# RESEARCH PAPERS

## THE DISTRIBUTION OF ERROR IN MOUSE INSULIN ASSAYS

By Peter A. Young

*The Wellcome Research Laboratories, Beckenham, Kent*

and G. A. Stewart

*The Biological Control Laboratory, The Wellcome Chemical Works, Dartford, Kent*

Received November 28, 1951

To specify the accuracy of a bio-assay technique some knowledge of the nature of the distribution of errors between tests of the same type is required, since it is the uniformity of variance from test to test as much as the mean variance which will determine the usefulness of the method. To investigate the distribution of errors in an assay involving a probit response, the mouse insulin test was chosen as being a well established technique on which many results were available for analysis. That the distribution of standard errors of log-potency in this test is not normal will be seen from the histogram shown in Figure 1. While some of the

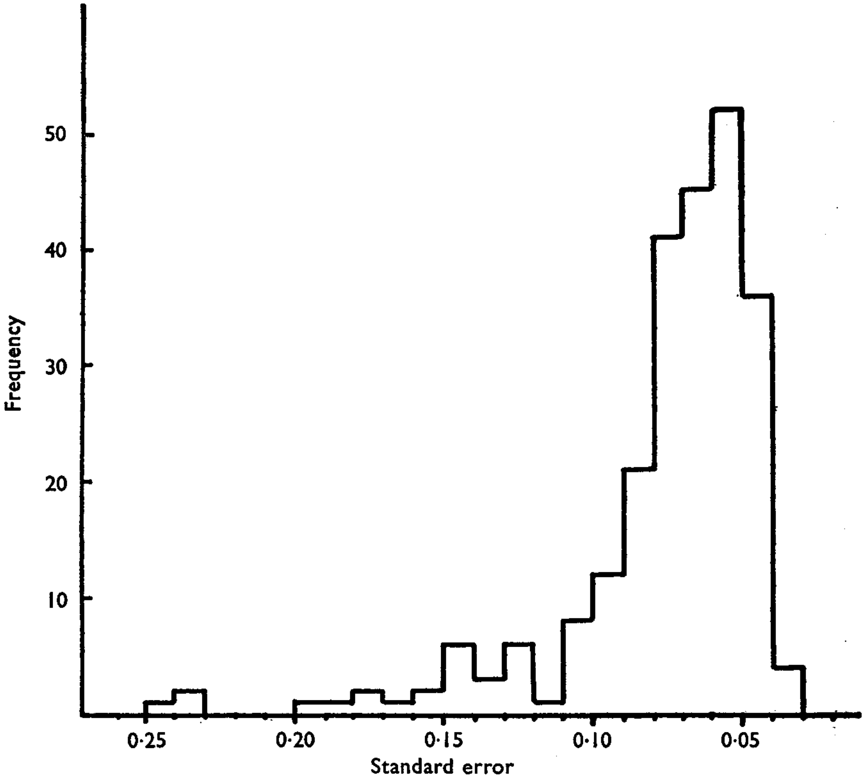

Fig. 1. Histogram showing the distribution of standard errors of log-potency for 257 assays.

conclusions reached in this study are peculiar to the insulin test, some of the more general findings may possibly be applicable to many other types of probit assay.

## The Test

A standard 4-point design was employed throughout, with a dose ratio high: low of 2:1·2, i.e., a log dose interval of 0·2219. A total of 96 mice were used for each assay, their extreme weights varying by not more than 1·5 g. The same mice were sometimes used up to 4 times, with at least 1 week between repetitions, but in one assay all the mice would have been employed the same number of times. The animals used for assays on one day would have been deprived of food since the previous afternoon, their normal diet consisting of an adequate supply of bread and milk. The criterion of response in the test depends on the production of hypoglycæmic convulsions by sufficiently large doses of insulin. To observe these the injected mice are placed in jars (3 mice in a jar) in a constant temperature cabinet (34° C.), and observed up to 1¼ hours after injection. Mice convulsing, or showing symptoms of convulsions, are removed and given an injection of glucose (they are not returned to the jars). After 1¼ hours the total responses to each treatment are counted up, and the computation of relative potency then made by a standard type of probit analysis.

## Results

The results took the form of data taken from the records of routine assays of crystalline insulin (a mixture of ox, pig and sheep insulins) carried out under the direction of one of the authors (G.A.S.). The first set of figures so obtained concerned 257 assays carried out between September, 1949, and February, 1950. These covered in all 28 samples of insulin, the number of tests per sample varying considerably. The values examined were:—(1) the responses out of 24 for each treatment, i.e., 257 values each for standard high and low, test high and low (SH, SL, TH, and TL respectively), (2) the weighted mean slope of standard and test for each assay, and (3) the weight assigned to the log potency estimate for each assay, this weight being thè reciprocal of the variance of the log potency. Some weight values were accidentally omitted from this series, leaving 249 estimates.

Two further sets of values were later extracted from the records. These will be described in the appropriate sections below.

