RUSSIAN VERSION OF THIS MANUSCRIPT IS PUBLISHED IN
Journal of Fundamental Biology (Moscow)
2006 vol. 67 no. 2 pp. 145-152

Pseudoreplication, chatter, and the international nature of science: A response to D. V. Tatarnikov

M. V. Kozlov[1] & S. H. Hurlbert[2]
[1]*Section of Ecology, University of Turku, Turku 20014 Finland*; e-mail: *mikoz@utu.fi*
[2]*Department of Biology and Center for Inland Waters, San Diego State University, San Diego, California, 92182-4614 USA;*
e-mail*: shurlbert@sunstroke.sdsu.edu*

The commentary by Tatarnikov (2005) on the design and analysis of manipulative experiments in ecology represents an obvious danger to readers with poor knowledge of modern statistics due to its erroneous interpretations of *pseudoreplication* and *statistical independence*. Here we offer clarification of those concepts – and related ones such as *experimental unit* and *evaluation unit* – by reference to studies cited by Tatarnikov (2005). We stress the necessity of learning from the accumulated experience of the international scientific community in order not to repeat the errors found in earlier publications that have already been analyzed and widely written about. (An English translation of the full article is available as a pdf file from either of the authors.)

"....[N]owhere in all of scholarship has the book or shorter contribution (the 'paper') become more thoroughly debased than in science ... the principal remedy is for everyone to write fewer and more significant works ... It seems to be a deeply held, quasi-philosophical position among contemporary scientists that publication, and lots of it, is an inalienable right ... it is no longer an honor to get a paper published ... publication of any and all results has become the norm ... the publication process has largely ceased to act as a quality control mechanism ... It is terribly important for students to appreciate the older literature in their field ... For scientists there is a danger that the vast tide of chatter in the current literature may isolate us from our intellectual underpinnings."

– Keith Stewart Thomson (1984), Dean, Graduate School, Yale University

Is Russian science to take a separate path?

We were pleased, surprised and alarmed by publication of the paper by Tatarnikov (2005). Pleased because its appearance indicates that the first goal of the paper by Kozlov (2003a; also published in abridged form, in English - Kozlov 2003b) – to attract attention of the Russian scientists to the problem of pseudoreplication[1] in ecological studies - was successfully achieved. We feel that even researchers who missed the publication by Kozlov (2003a) will now be intrigued enough to read it or search for the review by Hurlbert (1984) that first described the problem in extenso.

The surprise is, however, the weak basis of the criticism: Tatarnikov (2005) shared with the readers of the 'Journal of Fundamental Biology' his personal thoughts and impressions, but apparently without reading any of the numerous publications (other than Hurlbert, 1984) that have discussed the same problem. In particular, three analyses identified as simple pseudoreplication by Kozlov were claimed to be valid by Tatarnikov. Yet the design and statistical problems in those three studies are similar to ones that have been discussed by many, e.g. Underwood (1981, 1997), Hurlbert (1984, 2004), James & McCulloch (1985), Machlis et al. (1985), Kroodsma (1989), Mead (1988: Ch. 6), Hairston (1989), Hurlbert & White (1993), Wise (1993), Heffner et al. (1996), Lombardi & Hurlbert (1996), Riley & Edwards (1998), García-Berthou & Hurlbert (1999), Krebs (1999: Ch. 10), Morrison & Morris (2000), Ramirez et al. (2000), Kroodsma et al. (2001), Jenkins (2002), Cottenie & De Meester (2003), Hurlbert & Meikle (2003), Ruxton & Colegrave (2003: Ch. 3), Hurlbert & Lombardi (2004), Millar & Anderson (2004), and others. Identical errors are often present in statistical analyses of medical experiments and have been labeled by medical statisticians as "spurious replication", "trial inflation" or "the unit of analysis error or problem" (Whiting-O'Keefe et al., 1984, Andersen 1990: 147-156, Altman & Bland 1997). Although the label 'pseudoreplication' is used only in some of the above-mentioned studies, and our opinion on specific issues may differ from conclusions of some of these authors, they all offer serious discussions of the issue. The level of Tatarnikov's criticism is thus very superficial. Nearly every ecology student who has learned statistics in a university of North or South America, Western Europe, Australia, or some parts of Asia is familiar with the pseudoreplication problem. The problem was not recognized, however, by the referees or editors of a prestigious biological journal who placed this material in a section called 'Scientific discussions' – and this is indeed alarming.

