The Gain in Efficiency Resulting from the Use of Balanced Designs.

By F. Yates.

In a paper on experiments on human nutrition, published in this issue of the *Supplement*, Mr. Wake Simpson described a balanced design for testing four treatments on school-children during the three terms of a school year (his Table IV). He left to me the task of assessing how far this design had actually resulted in a gain in efficiency, as compared with various possible alternative designs of the randomized block type which might have been adopted. In my comments at the meeting I gave estimates of these gains, without saying how they were derived. It may be of interest to describe here the methods adopted, as the discussion will serve to throw light on the structure of these interesting and important balanced designs.*

Mr. Wake Simpson's design is reproduced below (Design I).

Design I.

Children.

Term	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
I	O	O	O	O	O	O	A	A	A	A	A	A	B	B	B	B	B	B	C	C	C	C	C	C
II	A	A	B	B	C	C	O	O	B	B	C	C	O	O	A	A	C	C	O	O	A	A	B	B
III	B	C	C	A	A	B	B	C	C	O	O	B	A	C	C	O	O	A	A	B	B	O	O	A

In order to assess the gain in efficiency, we must consider what alternative designs are available. One alternative is to keep the same child on the same treatment throughout the three terms, randomizing the treatments in blocks of four children chosen so as to be alike as possible. This gives a design of the type:

Design II.

Children

Term	1	2	3	4	5	6	7	8	. . . . . .
I	C	O	A	B	A	C	B	O	. . . . . .
II	C	O	A	B	A	C	B	O	. . . . . .
III	C	O	A	B	A	C	B	O	. . . . . .

The arrangement is analogous to the split-plot type of arrangement in agricultural field trials. There will be two separate errors, one

* See also Yates [2], [3], and Fisher and Yates [1].

Downloaded from https://academic.oup.com/jrssb/article/5/1/70/7026462 by IACR Rothamsted user on 27 February 2026

for the means over the three terms, and the other for the interactions of treatments with terms. In so far as the growth rate of the same child is similar in the three terms, the former error will be relatively larger than the latter. No proper assessment of cumulative or residual effects is possible, but the arrangement is such that the maximal cumulative effects will be included in the final assessment of the treatments.

A second alternative is to re-randomize the treatments within each block of four children for each term. This gives a design of the type:

DESIGN III.

Children.

Term	1	2	3	4	5	6	7	8	.	.	.
I	O	B	C	A	O	C	B	A	.	.	.
II	B	A	C	O	O	C	A	B	.	.	.
III	A	C	O	B	A	C	O	B	.	.	.

With this arrangement the mean effects of the treatments over the three terms, and the interactions of treatments with terms, are determined with the same relative precision, but again no simple or direct assessment of any cumulative and residual effects which extend beyond the duration of the holidays is possible.

With Design I, however, not only is the greatest precision obtained where it is most required—namely, in the comparison of the means of the treatments over the three terms—but direct and comparatively simple estimates of the residual effects which manifest themselves in the succeeding terms can be made. On the other hand, the cumulative effects have been deliberately eliminated; to investigate these, designs of types I and III would have to be combined. Moreover, the estimates of the average effects which are discussed in the following paragraphs are based on the assumption that residual effects persisting beyond the period of the previous holidays are in fact negligible. Some modification of the procedure here outlined will be necessary when this is not found to be the case.

Leaving out of account any residual or cumulative effects, the average effects of the treatments will be assessed by the comparison of the treatment means over the three terms. The precision of these comparisons for each design must therefore be assessed.

We may suppose the growth rate of each child to be made up of two parts, one constant over the three terms, and the other varying from term to term. Thus for any child the growth rate in the three terms may be represented by $g + t_1$, $g + t_2$, $g + t_3$, where we may assume g has a residual variance G (after eliminating the effects of

Downloaded from https://academic.oup.com/jrssb/article/5/1/70/7026462 by IACR Rothamsted user on 27 February 2026

any blocks or other restrictions that may be incorporated in the design), and t_1, t_2 and t_3 are independently distributed about zero with variance T. Note that in certain cases the growth rates of the same child in the three terms may be negatively correlated, but that this can be formally represented by allowing the variance T to assume a negative value, which must not be numerically greater than $3G$. There is, of course, also the possibility, which we have not considered here, that the growth rates of the same child for terms I and III may be less closely correlated than are those for terms I and II and for terms II and III.

In Design II the variance of the mean growth rate over the three terms is clearly $G + \frac{1}{3}T$ per child. In Design III it is $\frac{1}{3}(G + T)$ per child, since in each term the variance per child is $G + T$, and the three terms can be treated as independent, in virtue of the randomization process adopted.