*The distribution of responses.* Under ideal conditions the responses to high and low doses would be distributed independently according to the expansion of the binomial $(p + q)^{24}$, where $q$ is the true proportion reacting to the dose, and 24 the number of mice per treatment group. In practice, however, many uncontrolled factors influence the responses, and the distributions found (Table I) have variances much greater than binomial distributions with the same means. For example, the mean response to SH was 13·7 out of 24 (*ca.* 57 per cent.), the variance being 22·2. The variance of a binomial distribution with $q = 0{\cdot}57$ would be $24 \times 0{\cdot}43 \times 0{\cdot}57$, or 5·9. With the higher doses of both standard and test the responses out of 24 and the probits of these values could be fitted by a normal distribution (Table II). Since the responses to the low doses were

## TABLE I

FREQUENCY DISTRIBUTION OF RESPONSES TO HIGH AND LOW DOSES OF INSULIN IN 257 MOUSE ASSAYS

Ratio of high dose to low dose = 5 : 3

| Response | | Standard | | Unknown | |
|---|---|---|---|---|---|
| Out of 24 | Probit | High | Low | High | Low |
| 0 | 2·579 | 0 | 7 | 1 | 14 |
| 1 | 3·268 | 2 | 28 | 0 | 15 |
| 2 | 3·617 | 2 | 23 | 1 | 29 |
| 3 | 3·850 | 3 | 24 | 1 | 32 |
| 4 | 4·033 | 3 | 23 | 2 | 23 |
| 5 | 4·188 | 6 | 19 | 3 | 21 |
| 6 | 4·326 | 6 | 28 | 6 | 24 |
| 7 | 4·451 | 5 | 21 | 6 | 17 |
| 8 | 4·569 | 10 | 18 | 7 | 26 |
| 9 | 4·681 | 11 | 13 | 14 | 11 |
| 10 | 4·790 | 12 | 10 | 18 | 7 |
| 11 | 4·895 | 17 | 5 | 18 | 8 |
| 12 | 5·000 | 15 | 14 | 16 | 5 |
| 13 | 5·105 | 15 | 4 | 26 | 6 |
| 14 | 5·210 | 30 | 8 | 20 | 5 |
| 15 | 5·319 | 24 | 7 | 22 | 8 |
| 16 | 5·431 | 20 | 2 | 21 | 4 |
| 17 | 5·549 | 21 | 1 | 13 | 1 |
| 18 | 5·675 | 15 | 1 | 21 | 0 |
| 19 | 5·812 | 15 | 1 | 10 | 1 |
| 20 | 5·967 | 8 | 0 | 11 | 0 |
| 21 | 6·150 | 7 | 0 | 12 | 0 |
| 22 | 6·383 | 8 | 0 | 7 | 0 |
| 23 | 6·732 | 1 | 0 | 1 | 0 |
| 24 | 7·421 | 1 | 0 | 0 | 0 |

## TABLE II

THE NORMAL DISTRIBUTION FITTED TO FREQUENCIES OF RESPONSES AND PROBITS OF RESPONSES TO HIGH DOSES OF STANDARD AND UNKNOWN INSULINS

obs. = observed; Exp.$_r$ = expected from distribution of responses; Exp.$_p$ = expected from distribution of probits of responses

| Response /24 | Frequencies | | | | | |
|---|---|---|---|---|---|---|
| | Standard | | | Unknown | | |
| | Obs. | Exp.$_r$ | Exp.$_p$ | Obs. | Exp.$_r$ | Exp.$_p$ |
| 0 | 0 | 10·3 | 8·3 | 1 | 12·2 | 10·1 |
| 1 | 2 | | | 0 | | |
| 2 | 2 | | | 1 | | |
| 3 | 3 | | | 1 | | |
| 4 | 3 | | | 2 | | |
| 5 | 6 | | 5·8 | 3 | | |
| 6 | 6 | 5·6 | 7·5 | 6 | | 6·5 |
| 7 | 5 | 7·7 | 9·6 | 6 | 6·7 | 8·4 |
| 8 | 10 | 10·7 | 11·0 | 7 | 9·7 | 10·8 |
| 9 | 11 | 13·0 | 13·2 | 14 | 12·2 | 12·3 |
| 10 | 12 | 15·7 | 14·4 | 18 | 15·9 | 14·2 |
| 11 | 17 | 18·1 | 15·2 | 18 | 18·1 | 16·2 |
| 12 | 15 | 21·1 | 17·4 | 16 | 21·4 | 17·8 |
| 13 | 15 | 21·2 | 18·1 | 26 | 22·0 | 19·0 |
| 14 | 30 | 21·4 | 18·4 | 20 | 23·5 | 20·5 |
| 15 | 24 | 20·8 | 19·1 | 22 | 21·9 | 20·2 |
| 16 | 20 | 20·1 | 18·1 | 21 | 21·1 | 20·1 |
| 17 | 21 | 16·9 | 18·1 | 13 | 17·8 | 18·1 |
| 18 | 15 | 14·3 | 16·4 | 21 | 15·5 | 17·7 |
| 19 | 15 | 11·6 | 14·5 | 10 | 11·8 | 15·7 |
| 20 | 8 | 9·3 | 12·9 | 11 | 9·3 | 12·4 |
| 21 | 7 | 6·6 | 9·6 | 12 | 6·6 | 9·3 |
| 22 | 8 | 12·7 | 9·7 | 7 | 11·2 | 7·9 |
| 23 | 1 | | | 1 | | |
| 24 | 1 | | | 0 | | |
| $\chi^2$ | — | 15·8 | 16·6 | — | 13·6 | 11·9 |
| (d.f.) | | (17) | (18) | | (16) | (17) |
| $p$ | | 0·60 | 0·60 | | 0·56 | 0·80 |

