THE PLANNING OF EXPERIMENTS

J.H. MAINDONALD

Applied Mathematics Division, DSIR, Mount Albert Research Centre, Auckland

SUMMARY

Adequate planning is essential if expensive experiments are to achieve their objectives, yielding clear and interpretable results. The costing which is part of good planning allows a rational allocation of resources between competing areas of work. Both general and statistical aspects of planning are discussed. Most books on experimental design set limits to their subject which are too narrow.

INTRODUCTION

Good planning is sound economics. Can you afford to repeat the experiment? If not, then it is well to get it right the first time. Learning by doing may be acceptable if the experiment is cheap. Statistical planning is best set in the context of wider aspects of the planning of experiments. Hence the choice of title for this paper.

Nobel prizes call for more than merely good planning

Our discussion is concerned with experiments from which it is hoped to draw firm conclusions. Principles of experimental design do not show how to make the imaginative leaps which win Nobel prizes, but they are important to subsequent experimental work intended to confirm or reject a bold idea. Moreover imaginative leaps are likely to end in the ditch if the starting point is a collection of half-baked ideas based on the results of poorly conducted experiments.

Plan for profit

Planning, and the costing which goes with it, allows sensible allocation of resources. If experiments are intrinsically incapable of yielding the information required from them, they are a waste of resources, and effort is better concentrated elsewhere.

GENERAL PRINCIPLES

These remarks are primarily directed at field experimentation. Much of the discussion applies, with modification to fit the new circumstances, to any expensive or complex endeavour. If a computer network is to be installed it makes sense to verify, before committing oneself to the whole grand plan, that at least one local part of the system will work as intended. This is not too much different from the use of a pilot experiment as the first step in a large experimental investigation.

Watch out for the bargains

It's cheaper to use the results of other peoples' experiments than to repeat the work yourself. When it is necessary to repeat work (to perfect technique, or as a check on results) earlier results will help in planning the new experiment.

Look before you leap

A small pilot experiment is a cheap way of getting some initial information quickly, will give an indication of variability, will often draw attention to potential problems, and may be essential in order to get the information needed for the proper design of a large experiment.

Payoffs

One good experiment is worth any number of unsatisfactory experiments, which may mislead rather than enlighten. A well planned experiment may yield information which the experimenter had not anticipated, because the unusual and unexpected can be recognized for what it is.

Proc. 35th N.Z. Weed and Pest Control Conf.

Value should be understood quite broadly — to include any positive spinoff effects such as the widening of the experience of those involved. There is even value in learning that one needs to plan better next time. The discovery that two years' work have been wasted may concentrate the mind wonderfully. But it is a hard way to learn.

Two heads are better than one

There are two principles. Firstly the experiment should be checked out with at least one other person who is competent to comment. The second principle is that the individual who is to carry out the analysis should be involved in the planning. Discussion with a critical onlooker is always valuable. Discussion with the biometrician who will do the analysis gets him or her thinking about the analysis at the earliest possible stage. In addition it allows the experimenter to draw on the awareness which experience in analysing results brings of the difficulties which poor design poses for analysis and/or interpretation. The effort of explaining an experiment to a biometrician who is untrained in the scientific subject area has some incidental benefits. It is useful preparation for the writing of a report or paper which will explain the experiment to a wider public. The naive questions of someone from another specialism often highlight unstated assumptions.

Different times and different places

If experimental work is to be made the basis of advice to agriculture or horticulture generally then it is necessary to enquire whether results were obtained in a manner which makes them applicable next year, in another part of New Zealand. In some cases (e.g. variety trials) results found for one area will have no relevance to another locality. In other cases there may be seemingly compelling reasons why effects should be independent of geographic or climatic variation. Such claims should not be made, or accepted, lightly.

STATISTICAL PRINCIPLES

The main principles can be simply stated, without any mathematics. Putting them to use may involve mathematics.

Statisticians who strayed - confounding

A few years ago an article was published in a well-known statistical journal in which two methods for doing a calculation were compared for their precision. Method 1 was used on computer A, and gave very accurate results. Method 2 was used on computer B, and gave inaccurate results. The writers claimed that method 1 was better than method 2. In fact there was little difference in the methods. Computer A had much more precise arithmetic (about 14 decimal digits) than computer B (about 9 decimal digits). The two factors (machine and method of calculation) were, to use the statistical jargon, confounded.

