Austral Ecology (2000) **25**, 292–296

Comment

Pseudoreplication in experimental designs for the manipulation of seed germination treatments

DAVID A. MORRISON[1]* AND E. CHARLES MORRIS[2]
[1]*Department of Environmental Sciences, University of Technology Sydney, Westbourne Street, Gore Hill, New South Wales 2065, Australia (Email: david.morrison@uts.edu.au) and* [2]*Centre for Biostructural and Biomolecular Research, University of Western Sydney, Richmond, New South Wales, Australia*

Abstract In the published reports concerning the design of laboratory experiments examining signals that might trigger seed germination, there is currently a misunderstanding of what might constitute correct replication of the experimental treatment. This is particularly true for the study of dry heat, smoke and charcoal, where either an individual seed or a batch of seeds in a Petri dish or tray is being treated as the unit of replication of the experimental treatment, irrespective of whether or not those seeds were all subjected to the experimental manipulation simultaneously. Under these circumstances, the application of the treatment is unreplicated, while samples nested within that single application have been replicated, and the Petri dishes/trays are functioning solely as independent replicates of the variability in germination response within the seed batch used and variation within the pretreatment and post-treatment environments. Thus any observed difference in germination may be due to the germination treatment but, potentially, it could also be due to any chance event affecting the treated sample. There are a number of alternative experimental designs that avoid this problem. The essential point with these designs is that the application of the experimental manipulation to each replicate should be separated in space by the use of separate experimental equipment and/or in time by the repeated use of the same experimental apparatus.

Key words: charcoal, dry heat, germination signals, non-independence, pseudoreplication, scarification, smoke.

INTRODUCTION

In the published reports on signals that might trigger seed germination there is currently a misunderstanding of what might constitute correct replication of the experimental treatment for several types of manipulation. In this comment, we discuss the current experimental designs commonly used for examining the effect of germination signals such as scarification, dry heat, smoke and charcoal (there are many other seed treatments that could be discussed but we have not covered them here), the data analyses used in such experiments and the inferences subsequently drawn. We point out where we think the problems lie in the commonly used designs and suggest alternative designs that avoid these problems.

Our aim is thus to draw attention to ways of applying some of the experimental treatments that have been widely reported and yet are deficient in terms of correct experimental design. If this is not done then the standard methodology will become entrenched. Indeed, we have used the standard methodology ourselves (e.g. Morrison *et al.* 1998; Morris 2000), simply as a result of following precedent rather than questioning the logical basis of the designs. As Gibson *et al.* (1999) have noted: 'There are very few useless experiments (Cousens 1996). However, the inferences that can be validly drawn from a particular experiment depend on the design used, the measurements taken and the analysis of the data. If these logical limitations to inference are not fully appreciated…then biased assessments will result.' Thus we are not denying the potential usefulness of any particular experiment that we have reviewed. For this reason we have deliberately not nominated any of the reports that provide our descriptions of the currently used, but incorrect, experimental designs, because, following Anderson (1990), we have a 'reluctance to pillory the few for errors which many commit with impunity'.

*Corresponding author.
Accepted for publication May 1999.

THE PROBLEM

The main problem that we have encountered in the experimental designs is a lack of replication of the

experimental 'treatment'. A true replicate is the smallest unit to which an experimental manipulation is independently applied (Hurlbert 1984). That is, repeated application of the treatment is required. So, replicates derived from repeated sampling of a single application of an experimental treatment, for example, are not independent of each other and thus cannot act as true replicates. The non-independence means that chance events influencing a particular experimental unit will affect *all* of the samples from that unit, the effect of which will then be indistinguishable from the 'treatment effect'. This is true irrespective of whether that treatment is actually manipulated by the experimenter or whether it is a naturally occurring phenomenon.

Note that what we are discussing here is a problem that has long been recognised in field experiments under the name 'pseudoreplication' (e.g. Hurlbert 1984; Heffner *et al.* 1996) but it has often been tacitly ignored in laboratory experiments. It seems frequently to be assumed that independent replication of experimental treatments is not necessary in laboratory-based biological studies as the experimental conditions have been carefully controlled. This assumed uniformity applies both to the specific experimental treatments or manipulations used and to the general background or culturing conditions encountered. However, this assumption is incorrect for the experimental manipulations because, even if chance events may be less likely in a laboratory situation, their effects cannot be ignored. Just because the variation found in repeat applications of the experimental manipulations used in laboratory experiments is usually trivially small does not mean that this potential source of experimental variation can be ignored. Thus, both replication and control of background experimental conditions are essential components of all good experimental designs.