truncated at the zero response level, the estimation of mean and standard deviation was carried out by fitting a linear regression to the probits of cumulative frequencies at successive response levels. The "standard low" points are shown graphically in Figure 2. Using the parameters so obtained, the probits of responses gave a considerably better fit to a normal distribution than did the responses alone (see Table III). With the probit values significant deviations only occurred at one or other of the extremes, and it was therefore decided to carry out subsequent analyses on probits of responses. In this connection the empirical probits used in the original computation of results for 0 and 100 per cent. responses correspond to $\frac{1}{2}$ and $23\frac{1}{2}$ responses out of 24, that is, to the interval boundaries in a frequency table. We have therefore transformed these values arbitrarily to those shown in Table I, by the use of the table of working probits given by Finney.[1]

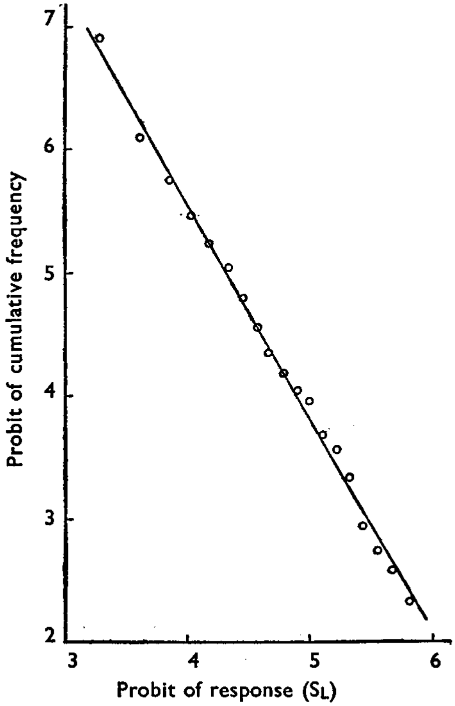

FIG. 2. Graphical method of estimating the mean and variance of the truncated distributions of responses to low doses of standard. The figure shows the plot for probit of response against probit of cumulative frequency.

To establish rather more firmly the normality of the probit responses, two further distributions were investigated:—(1) the 514 values of $b'$, the probit difference between the responses to high and low doses of each preparation, and (2) the 257 estimates of the difference in probits between the unweighted mean response to standard and that of test, that is $\frac{1}{2}$ (SH + SL − TH − TL), where SH, SL are the probit responses to high and low doses of standard, TH, TL to the test sample. Both these distributions could be fitted by the normal distribution with a probability of the order of 0·20. With this general conformity of the probit response to a normal variate in mind, an analysis of variance was carried out on the unweighted probit responses of the 257 assays (Table IV). There it will be seen that only the linear regression and "between assays" mean square are significant against the residual mean square. Thus in the overall picture, there is no detectable difference between standard and test solutions, nor is their interaction with assays (5) significant. The mean square indicating departure from parallelism (4) is large but not significant, nor is the doses × assays interaction. These two observations suggest

## TABLE III

THE NORMAL DISTRIBUTION APPLIED TO RESULTS WITH LOW DOSES OF STANDARD AND UNKNOWN

Coding as in Table II, means and variances estimated graphically by the method shown in Figure 2

| Response /24 | Frequencies | | | | | |
|---|---|---|---|---|---|---|
| | Standard | | | Unknown | | |
| | Obs. | Exp.r | Exp.p | Obs. | Exp.r | Exp.p |
| 0 | 7 | 21·2 | 17·5* | 14 | 12·5 | 30·6 |
| 1 | 28 | 10·0 | | 15 | 7·9 | |
| 2 | 23 | 12·9 | 22·7 | 29 | 10·7 | 33·2 |
| 3 | 24 | 15·8 | 25·3 | 32 | 14·9 | 33·4 |
| 4 | 23 | 19·5 | 26·9 | 23 | 18·7 | 30·3 |
| 5 | 19 | 20·9 | 24·9 | 21 | 22·1 | 26·4 |
| 6 | 28 | 22·2 | 23·5 | 24 | 23·5 | 22·9 |
| 7 | 21 | 22·5 | 21·0 | 17 | 25·5 | 18·8 |
| 8 | 18 | 21·7 | 19·6 | 26 | 25·1 | 14·8 |
| 9 | 13 | 20·0 | 15·9 | 11 | 23·2 | 12·3 |
| 10 | 10 | 17·5 | 13·9 | 7 | 19·4 | 9·4 |
| 11 | 5 | 14·6 | 11·1 | 8 | 16·6 | 7·4 |
| 12 | 14 | 11·6 | 9·1 | 5 | 12·7 | 5·6 |
| 13 | 4 | 8·8 | 7·2 | 6 | 9·2 | 12·0* |
| 14 | 8 | 6·4 | 5·8 | 5 | 6·0 | |
| 15 | 7 | 11·5 | 12·7 | 8 | 9·3 | |
| 16 | 2 | | | 4 | | |
| 17 | 1 | | | 1 | | |
| 18 | 1 | | | 0 | | |
| 19 | 1 | | | 1 | | |
| $\chi^2$ | | 72·9 | 30·3 | | 88·5 | 27·3 |
| d.f | | 15 | 14 | | 15 | 12 |
| $p$ | | <·001 | ·007 | | <·001 | ·007 |

