ASPECTS OF DESIGN FOR REPEATED MEASUREMENTS

R. MEAD

The design of trials and investigations is extremely important and includes the choices of experimental or observational units, the form and timing of measurements, the choice of applied or environmental treatments and the interrelationship of these three components. It became obvious later during the workshop that many difficulties of analysis could be attributed to design deficiencies (in a broad sense).

The general principles of design, applicable to all investigations are:

(a) efficient use of resources

(b) asking many questions

(c) reducing σ^2.

The consequent statistical principles are:

(i) Using appropriate amounts and forms of replication.

(ii) Using random allocation of treatments to units within stated design restriction to provide a valid basis for inferences and for estimating σ^2.

(iii) Using small "blocks" to control random variation.

(iv) Using factorial structure with the implied priority ordering of effects.

(v) Using the minimum necessary number of levels of quantitative factors.

The aspects discussed in this paper are (i), (iii) and (v).

1. REPLICATION AND RESOURCES

Replication for comparative mean values must provide adequate power; for random variance σ^2 and a difference that should be detected,

d, this implies a replication, n, where

$$\sqrt{2\sigma^2/n} < d/3.$$

Note that n may include implicit, as well as explicit, replication, derived from treatment structure or regression.

For a given replication level there will be a total amount of resources, represented by the total degrees of freedom $(N-1)$ in an analysis of variance. These resources are used in three ways:

(i) Asking treatment questions

(ii) Estimating σ^2

(iii) Controlling variation.

Only for (ii) are there limits on the number of degrees of freedom. The minimum *df* necessary for estimating σ^2 are about 10; the maximum *df* useful for estimating σ^2 are about 20 (for justification of these statements see t-tables). A trial with less than 10 *df* for Error is of doubtful value. A trial with many more than 20 *df* for Error is inefficient and should be redesigned by asking more treatment questions or improving the control of variation to utilise the surplus degrees of freedom.

2. REPEATED MEASUREMENTS AND INTERDEPENDENT INFORMATION

Increasingly more and more measurements are recorded and it is important to ask how much more information is provided by repeated measurement through time of a single variable. In extreme situations technology permits literally thousandfold repetitions. Of course the replication of units per treatment is not increased at all by repeated measurements, a fact frequently ignored by experimenters (see Mead (1988) chapter 6 for more discussion of this issue).

Two examples of repeated measurements are given below. In the data on brood areas in hives for honey bees it is clear that, for diet 3, hive 8 consistently gives much lower error than hives 7 and 9. The

repetition of this occurrence does not substantially increase the available information. The experiment included nine experimental hives, three hives of each of three diets. The brood area in each hive was measured on five occasions, giving the results shown.

Hive	Diet	Brood Area				
		Time 1	Time 2	Time 3	Time 4	Time 5
1	1	368	920	944	1084	1108
2	1	308	980	1014	1008	894
3	1	284	500	754	908	518
4	2	684	974	808	490	208
5	2	288	624	570	280	334
6	2	468	778	638	310	178
7	3	444	1330	1010	748	448
8	3	280	290	248	118	80
9	3	380	828	830	598	432

The data on daily rat growth are extracted from a much larger study in which daily weights were recorded for ten weeks. It is clear that the intermediate values between days 28 and 35 provide virtually no information additional to that for days 28 and 35, the occasional deviation from the linear trend giving doubt about the value showing the deviation, rather than additional information. Plainly there are fluctuations in an individual rat's weight both between and within days and we need to understand the general patterns of such fluctuations to be able to make the most informative measurements.