The basic idea in the paper by Tatarnikov (2005) is best described by an old Russian joke: if something is forbidden, but we want it very much – then it is allowed! Possibly useful in some contexts, this approach is dangerous in science, unless Russian ecologists wish to claim, in line with some politicians, that Russian science has its own rules and imperatives. Perhaps this is the case, but then one should be prepared for the multiplication table to be the next object of a 'scientific discussion'.

How many times will one step on the same rake?

The many critiques of the designs and analyses of ecological experiments published by Hurlbert (1984) and others following him have indeed been painful for many researchers who recognized pseudoreplication in their own past and present work. As the result, some weak

[1] The English word 'pseudoreplication' has no equivalent in Russian, because it describes a process – the commission of a particular type of statistical error. To avoid confusion, we provide <in a Russian version only> an English-Russian concordance for all key terms.

attempts at defense, including many misinterpretations of pseudoreplication, have been published. We do not review those here, but appearance of the paper by Tatarnikov (2005) clearly indicates that pseudoreplication merits further clarifying discussion. To usefully contribute to this, one should first familiarize himself with earlier publications on the topic. Tatarnikov (2005) missed this point – and therefore he did not advance the discussion. He made the same logical errors that others made earlier.

Perhaps we don't see the rake the first time, but our foreheads hope for adaptive behavior on our part after one or two intimate meetings with the rake handle. Let us follow the advice by Otto von Bismarck (known as the Iron Chancellor of Germany) and learn from errors made by other people, instead of making the same errors again and again. In research, it is necessary to take into account the experience accumulated by the international scientific community.

We began this commentary with a quotation from Thomson (1984), a distinguished American scientist who strongly indicts many aspects of western, especially American, science and the literature it is producing. Though written more than 20 years ago, its message has lost no relevance. Much of that indictment bears on the present discussion. In particular, "chatter minimization" is badly needed and should receive only applause from the international scientific community.

Clarifying 'experimental unit' and 'replication'

Although our comments are inspired by Tatarnikov's paper, we do not restrict ourselves to the analysis of his ideas concerning pseudoreplication. We also offer clarification of concepts and terminology relating to aspects of design and statistical analysis that are applicable to all disciplines that make use of manipulative experiments. In particular, we focus on the concepts of experimental unit, evaluation unit, and statistical independence, introducing Russian scientists to some ideas that are still being clarified in the international scientific literature.

To start at the beginning, a statement on similarity or dissimilarity of two groups of objects is valid only when between-group differences are compared with within-group variation. Very importantly, the level at which the variation is measured "within" a group is critical. Also it is obvious that assessment of within-group variation is only possible when the group consists of more than one object. A few situations exist where linear regression or multi-way analysis of variance allows valid testing for treatment effects for between-group differences in the absence of treatment replication (e.g., Kirk 1982:399, Mead & Curnow 1983:125, Milliken & Johnson 1989, Sokal & Rohlf 1995:292, 466, Hurlbert 2004:594).

Although in the context of manipulative experiments, the "object" of prime concern is the *experimental unit*, it is poorly or not at all defined in most statistics texts. We therefore offer the following definition of experimental unit, derived mainly from Cox (1958):

> *"The smallest system or unit of experimental material to which a single treatment (or treatment combination) is assigned by the experimenter* ***and*** *which is dealt with independently of other such systems under that treatment at all stages in the experiment at which important variation may enter. By 'independently' is meant that, aside from both receiving the same treatment, two systems or experimental units assigned to the same treatment will not be subject to conditions or procedures that are, on average, more similar than are the conditions or procedures to which two systems each assigned to a different*

treatment are subject" (Hurlbert, unpublished syllabus).

This may seem overly lengthy, but, on the evidence of dozens of textbooks, a shorter definition seems incapable of making explicit the key critical elements of the concept.