CURRENT DESIGNS

We start our discussion with a simple example where replication seems never to be a problem. As reported, scarification of seeds has usually been carried out by nicking each seed with a scalpel or dissecting needle or by rubbing each seed with sandpaper (the alternative is machine scarification). In both of these cases the experimental manipulation is actually applied to each seed individually, and so each seed can in fact act as a true replicate of the experimental treatment.

The application of dry heat to seeds, on the other hand, usually involves the use of an oven. Under these circumstances all of the seeds for any one experimental manipulation are often added to the oven simultaneously. This means that the experimental treatment has been applied once only, and that each seed can thus act exclusively as a replicate of that one manipulation. That is, the smallest unit to which each experimental manipulation was independently applied was the group of seeds that went into the oven together, and this batch of seeds thus constitutes a single true replicate. However, in many cases in the reports each of the seeds has been treated as a replicate of the experimental treatment, or different subsamples of the batch of seeds (e.g. those subsequently placed together for germination in a Petri dish) have been treated as replicates.

This means that it is impossible to distinguish between the effects on germination of the different experimental heat treatments and the effects of variation due to the different uses of the oven. For example, differences could arise as a result of temperature variation within a single oven at any one time, variation within an oven through time or variation among different ovens (e.g. a particular oven used may not be maintaining the desired temperature or may contain a volatile contaminant that affects germination). Alternatively, there could be variation in the handling of the seeds during the experimental manipulations. In reality, the number of sources of chance variation that could be identified is limited solely by the imagination of the researcher. While the possibility of within-oven variation might not be considered particularly likely (although this potential problem is not unique to ecology; see, e.g. Cayouette *et al.* 1998), the confounding of these sources of variation with the variation due to the experimental manipulations is obviously inappropriate if the objective of the experiment is to quantify the effects of the experimental treatments.

Most of these dry-heat experiments are studies of 'heat shock' as a result of sudden application of an increased temperature. However, it should be noted that the same experimental design problems can occur for the use of growth cabinets or chambers to study the effects of lower temperatures. For example, if all of the replicates of a particular experimental treatment are included in the same growth cabinet or chamber, then there is no true replication of the treatment as the effects of differences between the cabinets/chambers will be indistinguishable from the effects of the experimental treatments. The seeds are not independent of each other if they are all in the same 'container', and any chance events affecting a particular container will affect all of the seeds in that container.

A similar situation exists for the application of smoke treatments in the laboratory, although several different protocols for applying smoke have been reported. In many cases, smoke has been applied to the seeds in an enclosed chamber with smoke-laden air circulated through from an external combustion source. In the papers using this protocol, all of the seeds for any one experimental manipulation were often added to the chamber simultaneously, and thus the experimental treatment was applied once only. As an alternative

protocol, the smoke from the combustion source was bubbled through a container of water, thus producing 'smoked water'. This water was then added to the seeds as the experimental manipulation. However, if the smoked water is prepared only once for any one experiment, there is once again no true replication of the experimental treatment, and each seed can thus act exclusively as a replicate of that one manipulation only. Finally, several authors prepared 'smoked filter paper' by adding the paper to the smoking chamber (instead of adding the seeds directly), and these filter papers were then used as the substrate for subsequent seed germination. As before, if the smoked paper is prepared only once for any one experiment, then the seeds cannot act as replicates of the experimental treatment. So, once again, the variation due to the experimental treatments is confounded with an unknown number of other sources of variation. There is already evidence, for example, for differences in germination effect caused by variation in the plant species composition of the smoke source and also in the temperature of the smoke source (e.g. Baldwin *et al.* 1994; Baxter *et al.* 1995).

In the field, the experimental set-up for direct smoke application is usually similar to that for the laboratory smoking chamber, with a tent used as the container within which the smoke-laden air is circulated. In this case, the objective is to smoke the ground and to monitor subsequent germination of *in-situ* or added seeds. Under these circumstances, each plot of ground can potentially act as a true replicate of the experimental treatment if the tent is moved from place to place for each manipulation. However, this will only be strictly true if a new external combustion source is used each time. If the tent is merely moved while the same source of flammable material is burning, or if multiple tents are used with a single smoke source, then no true replication will result.