Treat	M/F	Daily weight of rats Day 28	29	30	31	32	33	34	35
0	1	45	49	49	54	58	60	65	67
0	1	44	47	45	52	54	57	61	63
0	1	30	32	31	36	37	40	42	46
0	1	39	43	46	48	50	54	57	60
0	1	54	57	59	62	66	68	70	74
0	0	44	48	50	54	57	61	66	70
0	0	50	54	58	62	64	68	71	75
0	0	44	48	52	54	56	58	63	65
0	0	51	60	64	69	72	77	81	85
0	0	47	57	56	58	61	65	68	71
1	1	45	51	54	58	63	68	70	75
1	1	38	42	46	48	51	55	57	60
1	1	47	51	56	59	62	66	67	73
1	1	50	55	60	62	67	70	72	78
1	1	44	47	52	55	58	62	65	68
1	0	37	41	44	48	51	55	59	63
1	0	54	57	61	64	69	73	78	83
1	0	56	61	65	69	74	78	84	89
1	0	60	63	71	72	78	83	88	90
1	0	48	54	60	63	67	71	75	78

3. GROUP TERMINATION STUDIES

In some repeated measurement trials, groups of experimental units are sacrificed (killed, terminated, destructively harvested) and there are two sets of measurements. Repeated measurements on live units and single measurements on recently dead units. In the rat trial referred to above groups of between 5 and 7 rats were killed after 5, 6, 7, 9, 11 and 14 weeks of the trial.

Particular design choices for such trials are:

(i) Decisions about group sizes.

(ii) Decisions about termination times.

(iii) Decisions about "blocking" groups, and adjusting for differences between groups.

4. CHOICE OF MEASUREMENT TIMES (NO SACRIFICE)

We assume that the interest is in the overall response pattern through time, that individual animals have varying response patterns, and that response curves will be fitted for each animal and the variation between the resultant response curves used as a basis for interpreting differences between mean response curves.

If the total number of measurements (times $\times$ units) is fixed then standard theory on the choice of levels of a quantitative factor indicates clearly that we should choose more animals and fewer observation times, the number of observation times being set at p or $p+1$, where p is the number of parameters in the response curve family to be used for fitting. Usually p will be two or three. However, this particular form of restriction on total number of measurements is not often realistic.

If the total number of animals (units) is fixed, then, in a trivial sense, more observations means more information. However, the increase may be negligibly small. Often there will be benefits of information from thinking about the patterns of variation between months, between days, between hours, between minutes and even between seconds. For example suppose that we know that there is a within-day variation of ± 3 and a linear trend of 20 over a seven-day period. The "obvious" procedure of one measurement per day may be considerably less efficient than three spaced measurements on the first and last days. The latter system would hope to average out, systematically, the within-day variation and also use the theoretical advantage of two well-located observation points.

5. CONTROLLING VARIATION BETWEEN TERMINATION GROUPS

When several groups of animals are to be killed at pre-determined times it should be hoped that the experimenter uses some form of blocking to attempt to control possible variation between the termination groups. Even if this blocking is carefully chosen it may become apparent as the experiment progresses that the groups are more different than had been hoped. Consider the following data from the rat growth experiment. The individual rat weights for the last three termination groups are shown at the beginning of the experiment, and at weekly intervals up to the point when group 4 are about to be sacrificed.

Group	Week 4	5	6	7	8
4	45	67	93	109	129
	55	79	98	114	130
	41	61	84	101	118
	44	69	90	105	124
	43	77	98	114	132
5	51	78	90	118	137
	44	69	92	105	127
	58	87	112	131	151
	58	85	112	127	142
	55	81	104	118	136
6	43	64	86	103	119
	40	60	80	100	119
	51	82	108	127	142
	53	82	111	132	155
	49	72	95	113	131

The differences between the three groups are not very distinct

at week 4, but by week 8 the separation of groups 4 and 5 is almost complete. Clearly the blocking control of variation for groups 4, 5 and 6 would be most effective if based on the week 8 information, or even better on the joint information of week 4 and the week 4 to 8 difference.

However blocking for groups 1, 2 and 3 must be determined earlier and in a blocking form that includes the animals for groups 4, 5 and 6. If we attempt sequential blocking we must consider how the resultant data might be analysed. One possible approach would be to start with a coarse blocking division (into perhaps just two blocks) and to produce one, or more, subsequent subdivision of each block into sub-blocks. This could be modelled by a sequence of (fixed) block effects

$$b_i + b_{ij} + b_{ijk}$$

with

$$\sum_i b_i = 0\,, \qquad \sum_j b_{ij} = 0\,, \quad \text{each } i\,, \text{ etc.}$$

6. ANALYSING LIVE MEASUREMENTS FOR TERMINATION GROUPS

There are two types of problem. The first is detecting if there are differences between the termination groups which would make the interpretation of the simple time means unreliable. The second is adjustment of time means.