The balanced design requires more careful consideration. The first point to notice is that there are two separate estimates of the treatment effects. One is given by the ordinary means, the other by adjusting these means so as to eliminate from the treatment comparisons any constant differences between children. The adjustment process is very simple: it consists of calculating quantities Q by deducting from each treatment total one third the sum of the totals over the three terms of the growth rates of all children receiving that treatment in any one term. Thus with the arrangement shown the Q for treatment O is:

$Q_O = T_O - \frac{1}{3}$ (Sum of totals of all children except 9, 12, 15, 18, 21, 24) $= T_O - \frac{1}{3}S_O$ say,

where T_O is the treatment total for O. T_O clearly contains 18 units of the effect of O on the growth rate, while S_O contains 18 units of O and 12 units of each of the other treatments. Thus Q_O contains 12 units of O and -4 units of each of the others, so that $Q_A - Q_O$, for instance, represents 16 ($= 12 + 4$) times the difference of A and O. Thus the differences between $\frac{1}{16}Q_O$, $\frac{1}{16}Q_A$, $\frac{1}{16}Q_B$ and $\frac{1}{16}Q_C$ represent estimates of the treatment effects from which constant differences in growth rate between different children have been eliminated. These may be called the *intra-child* or *adjusted* estimates. The variance of each difference will be found to be $2 \times \frac{1}{16}T$.

The comparisons of $\frac{1}{6}S_O$, $\frac{1}{6}S_A$, $\frac{1}{6}S_B$ and $\frac{1}{6}S_C$ also represent estimates of the treatment effects, which are independent of the Q estimates and depend on comparisons of child totals only. These may be called the *inter-child estimates*. The variance of each difference is $2 \times \frac{3}{2}(G + \frac{1}{3}T)$.

It is easily seen that the unadjusted means are weighted means of these two sets of estimates, the weights being in the ratio $8:1$.

Downloaded from https://academic.oup.com/jrssb/article/5/1/70/7026462 by IACR Rothamsted user on 27 February 2026

This is the correct weighting for maximal information if G is zero. If, however, G is not zero, the appropriate weights for maximal information will be in the ratio

$$\tfrac{3}{2}(G + \tfrac{1}{3}T) : \tfrac{1}{16}T = 8\,(1 + 3\,G/T) : 1$$

If G/T is known, or can be assessed from the experimental results with sufficient accuracy (as is the case in extensive experiments), the appropriate weighting of the two estimates can be determined. If G/T is large, the intra-child (adjusted) estimates contain practically all the information.

In the particular experiment under consideration, as Mr. Wake Simpson pointed out, the analysis of the increments in weight indicated that G was negligible for this measurement, and therefore the use of the unadjusted means was fully justified. In the case of the increments in height, on the other hand, G/T was found to have a value of approximately unity.* Leaving aside the possible reduction in the value of G/T which may result from the grouping of the children into groups of four instead of groups of 24 (no assessment of this reduction is possible from the results of the experiment), we arrive at the following table for the relative efficiencies of the various designs in assessing the mean treatment effects.

Design	Type of mean	Invariance	Units of Information
I ...	Unadjusted	$18/(\frac{1}{3}G+T)$	150
	Adjusted	$16/T$	178
	Weighted	$16/T+\frac{2}{3}/(G+\frac{1}{3}T)$	183
II	Unadjusted	$18/3G+T)$	50
III	Unadjusted	$18/(G\quad T)$	100

In this table we have expressed the amount of information of each design in terms of the amount given by Design III, taking this as 100 units. It will be seen that there is a gain of 83 per cent. resulting from the adoption of the balance design. Over 97 per cent. of the information is contained in the adjusted means, the remaining 3 per cent. being derived from the inter-child comparisons. The unadjusted means contain 82 per cent. of the total information.

Design II, the design with continuous treatments, furnishes only one half of the information of Design III and only a little over a quarter of the information of Design I. It is clear, therefore, that the evaluation of the cumulative effects will require much more experi-

* This value was obtained by comparing the error mean square for the unadjusted treatment means over the three terms (expectation $T + \frac{1}{3}G$) with the similar mean square for the interaction of treatments with terms (expectation $T + \frac{4}{3}G$). More efficient methods of estimation are obviously available, and should be used in less extensive experiments.

Downloaded from https://academic.oup.com/jrssb/article/5/1/70/7026462 by IACR Rothamsted user on 27 February 2026

mental material to attain the necessary accuracy than is necessary when cumulative effects can safely be ignored.

References

[1] Fisher, R. A., and Yates, F. "Statistical Tables for Biological, Agricultural, and Medical Research." Oliver and Boyd, 1938. (In the press.)

[2] Yates, F. "Incomplete Randomized Blocks." *Ann. Eugenics*, 1936, Vol. VII, pp. 121–140.

[3] Yates, F. "The Design and Analysis of Factorial Experiments." *Imperial Bureau of Soil Science*, 1937, Technical Communication No. 35.

Downloaded from https://academic.oup.com/jrssb/article/5/1/70/7026462 by IACR Rothamsted user on 27 February 2026