Journal of Chemical Ecology, Vol. 26, No. 6, 2000

PSEUDOREPLICATION AND ITS FREQUENCY IN OLFACTOMETRIC LABORATORY STUDIES

CLAUDIO C. RAMÍREZ,[1,]* EDUARDO FUENTES-CONTRERAS,[2] LUIS C. RODRÍGUEZ,[1] and HERMANN M. NIEMEYER[1]

[1]*Laboratorio de Química Ecológica, Departamento de Ciencias Ecológicas*
Facultad de Ciencias, Universidad de Chile
Casilla 653, Santiago, Chile

[2]*Departamento de Producción Agrícola, Facultad de Ciencias Agrarias*
Universidad de Talca
Casilla 747, Talca, Chile

(Received April 19, 1999; accepted February 5, 2000)

Abstract—The evaluation of behavioral responses of an organism to a particular stimulus normally implies the design of a bioassay. Measurements of the response in a number of replicates are necessary to perform inferential statistics and therefore accept or reject a hypothesis about the effect of the stimulus on the behavior of the organism under study. In the present article, we address the importance of pseudoreplication in studies of chemical ecology, particularly in laboratory experiments on olfactory responses of insects to semiochemicals in olfactometers and wind tunnels. Pseudoreplication may be caused by lack of independence in the stimulus or the experimental device, the reutilization of test insects, or the use of groups of test insects, without adequate statistical analysis addressing such dependency. Each and all of the cases reviewed (N = 105) lacked information in at least one of the factors listed above; hence no cases could be said with certainty to be free of pseudoreplication. Forty-nine cases (46.7%) contained explicit information revealing that pseudoreplication existed in terms of one or more of the criteria listed above; in only three of these cases did the authors address statistically the stated dependence of the samples. Pseudoreplication due to different factors ranged from 2% to 30% of the cases, with an average of 13%. The most frequent sources of pseudoreplication were the reuse of the device and the use of groups of test insects. The analysis showed the low importance given to obtaining independent replicates in bioassays involving olfactometric responses of insects to semiochemicals.

Key Words—Pseudoreplication, experimental design, olfactometer, wind tunnel, semiochemical, statistical independence.

*To whom correspondence should be addressed.

0098-0331/00/0600-1423$18.00/0

INTRODUCTION

The experimental approach is a basic element of biology. Irrespective of the particular attributes and the scale of analysis of different systems, the experimental approach involves five steps: hypothesis, experimental design, experimental execution, statistical analysis, and interpretation of results (Hulbert, 1984). The hypothesis is of primary importance; even well-designed, well-executed, well-analyzed, and well-interpreted experiments are meaningless if the hypothesis is inadequate. The first step of the experimental design is to specify the smallest division of the experimental material such that any two units may receive different treatments (Krebs, 1989). After identifying the experimental unit, the number and kind of treatments to be applied to the units, and the responses to be measured in them, should be considered. Then it is necessary to establish the number of experimental units to be used in each treatment (replicates), their spatial arrangement, the procedure followed in assigning the experimental units to each treatment, and the temporal sequence for applying the treatments and measuring the observed responses (Hulbert, 1984).

The execution of the experiment includes all procedures and steps taken to implement the experimental design selected. The experimenter should intend to reduce the random error and avoid the introduction of systematic errors (bias) that may produce invalid or inconclusive results. It is also important to decide on the amount of initial heterogeneity between replicates (e.g., age, sex, experience, or any other variable of the organisms under study), and the nature of environmental variables (e.g., light, temperature, relative humidity, etc.) to be regulated during the experiment and the extent to which they will or can be controlled. These decisions will affect the magnitude of the random error and therefore the sensitivity of the experiment to detect treatment effects. Statistical analyses are meant to improve clarity, conciseness, and objectivity in the interpretation of the experimental results. Since data can be reanalyzed or reinterpreted if a mistake in analysis or interpretation is detected, analysis and interpretation are not the most critical aspects of an experiment. On the other hand, failures in the design or execution of an experiment usually can be solved only with a new, improved experiment (Hulbert, 1984).