Detection of group differences can be achieved using models

$$y_{ijk} = \mu + g_i + \epsilon_{ijk}\,, \qquad \text{each time}\,,\ j\,,$$

or

$$y_{ijk} = \mu + g_i + t_j + \theta_{ij} + \epsilon_{ijk}\,,$$

where i, j, k represent group, time, and animal within group, and g_i, t_j and θ_{ij} represent group differences, time differences and group $\times$ time interactions.

Adjustment of time means to allow for group differences can be made using a model

$$y_{ijk} = \mu + g_i + t_j + \epsilon_{ijk}$$

or by a regression approach as used for supplementing partial information on one variable (parallel river flows, partial and whole animal carcase weights),

$$y_{ijk} = \alpha + \sum_{\ell<j} \beta_\ell \, y_{i\ell k} + \epsilon_{ijk} \, .$$

7. CHOICE OF GROUP SIZES FOR TERMINATION GROUPS

We have to consider the choice of group sizes in the context of two different sets of comparisons. For measurements on dead animals we have n_1 observations at time 1, n_2 at time 2, n_3 at time 3, and so on. For measurements on live animals we have n_1 observations for both times 0 and 1, n_2 for the three times 0, 1 and 2, n_3 for the four times 0, 1, 2 and 3, and so on. The gradual reduction of live animals results in an inevitably biassed distribution of information. If we consider a compromise criterion balancing the needs of both forms of measurement information, we should expect to find that we reduce the unevenness of the live information with a corresponding introduction of unevenness in the dead information.

We consider three specific small scale situations and a hypothetical general "rule".

A. TWO TERMINATION GROUPS: BOTH RESPONSES LINEAR

The number of observations are:

	Time 0	Time 1	Time 2
Death	–	n_1	n_2
Life	$n_1 + n_2$	$n_1 + n_2$	n_2

The variances for the linear regression coefficient are

Death $$\mathrm{Var}(\tilde{\beta}) = (\sigma^2 + \sigma_S^2)\left(\frac{1}{n_1} + \frac{1}{n_2}\right) + \sigma_\beta^2$$

Life $$\mathrm{Var}(\tilde{\beta}) = \sigma^2/(n_1 + 4n_2) + \sigma_\beta^2\,,$$

where σ_β^2 is the variance of true slopes over animals, σ_S^2 is the between subject (animal) variance component and σ^2 the within animal variance. In fact only the multiplying factors need to be considered.

The resulting variance factors for various alternative splits of a total of 20 animals, and the % increase in variance for death measurements compared with a 10/10 split, and the % increase of the reciprocal of variance for life measurements are shown below.

Group Sizes	Linear Response			
	Death		Life	
n_1/n_2	Var. factor	% Increase	Var. factor	% Increase (Var^{-1})
10/10	.2000		.0200	
9/11	.2020	+1%	.0189	+6%
8/12	.2083	+4.15%	.0179	+12%
7/13	.2198	+9.9%	.0170	+18%
6/14	.2381	+19.05%	.0161	+24%
5/15	.2667	+33.33%	.0154	+30%

If the two forms of information are assumed to be of equal importance then a reasonable compromise design would suggest

$$n_1 : n_2 = 3 : 5.$$

B. TWO TERMINATION GROUPS; DEATH LINEAR, LIFE QUADRATIC

The results for death are unchanged. For the life response we can consider the variance of the linear or quadratic coefficient or the generalised variance for the model

$$y = \alpha + \beta(t-1) + \gamma(t-1)^2 .$$

The variances are

$$\mathrm{Var}(\tilde{\beta}) = \sigma^2 \left(\frac{N + n_2}{4 n_2 N} \right) + \sigma_\beta^2$$

$$\mathrm{Var}(\tilde{\gamma}) = \sigma^2 \left(\frac{N + 5n_2}{4 n_2 N} \right) + \sigma_\beta^2$$