Alternatively, smoked water can be added to the soil in the field, to stimulate germination of *in-situ* or added seeds. Under these circumstances, the volume of water needed for the experiment is much greater than in the laboratory studies, and this dictates that the water must be smoked in batches. If the smoked water is prepared as several batches for any one experiment, then this avoids the confounding of the experimental treatment with chance events, irrespective of whether the batches are used sequentially for different replicate plots or are pooled before use.

A situation similar to that for dry heat and smoke also exists for the application of charcoal treatments, where a couple of different experimental protocols have been reported. In many cases, charcoal has been applied to the seeds as powdered ash produced from partially combusted (i.e. charred but not ashed) wood. In this situation, it is not possible to separate the treatment effects from unique chance events during the preparation of the charcoal, because the treatments have no independent replication. As an alternative protocol, the charcoal is thoroughly mixed in a container of water, thus producing 'charcoal water'. This water is then added to the seeds as the experimental manipulation. As described for smoked water, if the charcoal water is prepared only once for any one experiment then there is no true replication of the experimental treatment, and each seed can thus act exclusively as a replicate of that one manipulation only. There is already evidence, for example, for differences in germination effect caused by variation in the plant species composition of the charcoal source and also in the temperature of the charcoal (e.g. Keeley & Pizzorno 1986).

The basic problem identified for all of these experimental designs is thus that either an individual seed, or a batch of seeds in a Petri dish or tray, is being treated as the unit of replication of the experimental treatment, irrespective of whether or not those seeds were all subjected to the experimental manipulation *simultaneously*. However, if the seeds are treated simultaneously then neither the seeds nor the Petri dishes/trays are independent estimates of the treatment effects but are nested within the *single* application of the treatment. In the terms of Hurlbert (1984), the application of the treatment is unreplicated, while samples within that single application have been replicated. Thus, both the seeds and the Petri dishes/trays are functioning solely as independent replicates of the variability in germination response within the seed batch used and variation within the pretreatment and post-treatment environments – they are not acting as independent replicates of the experimental treatment. Hurlbert (1984) provides diagrams (his figs 1 and 5) that compare a pseudoreplicated experimental design, a properly replicated non-nested design and a properly replicated nested design.

Under these circumstances, if inferential statistics such as *t*-tests and analysis of variance are used to detect any effect of the germination treatment, then the assumption of independence of the replicates is not met. If the null hypothesis is rejected in such cases, the inference that the treatment has caused the difference can be challenged on the grounds that the operation of chance events in the application of the treatment has not been ruled out. The difference in germination may be due to the germination treatment but, potentially, it could also be due to any chance event affecting the treated sample. There is no estimate of the treatment effect *separately* from the error term. Any chance event that affected the single application of the treatment would thus appear as a treatment effect (Hurlbert 1984). If replicates were treated independently of each other, any chance event affecting one application would not affect the remaining replicates, and so would not be confused with the treatment effect.

SUGGESTED IMPROVEMENTS

Clearly, a number of steps could be taken to increase the independence of the data from the experimental manipulation of seed germination treatments. For smoking treatments, for example, the essential step is to move from the current protocol of smoking seed lots 'in parallel' (i.e. all replicates smoked simultaneously in the one chamber) to a protocol that ensures that each replicate is exposed to a separate batch of smoke. Thus for aerosol application, replicates could be smoked separately in time within the one smoking chamber (smoking 'in series' rather than in parallel), with the chamber cleared of smoke between each replicate treatment. Changing the fuel source for each replicate, so that each replicate gets its own smoke, would be a further step towards ensuring that the treatment applied to each replicate is independent. The degree of variability in the fuel source will depend on the questions being asked in the experiment. For instance, if the inferences are about the response of seeds from a given vegetation type to smoke exposure, then an independent sample of fuel batches drawn from that vegetation type should be used. A further step would be to have replicate smoking chambers in order to randomise any effects peculiar to individual chambers. For independent application of the smoke cue via other media (e.g. water, filter papers), separate batches of these media should be prepared, with the same procedure adopted for changing fuels as for aerosol smoke.