Hulbert (1984) reviewed the ecological literature and concluded that a widespread mistake in field experiments consisted in treating multiple samples from one experimental unit as multiple experimental units or using experimental units that are not statistically independent. He referred to this situation as false replication or pseudoreplication. Pseudoreplication is not a problem of experimental design itself; rather, it often is a combination of experimental design and statistical analysis that is inappropriate to evaluate the hypothesis under question. Pseudoreplication is not always understood, and some researchers may claim that all experiments are, at the extreme, dependent, since we are working on the same

planet Earth. These arguments fail to realize that pseudoreplication is the result of nonreplicated treatments or dependent experimental units at the relevant scale of analysis for the hypothesis being tested.

Pseudoreplication has been widely reported in environmental, ecological, and behavioral studies (Hulbert, 1984; Steward-Oaten and Murdoch, 1986; Searcy, 1989; Hulbert and White, 1993; Wiens and Parker, 1995; Heffner et al., 1996). However, its relevance in laboratory experiments has not been clearly acknowledged. Chemical ecology, among other disciplines, often deals with laboratory bioassays evaluating the behavioral responses of organisms toward chemical stimuli. The aim of the present work is to assess the frequency of pseudoreplication in studies published on olfactometric responses of insects and other terrestrial arthropods toward semiochemicals. Our intention is not to evaluate the performance of olfactometric devices (e.g., Giles et al., 1996) or wind tunnels, but to call to the attention of our colleagues the implications of not considering true replicates in their experiments. We have evaluated whether the experiments are inadequate due to failures in experimental design and statistical analysis. We hope that this study will increase awareness of the frequently overlooked importance of statistically valid replicates, from which insect responses to semiochemicals are inferred.

METHODS AND MATERIALS

Articles published between 1993 and 1997 in the *Journal of Chemical Ecology*, *Entomologia Experimentalis et Applicata*, *Journal of Applied Entomology*, *Biological Control*, and *Physiological Entomology* were reviewed. Based on an analysis by key words in the *BIOSIS preview* database, these journals cover approximately 80% of the articles published on olfactometry during the period analyzed. Articles were selected if they used olfactometric devices or wind tunnels and if they applied inferential statistics to their results. When more than one olfactometric device or wind tunnel was used per article they were analyzed separately. The independence of the replicates declared was evaluated in relation to aspects of the experimental set-up and utilization of test insects. The independence of the replicates may be violated if: (1) the same arena ("device" in Table 1 below), or connecting devices (connection) if applicable, or the same source of chemical stimulus (source), is used in more than one replicate within each treatment, or (2) the same test insect is used in more than one replicate (reused TI), or more than one insect is used at the same time and each insect is considered a replicate (grouped TI). When the information provided by the authors was not explicit, the article was considered as not reporting enough information to evaluate the occurrence of pseudoreplication.

The studies were separated in experiments performed with olfactometers or

wind tunnels. All experimental devices that allowed insects to walk were recognized as olfactometers, while experimental devices designed to permit insect walking and sustained flight to reach the stimulus were classified as wind tunnels.

RESULTS AND DISCUSSION

The study involved 59 reports with olfactometric devices of several types (four arms, Y tubes, linear tracks, linear T, etc.), and 42 reports with wind tunnels (three reports involved two different experiments with wind tunnels). One report tested more than one type of olfactometer. Thus, the study comprised 60 cases involving olfactometers and 45 involving wind tunnels. The overall analysis showed that all of the cases studied did not provide sufficient information in the description of the bioassays to assert whether replicates were independent or not. Table 1 shows the number of pseudoreplicated, correctly analyzed, or insufficiently detailed cases, for each of the aspects studied in the experimental designs involving olfactometers and wind tunnels.