$$\mathrm{G.V.} = \sigma^2 / (4 n_2 N^2),$$

where $N = n_1 + n_2$. For the same n_1/n_2 splits the increase in information (= variance^{-1}) are

n_1/n_2	$\mathrm{Var}(\tilde{\beta})$	$\mathrm{Var}(\tilde{\gamma})$	G.V.
9/11	+6.5%	+2.7%	+10%
8/12	+12.5%	+5.0%	+20%
7/13	+18.3%	+7.1%	+30%
6/14	+23.5%	+8.9%	+40%
5/15	+28.7%	+10.6%	+50%

C. THREE TERMINATION GROUPS; BOTH RESPONSES QUADRATIC

The numbers of animals are

	Time 0	Time 1	Time 2	Time 3
Death	-	n_1	n_2	n_3
Life	$n_1 + n_2 + n_3$	$n_1 + n_2 + n_3$	$n_2 + n_3$	n_3

For the model,

Death $\quad y = \alpha + \beta(t-2) + \gamma(t-2)^2$

Life $\quad y = \alpha + \beta(t-1) + \gamma(t-1)^2$.

The variance-covariance matrices are

Death

$$\begin{pmatrix} n_1+n_2+n_3 & n_3-n_1 & n_1+n_3 \\ n_3-n_1 & n_1+n_3 & n_3-n_1 \\ n_1+n_3 & n_3-n_1 & n_1+n_3 \end{pmatrix} \qquad \text{G.V.} = (4n_1n_2n_3)^{-1}.$$

Life

$$\begin{pmatrix} 4n_3+3n_2+2n_1 & 2n_3-n_1 & 6n_3+2n_2+n_1 \\ 7n_3-n_1 & 6n_3+2n_2+n_1 & 8n_3-n_1 \\ 6n_3+2n_2+n_1 & 8n_3-n_1 & 18n_3+2n_2+n_1 \end{pmatrix}.$$

The values of the determinants (= G.V.$^{-1}$) for life and death information, for varying n_1 and n_3 ($n_2 = 30 - n_1 - n_3$) are

Death		n_1 10	8	6	4
	20				192
	18		230	259	230
	16	256	314	314	256
n_3	14	336	358	336	269
	12	384	384	346	269
	10	400	384	336	256

Life		10	8	6	4
	20				1386
	18		1138	1188	1258
	16	974	1020	1066	1131
n_3	14	862	902	943	1004
	12	749	785	821	877
	10	636	667	698	730

The Products of the two determinants are

		n_1				
	10	8		6		4
	20					266
	18	262		308	306	294
		293	315	322	313	
	16	249 320	327	334	313	290
n_3		323	330	325	308	
	14	290 323	326	319		270
		316				
	12	288 301		284		235
	10	255 256		235		187

The optimum compromise information appears to be in the range n_1 (5–8) n_2 (6–10) n_3 (13–17).

A possible rule of thumb for optimum compromise design group numbers could be to equalise the products of the Life and Death numbers across times. That is choose $n_1, n_2, n_3, \ldots$ to make

$$n_1 N\,,\ n_2(N-n_1)\,,\ n_3(N-n_1-n_2)\ldots n_5^2$$

as nearly equal as possible. The resulting designs would require:

Two groups	n_1/n_2	7.64/12.36
Three groups	$n_1/n_2/n_3$	6.83/8.85/14.32
Four groups	$n_1/n_2/n_3/n_4$	6.42/7.65/9.90/16.02
Five groups	$n_1/n_2/n_3/n_4/n_5$	6.17/7.04/8.38/10.85/17.56

The effect of autocorrelation of observations might lead to a possibly even greater value for the last group size because of the increased dependence on the extremes for the Life measurement, but this effect would be small. There is no effect, of course, on the Death measurements.

REFERENCE

Mead, R. (1988) The Design of Experiments: Statistical Principles for Practical Application. C.U.P.

Department of Applied Statistics
University of Reading
Whiteknights, PO Box 217
Reading RG6 2AN
UK