* $\chi^2$ values omitting these entries:—
Standard $\chi^2$ = 12·8, 13 d.f., $p$ = 0·50.
Unknown $\chi^2$ = 13·1, 11 d.f., $p$ = 0·30.

## TABLE IV

ANALYSIS OF VARIANCE OF UNWEIGHTED PROBIT RESPONSES IN 257 ASSAYS

| Source of variation | d.f. | Mean square | $p$ |
|---|---|---|---|
| 1. Between std. and test (S–T) .. .. | 1 | 0·0266 | >0·20 |
| 2. Between doses (H–L) .. .. | 1 | 265·6719 | <0·001 |
| 3. Between assays .. .. .. | 256 | 1·1360 | <0·001 |
| 4. (S–T) X (H–L) .. .. .. | 1 | 0·2783 | 0·20–0·05 |
| 5. (S–T) X Assays .. .. .. | 256 | 0·1222 | >0·20 |
| 6. (H–L) X Assays .. .. .. | 256 | 0·1299 | >0·20 |
| 7. Residual inter-action (S–T) (H–L) (Assays) .. .. .. .. | 256 | 0·1147 | — |
| 8. Pooled error, items 4–7 .. .. | 769 | 0·1225 | — |

that the standard and test preparations produce parallel regressions of probit response on dose, and that their common slope does not vary significantly from assay to assay. From these findings, it may be inferred that the four responses in any one assay, are mutually correlated. The correlation coefficients were therefore computed between

(*a*) the pairs of values (SH + SL) and (TH + TL), and
(*b*) the pairs of values (SH + TH) and (SL + TL),

the first giving the correlation between the sums of standard and test probit responses, the second between the sums of high and low dose responses. Both values for $r$ were highly significant, that for (*a*) being +0·8058, and for (*b*) +0·8003, each based on 257 pairs of results.

Using the values of $b'$, the probit differences of high and low dose responses, as estimates of the slopes, it was found that there was no correlation between the slope of standard and that of test within individual assays. For 257 pairs of values $r$ was found to be $+0{\cdot}0380$.

Thus in comparing the responses obtained in different assays only one factor appears to vary. That is the absolute sensitivity of the mice to the treatments; this is confounded with variations of the absolute doses administered. In a given group of mice, assuming uniform sensitivity throughout the group, the absolute probit responses are correlated equally between standard and test as between high and low doses, the order of their values being determined by the sensitivity of the mice (and by the dose given), the slopes being governed by the intrinsic regression coefficient of the technique. The variance of any one probit response under these conditions would be given by the error mean square of Table IV.

TABLE V

DISTRIBUTION OF 257 ESTIMATES OF THE WEIGHTED MEAN SLOPE FOR STANDARD AND UNKNOWN COMPARED WITH A NORMAL FITTED CURVE

| | Frequency | |
|---|---|---|
| Slope intervals | Observed | Expected |
| <1·8 | 9 | 10·3 |
| 1·8–2·2 | 8 | 7·5 |
| 2·2–2·6 | 5 | 11·2 |
| 2·6–3·0 | 15 | 14·9 |
| 3·0–3·4 | 24 | 19·0 |
| 3·4–3·8 | 25 | 23·7 |
| 3·8–4·2 | 35 | 25·5 |
| 4·2–4·6 | 22 | 27·7 |
| 4·6–5·0 | 25 | 25·8 |
| 5·0–5·4 | 19 | 23·5 |
| 5·4–5·8 | 24 | 20·6 |
| 5·8–6·2 | 14 | 15·7 |
| 6·2–6·6 | 9 | 11·6 |
| 6·6–7·0 | 10 | 8·3 |
| 7·0–7·4 | 10 | 5·1 |
| ⩾7·4 | 3 | 6·6 |

$\chi^2$ for deviations 19·0, 15 d.f., $p = 0{\cdot}20$

In practice, of course, the evaluation of individual tests is complicated by the introduction of weighting coefficients, which dismisses the possibility of the use of a common error variance for the probit responses. How the weighting of responses is reflected in the results obtained is indicated below.

*The slope distribution.* The slope of any pair of responses is given by $b'$, their probit difference, divided by 0·2219, the constant log dose interval. The variance of the $b'$ values was 0·2449, equivalent to twice the error variance for a single estimate, and the mean $b'$ was 1·017 probits. Since there was no correlation between the slope of standard and test in any one assay, the variance of their unweighted means would be half the overall variance. Thus in true slope units the overall mean value would be $\frac{1{\cdot}017}{0{\cdot}2219} = 4{\cdot}583$ with a variance between assays of $\frac{0{\cdot}2449}{(0{\cdot}2219)^2} \times \frac{1}{2} = 2{\cdot}487$.

