DESIGN: A CRITICAL NEED IN PEST-DAMAGE CONTROL EXPERIMENTS

CHARLES R. INGRAM, Ohio Field Station, Patuxent Wildlife Research Center, Sandusky, Ohio 44870

ABSTRACT: The manner in which an experiment is conducted determines the inferences that can be made from the results of the analysis of the experiment. This paper emphasizes the critical need in pest-damage control (PDC) experiments for a detailed planning process (i.e., the design of experiments) by exampling improper designs that prohibit a researcher from making valid inferences about his hypotheses of interest. Emphasis is placed on identification of experimental units, determination of restrictions on the randomization procedure, and specification of treatment forms of pest control materials. A list of some specific actions to strengthen PDC experiments is given.

INTRODUCTION

Design of experiments is the planning process that allows researchers to efficiently and objectively direct their efforts toward gathering information pertinent to the hypothesis under test. Unfortunately, many authors of texts on statistical design and analysis concentrate on analysis at the expense of design. These authors unwittingly encourage analysis by rote. Vague terminology (e.g., "cell" and "cross classification") often obscures the manner in which the experiment was conducted. Users of statistics obtain the erroneous impression that, to analyze an experiment correctly, they must first arrange their data in a standard tabular format. A similar tabular format is then observed in a statistical analysis text and the analysis performed on the researcher's data is the analysis associated with tabular format. Consequently, the inferences made from the results of many pest-damage control (hereafter referred to as PDC) experiments are incorrect because the researcher, as well as the statistician, have failed to recognize the hypothesis under test.

Recognition is the key to design of any experiment. Unless the researcher recognizes the hypothesis under consideration, he cannot select treatments that address this hypothesis. Unless he recognizes the sources of variability present in the experiment, he cannot select an experimental design that will increase the efficiency and sensitivity of his treatment comparisons. However, recognition also is the key to the analysis of an experiment and to an understanding of the inferences that can be drawn from the results of an experiment.

In this paper, I attempt to show how improper design can prohibit a researcher from making valid inferences about the hypotheses of interest. The approach is somewhat backward because the ideal procedure is to adequately design an experiment and then to verify the validity of the experiment through the use of a mathematical model and an outline of the analysis. Unfortunately, there are many cases in PDC research where a less than perfect design is forced on the researcher or where a good design is inadvertently modified in the field. In these circumstances, it is the responsibility of the statistician and the researcher to identify the deficiencies associated with the experiment and to determine how these imperfections could cloud inferences. The researcher has one additional responsibility; he must assess the biological importance of the imperfections.

Three topics in the design of PDC experiments will be considered in this paper. The first, recognition of the randomization procedure, is presented through incomplete examples of the type that appear in many statistical texts. The purpose of this presentation is to inform the reader that, while the mechanical computations involved in obtaining sums of squares and mean squares for an analysis of variance table are identical, the inferences that can be made concerning the effects of interest are highly dependent on the randomization procedure. Second, we consider specification of treatment forms of a pest control material (hereafter called PCM) investigated in PDC experiments. The purpose of this discussion is to define the parameters of a treatment form and to illustrate the need for well-defined commercially-realistic treatment forms. Finally, we deviate from the general approach and list specific actions to strengthen PDC experiments.

DISCUSSION

Randomization Procedure

The randomization procedure utilized by a researcher uniquely defines the experimental plan. Randomization may be either restricted or unrestricted. Restrictions designed into an experiment are attempts to control extraneous sources of variation, thereby increasing the sensitivity of the experiment. These restrictions affect the analysis of an experiment and limit the scope of inferences that can be made from the analysis.

There are three basic experimental plans: The completely randomized, the randomized block, and the Latin square. The differences between these plans are the number of restrictions (0, 1, and 2, respectively) placed on the randomization procedure.

To recognize the randomization procedure used for a study, the researcher must be able to identify the experimental unit and to determine the restrictions on randomization involved in the study.

Identification of Experimental Units

The experimental unit is the smallest unit to which a treatment is assigned within the restrictions imposed by the randomization procedure. An experimental unit may contain several observational units; e.g., a field of planted corn that is allocated a certain treatment form of a PCM may be the experimental unit, but a row plot of 20 consecutive ears of corn in that field is the observational unit.