Olfactometers. Thirty-four cases (56.7%) contained explicit information revealing that pseudoreplication had occurred; in only two of these cases did the authors address statistically the stated dependence of the samples.

Fifteen percent of the cases explicitly declared no change in the semiochemical source or change after a given number of "replicates" (e.g., every three replicates or every quarter of total replicates), but failed to include such depen-

TABLE 1. NUMBER AND PERCENTAGE OF CASES QUALIFIED AS PSEUDOREPLICATED, CORRECTLY ANALYZED, NOT PROVIDING ENOUGH INFORMATION TO DECIDE IF PSEUDOREPLICATION OCCURRED, AND NOT APPLICABLE TO ANALYSIS[a]

Category	Source	Device	Connection	Reused TI	Grouped TI
Olfactometer					
Not enough information	44 (73.3%)	30 (50%)	42 (70%)	36 (60%)	3 (5%)
Pseudoreplicated	9 (15%)	18 (30%)	8 (13.3%)	5 (8.3%)	13 (21.7%)
Correctly analyzed	7 (11.7%)	12 (20%)	7 (11.7%)	19 (31.7%)	44 (73.3%)
Not applicable	0	0	3 (5%)	0	0
Wind tunnel					
Not enough information	42 (93.3%)	37 (82.2%)	4 (8.9%)	30 (66.7%)	2 (4.4%)
Pseudoreplicated	2 (4.4%)	3 (6.7%)	1 (2.2%)	2 (4.4%)	11 (24.4%)
Correctly analyzed	1 (2.2%)	4 (8.9%)	0	13 (28.9%)	32 (71.1%)
Not applicable	0	1 (2.2%)	40 (88.9%)	0	0

[a] The cases were dissected into those involving an olfactometer or wind tunnel. The analysis to detect pseudoreplication was focused on the source of semiochemicals, experimental device (if applicable), connections (if applicable), reuse of test insects (Reused TI), or use of groups of test insects (Grouped TI).

dence between replicates in the statistical analysis. Only 11.7% of the cases declared the use of independent odor sources for each replicate, and the largest proportion (73.3%) of the cases did not clearly report whether they did or did not use independent semiochemical sources to run the replicates.

The change or cleaning of the olfactometric device was not performed or was performed after a given number of "replicates", and this fact was not taken into account in the statistical analysis in 30% of the cases analyzed. In 20% of the cases, the olfactometer was stated as being appropriately changed or cleaned between replicates, while 50% of cases did not give enough information to judge whether pseudoreplication had occurred. The change or cleaning of other connecting devices in the set-up was largely not mentioned by the authors (70%).

Pseudoreplication due to the reutilization of test insects with no inclusion of this dependence factor in the statistical analysis was found in 8.3% of the cases, while the experimental design and analysis were correct in 31.7%, and the information was insufficient for accurate evaluation in 60% of the cases. The inclusion of individual insects within a group in "independent" replicates (pseudoreplicates) occurred in 21.7% of the cases analyzed. Contrastingly, 73.3% of the cases were correctly analyzed with regard to pseudoreplication. Indeed, several of these studies used groups of insects to test their response to semiochemicals and regarded the whole group as part of an experimental unit, hence avoiding pseudoreplication. A rather low percentage (5%) of cases did not provide explicit information about the use of groups or single insects in replicates.

Wind Tunnels. Fifteen cases (33.3%) contained explicit information revealing that pseudoreplication existed in terms of one or more of the criteria listed above. In only one of these latter cases did the authors address statistically the stated dependence of the samples.

The analysis showed that a very large proportion of the cases did not mention either the change of sources of semiochemicals (93.3%) or proper cleaning of the flying chamber of the wind tunnel between "replicates" (82.2%). The reutilization of insects was not explicit in 66.7% of the cases, while a proper design and analysis was found in 28.9%. Explicit pseudoreplication by reuse of insects was found in only 4.4% of the cases. The use of groups or single individuals of test insects was explicit in the majority of cases and appropriate statistics applied (71.1%). However, 24.4% of cases utilized groups of test insects and did not include this dependency in the analysis.

