PLANNING AND INTERPRETING EXPERIMENTS

By N. MOUNTIER, Department of Agriculture, Wellington

To conduct an experiment is to standardise all the factors in our control except the ones we are testing.

It might be thought that one way of finding which is the most effective weedkiller for Californian thistle is to find a farmer who has used one product, another who has used another product, and so on, and go and see what effect each has had.

This method is quite valueless. The different farmers may have applied the weedkillers at different rates or under different conditions: one may have obtained good foliage cover, another poor. It is obvious therefore that one person should apply all treatments at the same rates and in the same conditions. Furthermore, it is no use applying one treatment on one farm and another on another farm, because the soil type, the growing conditions, and the vigour of the weeds may be different. Treatments should be applied to blocks of the same paddock.

That is, the factors we are not actually testing are standardised as far as they possibly can be. If this is not done, it can hardly be called an experiment. Of course, it is not possible to standardise *completely* on all things—two adjacent plots are bound to have slight differences in number of weeds, in soil fertility, and so on. It is to take account of these differences that we have replicated, randomised trials and I shall deal with this in more detail later.

Let us deal with a simple experiment first. Suppose that all we want to do is to test a new preparation on an annual thistle; and suppose that we already know the most effective time to spray and the optimum rate and that we are not concerned at present with the effects on pasture. Then the problem is simple and can be dealt with by a *series* of simple observational trials with one plot each of control, the standard treatment, and the new preparation at the same rate. Or better, two sets of plots could be laid down.

Treat. A Control Treat. B

Now, it cannot be emphasised too strongly that from *one* such trial we can conclude virtually *nothing*. In one trial the conditions may be, unknown to us, very favourable for killing the thistles on treatment A's plot and unfavourable for killing the thistles on treatment B's plot. But it is extremely unlikely that this will happen in the same way in 10 or 20 trials at different places. Therefore, it is only from a *series* of such trials that we can draw any conclusions at all. If we find that A consistently kills thistle better than B does in each of 10 trials, then we are able to say that this is a real difference.

This is the sort of trial that the Department of Agriculture has run (although usually with more treatments) on weeds such as annual thistle, Californian thistle, and gorse. The methods of observation differ. With annual thistle a brief trial is probably all that is needed

and with a perennial the trial must run at least a year, but the general form of the trials is similar.

The trials just described are essentially *observational* trials and no weights or counts would normally be taken on them. Now, though observational trials are very useful for some problems, they have severe limitations. When we want to find the effect of weedkillers on crops, for instance, we may be anxious to detect a 10 to 20 per cent. depression in yield, and this could pass quite unnoticed in an observational trial. The trials that the Department conducts on tolerance of crops to weedkillers are therefore trials from which measurements (usually weights) are taken.

If a trial is worth weighing, then it is worth replicating and designing for statistical analysis.

A replicated trial is one in which several plots of each treatment occur instead of just one; a trial designed for statistical analysis is normally a replicated trial in which the treatment for each plot in a replication is chosen at random. A simple case is a trial with 4 treatments replicated 4 times. There are then 16 plots with treatments allocated in a manner shown below:—

4 2 1 3 1 3 2 4 2 1 4 3 1 4 2 3

Let us say the crop is potatoes, and that treatments 1,2,3,4 are control, low, medium, and high rates of a weedkiller respectively.

It is clear that the average of the 4 plots of a treatment gives us a better estimate of yield than would one plot. That is the point of replication. The reason for randomisation, however, is not primarily to get a better estimate (although it may do this), but to find out just how accurate our estimate is. With a randomised trial we can tell whether the differences between treatments can easily be accounted for by such things as soil variability or whether they cannot. In the latter case we say that the differences are *significant*.

The two examples dealt with so far are ones in which the problem to be dealt with was already clearly stated. In practice, the first step to take when planning an experiment is to define the problem; and defining the problem leads to three main considerations:—

1. The treatments to be used.
2. The accuracy to be aimed at.
3. The limitations of the experiment.

1. TREATMENTS

Let us take now a more complex problem and suppose that we want to kill couch grass in a field to be planted in potatoes. Then we need to find:—

- (a) The effect on couch grass, and
- (b) The effect on potatoes of the treatments used. We might deal with this by having one trial (or set of trials) on couch and one on potatoes or we might combine the two. We may want to know in this case:—

- (a) The best preparation to use
- (b) The best rate of application
- (c) The best time of application.