To compare with these estimates we have the 257 values of the weighted mean slope of standard and test, the frequency distribution of which is given in Table V. The unweighted mean of this series was 4·439, with a variance between assays of 2·300, both these values being less than those derived from the previous unweighted data.

The frequencies of Table V may be fitted to a normal distribution with a probability of about 0·20, and although this fit may be rather fortuitous in the light of further evidence (see below), it may be assumed that the weighted mean slopes are distributed approximately normally.

*The standard error of log potency.* The standard error of a single estimate of log-potency in these assays in its uncorrected form is given by:

$$s.e. = \pm \frac{1}{b}\sqrt{\left(\frac{1}{n\mathrm{S}w_{\mathrm{S}}} + \frac{1}{n\mathrm{S}w_{\mathrm{T}}}\right)}$$

Where $b$ is the weighted mean slope of standard and test; $n = 24$, the number of animals per treatment group; and $\mathrm{S}w_{\mathrm{S}}$, $\mathrm{S}w_{\mathrm{T}}$, the sums of the weighting coefficients for high and low dose response probits for standard and test respectively. If it be assumed that standard and test in any one assay are administered in equipotent doses, producing equal responses, then their weighting coefficients will also be equal. The standard error may then be written in the form $\pm \frac{1}{b}\sqrt{\left(\frac{2}{n\mathrm{S}w}\right)}$. The reciprocal of the standard error will therefore be $\pm b\sqrt{(\mathrm{S}w)} \times \sqrt{(12)}$; from the point of view of investigating the nature of the distribution of this expression we may further eliminate the constants to reduce it to $b'\sqrt{(\mathrm{S}w)}$ where $b'$ is the probit difference equivalent to the slope $b$. It has been shown that $b$ and $b'$ are normally distributed, hence the distribution of $b'\sqrt{(\mathrm{S}w)}$ will only be normal if it is linearly related to $b'$. Any given value of $b'$ however, may be associated with a range of values of $\mathrm{S}w$ according to the degree of asymmetry of the responses relative to the 50 per cent. point. Thus $\mathrm{S}w$ is maximal when the responses are symmetrically disposed, and minimal when they are least symmetrical, i.e., when either response is 0 or 100 per cent. Figure 3 shows these limiting values of $b'\sqrt{(\mathrm{S}w)}$ over the range of positive values of $b'$ encountered in the observations. The values of $b'\sqrt{(\mathrm{S}w)}$ actually obtained were distributed over the area between these limiting curves, and their regression analysis on $b'$, Table VI, shows that the linear regression term is by far the greatest. We could therefore infer that the relationship over the whole range is sufficiently near to direct proportionality for the distribution of $b'\sqrt{(\mathrm{S}w)}$ not to deviate significantly from the normal. Since this function was derived from the reciprocal of the standard error by the elimination of constants and the assumption only of equal potency of standard and test, which we know to be the case in the over-all picture, it may be inferred that the reciprocals of the standard errors will be normally distributed. A histogram of these values, which correspond to the square roots of the weights assigned to log potencies, is shown in Figure 4 together with a fitted normal curve.

TABLE VI

THE REGRESSION ANALYSIS OF $b'\sqrt{(\mathrm{S}w)}$ ON $b'$ (GROUPED DATA)

$b'$ = probit difference of high and low dose responses
$\mathrm{S}w$ = sum of the weighting coefficients corresponding to high and low responses

| Source of variation | d.f. | Mean square |
|---|---|---|
| Linear regression .. .. .. | 1 | 86·190 |
| Deviations from linearity .. .. | 12 | 0·186 |
| Residual error .. .. .. .. | 500 | 0·021 |

Linear regression coefficient = 0·807

Deviations from the normal produce a value for $\chi^2$ of 23·01 which with 19 degrees of freedom corresponds to a probability of 0·24. The mean value was 15·0, standard deviation $\pm 5{\cdot}085$.

To confirm this distribution a further series of 424 values were extracted from the records. Deviations from normality gave $\chi^2 = 19{\cdot}63$, which with 19 degrees of freedom, gave $p = 0{\cdot}40$. On the other hand while the variance in this second series of $\sqrt{}$(weight) values was not significantly

