Brit. J. Pharmacol. (1963), **21**, 67–77.

BALANCED INCOMPLETE BLOCK DESIGNS IN BIOLOGICAL ASSAY ILLUSTRATED BY THE ASSAY OF GASTRIN USING A YOUDEN SQUARE

BY

D. COLQUHOUN

From the Department of Pharmacology, University of Edinburgh Medical School

(*Received March* 19, 1963)

General equations are given for computing the results of any $(j+j)$ dose parallel line biological assay based on an incomplete block design. Only the information from comparisons within blocks is used. The computation is illustrated numerically by means of a (2+2) assay of gastrin devised and performed by Lai (1962), in which only three of the four doses could be given to each animal.

The aim of this paper is to illustrate the use of balanced incomplete block designs for biological assay.

The method of computation is illustrated using results obtained by Lai (1962), who used a (2+2) dose Youden square design for the assay of gastrin. The equations given are, however, generally valid for any $(j+j)$ dose biological assay based on a balanced incomplete block design. Catalogues of such designs are given by Cochran & Cox (1957) and by Fisher & Yates (1957).

Only that part of the computation based on comparisons within blocks is given, partly for simplicity and partly because, as explained below, more elaborate calculations would frequently yield little extra information (as in the present example).

THE NATURE OF THE DESIGN

The assay to be described is of the (2+2) type, that is it uses two doses of the standard preparation of gastrin and two of the preparation of unknown potency. Obviously it would be best to give all four doses to each rat (using as many rats as are necessary for the accuracy required) so that the estimate of potency of the unknown preparation is not biased by differences in sensitivity to gastrin between the individual rats. Thus a randomized block or Latin square design would normally be used. However, Lai (1962) found it impracticable to obtain from each animal responses to more than three of the four treatments. The duration of action of each dose was such that if all four were given to the same animal responses would have had to be so close together that they would interact, or else the assay would have had to be impossibly long. Consequently the fourth dose had to be carried over to a different animal.

The aim of the method of calculation used is to give, with as little computation as possible, an estimate of the potency of the unknown preparation unbiased by differences between rats, limits of error for this estimate, and tests of the validity of the assay.

The Youden square (Youden, 1937, 1940) is a member of the class of experimental designs known as balanced incomplete blocks (Yates, 1936, 1940). The blocks (in this instance individual rats) are incomplete because each contains fewer than the total number of treatments; only three out of four in the present analysis. The design is said to be balanced when, as in the present example, each treatment is given an equal number of times and each possible pair of treatments is given to the same number (λ) of rats (for the design in Table 1, two rats).

TABLE 1
THE GENERAL DESIGN (*A*) AND THE PARTICULAR CASE (*B*) OF THE GENERAL DESIGN

(*a*)

Block	Treatment *A*	*B*	*C*	*D*
1	X	X	X	
2	X	X		X
3	X		X	X
4		X	X	X

(*b*)

Rat	Order 1	2	3
1	*A*	*B*	*C*
2	*D*	*A*	*B*
3	*C*	*D*	*A*
4	*B*	*C*	*D*

The characteristic feature of Youden squares (which are really rectangles) is that, as in Latin squares, every treatment occurs in each column, and in any given column each treatment occurs equally often. In fact the designs are incomplete Latin squares and the example in Table 1 can be formed by omission of one column from a 4×4 Latin square. As the columns represent the *order* in which doses are given, the means of the columns can be used to judge whether there is any systematic difference between responses to the first, second and third doses of the day.

Table 1*a* shows the basic design employed by Lai (1962), using four rats. Table 1*b* gives a particular case of Table 1*a*. It should be randomized by re-arranging the order of the rows, and then of the columns in a sequence taken from a table of random numbers (for example Fisher & Yates, 1957). The resulting design will still be a particular case of that in Table 1*a*.