There are numerous examples of studies where biologists have confused experimental units with observational units and have made serious inference errors. Example 1 and the ensuing discussion will illustrate this type of study.

The presentation for this section will include the use of a mathematical model to specify the manner in which an experiment is conducted. The reader unfamiliar with this statistical tool should not dwell on the models and the discussion of same, but should proceed with the text.

Example 1. An experimenter has completed a study to evaluate the effectiveness of a repellent treatment for protecting sweet cherries from bird damage. He has tabulated the data for analysis. This table lists two treatments that differ only in the amount of repellent applied (i.e., a positive-level repellent treatment ["treated"] and a zero-level repellent treatment ["untreated-control"]). There are s responses under each treatment. Each response corresponds to a damage assessment made on an individual tree.

This appears to be a textbook example of a two-treatment, completely randomized experiment which is detailed by the mathematical model:

$$Y_{ik} = \mu + R_i + \varepsilon_{ik};\ i = 1, \ldots, t = 2;\ k = 1, \ldots, s; \qquad (1.1)$$

where

Y_{ik} is the response (damage) measured on the k^{th} tree to receive repellent treatment level i;

μ is the overall mean;

R_i is the fixed effect of repellent treatment level i; and

ε_{ik} is the experimental unit error which is normally and independently distributed about a mean of 0 and variance σ^2_ε, NID $(0, \sigma^2_\varepsilon)$.

The corresponding analysis of variance (AOV) for this model is given in Table 1.

Table 1. AOV summary and expected mean squares associated with Model 1.1

Source	Degrees of freedom	Expected mean squares
Treatment:		
Repellent level	$(t-1) = 1$	$\sigma_\varepsilon^2 + \phi_R$ (where $\phi_R = s\Sigma_i(R_i)^2$)
Error:		
Experimental unit error (trees within repellent level)	$t(s-1) = 2(s-1)$	σ_ε^2

The information given in Example 1 is incomplete. The study area consisted of a single row of mature cherry trees. Because the branches of the trees intertwined, trees could not be treated individually. Therefore, two discrete groups of s contiguous trees each were established in the row separated by a buffer of nontreated trees. Selection of the group to receive the positive-level treatment was random.

Model 1.1 assumes that individual cherry trees are the experimental unit and that there is only one observational unit per experimental unit. For the study conducted, trees were in fact observational units belonging to an experimental unit of s contiguous trees (remember that repellent levels were randomly assigned to groups of contiguous trees and not to individual trees). Thus, there are two sources of variability in the experiment; experimental (unit) error and observational (unit) error. If a term representing the variability between observations from the same experimental unit is added to Model 1.1, the model becomes:

$$Y_{ijk} = \mu + R_i + \varepsilon_{ij} + \delta_{ijk};\ i = 1, \ldots, t = 2;\ j = 1, \ldots, r = 1;\ k = 1, \ldots, s; \quad (1.2)$$

where

δ_{ijk} is the observational error associated with observation k within experimental unit ij, NID $(0, \sigma_\delta^2)$ and independent of the $_{ij}$'s.

The AOV summary for Model 1.2 is given in Table 2.

Table 2. AOV summary and expected mean squares associated with Model 1.2

Source	Degrees of freedom	Expected mean squares
Treatment:		
Repellent level	$(t-1) = 1$	$\sigma_\delta^2 + s\sigma_\varepsilon^2 + \phi_R$
Error:		
Experimental error (experimental units within repellent level)	$t(r-1) = 2(1-1) = 0$	$\sigma_\delta^2 + s\sigma_\varepsilon^2$
Observational error (trees within experimental units)	$tr(s-1) = 2(1)(s-1) = 2(s-1)$	σ_δ^2

The expected mean squares in Table 2 reveal that the appropriate test for a repellent-level effect is (repellent level mean square)/(experimental error mean square). This test cannot be performed because there is no estimate for experimental error (i.e., the degrees of freedom are zero).