We may also want to find the effect of a cultivation treatment and perhaps by different varieties of potato. (In the previous examples we assumed that we knew the answer to all except one of these factors.)

Where we have only a slight idea of the answers to these questions it is not sound practice to put down one trial to test different preparations, another trial to test rates, and another to test times; because to conduct a trial on types of weedkiller we must choose a rate and a time of application, and if the rate is far too high, say, we shall get no worthwhile result. Similarly, if we want to test only rates of weedkiller, then we must choose a particular preparation and a time of application.

It is much better to put down what is called a factorial trial in which all combinations of, say, 3 weedkillers, 3 rates, and 2 times of application are compared. This would give $3 \times 3 \times 2 = 18$ treatment combinations. It is always worth while including all combinations of treatments, because only then can interactions be examined. (An interaction occurs if, for instance, 2 weedkillers are equally effective at one rate, but one is much more effective at another rate.) We might also include 2 varieties or 2 cultivation treatments in the same trial.

2. ACCURACY

We must also decide on the accuracy we require. In this case we shall almost certainly want to take weights of potatoes and perhaps counts of couch. We shall therefore use a design that can be analysed statistically. A replicated trial with 18 treatments may sound fearsome at first, but it is one of the features of the factorial design that much less replication is necessary than with other trials, and two or three replicates would give an excellent trial. (A good example of a factorial trial is that on wheat tolerance described by Mr. Ward in another paper.)

3. LIMITATIONS

The limitations of the trial will easily become apparent: One trial is conducted in only one district, on one soil type, at one time of the year, perhaps on one variety of potato. If we want to generalise we must put down experiments in other districts, on other soil types, and so on.

OTHER PROBLEMS

The problem to be dealt with can vary in many ways from this one on potatoes. I shall mention two examples briefly:—

1. In a clover tolerance trial we may want to find the effect on clover more accurately than an observational trial will tell us, but exactly what we should measure is not straightforward. In potato trials yield is the obvious thing to measure, but with clover it could be the number killed, the number surviving, the percentage surviving, or the weight produced.
2. If we are dealing with a new weed problem we may want to try a dozen different preparations and have practically no idea of the best rate of application. Here a preliminary observational trial followed by more precise trials may be the best method of attack. The preliminary trial which would contain all the weedkillers at two or three widely spaced rates could sort out the possibly suitable preparations from the hopeless and give a general idea of the best rate.

In conclusion I would emphasise that without careful and accurate field work—the even application of the correct amounts of weedkiller,

the intelligent selection of a site, careful observation, and accurate weighing or counting—without these things the best planned experiments will tell us little.

DISCUSSION

Q.—It has been suggested that with the use of two to three replications, a little experience, and a little intelligence more information is obtained from such an experiment than from one with a lot of replications which becomes confounded in detail.

A.—No necessity to be confounded in detail by a properly replicated experiment. The number of replications partly depends on the number of treatments being used. Usually it is not likely that an experiment which is well replicated is laid down with less than 20 plots nor with more than 64 plots. Intelligence is useful in experimental work, but by starting to rely on personal judgment the results become biased. This is one variable that we want to be free from in experimental work.

Comment: If a person is intelligent, he will use a well-replicated trial. If he is not intelligent, he will try to get away without a well-replicated trial. Therefore, the answer is obvious; if he is not intelligent enough to use a replicated trial, the results would probably be of little value.

Q.—The experiments on dock control were not replicated. Is the information from a large number of simple trials more or less efficient than the information from a smaller number of more detailed replicated trials?

A.—Depends on whether one needs only observations on the effects of the treatments or requires to weigh and measure results. A series of observational trials is better if big differences between treatments are likely.

Q.—If finer differences need to be determined, is it then necessary to put down replicated trials?

A.—A series of observational trials suffers from the limitations of accuracy it can measure. On the other hand, a replicated trial usually suffers from the limitation of only being carried out in one place. Therefore more than one replicated trial is needed to measure finer differences.

Q.—It is suggested that plot size should be larger where much variation occurs in soil fertility and soil moisture. How is the size of plot determined in relation to soil variation?

A.—If variation in soil fertility and moisture occurs from one yard to the next, the plot size should be made several yards square, to avoid one plot being situated on only a highly fertile area and another plot on only an area of low fertility. Size of plot is governed by other factors such as method of sowing and method of spraying used.

Comment: The use of large plot size is very desirable, especially for weed control trials which are usually observational trials. Small plots used in simple observational trials could be very misleading.