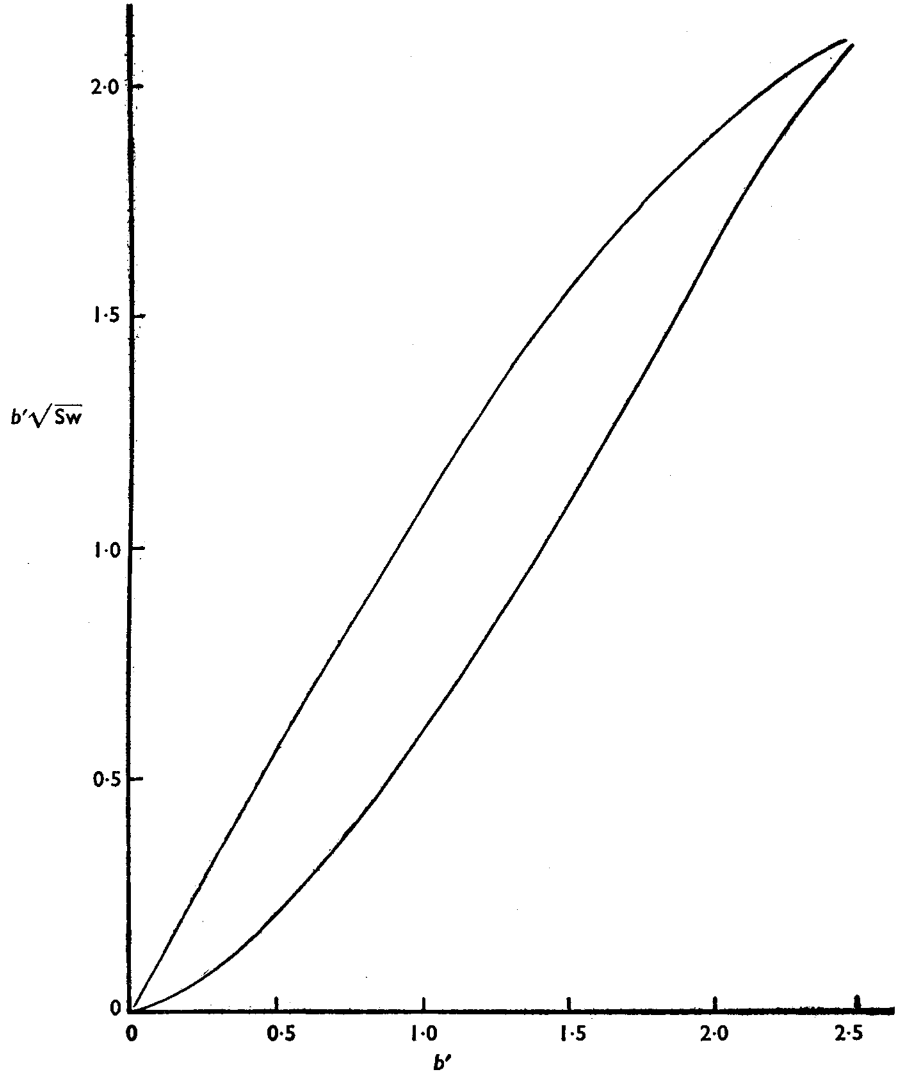

FIG. 3. Limiting values of $b'\sqrt{(Sw)}$ related to $b'$. The upper line of maximum values corresponds to pairs of responses symmetrical about the 50 per cent. point, the lower line of minimum values to pairs of responses, one of each pair being 0/24 or 24/24. In practice of course the relationship is discontinuous.

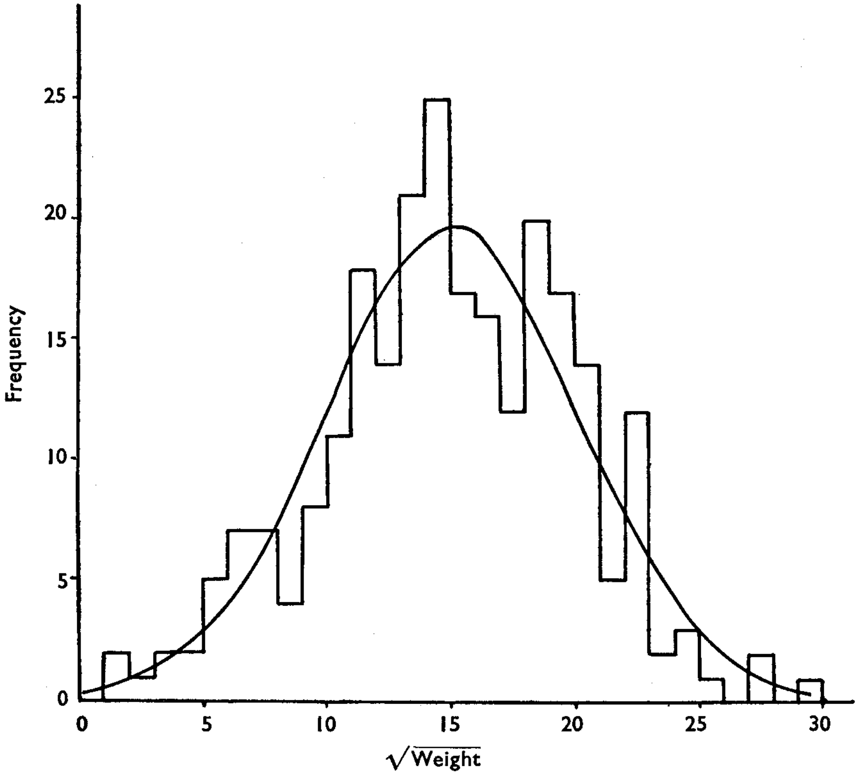

FIG. 4. Histogram of the distribution of the reciprocals of standard errors of log potency (√weight), with a normal curve with the same mean and variance.

reduced (19·66 compared with 25·86), the mean was significantly greater at 16·92. Thus it appeared that the internal accuracy of the second series of tests was greater than that of the first series. Since the internal error might be reduced either by more symmetrical responses about the 50 per cent. point, or by an increase of the regression coefficient, it seemed desirable to decide in which manner the change had been produced.