The treatments are next assigned randomly to the letters. Say, in the present example, $A\equiv HS$, $B\equiv LS$, $C\equiv HT$ and $D\equiv LT$, where *HS* and *LS* are the high and low doses of the standard preparation and *HT* and *LT* are the high and low doses of the test (unknown) preparation. Thus according to the scheme in Table 1*b* the first rat is given first *HS*, then *LS* and finally *HT*. The second rat is given *LT*, *HS* and *LS* in that order: and so on.

The symbols used are:

$t=2j$	Number of treatments
k	Number of experimental units per block (responses per rat in the present example)
r	Number of replicates of each treatment
b	Number of blocks (rats)
$N=tr=bk$	Number of observations
G	Grand total of all responses
D	Ratio between successive doses, the same for both standard and unknown
z	Dose
x	Log dose.

TABLE 2

APPLICATION OF THE DESIGN SHOWN IN TABLE 1

(*a*) Results of an assay using the design in Table 1*b*. Responses are expressed as the mean rate of acid secretion (μequiv./10 min) (Lai, 1962)

		Order						
		1		2		3		Total
Rat	1	*HS*	2·190	*LS*	0·975	*HT*	1·700	4·865
	2	*LT*	1·570	*HS*	3·130	*LS*	1·850	6·550
	3	*HT*	2·570	*LT*	1·680	*HS*	3·000	7·250
	4	*LS*	1·150	*HT*	2·275	*LT*	0·730	4·155
Total			7·480		8·060		7·280	22·820

(*b*) Total and mean responses to each treatment

Treatment	Total response	Uncorrected mean
LS	3·975	1·325
HS	8·320	2·773
LT	3·980	1·327
HT	6·545	2·182
	22·820	

For the results in Table 2: $t=4, j=2, r=3, k=3, b=4$ and $N=12$. The standard doses were 11.0 and 5.5 μg of a purified preparation of hog gastrin, so $D=2$. The "unknown" solution was actually made from the same preparation in this assay, and contained 1,100 μg/ml.; it was given in doses of 0.00750 and 0.00375 ml. The response was the mean rate of acid secretion (μequiv./10 min).

The computations can be divided into four parts:

Corrected mean responses

To compute, for each treatment, the mean response corrected for the bias introduced by the differences between rats, the quantity Q is first formed for each treatment. Q_i, the Q for the ith treatment, is defined as

$Q_i=k$ (sum of responses for the ith treatment) –
(sum of the block sums for those blocks that contain the ith treatment)

Use of the values in Table 2 gives

$$Q_{LS}=(3\times 3.975)-(4.865+6.550+4.155)=-3.645$$
$$Q_{HS}=(3\times 8.320)-(4.865+6.550+7.250)=\ 6.295$$
$$Q_{LT}=(3\times 3.980)-(6.550+7.250+4.155)=-6.015$$
$$Q_{HT}=(3\times 6.545)-(4.865+7.250+4.155)=\ 3.365$$

The results can be checked by making sure that the sum of the Q's is zero.

The corrected mean response to the ith treatment is, in general,

$$(\bar{y}_i)\,corr=\frac{(t-1)\,Q_i}{N(k-1)}+\frac{G}{N}$$

where G/N is the grand mean of all the responses. This expression can be derived from the model for the analysis by a least squares multiple regression method (see, for example, Brownlee, 1960). In the present example

$$(y_{LS})corr=\frac{(4-1)}{12(3-1)}(-3.645)+\frac{22.820}{12}$$
$$=\frac{-3.645}{8}+1.9017=1.4460$$

Similarly

$$(\bar{y}_{HS})corr=\frac{6.295}{8}+1.9017=2.6885$$
$$(\bar{y}_{LT})corr=\frac{-6.015}{8}+1.9017=1.1498$$
$$(\bar{y}_{HT})corr=\frac{3.365}{8}+1.9017=2.3223$$

These corrected means, and the uncorrected means given in Table 2*b*, are plotted in Fig. 1. It can be seen that in this assay the uncorrected response/log dose lines are seriously non-parallel, but after correction the deviation from parallelism is insignificant (as indicated by the analysis of variance below) so a valid estimate of relative potency can be made.