When there is only a single experimental unit under each treatment, however, and within-group variation is calculated from measurements made on multiple *samples* or *evaluation units* within a single experimental unit, then *simple pseudoreplication* is committed. This was exactly the error in Rudneva & Zherko (2000) and Smirnow (2001), two of the examples selected by Tatarnikov (2005) for discussion. The error is quite common: in four reviews alone, 59 articles containing simple pseudoreplication have been cited and several specific cases discussed (Hurlbert 1984, 14 papers; Hurlbert & White 1993, 10 papers; Hefner et al. 1996, 11 papers; Kozlov 2003a,b, 24 papers). So it is somewhat surprising to discover that some still regard this kind of statistical analysis as acceptable.

An *evaluation unit* is defined as "the unit of research material on which a response is evaluated" (Urquhart 1981) or "that element of an experimental unit on which an individual measurement is made" (Hurlbert 1990, Hurlbert & White 1993). The concept was perhaps first clearly articulated by Kempthorne (1952:163, 1979:163):

> *"The experimental unit can contain several observational units; for instance, a class of students that receive a certain method of teaching in common can be an experimental unit, while the individual students are the observational units. The distinction is ... very important, because, from the point of view of inference on the effects of treatments, the experimental unit must be considered as a whole, and the variation between the observational units within an experimental unit is usually of little value in assessing the errors of estimates of treatment effects."*

Urquhart's term *evaluation unit* seems preferable to *observational unit*, as the latter has very general connotations. Its use could further confuse the distinction between manipulative experiments and observational studies. This distinction is increasingly blurred in some recent books on design and analysis of experiments (see Hurlbert 1994, 1997, Hurlbert & Lombardi 2003, Mead 2003).

Clarity of language is critical here, and some aspects of English terminology have been commented on in several publications (Hurlbert 1990, Hurlbert & White 1993, Hurlbert & Meikle 2003). Portions of these discussions are hardly applicable to Russian terminology on this topic, which is still at an incipient stage of development. But it is important to distinguish between 1) *designs* involving unreplicated treatments (sometimes necessary and not an error in itself) and 2) strong *claims* of treatment effects based on statistical analyses where variation among multiple samples or evaluation units *within* experimental units is used as a surrogate for variation *among* experimental units. Such strong claims are the essence of pseudoreplication. Last but not least, descriptions of experimental designs and data analyses should always indicate *the level(s)* at which replication is implemented, e.g. "replicate experimental units" or "replicate samples of an experimental unit" or "replicate evaluation units". The adequacy of replication cannot be evaluated – nor can pseudoreplication be discovered! – if the structure of an experiment remains unclear. Neither replication nor replicates can be 'true' or 'false' *per se*; instead, it is the manner in which replicate values are used in a statistical analysis that is either correct or incorrect.

Let us consider the four examples mentioned by Tatarnikov (2005) and analyze the reasoning and language issues underlying his refusal to acknowledge that their statistical analyses constitute pseudoreplication.

Replicate fish vs replicate aquaria

Rudneva & Zherko (2000) tested effects of a toxicant on rockfish by comparing fish in a treated aquarium with those in a control aquarium, exemplifying simple pseudoreplication, as correctly diagnosed by Kozlov (2003a,b). The experimental unit is an aquarium and all the fish it contains, but Tatarnikov (2005) argues, in effect, that the experimental unit is a fish because the individual fish yield "independent responses of living beings to environmental factors" and that pseudoreplication would be present only if multiple measurements on an individual fish were made and, he implies, treated as representing multiple experimental units. He nevertheless admits the design was "imperfect" because differences among aquaria independent of treatment effects "may have occurred" in "physical and chemical parameters" despite attempts to eliminate them. This indeed means that any two fish in the same aquarium shared more similar environmental conditions compared to any two fish from different aquaria. The high likelihood that two fish in the same aquarium interact with and influence each other - physiologically, behaviorally, etc. - further contradicts the notion that measurements on them could be treated as statistically independent. By the definitions above, individual fish are evaluation units, not experimental units.

