Journal of the American Mosquito Control Association, 13(1):102–103, 1997

LETTERS TO THE EDITOR

PSEUDOREPLICATION: WHAT DOES IT MEAN, AND HOW DOES IT RELATE TO BIOLOGICAL EXPERIMENTS?

Consider the following hypothetical experiment: a study was conducted to determine temporal residual effectiveness of permethrin against adult *Culex salinarius* Coquillett, when applied as an ultra-low volume (ULV) mist, to the interior of horse barns. In the study design, 4 horse barns were treated with the insecticide (referred to as treatment). Mosquito mortality was evaluated weekly in barns by exposing 4 cages of 25 live *Cx. salinarius* females each to one of the treated walls of each barn for 2 h. One nontreated barn (referred to as control) was used to compare treatments but used 4 cages of *Cx. salinarius* females that were also placed against the interior walls. Mortality in the control was assessed as in the treatment. After the exposure time ended, data on mean percent mortality at 24 h postexposure in treatments were compared with that of controls. Caged mosquitoes were exposed weekly for 15 wk. Treatment and control mean mortality data were recorded and compared at each date using a *t*-test. Significant differences for each time interval (day of exposure) were determined at $P < 0.05$.

In the above hypothetical study, a serious mistake was made, relative to experimental design and statistical inference, referred to as pseudoreplication. A valuable discussion of this subject has been provided by Hurlbert (1984) and the reader is urged to examine that article in length. However, I would like to review some of the pertinent points of pseudoreplication outlined in that paper using my hypothetical study as an example.

Pseudoreplication "results from the use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated (though samples may be) or replicates are not statistically independent" (Hurlbert 1984). In my study, the control was not replicated, although treatments were. The experimental units in this study were horse barns NOT number of mosquito cages. Ideally, there should have been 4 control barns (or at the very least 2). Replication provides an estimate of experimental error; this improves precision of the experiment by reducing the standard deviation of a treatment mean. Furthermore, replication increases the scope of inference of an experiment. Lastly, replication restrains, to a certain extent, the error variance. Reduced variation results in increased precision estimates for parameters such as treatment means or difference between two means (Hurlbert 1984). An experiment that lacks replication between experimental units does not give the probability (P value determined by variation between treatments) of rejecting the null hypothesis when true (referred to as type I error).

Pseudoreplication is not a defect in experimental design, rather it is a problem of sampling coupled with inappropriate statistical analysis for testing a hypothesis. Hurlbert (1984) covered 3 instances of pseudoreplication that I will summarize here: simple, temporal, and sacrificial. Simple pseudoreplication would occur if treatments were not segregated in space or time but were somehow interconnected so that "replicates" were actually subsamples from a single experimental unit (Fig. 1A). This was the instance in my study. Four mosquito cages were used in the one control barn so it looked like I had 4 replications (cages) for analysis but cages, as stated earlier, were not the experimental units.

Another common mistake researchers can make is temporal pseudoreplication, when multiple samples are taken sequentially over time rather than simultaneously. This was another flaw in my study. Instead of replicating barns, I chose to use a single barn but used time (i.e., week) as my replication (Fig. 1B). Also, it was inappropriate for me to use a significance test that compared treatment mortality data with that of the control because *successive dates were treated as if they were independent replicates of the control, which they were not.*

Sacrificial pseudoreplication results when 2 or more samples taken from an experimental unit are treated as independent replicates. In the example from Fig. 1C, the experimenter had set up his study so that there were 4 total samples (which he called "replicates") from a treatment category and 4 samples from a control category. (For the sake of simplicity we will assume that the analysis of variance table showed a significant F value for the interaction term of week/treatment.) A *t*-test was then employed to compare the pooled data points from each category on the assumption that they came from 4 independent experimental units (i.e., replicates). In reality, the experimenter in Fig. 1C had 2 data points (or replicates) per category (i.e., 2 treatments and 2 controls). In this case, the *assumption of independence between data points for each control, as well as each treatment replicate, was violated.* In this example, *information on the variance among treatment replicates was confounded with variance among samples within replicates* (Hurlbert 1984).

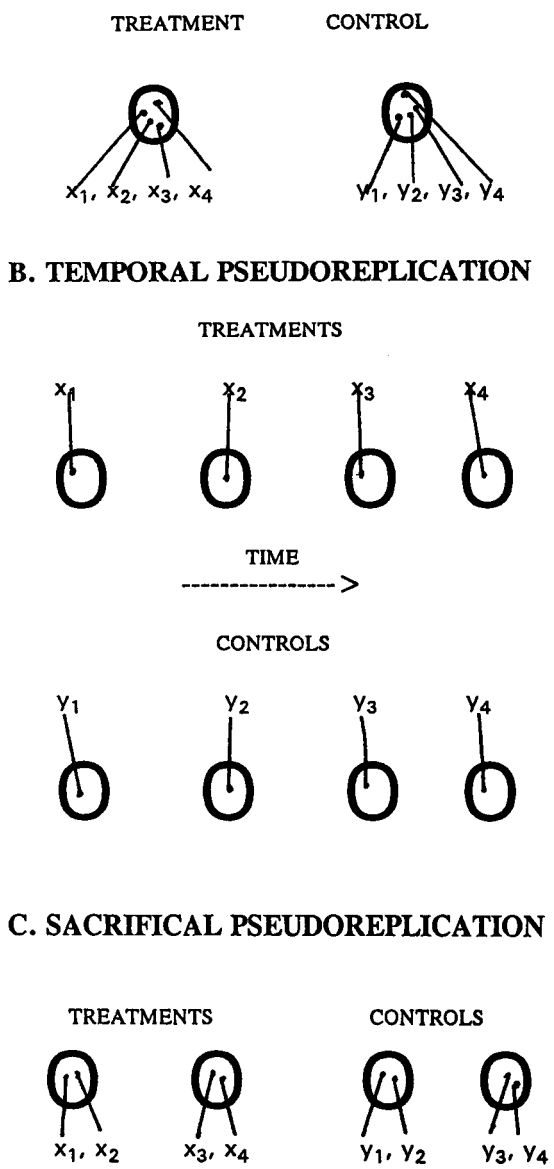

Fig. 1. Various types of pseudoreplication.

My study used 4 mosquito cages as separate independent replicates within the single control and, as a result, violated an assumption of independence.

However, Hurlbert (1984) does concede there are instances where unreplicated experiments are valid, but they depend on the "*experimental units being identical at the time of manipulation and on their remaining identical to each other after manipulation, except [for] . . . treatment effect.*" However, the lack of significant effects prior to the experimental treatment *cannot* be interpreted as evidence for nonreplication and I refer the reader to Hurlbert's (1984) continued discourse on this topic.

So what have I learned from the above discussion? My original study should have started out with, optimally, 4 horse barns for treatment and the same number of barns for control, **as barns were the experimental units, not the number of mosquito cages used per barn.** Therefore, my study would have been replicated adequately and a statistically appropriate approach using, for example, a comparison test (such as a *t*-test) that compared mosquito treatment mortality data with controls, at each time interval (assuming the analysis of variance table showed a significant F value for the interaction term of week/treatment in the model), would have been totally adequate.

REFERENCE CITED

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecol. Mongr. 54:187–211.

J. E. Cilek
John A. Mulrennan, Sr.
Arthropod Research Laboratory
Florida A & M University
4000 Frankford Avenue
Panama City, FL 32405