Further inspection of results obtained over the same period as the second series described above showed that the mean slope of 464 assays was 4·906 compared with the mean slope of the original series of 4·439. Thus the higher mean of the √(weight) values of the second series may be accounted for by an increase in the mean slope. Since this was the first indication in the data that the slopes were not simple estimates of some true mean value, it became necessary to investigate the variables which might effect such a change.

The only factors which could be demonstrated to influence the slope were the weight of mice used and their previous usage. Table VII gives the frequencies of different slope values according to the weight range and previous history of the mice. In the case of normal mice, i.e., those used

TABLE VII

THE FREQUENCY DISTRIBUTION OF SLOPE VALUES RELATED TO BODY WEIGHT OF MICE AND PREVIOUS USAGE

| Key to weight groups: | | |
|---|---|---|
| | 1 | 14·5–16·0 |
| | 2 | 16·0–17·5 |
| | 3 | 17·5–19·0 |
| | 4 | 19·0–20·5 |
| | 5 | 20·5–22·0 g. |

| | Number of times previously used: | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | None | | | | | Once | | | | | Twice or more | | | | |
| Wt. group: | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Slope values: | | | | | | | | | | | | | | | |
| 1·8–2·1 | | | 1 | | | | | 1 | | | | | | | |
| 2·1–2·4 | | 3 | 1 | | | | | | 1 | 1 | | | | | |
| 2·4–2·7 | 5 | 2 | 2 | | | | 1 | | 1 | 1 | | | | | |
| 2·7–3·0 | 5 | 8 | 1 | 1 | | | 1 | | | 1 | | | | | |
| 3·0–3·3 | 3 | 3 | 5 | 3 | | | | 1 | 2 | 1 | | | | | |
| 3·3–3·6 | 3 | 3 | 6 | 3 | | | | 2 | 4 | | | | | 1 | |
| 3·6–3·9 | 5 | 8 | 10 | 4 | 2 | | 3 | | | | | | | 2 | 1 |
| 3·9–4·2 | 3 | 5 | 11 | 10 | 2 | | 2 | 5 | 4 | 4 | | | | 1 | |
| 4·2–4·5 | 2 | 6 | 7 | 11 | 1 | | 3 | 2 | 3 | 1 | | | | | 1 |
| 4·5–4·8 | 5 | 10 | 4 | 5 | 2 | 1 | | 3 | 1 | 2 | | | 1 | | |
| 4·8–5·1 | 6 | 4 | 3 | 5 | 2 | 2 | | 4 | 1 | 5 | | | | | |
| 5·1–5·4 | 3 | 7 | 11 | 8 | | | | | 3 | 2 | | 1 | 1 | | 1 |
| 5·4–5·7 | 6 | 8 | 7 | 8 | | | 1 | 1 | 1 | 1 | | | | | 1 |
| 5·7–6·0 | 1 | 5 | 11 | 11 | 2 | | | 3 | 1 | 1 | | | 1 | 1 | 1 |
| 6·0–6·3 | 1 | 4 | 5 | 7 | 3 | | | 1 | 2 | | | | 1 | | 1 |
| 6·3–6·6 | 5 | 3 | 6 | 3 | 1 | | | | | 1 | | | | | |
| 6·6–6·9 | | 3 | 2 | 2 | 1 | | 1 | | 2 | | | | | 1 | 1 |
| 6·9–7·2 | 2 | 1 | 2 | 4 | | 1 | | | 1 | | | | | 1 | |
| 7·2–7·5 | | 2 | 2 | 4 | | | | 1 | | | | | | | |
| 7·5–7·8 | | | 2 | 1 | | | | | | | | | | | |
| 7·8–8·1 | | 2 | | 1 | | | | | | | | | | 1 | |
| 8·1–8·4 | | | 1 | | 1 | | | | 1 | | | | | | |
| ⩾8·4 | | | 2 | 2 | 1 | | | | | | | | | | |
| Total frequencies .. | 55 | 87 | 102 | 93 | 18 | 4 | 12 | 24 | 28 | 21 | 0 | 1 | 4 | 8 | 7 |
| Mean slopes .. .. | 4·49 | 4·71 | 5·01 | 5·30 | 5·55 | 5·40 | 4·17 | 4·64 | 4·71 | 4·46 | — | 5·25 | 5·48 | 5·33 | 5·38 |

for the first time, a significant regression of slope on mouse weight may be demonstrated (Fig. 5). This does not hold however in the case of mice which have undergone previous testing, although admittedly the numbers of assays are somewhat lower in these groups. The mean slope given by mice previously used once is considerably less than would be given by normal mice of the same average weight, yet that for mice previously used twice or more, while significantly greater than the mean of the once used group, does not differ significantly from the slope for normal mice of the same average weight.