The consequences of pseudoreplication on the conclusions that are drawn from an experiment are best seen through an example. Suppose a scientist wishes to test the hypothesis that volatiles from a plant attacked by a herbivore attract natural enemies of that herbivore. An olfactometer is employed consisting of a stimulus area, a control area, and a decision-making area; natural enemies are individually introduced into this latter area, and the time they spend in each area is recorded for a total of 450 sec. Volatiles from an attacked plant are con-

ducted to the stimulus area of the olfactometer, while the control area receives volatiles from an unattacked plant. Fifty observations are performed ($N = 50$). After each observation the olfactometer is cleaned and rotated by 180°, and the plants (attacked and unattacked) are changed every five observations. The results of this hypothetical experiment are shown in Table 2. If all 50 observations are taken as replicated, a simple t test for paired data shows that the difference between treatment is significant ($t = -2.69$, $df = 49$, $P = 0.0096$), leading to the conclusion that an attacked plant attracts the natural enemy of the herbivore.

The above conclusion is based on a statistical test that assumes that replicates are independent. However, this was not the case. Replicates were interdependent because the scientist did not change the plants after every observation, but only after every fifth observation. As noted in Table 2, interdependency of replicates affected groups of five observations (observations 1–5, 6–10, and so on). Therefore, within each of these groups, observations were not independent. The 10 groups constitute, however, legitimate replicates, and each is represented by the mean value of the five observations performed within it. A t test applied to these paired means indicates no significant differences between treatments ($t = -1.59$, $df = 9$, $P = 0.14$). The same conclusion is obtained when a nonparametric test is performed, namely the Wilcoxon matched pair test ($Z = 0.45$, $P = 0.64$), which can be performed when normality assumptions are violated, for example, due to a small number of replicates. Alternatively, an ANOVA for repeated measures, which incorporates the dependence of observations in each treatment, may be applied. In this case also no treatment effect is observed ($F = 0.49$; $df = 40.9$; $P = 0.87$). Thus, the correct conclusion from this hypothetical experiment should be that differences in attractivity to the natural enemy between attacked and unattacked plants are not significant. It is to be noted that the same contrasting results would have been obtained if the same devices, connectors, or test insects had been used for a group of five consecutive observations.

In spite of the fact that statistical methods are available that may be used to draw apparently meaningful conclusions from most data sets, it should be a goal of the experimenter to establish independence between replicates since results from dependent observations may be rendered inaccurate on biological or chemical grounds. Thus, the neutrality of a device may suffer significant changes because the previously tested insect may release semiochemicals that affect the response of the following test insect; a test insect may be stressed by an observation, or its response to a given stimulus may be age- or experience-dependent, and it may respond in a different way in the next observation. A chemical may be altered by exposure to conditions in the device, or its release rate from the biological source and even its composition, in the case of mixtures of semiochemicals, may change with time. Finally, the use of a group of test insects in a bioassay should not be used to obtain several "replicates" from the individual responses of insects in the group because results may be influenced by social

TABLE 2. HYPOTHETICAL EXAMPLE OF EXPERIMENT TO SHOW CONSEQUENCES OF PSEUDOREPLICATION[a]