We usually strive for a high degree of homogeneity among experimental units in order to increase the power or sensitivity of an experiment. But variation *always* occurs among aquarium systems, growth chambers, field plots, cages with white mice, or any other entities established and treated as separate experimental units, and estimates of the magnitude of that variation are required for valid statistical analyses. Tatarnikov claimed that the only design yielding such estimates would involve placing "each fish into a separate aquarium." But that is not true. That experimental design is indeed possible, but it would not be an efficient one. The researcher could more reasonably divide the fish available for each treatment between two aquaria; two-fold replication of treatments is sufficient to yield the error term needed for a valid statistical test. And if aquarium differences were as small as Tatarnikov implies, then the test would be not only valid but sensitive as well.

Differences between two well-known statistics texts on this point illustrate both the issue and the confusion surrounding it. Steel & Torrie (1980:125), confirming Kempthorne (1952), correctly state that, "if 50 hens are penned together and fed the same ration, the experimental unit consists of the 50 hens. Other pens of 50 hens are needed before we can measure variation among units treated alike." In contrast, Sokal & Rohlf (1969:438, 1981:488) presented an example where a single tank of fish was set up at each of four treatments, e.g. each tank served an experimental unit. In spite of that, Sokal & Rohlf carried out an ANOVA that treated the individual fish (which were, in fact, evaluation units) as the experimental units, thus committing – and advocating! – pseudoreplication (Hurlbert 2004). In 1985 a student of SH challenged Sokal on the matter in a letter (Lisa Wood, in litt. to R. Sokal, 20 May 1985), and the entire multi-page example was omitted – without comment – from their third edition (Sokal & Rohlf 1995). The literature of fish physiology and aquaculture is rife with simple pseudoreplication, and the widely used first two editions of this popular reference work may be partly responsible.

One exclosure versus the outside?

Smirnov (2001) studied effect of fencing (experimental exclusion of large herbivores) on vegetation. This effect was evaluated by comparing vegetation on 35 $5m^2$ plots within one $450m^2$ fenced site with 35 plots, possibly in two groups (the arrangement is not clear), just outside the fenced area. P values were reported for differences between the enclosed and unenclosed site, but no information is given as to what test was used or how data were treated. In defense of the analysis applied by K.A. Smirnov (2001), Tatarnikov (2005) claims that merely "because any two spots in the forest differ from each other.... they are independent replicates."

The physical conduct of this experiment defined a 450m^2 site as the experimental unit here, and therefore we need to estimate the variation among such sites under the same treatment in order to reveal the effects of exclosure. Variation among plots within sites represents another level, one step down in the hierarchical analysis; therefore it cannot validly be used to reveal effects of fencing. This variation can be used to test for difference between sites, but such a test can provide no *statistical* grounds for concluding that the differences are attributable to fencing.

The cost and logistical difficulties of creating and maintaining large exclosures obviously imposes a severe constraint on treatment replication in such experiments. However, there often exists a low-cost solution that, while not having great power, is definitely better than complete lack of treatment replication. Control sites are much less costly to define and establish, because these usually require only imaginary boundaries. Thus, even if only one site can be treated or experimentally manipulated, multiple control sites can be defined, each preferably of the same size as the manipulated site to avoid complexities that would be produced by experimental units defined at more than one spatial scale. Note that the researcher should first define the set of sites and then make a random choice of the site to be fenced. Such designs, with replication of at least one of the treatments, permit valid statistical tests for treatment effects and more accurate estimation of effect sizes (e.g. Schindler et al. 1985, Frost et al. 1988, Underwood 1994). In the two-treatment case, all that is required is a t-test checking whether a single observation and a set of observations can be reasonably assumed to have been drawn from the same population. However, this analysis is based on the assumption that variation among fenced sites, if established, would be the same as among control sites.

Trees vs plots with trees

Referring to a hypothetical example presented by Kozlov (2003a) concerning effects of fertilization on Scots pine, Tatarnikov (2005) claimed that the multiple trees on a single fertilized plot could, in essence, be treated as independent experimental units. His reasoning was that "each tree responds to fertilization statistically independently ...[so the assumption of statistical independence would be violated only] if the response of the same organism [i.e. tree] is measured more than once." This accords with his view of the individual fish in the study of Rudneva & Zherko (2000). However, keeping in mind the definition of experimental unit, we emphasize that two trees within the same plot (whether fertilized or control) experienced, on average, more similar environments than did two trees from different plots under the same treatment. Trees within a plot are evaluation units, while the plots are experimental units.