Table 2 further reveals that the only significance test that can be performed on the data obtained from this study is (repellent level mean square)/(observational error mean square) with 1 and 2(s-1) degrees of freedom for the numerator and denominator, respectively. Comparison of the expected mean squares for this test indicates that the ratio would measure the effects of both the repellent level and experimental error (i.e., $s\sigma^2_\varepsilon + \phi_R$). This is precisely the test that would be performed if the researcher had analyzed the data by the incorrect Model 1.1. Since it is unreasonable to assume that experimental error (σ^2_ε) is zero, one would expect the Model 1.1 test to be highly significant. Thus, the researcher who confuses observational units and experimental units is, in his naivete, guaranteeing himself a significant difference due to an incorrect analysis of a poorly designed experiment.

Example 1 shows that specification of the mathematical model and outline of the AOV are valuable aids to the researcher because these aids identify the manner in which the experiment was conducted and also identify the inferences that can be made from the results. In addition, the researcher's deductive reasoning often can help him identify problems with his experimental design: The example study contains only two experimental units; if the units remained untreated, a full measurement of the damage to entire units almost surely would have revealed the bird damage to be at least one cherry different. The random assignment of a positive-level treatment form to one of the experimental units could affect the degree to which the two units differ, and any difference observed could be due to either experimental unit difference or a combination of experimental unit and treatment level differences. Thus, the study confounds treatments with experimental units.

For this study and for many PDC experiments, observational units are an additional source of variation that was forced into the study by the logistics of the experimental situation. The use of observational units does not salvage a study that was not designed with sufficient experimental units (replication of repellent levels) to give an estimate of experimental error.

Determination of the Restrictions on the Randomization Procedure

Identification of the restrictions placed on the randomization is an aid to the recognition of the design of an experiment because these restrictions can affect the treatments under investigation in the experiment. A treatment is a particular combination of factors (i.e., variables of interest) under the null hypothesis that is imposed on experimental units by the researcher. In PDC research we are primarily concerned with treatment forms of a PCM; however, the researcher may be interested in additional factors (e.g., variety of corn). The determination of whether or not inferences can be made about a variety effect depends on the procedures used to assign variety to unplanted fields. This situation is illustrated by Example 2.

Example 2. An experimenter is interested in evaluating the effectiveness of t repellents in protecting v varieties of corn. He has available to him tr fields of each v varieties of corn. The r replicates of each repellent are randomly assigned to fields of each variety. There are s observational units from each field.

Two variables are defined in Example 2, namely, repellent and variety. The random assignment of repellents to fields identifies it as a treatment factor and permits valid inferences to be made concerning repellent effects. The information on the variety variable is incomplete. Yet, many texts on statistical analysis would describe this experiment by the statement "a two-way, cross-classification with r observations per cell." This statement describes the computations involved in the analysis of the data, but fails to specify the manner in which the experiment was conducted, or the inferences that can be made as to variety and treatment effects.

The method used to assign varieties to unplanted fields determines the experimental unit and whether or not the researcher can make inferences about a variety effect. Two methods will be considered in Experiments A and B.

Experiment A: Random assignment of varieties to experimental units (unplanted fields) implies that the researcher is interested in the variety effect. He has planned his experiment to investigate not only variety and repellent effects but also the possibility of a variety-repellent interaction.

Experiment A involves a completely randomized experimental plan with r replications of each of the vt factor combinations (treatments). The mathematical model for this experiment is:

$$Y_{ijkl} = \mu + R_i + V_j + (RV)_{ij} + \varepsilon_{ijk} + \delta_{ijkl}; \; i = 1, \ldots, t; \; j = 1, \ldots, v; \; k = 1, \ldots, r; \; l = 1, \ldots, s; \qquad (2.1)$$

where

Y_{ijkl} is the response measured on the observational unit l in field k which receives treatment ij;

μ is the overall mean;

R_i is the fixed effect of repellent level i;

V_j is the fixed effect of variety level j;

$(RV)_{ij}$ is the interaction effect associated with repellent level i and variety level j;

ε_{ijk} is the experimental unit error, NID $(0, \sigma^2_\varepsilon)$;

δ_{ijkl} is the observational unit error, NID $(0, \sigma^2_\delta)$ and independent of the ε_{ijk}'s.

The AOV summary for Model 2.1 is given in Table 3.