## Discussion

It has been shown that the reciprocals of standard errors of log potency obtained in a number of assays are distributed approximately normally. Some factors which make it impossible for the distribution to be truly normal have been described. It is interesting to note that Hemmingsen[2] reported that the variation of slope values from test to test exceeded that to be expected from consideration of the binomial distribution of responses. Thus in his data either the slope did in fact vary, or the mice within each assay were so heterogeneous as to render the application of the binomial sampling rule invalid. It is likely that both of these conditions should equally well apply to his results as well as to our own. On the other hand, Smith[3] could not detect significant differences of slope associated with the weight of mice or their previous usage. His examination of the results of 231 assays was made by isolating each variable in turn, however. Thus the results obtained in the different weight groups were confounded with the previous usage of the mice, and *vice versa*. Both Hemmingsen and Smith employed a far coarser range of weight for the mice included in any one test, while the latter author used mice up to 30 g. in weight, 8 g. higher than our maximum. The limitation we have imposed on the weight of mice used for testing, particularly with regard

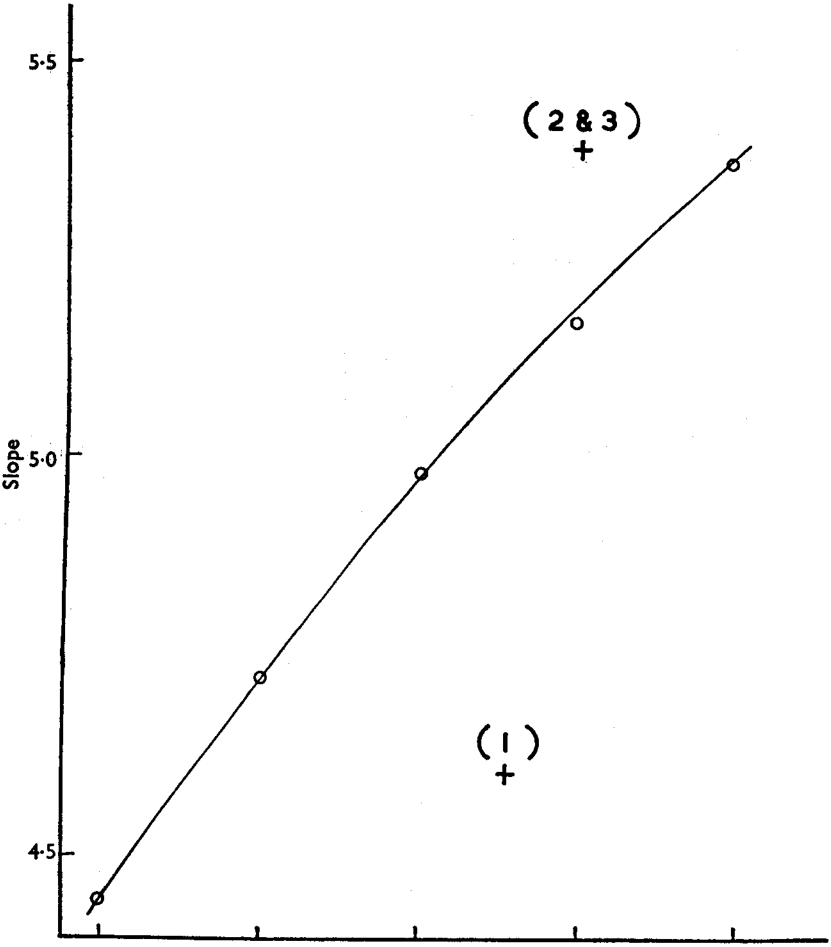

FIG. 5. The relationship between the slope of the log. dose-response line and body weight of normal mice, i.e., those used for the first time. The mean slopes for mice used once (1), and two or more times (2 and 3) previously are shown plotted against their respective mean weights.

to the upper limit of 22 g., entails a complicated system of selection for mice used more than once. Thus those which fell in the heavier groups at their first usage, and subsequently gained weight at the normal rate, might be too heavy for further use, while similar animals failing to gain in weight would be pooled for assay purposes with previous "lightweights" which had grown normally. Further experimental work would be necessary to investigate this problem satisfactorily, but so far as the theme of this paper is concerned it is sufficient to record that the slope of the assays cannot be considered constant.

## Summary

1. The distribution of probits of responses out of 24 to high and low doses of insulin in routine mouse assays is approximately normal.

2. Equally significant positive correlations hold between the responses to standard and unknown samples as between high and low doses. No significant correlation could be detected between the slopes of standard and test within assays, although in the overall picture the two were parallel.

3. The mean slopes for standard and test within assays were normally distributed.

4. The reciprocals of the standard errors of log-potency do not differ significantly in their distribution from the normal, but the effect of using weighting coefficients for different responses is to make this distribution approximate if the slopes are distributed in a truly normal manner.

5. The slope of the assays has been shown to be directly proportional to the body weight of the mice, when these have not been used previously. This relationship did not appear to apply to animals used for the second, third, or fourth time.

## References

1. Finney, Probit Analysis, University Press, Cambridge, 1947.
2. Hemmingsen, *Quart. J. Pharm. Pharmacol.*, 1933, **6**, 39, 187.
3. Smith, Hormone Assay, Chap. II, "Insulin," Academic Press Inc., New York. 1950.