For dry-heating of seeds, the same series of steps would apply – heating should be separated in time and space within the one oven, or separated in time and space in replicate ovens. For the examination of charcoal effects, independent charcoal sources would need to be prepared for each replicate. For example, Jones & Schlesinger (1980) placed individual charred stem segments in each Petri dish rather than preparing a stock of ash from which subsamples were taken; this goes a long way towards producing independent replication of the experimental treatment.

Clearly, these suggested steps could create enormous practical problems in that the experiments would become much more unwieldy; however, innovative approaches to the methodology can overcome many of these potential difficulties. Time can be saved by using the concept of blocking the experimental treatments (e.g. Mead 1988). For example, rather than putting six Petri dishes of seeds of the same species in an oven or smoking chamber at the one time and then repeating this for six different species, it would be no more work to place one Petri dish from each of six species in the oven/chamber at each time and then repeating this six times. The former approach creates no true replication, while the latter approach would create six blocks of dishes each of which can function as a replicate for each species. For preparing smoked water, several smaller volumes, each smoked for shorter periods, could be prepared, rather than preparing one big volume smoked for a longer period; and similarly for the preparation of charcoal water.

Another approach to save time is to combine fewer independent applications of the treatments with nested replicates within those applications (e.g. three applications of the treatment with two nested replicates for each application, rather than six applications of the treatment). Correct nested analysis of the data (e.g. Hurlbert 1984) will detect the variability due to independent applications, as well as the variability between replicates within the applications. If variation between independent applications is found to be much less than the variation within applications, then pooling procedures can be used to increase the statistical power of the analysis (e.g. Underwood 1997). For example, Morris (2000) found this to be so in a trial study of independent smoke application, and Baldwin *et al.* (1994) found no significant difference in the stimulatory activity of different commercial brands of 'liquid smoke'. It is likely that the same situation would be found for application of dry-heat in ovens.

It may well be that any observed differences between independent applications of smoke, charcoal or dry-heat are inconsequential. For example, if an oven is functioning correctly then such differences may be trivial. However, demonstrating that differences between independent germination treatments are inconsequential is an entirely different matter from *assuming* that they are so. Nor should previous demonstrations of insignificant differences due to smoking or heating apparatus be relied upon without question, as to do so is naive inductionism because chance events are not being taken into account. The changes to the current experimental protocols that we are suggesting, while perhaps entailing more work, ensure that the germination responses observed are truly treatment effects and not simply artefacts of chance events, and can therefore be more reliably inferred to have some ecological importance.

We should point out that most of the papers we have reviewed involved the study of multiple species. It is thus also possible to use the species as replicates in a statistical analysis, provided that the seeds of each species have been treated separately, thus avoiding the problems of lack of replication. For example, if the study species are all from the same taxonomic group (e.g. a genus/family) and the objective of the experiment is to estimate the parameter for that group (e.g. to find out whether that genus/family generally responds to that germination treatment), then it is appropriate to use those species as true replicates, irrespective of whether there is true replication of the experimental treatment within those species (statistically, the species are nested within the treatments). A similar argument applies to comparisons of non-heat

treatments with several temperature treatments, as each level of temperature can act as an independent replicate, and comparisons of non-smoke treatments with several times of smoking allow each level of time to act as an independent replicate. However, these are less powerful experimental designs than one in which there is also true replication within each of the species/temperatures/times, and the use of true replication in all aspects of an experimental design is thus to be encouraged.

As a final note, we point out that we have dealt only with the problems of pseudoreplication with respect to manipulative treatments not being properly replicated. However, there are other possible sources of non-independence that can affect experiments on seed germination. The first of these relates to the problems created by doing experiments that are conducted in only one laboratory and/or using only one batch of seeds. Variation in laboratory conditions may be trivial, but variability in experimental results has been recorded both within (e.g. Dirilgen & Inei 1994; Krassoi & Julli 1994) and between (e.g. Damron *et al.* 1986; Dorn *et al.* 1987; Pelloux *et al.* 1998) laboratories, so this is a source of variation that should not be ignored. Potentially more important is variation due to the source of the seeds, e.g. their provenance, the time of collection or the number of plants from which the seeds are collected. These are not necessarily easy issues to deal with, but it is important to recognise the limitations imposed by the potential nonindependence and to be careful about exactly what hypotheses are being tested and then not generalising beyond the limitations imposed by the sampling procedure.