Table 3. AOV summary and expected mean squares associated with Model 2.1.

Source	Degrees of freedom	Expected mean squares
Treatments:		
Repellent (R)	(t-1)	$(\sigma^2_\delta + s\sigma^2_\varepsilon + \phi_R)$
Variety (V)	(v-1)	$(\sigma^2_\delta + s\sigma^2_\varepsilon + \phi_V)$
R * V (Interaction)	(t-1) (v-1)	$(\sigma^2_\delta + s\sigma^2_\varepsilon + \phi_{RV})$
	tv-1	$\sigma^2_\delta + s\sigma^2_\varepsilon + \phi_{TREAT}$
Error:		
Experimental error	tv(r-1)	$\sigma^2_\delta + s\sigma^2_\varepsilon$
Observational error	tvr(s-1)	σ^2_δ

Experiment B: The researcher must conduct his test in an area where farmers have already planted their fields. The experimenter expected that susceptibility to damage would be different across variety levels; thus, he randomly assigned r replicates of each repellent level to each of v varieties.

Obviously the randomization in this experiment differs from Experiment A. The researcher has no control over the assignment of variety to a planted field. The farmer's selection of variety may have been influenced by many factors that affect the bird damage in a field (e.g., growing conditions, anticipated bird numbers). Therefore, the variety effect cannot be measured directly since it is confounded with a farmer's selection effect (which I will refer to as the location effect).

The researcher in Experiment B has restricted the randomization of repellent levels. Repellents have not been randomly assigned to experimental units (planted fields); instead, the variety of a field has been determined and r replicates of each of the t repellents

have been randomly assigned to fields of a given variety. The restriction on randomization represents an attempt by the researcher to control extraneous sources usually referred to as blocking. (The reader should note that if repellents had been randomly assigned to planted fields [without regard to variety], the study would be a completely randomized experiment similar to the one given in Model 1.2.)

The experimental design used is a generalized randomized block with r replicates of the t repellent treatments in each of v blocks. The mathematical model for this experiment is:

$$Y_{ijkl} = \mu + R_i + B_j + (RB)_{ij} + \varepsilon_{ijk} + \delta_{ijkl}; \qquad (2.2)$$

where

B_j is the fixed effect associated with block j. In actuality this effect is the sum of two effects--the variety effect and what I have called the location effect;

$(RB)_{ij}$ is the block by repellent interaction.

The AOV summary for this experiment is given in Table 4.

Table 4. AOV summary and expected mean squares associated with Model 2.2.

Source	Degrees of freedom	Expected mean squares
Treatments:		
Repellents (R)	t-1	$\sigma^2_\delta + s\sigma^2_\varepsilon + \phi_R$
Control:		
Blocks (B) (variety and location effects)	v-1	$\sigma^2_\delta + s\sigma^2_\varepsilon + \phi_B$
R * B	(v-1) (t-1)	$\sigma^2_\delta + s\sigma^2_\varepsilon + \phi_{RB}$
Error:		
Experimental error	tv(r-1)	$\sigma^2_\delta + s\sigma^2_\varepsilon$
Observational error	tvr(s-1)	σ^2_δ

The analyses given in Tables 3 and 4 are mechanically identical; i.e., AOV tables, including tests of significance, are computed in the same manner. Yet, as the mathematical models indicate, the results of the two experiments are different. Experiment B does not permit valid inferences about the variety effect since any significance of the block mean square may be due to either variety or location, or a combination of the two effects. In other situations, the confounding that is present in the block effect may not be as easily identified. Nevertheless, the restrictions placed on the randomizations procedure reflect the researcher's attempt to control the magnitude of the experimental error in his experiment. The institution of control restricts the inferences that can be made about the controlling variable (in this case, variety).

Example 1 and Example 2 emphasize the importance of understanding the randomization procedure used in an experiment. This understanding allows the observer to recognize the experimental plan and the treatments under investigation. The process of recognition involves the identification of: (1) experimental units, (2) observational units, (3) restrictions on randomization, and (4) the treatments under consideration.

We now consider a final example of a poorly designed experiment that will incorporate all that has been discussed in this section.

