more widely from our calculated dosage-mortality curve than could be expected by chance. If χ^2 is smaller than the value in the column for P equal to 0·05, the data may be considered consistent with the straight line that has been fitted. If the χ^2 is greater than the value corresponding to this probability (P), either the observations depart significantly from a straight-line relationship, or some uncontrolled condition in the experiment is causing a greater variation about the line than could be expected from simple fluctuations in sampling. Since systematic departures from rectilinearity were eliminated at the start, the second of these causes is more likely to be involved. Heterogeneity of this type does not necessarily invalidate the procedures described in the present paper.

(2) *The variances of position and slope.* The two parameters determined from an experimental series in calculating the regression line are those giving its position, a (or $\bar{y}$), and slope, b; from the variance of a and of b we may determine how accurately they have been estimated. The square root of the variance of any statistical constant is its standard error, but since the variance must be computed in order to determine the standard error and is here much the more useful, we will deal with the variances directly rather than with their square roots, the standard errors. Since $\bar{x}$ in the regression equation (2) is the independent variable, the average of the dosages selected by the experimenter for testing, it is not a "sample" from a "population" of dosages and is not subject to sampling error in the ordinary sense.

The regression line is calculated so as to intersect the point fixed by the average dosage and the average probit, so that the term a is numerically equal to $\bar{y}$, but since a is defined as a value on the regression line, its variance, $V(a)$, will be that about the regression line at a single dosage at or near the mean dosage, and hence considerably smaller than the variance of the observed probits for all dosages. The equation for the variance of a is

$$V(a) = s_a^2 = \frac{\chi^2}{nS(w)}, \qquad \text{.....(8)}$$

where the symbols have the same significance as before.

The variance of the regression coefficient, b, is given by the equation

$$V(b)=s_b^2=\frac{\chi^2}{nA}. \qquad \text{.....(9)}$$

The formulae for the variance of a and of b given in equations (8) and (9) represent the errors involved in the particular series of records from which they were calculated and are valid however great χ^2 may be. This comparison of χ^2 with its mean value n is a comparison of actual deviations with those theoretically to be expected from the numbers of units observed. If observation and computed curve agree satisfactorily within the limits of sampling error as tested by χ^2 (P greater than 0·1), the errors observed in such a specific experimental series may be replaced by a simpler form which will give the expected error for all similar tests involving the same dosages and numbers of organisms. The theoretical form for the sampling errors in a and b may be obtained from the fact that the mean value of χ^2/n is equal to 1. When the errors in a and b arise solely from the chance distribution of susceptibilities from one test to another, the calculation of their variances may be simplified to

$$V(a)=s_a^2=\frac{1}{S(w)}, \qquad \text{.....(10)}$$

and

$$V(b)=s_b^2=\frac{1}{A}. \qquad \text{.....(11)}$$

(3) *The zone of error of the regression line.* The best available estimate of the true dosage-mortality curve is the calculated regression line. The experience of statisticians indicates that if we can determine limits on either side of the regression line, such that there are 19 chances in 20 of their enclosing the true dosage-mortality curve, we will have a reasonable standard for prediction. Our next problem, therefore, is to determine the accuracy or "sensitivity" of the dosage-mortality curve which we have computed, using the margin of safety represented by 19 chances in 20 or $P=0{\cdot}05$.

From the variance, $V(a)$, we can determine by how much the true regression line may lie above or below the most likely position as fixed by a, and from the variance, $V(b)$, we can find how much more or less it may be tilted. At the average dosage, $\bar{x}$, an error in b could have no influence upon the sensitivity with which a is an index to the true regression line, but as the dosage differs more or less widely from the average, both errors are of importance and will modify the accuracy of estimate of the true mortality corresponding to any given dosage. As

shown by Working and Hotelling(14), the formula for the regression equation and its error may be written as

$$Y = a + b\,(X - \bar{x}) \pm t\sqrt{V(a) + (X - \bar{x})^2\,V(b)}. \qquad \ldots\ldots(12)$$

The value of t is not calculated but is taken from a table of "Student's" integral, such as Table IV of Fisher's text, from the column for $P = 0{\cdot}05$ at the value of n equal to the number of degrees of freedom for the curve. From equation (12) we may calculate the probit of kill and its error of estimate for a series of dosages covering the same range as our original experimental observations; from the plus errors draw a line above, and from the minus errors a line below the dosage-mortality curve such that there are 19 chances in 20 of these two boundaries, the branches of a hyperbola, enclosing the true dosage-mortality curve when transformed to the logarithmic-probit diagram. If it is preferred that the boundaries represent odds of 1 in 2, as in the familiar probable error, t is read from the column for $P = 0{\cdot}5$.