Given his stance on the Scots pine example, we wonder how Tatarnikov would view a comparable fertilization experiment with a cereal crop involving one control plot and one fertilized plot. If it would be valid to test for a fertilizer effect using yield data gathered from multiple 4m^2 quadrats in each of the two plots, then agronomists around the world have been wasting tremendous amounts of resources in setting up experiments with multiple plots assigned at random to each treatment.

Two half-fields: manure vs no manure

Tatarnikov (2005) gives us no reason to change our opinion concerning the early study by Fisher & Mackenzie (1923). Indeed, he concludes – as did Box (1978), Cochran (1980), and Hurlbert (1984) – that with the design used "it was impossible to separate the effect of manure from differences between two halves of the field." For unclear reasons, he disagrees with Hurlbert (1984) and Kozlov (2003a) that the erroneous statistical analysis and contrary claim in the original work is appropriately labeled pseudoreplication. He notes that "variation existed within each half of the field," but seems not to recognize that within-experimental unit variation is universal and that, at the scale of the manure treatments in this split-split-unit design, the experimental unit was the 'half-field' and the separate plots in each half-field were, in effect, evaluation units.

Statistical independence: a core issue

Tatarnikov's (2005) misunderstanding of pseudoreplication derives from failure to distinguish between the concepts of experimental unit and evaluation unit and their differing implications for statistical independence and statistical analysis (Kempthorne 1952:163, Steel and Torrie 1980:125, Urquhart 1981, Hurlbert 1984, Whiting-O'Keefe et al., 1984, Hurlbert & White 1993). As his abstract and examples make clear, he believes pseudoreplication would result only if multiple measurements made on a single evaluation unit - a fish in Rudneva & Zherko (2000), a $5m^2$ plot in Smirnow (2001) - were to be treated as statistically independent for the purpose of testing for a treatment effect. In fact, however, pseudoreplication results whenever multiple measurements on the same *experimental* unit are treated in this manner, regardless of whether they are made on one evaluation unit or on several different ones, and regardless of whether they are made at essentially the same moment or at intervals over time.

Such confusion is widespread. Few books on experimental design use a common terminology and give clear discussions of the notion of statistical independence. This is true of books written by statisticians as well as those written by biologists, psychologists, and other scientists. One widespread misconception is that statistical independence is an inherent property of a sample or measurement or set of them, i.e. a property determined solely by how and where the samples were taken or measurements made. In fact, however, statistical independence:

> *"can be evaluated only in reference to both a data set and a specified hypothesis. If we take a set of random samples of bug density from each of two plots, the "errors" (epsilons) will possess the statistical independence needed for testing the H_0: no difference between plots. But, in the case where one plot has been sprayed with an herbicide and the other kept as a control, these errors will not possess the statistical independence required for testing the H_0: no difference between treatments. "* (Hurlbert 1997).

Last but not least: inability to replicate study systems or treatments does not necessarily diminish the validity or scientific value of a study (Hurlbert 1984, Carpenter et al. 1995, Schindler 1998, Kozlov 2003a,b). However, scientists need to understand fully the concept of statistical independence and the limitations imposed by lack of treatment replication; incorrect use of statistics is more misleading than is complete absence of statistical analysis.

Conclusion

Tatarnikov (2005) concluded his paper with the advice 'not to follow unthinkingly [!] all the recommendations of professional statisticians'. Honestly speaking, we are not professional statisticians but only biologists who appreciate the value of statistics for modern ecology. The experience accumulated by the international scientific community clearly indicates that manipulative experiments with unreplicated treatments generally provide less information than do those with such replication, and that results of the former are more difficult to interpret and report. Pseudoreplication, which can occur even in analyses of experiments *with* replicated treatments, is the result of statistical ignorance, more specifically of failure to distinguish between the concepts of experimental unit and evaluation unit and their differing implications for statistical independence and statistical analysis. Tatarnikov (2005) suggested that ecologists need to improve their knowledge on statistics, and we strongly support this point. Better understanding of statistical issues associated with the design and analysis of manipulative studies is prerequisite to integration into the international scientific community and, in particular, to getting studies published in respected international journals.