Example 3. A researcher wishes to determine the effects of a fixed amount of simulated bird damage to corn. He has selected a 4-row plot of corn with 80 ears in each row. When the corn is in early milk stage (maturity level 1), the researcher inflicts the fixed amount of damage to every other ear in an exterior row of the plot. At the appropriate stage of maturity (levels 2-4), the researcher proceeds to damage every other ear in the plot-row that is adjacent to the row containing the ears that were damaged during the previous maturity level. Measurements are obtained from individual ears.

Clearly, individual ears are the observational units for this study. However, the experimental unit is not ears but the every-other-ear systematic pattern of 40 ears in a row. Maturity levels were not randomly assigned to the eight experimental units, but were sequentially assigned to rows thereby confounding the two effects (row and maturity). The single treatment factor is damage at two levels (damaged and undamaged). Thus, under the assumption that damage levels were randomly assigned under the restriction that both levels must appear in every row-maturity level block, the experimental plan is a randomized block. The experiment can be specified by a mathematical model similar to Model 2.2 where $t = 2$ treatments, $v = 4$ blocks, $r = 1$ experimental unit per block-treatment combination, and $s = 40$ observational units per experimental unit. Substitution of these values into the degrees of freedom column of Table 4 indicates that the effect of the simulated damage cannot be tested under the present model because there is no estimate of experimental error. Thus, the researcher, in a poorly designed attempt to gain information on a maturity effect, destroyed the replication in his experiment and lost the ability to make inferences about either a damage or a maturity factor, or perhaps more importantly, the interaction between these factors.

Specification of Treatment Forms

The most significant discrepancy in PDC research occurs in the failure of many researchers to specify sufficiently the treatment forms being investigated in an experiment. This discrepancy has apparently arisen because economic and cultural constraints severely limit the size (i.e., number of experimental units) of any experiment we may wish to conduct. Thus, PDC experiments routinely compare only one positive-level treatment form against an untreated control (zero-level), and this positive-level treatment form is often identified solely by the PCM involved. The failure to distinguish treatment forms permits results from experiments involving different positive-level treatment forms of the PCM to be combined and erroneously extrapolated to infer efficacy for a PCM.

Extrapolation is a serious inference error. Efficacy can only be established for well-defined treatment forms of PCM's. However, we in PDC research have allowed the extrapolations to come full circle to influence selection of treatment forms investigated in an experiment. Thus, many studies have been conducted to evaluate ill-defined treatment forms. As will be discussed in this section, the results of these studies must be viewed as suspect and cannot be considered to directly support the efficacy of a particular treatment form.

This section will address: (1) the specification of treatment forms of PCM's, and (2) the pitfall of subjective specification.

Specification of Treatment Forms of a PCM

The objective of PDC experiments is to develop commercially and ecologically realistic treatment forms of a PCM. These treatment forms must restrict the damage activity of pest species without causing undue hazard to other wildlife. A treatment form is specified by the following parameters, which detail the product used to carry the material to the pest and the method of delivery.

A. Product specification parameters

(1) Formulation
(2) Carrier
(3) Dilution

B. Delivery specification parameters

(1) Type (ULV, LV, etc.)
(2) Equipment
(3) Method
(4) Sticker material
(5) Rate per application
(6) Maximum number of applications
(7) Distribution of product
(8) Timing of application criteria
    (a) Initial
    (b) Subsequent

Distinct treatment forms differ in the levels of at least one of the above parameters. Without knowledge as to how the parameters relate to affect performance, one is forced to assume that each parameter is important and that information obtained about one treatment form cannot be extrapolated to imply efficacy of other treatment forms. It is, of course, desirable that this knowledge be obtained through well-planned research; however, there will always be cases where biological insight or common sense should be invoked to maintain practicality as long as objectivity is not sacrificed. For example, it would be absurd to require that an aerial application treatment form be evaluated for every type of aircraft that could be used to apply the product. Conversely, broadcasting by hand could be greatly different from broadcasting by aircraft. Thus, to make inferences, the researcher must identify the parameters of treatment forms involved in his experiment and objectively assess how these parameters affect performance.