Tests for invalidity. The analysis of variance

Before the result is calculated it should be shown that the assay is not demonstrably invalid, as is done in the usual analysis of bioassays. Obviously it is impossible to test for deviation from linearity in a (2+2) assay. However, it has been shown (Lai, 1962) that, in his gastrin assay, deviation from linearity is not significant over a range of response wider than that used in the present assay.

The general form of the part of the analysis of variance to be described is shown in Table 3. The results for the present assay are given in Table 4. The sum of squares for each source of variability is computed first, and entered in Table 4.

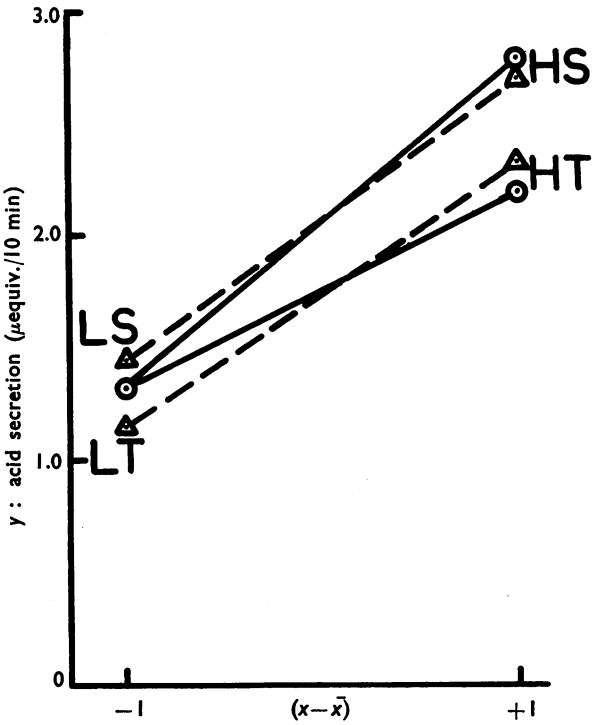

Fig. 1. Result of gastrin assay on four rats. The abscissa is $(x-\bar{x})$ where $x=\log_{\sqrt{D}}$ (dose). The ordinate is response expressed as mean rate of acid secretion (μequiv./10 min). ⊙——⊙ uncorrected treatment means; △ - - - - △ corrected treatment means.

TABLE 3

GENERAL FORM OF PART OF THE ANALYSIS OF VARIANCE

Source of variation	d.f.	Sum of squares
Linear regression	1	K
Deviation from parallelism	1	H
Between preparations	1	F
Deviation from linearity	$t-4$	M
Between treatments (eliminating blocks)	$t-1$	D
Between blocks (ignoring treatments)	$b-1$	C
Between columns	$k-1$	B
Error (intrablock)	$N-b-t-k+2$	$E=A-(B+C+D)$
Total	$N-1$	A

TABLE 4

RESULTS OF ASSAY

Source of variation	d.f.	Sum of squares	Mean square	Variance ratio	P
Linear regression	1	3·8881	3·8881	150·0	<0·001
Deviation from parallelism	1	0·0033	0·0033	0·1273	Not signif.
Preparations	1	0·2926	0·2926	11·29	0·05–0·01
Between treatments (eliminating rats)	3	4·1840	1·3947	53·81	<0·001
Between rats (ignoring treatments)	3	2·0697	0·6899		
Order of administration	2	0·0821	} 0·02592		(>0·2)
Error (intrablock)	3	0·0475			
Total	11	6·3833			