The second source of potential non-independence relates to appropriate methods of replicating the control treatment in an experiment. Technically, a control treatment should be identical to the manipulative treatment in all ways except for the variable(s) being tested if it is to act as a useful means of identifying causal relationships with signals that might trigger germination. This implies that for dry-heat treatments, for example, a suitable control treatment would involve putting seeds into an oven that has been turned off, rather than merely attempting to germinate 'untreated' seeds. Similar protocols need to be applied for scarification, smoke and charcoal treatments as well. Under these circumstances, all of our caveats for the independent replication of manipulations must apply to the control treatment as well.

REFERENCES

Anderson B. (1990) *Methodological Errors in Medical Research. An Incomplete Catalogue.* Blackwell Science, Oxford.

Baldwin I. T., Staszak-Kozinski L. & Davidson R. (1994) Up in smoke I. Smoke-derived germination cues for the postfire annual, *Nicotiana attenuata* Torr. ex Watson. *J. Chem. Ecol.* **20**, 2345–71.

Baxter B. J. M., Granger J. E. & van Staden J. (1995) Plant-derived smoke and seed germination: is all smoke good smoke? That is the burning question. *S. Afr. J. Bot.* **61**, 275–7.

Cayouette M., Moores J., Capodanno J. & Rahilly M. (1998) Effects of thermal cycler well-to-well temperature uniformity on PCR results. *Strategies* **11**, 18–19.

Cousens R. (1996) Design and interpretation of interference studies: are some methods totally unacceptable? *N.Z. J. Forest. Sci.* **26**, 5–18.

Damron D. J., Warren J. W., Chippendale G. R. & Tenney J. H. (1986) Do clinical microbiology laboratories report complete bacteriology in urine from patients with long-term urinary catheters? *J. Clin. Microbiol.* **24**, 400–4.

Dirilgen N. & Inei Y. (1994) Effects of zinc and copper on growth and metal accumulation in duckweed, *Lemna minor*. *Bull. Env. Contam. Toxicol.* **53**, 442–9.

Dorn P. B., Rodgers J. H., Jop K. M., Raia J. C. & Dickson K. L. (1987) Hexavalent chromium as a reference toxicant in effluent toxicity tests. *Env. Toxicol. Chem.* **6**, 435–44.

Gibson D. J., Connolly J., Hartnett D. C. & Weidenhamer J. D. (1999) Designs for greenhouse studies of interactions between plants. *J. Ecol.* **87**, 1–16.

Heffner R. A., Butler M. J. & Reilly C. K. (1996) Pseudoreplication revisited. *Ecology* **77**, 2558–62.

Hurlbert S. H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecol. Monog.* **54**, 187–211.

Jones C. S. & Schlesinger W. H. (1980) *Emmenanthe penduliflora* (Hydrophyllaceae): further consideration of germination response. *Madroño* **27**, 122–5.

Keeley S. C. & Pizzorno M. (1986) Charred wood stimulated germination of two fire-following herbs of the California chaparral and the role of hemicellulose. *Am. J. Bot.* **73**, 1289–97.

Krassoi F. R. & Julli M. (1994) Chemical batch as a factor affecting the acute toxicity of the reference toxicant potassium dichromate to the cladoceran *Moina australiensis* (Sars). *Bull. Env. Contam. Toxicol.* **53**, 153–7.

Mead R. (1988) *The Design of Experiments. Statistical Principles for Practical Applications.* Cambridge University Press, Cambridge.

Morris E. C. (2000) Germination response of seven east Australian *Grevillea* species (Proteaceae) to smoke, heat exposure and scarification. *Aust. J. Bot.* **48** (in press).

Morrison D. A., McClay K., Porter C. & Rish S. (1998) The role of the lens in controlling heat-induced breakdown of testa-imposed dormancy in native Australian legumes. *Ann. Bot.* **82**, 35–40.

Pelloux H., Guy E., Angelici M. C. *et al.* (1998) A second European collaborative study on polymerase chain reaction for *Toxoplasma gondii*, involving 15 teams. *FEMS Microbiol. Lett.* **165**, 231–7.

Underwood A. J. (1997) *Experiments in Ecology. Their Logical Design and Interpretation Using Analysis of Variance*. Cambridge University Press, Cambridge.