It is nowadays well understood that an experiment in which treatment A is applied to trees in one orchard, and treatment B to trees in another orchard, is intrinsically incapable of yielding evidence of treatment differences which will stand scientific scrutiny. Any differences which are observed may be differences between orchards. Similar comments apply if the experimental unit is a tree or a group of trees, and each treatment is applied once only, to the next available experimental unit. Of course confounding often arises in much more subtle ways. Two different water supplies, one of them contaminated, may be used in mixing two different sprays. It is a nice matter of scientific judgement to decide what is a possible source of confounding, and what is not. It is unlikely to matter if the technician wears a different coloured hat for each different spray which is applied.

Randomization

Everyone knows what happens if a small boy's fingers are left to follow their natural bent to the next piece of cake. Subtler influences still are liable to guide experimenters in their allocation of experimental units to treatments, unless the matter is made genuinely the result of chance. It is not sufficient to line plants up along a bench, and take the next one that comes. The way in which plants are lined up along the bench will correspond in some rough manner to the way in which they were laid out in

the nursery. Exposure to an infection which spreads out from one point will lead to an infection gradient. Other gradients may arise from differential exposure to light or heat. The haphazard choice of experimental units (perhaps taking them as they come) is NOT random allocation.

If one knows what environmental effects are important allowance should be made for them in the experimental design. Randomization is the best one can do to cope with possible effects which cannot be handled in other ways.

Blocking

Suppose, to return to the small boy analogy, that a number of different kinds of cake are to be compared. Ignoring problems of order of eating (which are not trivial) the comparisons will be made with better precision if each boy is able to compare all the different cakes, and this is repeated a number of times over with a number of different boys. The boys are the blocks, and the different cakes are the different treatments.

Modern fertilizers even out most major fertility differences between one part of a field and another, and make the choosing of blocks on the basis of such differences less important than in the past. Differential exposure to disease or pests is nowadays more likely to be important to the choice of blocks.

Blocks are by no means necessarily blocks of land. A tray of apples in cold storage may be a block, with the apples on the tray divided up between the different treatments. In this way treatments can be compared without differences between trays entering to confuse the comparison. If differences between trays are much bigger than differences within trays then the precision of treatment comparisons will be improved. Improvement of precision by blocking is, when it is possible, usually cheaper than improvement from increasing the number of experimental units.

Other relevant information

If one cannot take account of an obvious influence by blocking it may be possible to adjust for it in the analysis. A spray, whose maximum effect is upon trees caught at king blossom, may catch some trees at earlier or later stages. In that case a rough assessment of the stage of blossom at spraytime should be recorded for each tree, and account taken of this assessment in the analysis. Past disease levels may be a strong influence on what is observed in the current season, and this information should be available at both the planning and the analysis stages of any new experiment. Quite rough information on past levels may allow worthwhile gains in experimental precision.

LET'S GET THE BIOMETRICIANS ON THE DEFENSIVE

It is sometimes alleged that biometricians exist to give respectability to poor experiments. This may be balanced against the occasional complaint that the biometrician's methods are inadequate to reveal the real effects which, the experimenter is convinced, are demonstrated by the data.

An ideal is well-designed experiments whose results stand out so clearly that little is required in the way of subsequent analysis. This is to emphasize that it is far more useful to contribute to careful design than to spend time in rescuing doubtful enterprises. Unfortunately the variability of biological material is such that the best one can do, even with a careful design, may be to turn up effects which are significant at the 5% level. Confidence in the results may come from repeating the experiment, and from other results which tend in the same direction, so that in due course the evidence becomes an adequate basis for recommending changes in agricultural practice.

If you can see it, it's real

Differences which are obvious to field observation may fail to show up as statistically significant because the recorded measurements do not adequately capture what the eye can see. But what of the claim that the biometrician's methods are inadequate to reveal effects which are clear from the data? It may turn out that the analysis has been wrong or inappropriate. The wrong error estimate may have been used in making treatment comparisons, or assumptions may have been made which did not fit the experimental setup. Or an outlier, or a mistake in the data, may have gone unnoticed. Otherwise the conclusion is inescapable that claimed effects may be due to chance, to the way that the dice fell. If the experimenter has other reasons independent of the

results which have been analysed for believing that effects are real, then these other considerations should stand or fall on their own merits.