(1) *The total sum of squares.*

$$A = \Sigma y^2 - \frac{G^2}{N}$$

giving, for the results in Table 2,

$$(2.190^2 + 1.570^2 + \ . \ . \ . \ . \ + 0.730^2) - \frac{(22.820)^2}{12}$$
$$= 49.7793 - 43.3960 = 6.3833$$

(2) *Sum of squares for variation of response with order of administration.*

$$B = \frac{\Sigma(\text{column totals})^2}{b} - \frac{G^2}{N}$$

This can only be calculated if the design is a Youden square as well as a balanced incomplete block. The results in Table 2 give

$$\frac{(7.480)^2 + (8.060)^2 + (7.280)^2}{4} - 43.3960 = 0.0821$$

(3) *Sum of squares between rats (ignoring treatments).*

$$C = \frac{\Sigma(\text{rat totals})^2}{k} - \frac{G^2}{N}$$

giving, for the data of Table 2,

$$\frac{(4.865)^2 + (6.550)^2 + (7.250)^2 + (4.155)^2}{3} - 43.3960 = 2.0697$$

(4) *Sum of squares between treatments (eliminating effect of differences between rats).*

$$D = \frac{(t-1)\,\Sigma Q^2}{Nk(k-1)}.$$

giving, using the values of Q calculated above,

$$\frac{(4-1)}{12 \times 3 \times (3-1)} [(-3.645)^2 + (6.295)^2 + (-6.015)^2 + (3.365)^2] = 4.1840$$

This sum of squares can now be split into independent components which provide the tests for invalidity, as indicated in Table 3. The individual sums of squares are found via the *contrasts* which are so defined that, on the null hypotheses that the slope, deviation from parallelism, etc., are zero, the contrasts would, on the average, be zero.

The contrasts are computed using the coefficients given, for example, by Finney (1952, Tables 5.3, 5.4 and 5.7) in combination with the values of Q already found (rather than, as in ordinary complete block assays, in combination with the total responses).

(a) *Sum of squares due to linear regression.* First the regression contrast, L_1, is computed. Using the values of Q found in the present example

$$L_1 = -Q_{LS} + Q_{HS} - Q_{LT} + Q_{HT}$$
$$= 3.645 + 6.295 + 6.015 + 3.365 = 19.320$$

The sum of squares is now found, in general for a $(j+j)$ dose assay, as

$$K=\frac{(t-1)(L_1)^2}{cjNk(k-1)}$$

where

$$c=\frac{2}{3}(j^2-1) \text{ if } j \text{ is even} \qquad (1)$$

$$=2 \text{ in the present example,}$$

$$\text{or} \quad c=\frac{(j^2-1)}{6} \quad \text{if } j \text{ is odd.} \qquad (2)$$

In the present example ($j=2$)

$$K=\frac{(L_1)^2}{96}=\frac{(19.320)^2}{96}=3.8881$$

(b) *Sum of squares due to deviation from parallelism.* The deviation from parallelism contrast is, in the present example,

$$L'_1=Q_{LS}-Q_{HS}-Q_{LT}+Q_{HT}$$
$$=-0.560$$

The sum of squares is found using the same equation as for the sum of squares for regression, but using L'_1 instead of L_1. Thus, in the present example,

$$H=\frac{(-0.560)^2}{96}=0.0033$$

(c) *Sum of squares due to difference in response to standard and unknown preparations.* The between preparations contrast is

$$L=-Q_{LS}-Q_{HS}+Q_{LT}+Q_{HT}$$
$$=-5.300$$

and the sum of squares is, in general,

$$F=\frac{(t-1)(L_p)^2}{2jNk(k-1)}$$

Thus, in the present example,

$$F=\frac{(-5.300)^2}{96}=0.2926$$

(d) *Sum of squares for deviation from linearity.* This can be calculated as the difference between the total sum of squares between treatments and the sum of the preceding components. When $j=2$, as in the present example, the difference should be zero and this provides a check for arithmetical accuracy.