The Pitfall of Subjective Treatment Forms

PDC studies often have involved subjectively defined treatment forms. Field trials have been conducted in which a group of fields (experimental units) are selected to receive applications of a PCM product formulation on a "when needed" basis. Each field is observed (schedule not defined) and when bird activity reaches a subjectively determined, but normally undefined, level, the field receives an immediate application of the product.

Researchers should require that treatment forms be specified by quantitative, objective criteria. For example, the criteria for applications of a particular treatment form could be objectively specified as follows:

> "Initial application will be made 20 days prior to projected harvest date. A subsequent application will be made 5 days prior to projected harvest date. A third application will be made, in the period 18-7 days before projected harvest date, if 0.5 in. of rain occurs in a 24-hour period or if 100 blackbirds or more are seen (in that treated experimental unit) during a scheduled observation period. Note: If the projected harvest date is revised during the course of the experiment, applications will be governed by the revised date."

Conversely, subjective "apply as needed" criteria might be given as follows:

> ". . . will be closely observed for bird activity. Initial treatment will be made as soon as damage is noted. Subsequent treatments will be made when it is apparent that previous treatments are becoming less effective or ineffective, or if considerable rain occurs. No treatment will be made within 5 days of harvest." (Italics [underscore] indicate subjectiveness.)

These latter criteria do not specify a single well-defined treatment form, but instead, permit the use of many treatment forms that cannot be either related or distinguished objectively. Two research principles are violated: First, the scientific method is violated because it is impossible for an independent research team to reproduce the treatments for another experiment. Second, the subjectiveness of the application (incorrectly referred to as "treatments") criteria gives maximum advantage to the PCM without yielding information as to the efficacy (including valid estimates of hazards to nonpest species) of a commercially-realistic treatment form. The reader should note that, under the subjective criteria given above, a treated experimental unit may, in actuality, never receive a single application of the PCM.

Studies using subjective criteria are not experiments, but, instead, are demonstrations. For an experiment, the researcher establishes objectives, formulates a null hypothesis, and selects treatments pertinent to this hypothesis. These well-defined treatments are then randomly assigned to experimental units within the restrictions on randomization (e.g., blocking) imposed by the experimental plan. The experiment must include sufficient replications to provide a sensitive test of treatment differences.

There are occasions (e.g., a paucity of experimental units or experimental material) when demonstrations are the only appropriate way to approach problems experimentally. If a demonstration must be used, the treatment forms investigated must be well defined and objective. Furthermore, the results of these demonstrations must be qualified to point out the deficiencies in design that make a study a demonstration and not a true, replicated experiment.

Actions to Strengthen PDC Experiments

Specific actions that can be taken to upgrade the quality of PDC experiments and ensure the appropriateness of inferences made from analysis of data from them are listed below.

1. Use proper and adequate experimental design.
2. Size experimental units to be commercially realistic.
3. Acquire advance information on pest populations (including damage activities) in proposed experimental areas.
4. Identify observational units as samples from within experimental units.
5. Acquire data on cultural practices that would affect the commercial use of the PCM.
6. Evaluate well-defined and commercially-realistic PCM treatment forms.
7. Use more replications of each PCM treatment form.
8. Use screening experiments to determine which factors affect the performance of a treatment form.
9. Ensure the statistical analysis of an experiment is appropriate to the manner in which the experiment was conducted.
10. Avoid extrapolations:

    a. As to the universal efficacy of a chemical compound (i.e., realization that the efficacy of one treatment form of a PCM cannot generally be used to infer that another treatment form of the PCM produces the same results).

    b. As to the efficacy of a treatment form over the whole damage period when the experiment was conducted to compare treatments in a shorter period.

11. Use well-defined procedures to specify the manner in which bird observations are to be conducted.
12. Evaluate hazards to nontarget populations.
13. Publish well-defined procedures so independent researchers could duplicate the experiment.

CONCLUSION

The manner in which an experiment is conducted determines the inferences that can be made from the results of the analysis of the experiment. Design is the planning process that permits the researcher to verify, in advance, that his work will permit valid inferences concerning the hypothesis under test. No amount of statistical sophistication or maneuvering in the analysis of an experiment can extract information from data that were not wrought into them by an adequately designed and competently executed experiment.