My investigations are not amenable to planned experimentation

Exploratory work which precedes carefully planned experiments has a legitimate place. The penalty is that no sure conclusions can, in most instances, be drawn from what is observed. Difficult problems call for imaginative solutions. Thus spacing trials may become impossibly large and unwieldy if the attempt is made to lay out replicate groups of plants with all plants in a group at one of a few fixed spacings. A design in which spacing is varied systematically along each of a number of rows may give better information for acceptable effort. Such designs exploit an ordering in treatment levels to allow a trend analysis.

Sometimes difficulties with replication arise because alternative choices of experimental units have not been considered. Instead of making a tray of fruit the experimental unit, why not divide fruit on a single tray up among the alternative treatments?

BUT HOW WAS THE EXPERIMENT DONE?

This is the title of a paper (Ryder 1980) which discusses the topics of this section in more detail. Too often experimenters are vague in their description of the experimental setup, and of the nature of the randomization. The claim may be made that treatments were assigned at random to experimental units when the assignment was merely haphazard. A biometrician analysing results must be given details on how the data matches up with experimental units. The order in which plots were harvested may be important. Details of damage or confusion caused by pests, disease, weather, or marauding cows or other accident should be recorded.

Field staff responsible for an experiment will often have information relevant to the design or analysis of an experiment. It is important to consult with them. An understanding of the experimenter's requirements makes their work more satisfying, and allows them to contribute their own perspectives. This is more than ever important when experiments are conducted at a site which the superintending scientist visits infrequently.

OBSERVATIONAL STUDIES

Observational studies are sometimes a recourse when other approaches have failed. Careful observation, particularly when it yields small and manageable amounts of data, is fundamental to all scientific work. Designed experiments are set up so that the data which they yield, even if large in volume, can be interpreted by a trained analyst. By contrast the data obtained from an observational study involving a large number of variables brings with it a whole variety of possibilities for misinterpretation. Only in those fortunate cases where a number of different lines of evidence tend to the same conclusion can the investigator proceed with any confidence on the basis of such positive results as observational data may appear to yield. Naive attempts to draw firm conclusions from associations found in observational data leaves the investigator open to the charge that it is all "lies, damned lies, and statistics".

As an example of the spurious relationships which are the bane of anyone who has to work with observational data, consider the yarn factory where it was found that yarn quality increased with the amount of soft drink consumed at the factory cafeteria. It did not follow that the way to get better yarn was to urge employees to consume ever larger amounts of soft drink. Both changes were, as it turned out, associated with changes in humidity.

Intractable problems call, finally, for an understanding of the underlying biological mechanisms. It is a nice question where resources are best concentrated to achieve this.

REPORTING THE RESULTS

Data

In order to keep errors to a minimum data recording and entry should be kept as simple as possible. Hand copying of data should as far as possible be avoided, as this will introduce errors. Checksums are useful. Percentage components which sum to 100% should all be recorded, to allow a check that the numbers as entered into the computer add up to 100%. An ideal is to record data directly in machine-readable form, for immediate printing out and checking on the site. Devices which will allow this are just now becoming available. Further checks (particularly graphical checks) should accompany any analysis.

Data too often disappears once it has formed the basis of a published paper. This makes checking impossible in cases where a mistaken or inappropriate analysis, or an error or aberration in the data, is later suspected.

Analysis

Remember that the aim is to show the picture which the data presents. Authors too often use statistical analysis to add complication, rather than to summarize and give insight. Statistical tests provide a check on the reality of effects which may be present. Once that is established, interest centres on their magnitude, and on the patterns which they present. If it is obvious that any effects present are too small to be of any biological interest, the effort of doing a statistical analysis may not be justified.

REFERENCE

Ryder, K., 1980. But how was the experiment done? *Bias* 7(1), 22-23.

OTHER READING

Heath, O.V.S., 1972. *Investigation by Experiment*. Edward Arnold, London.

Pearce, S.C., 1965. *Biological Statistics: An Introduction*. McGraw-Hill, New York.

Pearce, S.C., 2nd. edn., 1975. *Field Experimentation with Fruit Trees and other Perennial Plants*. Commonwealth Agricultural Bureaux, Farnham Royal, Slough.

Symposium Insert (a collection of 5 papers), 1981. *Statistics: a tool for the horticultural scientist*. *Horticultural Science* 16, 621-640.

Note: The proper response to "Lies, damned lies, and statistics" is "Figures do not lie, but liars often figure!"