The analysis of variance can now be completed in the usual way. The results are set out in Table 4. In the present example the variance ratio for " order of administration " is 0.0410/0.0158=2.59 with 2 degrees of freedom in the numerator and 3 in the denominator. Although this corresponds to $P>0.2$ it is not a small enough variance ratio to comply with the criterion of Paull (1950) for the pooling

of variances. However, Lai (1962) has only once detected any apparent difference between columns in thirty-four assays similar to that described here, and over half these assays gave variance ratios for " order " of less than 1.0. Also in two larger (12-rat) assays the variance ratios for " order " were F(2,19)=1.066 and 0.334, easily complying with Paull's recommendation. In the light of these values it was decided that there was no reason to believe that response depended at all on the order of administration of doses and that, in order to increase the power of the analysis, the sums of squares for columns and for error should be pooled. The pooled estimate of error thus becomes $(0.0821+0.0475)/(2+3) = 0.02592$ with $(2+3)=5$ degrees of freedom. It is this error variance which is now used to form the variance ratios in Table 4.

It is seen from Table 4 that there is no reason to consider the assay invalid, so the potency ratio and its limits of error can now be computed.

The potency ratio

This is calculated from the corrected means for treatments. When the contrasts L_1 and L_p are defined as above, the equation for the potency ratio reduces to the usual form for a $(j+j)$ assay (see, for example, Finney, 1952):

$$R=\frac{z_S}{z_T}\text{antilog}_{10}\left[\frac{cL_p}{L_1}\log_{10}\mathbf{r}\right]$$

where R is the potency ratio, z_S/z_T is the ratio of standard dose to unknown dose, c is defined in equations (1) and (2), and

$$\mathbf{r}=\sqrt{D} \text{ if } j \text{ is even} \qquad (3)$$
$$=\sqrt{2} \text{ in the present example,}$$
$$\text{or} \quad \mathbf{r}=D \text{ if } j \text{ is odd.} \qquad (4)$$

The factor $\log_{10}\mathbf{r}$ occurs because the calculations have implicitly used the logarithms to base $\mathbf{r}$ of the doses. In the present example this makes both $(x_S-\bar{x}_S)$ and $(x_T-\bar{x}_T)$ take the value -1 for the low doses and $+1$ for the high doses (see Fig. 1).

The units of the potency ratio depend on the units used for z_S and z_T. If z_S is in mg (or units) and z_T in ml. then the equation above gives the potency of the unknown solution directly in mg (or units) per ml. For the present example, remembering that $2\log\sqrt{2}=\log 2$,

$$R=\frac{11.0}{0.0075}\text{antilog}_{10}\left[\frac{-5.300}{19.320}\log_{10}2\right]$$
$$=1{,}213\ \mu\text{g/ml.}$$

As the true potency of this solution was 1,100 μg/ml. this estimate is about 10% high.

The fiducial limits of the potency ratio

The limits of error computed for the potency ratio are such that if it is consistently asserted that the true potency of the unknown solution lies between the limits then, on the average, 95% (or any other chosen proportion) of such assertions will be correct.

The limits are computed by means of Fieller's theorem (Fieller, 1944; Finney, 1952, p. 27) which is applied, as usual, to the ratio $(\bar{y}_T-\bar{y}_S)/b_c$. For a $(j+j)$ dose incomplete block assay this ratio is equal to cL_p/L_1, when c, L_p and L_1 are defined as in the previous sections. The theorem can be written:

$$\text{Limits of } R = \frac{z_S}{z_T}\,\text{antilog}_{10}\left[\left\{\frac{cL_p/L_1}{(1-g)} \pm \frac{s\mathbf{t}}{b_c(1-g)}\sqrt{(1-g)v_{11}+\left(\frac{cL_p}{L_1}\right)^2 v_{22}}\right\}\log_{10}\mathbf{r}\right]$$

where $g=\dfrac{s^2\mathbf{t}^2v_{22}}{b_c^2}$, an index of significance of the slope, b_c;

c is defined by equations (1) and (2);

$\mathbf{r}$ is defined by equations (3) and (4);

$$v_{11}=\frac{2k(t-1)}{jN(k-1)}$$

$=3/8$ in the present example;

$$v_{22}=\frac{k(t-1)}{cjN(k-1)}$$

$=3/32$ in the present example.