Obs.	Time spent in arm receiving volatiles from unattacked plants	Average of each group	Time spent in arm receiving volatiles from attacked plants	Average of each group	Time spent in volatile neutral area	Average of each group
1^1	94		322		34	
2^1	112.5		292.5		45	
3^1	104.5	99.8	290.5	305.4	55	44.8
4^1	107.5		311.5		31	
5^1	80.5		310.5		59	
6^2	213.5		195.5		41	
7^2	188.5		228.5		33	
8^2	213.5	202.8	181.5	199.6	55	47.6
9^2	210		180		60	
10^2	188.5		212.5		49	
11^3	89		313		48	
12^3	45		361		44	
13^3	44.5	57	314.5	337.8	91	55.2
14^3	38		358		54	
15^3	68.5		342.5		39	
16^4	213.5		199.5		37	
17^4	207		189		54	
18^4	193.5	205.1	205.5	199.9	51	45
19^4	188		281		44	
20^4	223.5		187.5		39	
21^5	206.5		196.5		47	
22^5	212.5		204.5		33	
23^5	214	206.2	182	200.6	54	43.2
24^5	182		222		46	
25^5	216		198		36	
26^6	248		154		48	
27^6	81		309		60	
28^6	150.5	198.2	248.5	206.2	51	45.6
29^6	201		213		36	
30^6	310.5		106.5		33	
31^7	274		146		30	
32^7	247		145		58	
33^7	203.5	203	215.5	201.4	31	45.6
34^7	196.5		202.5		51	
35^7	94		298		58	
36^8	111		285		54	
37^8	220.5		180.5		49	
38^8	84	153.6	318	250.4	48	46
39^8	219		187		44	
40^8	133.5		281.5		35	

TABLE 2. (CONTINUED)

Obs.	Time spent in arm receiving volatiles from unattacked plants	Average of each group	Time spent in arm receiving volatiles from attacked plants	Average of each group	Time spent in volatile neutral area	Average of each group
41[9]	300		114		36	
42[9]	209		197		44	
43[9]	189	215.4	213	191.4	48	43.2
44[9]	230.5		168.5		51	
45[9]	148.5		264.5		37	
46[10]	227.5		167.5		55	
47[10]	356.5		36.5		57	
48[10]	122.5	207.1	280.5	192.7	47	50.2
49[10]	127		265		58	
50[10]	202		214		34	
Mean	174.82	174.82	228.54	228.54	46.64	46.64
SD	71.84	54.44	69.49	52.34	11.04	3.63

[a] The device is assumed to consist of three areas: stimulus, control, and neutral (decision-making area). The same superscript in numbers in the first column (Obs.) indicates hypothetical interdependent replicates.

interactions (e.g., interference, aggression, attraction, etc.) between the individual insects.

Hence, it is strongly suggested that devices should be thoroughly cleaned, and test insects and chemical sources changed between observations. A test insect may be reused if its biology is sufficiently well known to ensure that the insect will respond consistently at every observation it is used. Most importantly, a detailed description of experimental conditions and procedures must be given in order to judge the adequacy of the conclusions drawn from the results.

Acknowledgments—This work was supported by the Cátedra Presidencial en Ciencias awarded to H.M.N., and the Latin American Network for Research on Bioactive Natural Compounds (LANBIO) through a fellowship to L.C.R. We gratefully acknowledge the comments of one of the referees, which helped clarify some of the ideas presented.

REFERENCES

GILES, D. K., HEINZ, K. M., and PARRELLA, M. P. 1996. Quantitative assessment of insect olfactometer performance by experimental flow analysis. *Biol. Control* 7:44–47.

HEFFNER, R. A., BUTLER, M. J., IV, and REILLY, C. K. 1996. Pseudoreplication revisited. *Ecology* 77:2558–2562.

HULBERT, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54:187–211.

HULBERT, S. H., and WHITE, M. D. 1993. Experiments with freshwater invertebrate zooplanktivores: quality of statistical analyses. *Bull. Mar. Sci.* 53:128–153.

KREBS, C. H. 1989. Ecological Methodology. Harper Collins, New York, 654 pp.

SEARCY, W. A. 1989. Pseudoreplication, external validity and the design of playback experiments. *Anim. Behav.* 38:715–717.

STEWART-OATEN, A., and MURDOCH, W. W. 1986. Environmental impact assessment: "pseudoreplication" in time. *Ecology* 67:929–940.

WIENS, J. A., and PARKER, K. R. 1995. Analyzing the effects of accidental environmental impacts: Approaches and assumptions. *Ecol. Appl.* 5:1069–1083.