These different operations may now be illustrated from our example in Table IV, the computations for the main curve being summarised at the end of the table. For this range of higher dosages, $\chi^2 = 5{\cdot}556$. Although the curve is based upon 12 separate determinations of mortality, the total number of survivors expected from the four tests at the two highest concentrations of carbon disulphide was only 1·36 beetles (1 survivor observed). Therefore these will count as 1 instead of as 4 in determining n', and since $n = n' - 2$, the number of degrees of freedom will be $9 - 2 = 7$. From a table of χ^2, such as Table III in Fisher's text, the corresponding value of P lies between 0·5 and 0·7, so that the data may be considered consistent with the regression line which has been fitted to them. When the same test is applied to the line fitted to the range of smaller dosages, the χ^2 test again indicates satisfactory agreement

$$(\chi^2 = 1{\cdot}404,\ n = 4,\ P = 0{\cdot}84).$$

Since χ^2 indicates a satisfactory agreement between observation and fitted curve, the generalised form of the variances in the position and slope of the regression line may be used (equations (10) and (11)), when $V(a) = 0{\cdot}007758$ and $V(b) = 6{\cdot}6683$. We now have all the terms for computing the regression line and its errors (equation (12)) with the exception of t. For $n = 7$, at odds represented by $P = 0{\cdot}05$, the value of t is given by Table IV in Fisher's text as 2·365. The equation for estimating the mortality in probits, Y, and its error within odds of 19 to 1, at any desired log. dosage, X, above a concentration of 57·8 mg. per litre, is

$$Y = 5{\cdot}450 + 25{\cdot}51\,(X - 1{\cdot}7967) \pm 2{\cdot}365\sqrt{0{\cdot}007758 + (X - 1{\cdot}7967)^2\,6{\cdot}6683}.$$

The limits shown as curved dotted lines in Fig. 3 have been computed from this equation for the range of higher dosages and from a similar equation for the lower dosages. These boundaries define the accuracy with which the two solid regression lines have been determined by the experiment.

If the two series of tests had been combined, either when the experiments were made originally or in computing the percentages, that part of the error under the

square root would remain as it is in the longer form, since the generalised errors in position and slope depend only upon the sum of the weights and the variance of the log. dosage. The number of degrees of freedom would have dropped, however, from 7 to 3, so that t would have been increased from 2·365 to 3·182, and the limits of the estimated error increased proportionately.

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

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

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

and

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

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

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

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

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

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

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

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

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

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

Now when both s and $n-s$ are so large that the distribution of s may be treated as normal, the factor $(qn-s)$, which is n times the difference

between the proportion of survivors expected and observed, is taken to be proportional to the difference between the probit values expected and observed, according to the formula

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

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

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

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

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

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

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

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

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

These values are shown in Table II.

VI. Summary.

The sigmoid dosage-mortality curve, secured so commonly in toxicity tests upon multicellular organisms, is interpreted as a cumulative normal frequency distribution of the variation among the individuals of a population in their susceptibility to a toxic agent, which susceptibility is inversely proportional to the logarithm of the dose applied. In support of this interpretation is the fact that when dosage is inferred from the observed mortality on the assumption that susceptibility is distributed normally, such inferred dosages, in terms of units called probits, give straight lines when plotted against the logarithm of their corresponding observed dosages. It is shown that this use of the logarithm of the dosage can be interpreted in terms either of the Weber-Fechner law or of the amount of poison fixed by the tissues of the organism. How this transformation to a straight regression line facilitates the precise estimation of the dosage-mortality relationship and its accuracy is considered in detail. Statistical methods are described for taking account of tests which result in 0 or 100 per cent. kill, for giving each determination a weight proportional to its reliability, for computing the position and slope of the transformed dosage-mortality curve, for measuring the goodness of fit of the regression line to the observations by the χ^2 test, and for calculating the error in position and in slope and their combined effect at any log. dosage. The terminology and procedures are consistent with those used by R. A. Fisher, who has contributed an appendix on the case of zero survivors. Except for a table of common logarithms, all the tables required to utilise the methods described are given either in the present paper or in Fisher's book. A numerical example selected from Strand's experiments upon *Tribolium confusum* with carbon disulphide has been worked out in detail.

It is a pleasure to record my indebtedness to Prof. R. A. Fisher, not only for the note appended to the paper, but also for invaluable advice throughout its preparation and for the facilities of the Galton Laboratory which have so generously been placed at my disposal. Among others who have been kind enough to read and criticise my manuscript, I wish especially to thank Prof. A. J. Clark, Dr F. Tattersfield, Dr J. O. Irwin, Dr A. B. P. Page, Mr H. H. S. Bovingdon, and Dr A. E. Brandt.

REFERENCES.

(1) BLISS, C. I. (1934). *Science*, LXXIX, 38 and 409.
(2) CLARK, A. J. (1933). *The Mode of Action of Drugs on Cells*. Arnold.
(3) FISHER, R. A. (1924). *Proc. Inter. Math. Congress, Toronto*, p. 805.
(4) —— (1932). *Statistical Methods for Research Workers*. 4th ed. Oliver and Boyd.
(5) GADDUM, J. H. (1933). *Med. Res. Council Spec. Rep. No.* 183. His Majesty's Stationery Office.
(6) HEMMINGSEN, A. M. (1933). *Quart. J. Pharm.* VI, 39, 187.
(7) KELLEY, T. L. (1923). *Statistical Method*. Macmillan.
(8) MCCALLAN, S. E. A. and WILCOXON, F. (1933). *Contr. Boyce Thompson Inst.* V, 173.
(9) PEARSON, K. (1924). *Tables for Statisticians and Biometricians*. Part I, 2nd ed., Cambridge.
(10) PRATT, F. S., SWAIN, A. F. and ELDRED, D. N. (1931). *J. Econ. Ent.* XXIV, 1041.
(11) STRAND, A. L. (1930). *Ind. Eng. Chem. Analyt.* Ed. II, 4.
(12) THOMPSON, G. H. (1919). *Biometrika*, XII, 216.
(13) TREVAN, J. W. (1929). *J. Path. Bact.* XXXII, 127.
(14) WORKING, H. and HOTELLING, H. (1929). *Proc. Amer. Stat. Assn.* p. 73.

(*Received June 9th*, 1934.)