The estimate of slope, combining data for both preparations and using corrected responses, is given by

$$b_c=\frac{(t-1)L_1}{cjN(k-1)}$$

$=19.320/32=0.60375$ in the present example.

s^2 is the error variance from the analysis of variance, that is 0.02592 with 5 degrees of freedom in the present example.

$s=\sqrt{0.02592}=0.16099$

$\mathbf{t}$ is Student's t (tabulated in, for example, Fisher & Yates, 1957) for the level of confidence wanted and with the number of degrees of freedom of the error variance. For $P=0.95$ and 5 degrees of freedom, $\mathbf{t}=2.571$.

$$\frac{z_S}{z_T}=\frac{11.0}{0.0075}=1466.67$$

$$\frac{cL_p}{L_1}=\frac{2(-5.300)}{19.320}=-0.54865$$

$$\text{Thus } g=\frac{2.571^2\times 0.02592}{(0.60375)^2}\times\frac{3}{32}=0.04407$$

and $(1-g)=0.9559$

As g is small a simplified equation could be used for the limits (Finney, 1952). However, it is almost as quick to use the full equation. The part in square brackets of the equation for the limits is

$$\left[\left\{\frac{-0.54865}{0.9559}\pm\frac{0.16099\times 2.571}{0.60375\times 0.9559}\sqrt{\left(0.9559\times\frac{3}{8}\right)+\left((-0.54865)^2\times\frac{3}{32}\right)}\right\}0.1505\right]$$

$$=-0.15352,\ -0.01926$$

and thus

Upper fiducial limit of R $=1466.67$ antilog (-0.01926)
$=1{,}403$ μg/ml.

Lower fiducial limit of R $=1466.67$ antilog (-0.15352)
$=1{,}030$ μg/ml.

Summarizing, the result of the assay is

Estimated concentration of " unknown " solution	1,213 μg/ml.
Error of this estimate	+10.3%
Fiducial limits (P=0.95) of estimated concentration	1,030 to 1,403 μg/ml.
that is, expressed as a percentage,	−15.1 to +15.7%

DISCUSSION

Replication. In Table 1*b* each dose is given three times and four rats are used. If greater accuracy is needed eight rats may be used (giving six replicates of each dose) simply by placing another separately randomized square directly below Table 1*b*. Every column will now contain each treatment twice. Similarly, by placing three squares one below the other, a design using twelve rats, and with nine replicates of each dose, is produced.

When twelve rats are used it is possible to eliminate any (additive) residual effects of one response on the next by having each dose given second in order preceded by each other dose given first, and similarly for the dose given third in order and second (D. J. Finney, personal communication). Such a modification of the present design is produced by omitting the last column of a design given by Finney (1952, Table 10.7, p. 275). In the case of the assay of purified hog gastrin and crude human gastrin Lai (1962) has shown that residual effects are negligible and that four rats give adequate accuracy for most purposes. However, in responses to crude hog gastrin residual effects were quite noticeable, and if it were required to assay this material the 12-rat balanced design might be useful.

With four rats Lai found the 95% fiducial limits of the potency ratio were usually between ±15 and ±20%. In a 12-rat assay, using a preparation of known potency, the potency estimate was only 1.4% in error with fiducial limits of ±12.3%.

Efficiency of the design. When not all of the treatments can be compared in the same block the comparisons between block totals will contain a certain amount of " interblock " information about treatment effects, which has been ignored in the simple analysis above. The efficiency of a balanced incomplete block design is, in general, defined as $[t(k-1)/k(t-1)]$ (Yates, 1936), that is about 89% in the present example. This means that if the error variance in the present example were the same as the error variance which would be found if all four treatments could be given to each rat, the variance of comparisons between treatments based on the latter design would be about 89% of the variance of comparisons based on the present design. However, since the four treatments cannot be given to each rat, the incomplete block design is much more efficient than any available alternative.

Recovery of interblock information. When the number of blocks is greater than the number of treatments (as, in the present example, when more than four rats are used) an independent estimate of potency can be obtained from the interblock information about treatment effects. This estimate is combined with the intrablock estimate already obtained to form a weighted mean potency estimate.

Data from assays based on eight and twelve rats have been further analysed by splitting the sum of squares for blocks (ignoring treatments) into a treatment component with $(t-1)$ degrees of freedom and an interblock error with $(b-t)$ degrees of freedom (see, for example, Fisher & Yates, 1957, Introduction). The treatment component can then be divided into contrasts in a way analogous with that used for the sum of squares for treatments eliminating rats. From these contrasts the interblock potency estimate is computed.

However, because of the initial high efficiency of the particular incomplete block design used for the gastrin assay, and because of the considerable differences between rats, no significant amount of information was recovered by the interblock analysis even when twelve rats were used. Even in far larger assays the interblock estimate would have so much less weight than the intrablock estimate that it is doubtful whether its use would, in the case of the gastrin assay, result in any worthwhile improvement on the simple intrablock estimate of potency described above.

It is possible that in assays based on less efficient designs than that used in the present example, or in assays where there is not much difference between blocks, the interblock information might be worth recovering even though its use means that the fiducial limits for the potency ratio must be based on less universally accepted theory than that used in the simple analysis. Finney (1952) describes the calculations for a (3+3) assay with two treatments per block. Bliss (1947) gives the analyses for two balanced incomplete block designs which are also cross-over designs, with recovery of interblock information by the more elaborate method of Yates (1940).

I am very grateful to Dr R. C. Elston for his invaluable advice and criticism of the typescript, to Dr D. J. Finney for his advice in the initial stages and to Dr K. S. Lai for supplying the problem.

REFERENCES

BLISS, C. I. (1947). 2×2 Factorial experiments in incomplete groups for use in biological assays. *Biometrics*, **3**, 69–88.

BROWNLEE, K. A. (1960). *Statistical Theory and Methodology in Science and Engineering.* New York: Wiley.

COCHRAN, W. G. & COX, G. M. (1957). *Experimental Designs.* New York: Wiley.

FIELLER, E. C. (1944). A fundamental formula in the statistics of biological assay, and some applications. *Quart. J. Pharm.*, **17**, 117–123.

FINNEY, D. J. (1952). *Statistical Method in Biological Assay.* London: Griffin.

FISHER, R. A. & YATES, F. (1957). *Statistical Tables for Biological, Agricultural and Medical Research*, 5th ed. London: Oliver & Boyd.

LAI, K. S. (1962). *A Method of Biological Assay of Gastrin and its Application to the Study of Human Tissues.* Ph.D. Thesis at the University of Edinburgh.

PAULL, A. E. (1950). On a preliminary test for pooling mean squares in the analysis of variance. *Ann. math. Statist.*, **21**, 539–556.

YATES, F. (1936). Incomplete randomized blocks. *Ann. Eugen. (Lond.)*, **7**, 121–140.

YATES, F. (1940). The recovery of inter-block information in balanced incomplete block designs. *Ann. Eugen. (Lond.)*, **10**, 317–325.

YOUDEN, W. J. (1937). Use of incomplete block replications in estimating tobacco-mosaic virus. *Contr. Boyce Thompson Inst.*, **8**, 41–48.

YOUDEN, W. J. (1940). Experimental designs to increase accuracy of greenhouse studies. *Contr. Boyce Thompson Inst.*, **11**, 219–228.