Acknowledgements

Authors are grateful to L. McDonald and E. Zvereva for commenting on an earlier version of the manuscript, and to Z. Tsyrlina and V. Zverev for discussion on the correct translation of terminology.

References

Altman D.G., Bland, M.J., 1997. Statistics notes: units of analysis // British Medical J. V. 314. P. 1874.

Andersen B., 1990. Methodological errors in medical research. Oxford: Blackwell. 270 p.

Box J.F., 1978. R.A. Fisher: the life of a scientist. New York: Wiley. xiii + 512 p.

Carpenter S.R., Chisholm S.W., Krebs C.J., Schindler D.W., Wright R.W., 1995. Ecosystem experiments // Science. V. 269. P. 324-327.

*Cochran W.G.,*1980. Fisher and the analysis of variance // Fienberg E., Hinckley D.V., eds. R.A. Fisher: an appreciation (Lecture Notes in Statistics, V. 1). New York: Springer. P. 17-34.

Cottenie K., De Meester L., 2003. Comment to Oksanen (2001): reconciling Oksanen (2001) and Hurlbert (1984) // Oikos. V. 100. P. 394-396.

Cox D.R., 1958. Planning of experiments. New York: Wiley. vii + 308 p.

Fisher R.A., Mackenzie W.A., 1923. Studies in crop variation. II. The manurial response of different potato varieties // J. Agric. Sci. V. 13. P. 311-320.

Frost T.M., DeAngelis D.L., Allen T.F.H., Bartell S.M., Hall D.J., Hurlbert S.H., 1988. Scale in the design and interpretation of aquatic community research // Carpenter S.R., ed. Complex interactions in lake communities. New York: Springer. P. 256-282.

García-Berthou E., Hurlbert S.H., 1999. Pseudoreplication in hermit crab shell selection experiments: comment to Wilber // Bull. Marine Sci. V. 65. P. 893-895.

Hairston N.G. Sr, 1989. Ecological experiments: Purpose, design, and execution. Cambridge: Cambridge Univ. Press. 370 p.

Heffner R.A., Butler M.J.IV, Reilly C.K., 1996. Pseudoreplication revisited // Ecology. V. 77. P. 2558-2562.

Hurlbert S.H., 1984. Pseudoreplication and the design of ecological field experiments // Ecol. Monogr. V. 54. P. 187-211.

Hurlbert S.H., 1990. Pastor binocularis: now we have no excuse [review of *Design of Experiments* by R. Mead] // Ecology. V. 71. P. 1222-1223.

Hurlbert S.H., 1994. Old shibboleths and new syntheses [review of *Design and Analysis of Ecological Experiments*, ed. by S.M. Scheiner and J. Gurevitch] // Trends Ecol. Evol. V. 9. P. 495-496.

Hurlbert S.H., 1997. Experiments in ecology [Review of book by same title by A.J. Underwood] // Endeavour. V. 21. P. 172-173.

Hurlbert S.H., 2004. On misinterpretations of pseudoreplication and related matters: A reply to Oksanen // Oikos. V. 104. P. 591-597.

Hurlbert S.H., Lombardi C.M., 2003. Design and analysis: uncertain intent, uncertain result [review of *Experimental design and data analysis for biologists* by G. Quinn and M. Keough] // Ecology. V. 83. P. 810-812.

Hurlbert S.H., Lombardi C.M., 2004. Research methodology: experimental design sampling design, statistical analysis // Bekoff, M.M. ed. Encylopedia of Animal Behavior, V. 2. London: Greenwood Press. P. 755-762.

Hurlbert S.H., Meikle W.G., 2003. Pseudoreplication, fungi, and locusts // J. Econ. Entomol. V. 96. P. 533-535.

Hurlbert S.H., White M.D., 1993. Experiments with freshwater invertebrate zooplanktivores: Quality of statistical analyses // Bull. Marine Sci. V. 53. P. 128-153.

James F.C., McCulloch C.E., 1985. Data analysis and the design of experiments in ornithology // Current Ornithology. V. 2. P. 1-63.

Jenkins S.H., 2002. Data pooling and type I errors: a comment on Leger & Didrichson // Animal Behaviour. V. 63. P. F9-F11

Kempthorne O., 1952, 1979. The design and analysis of experiments, orig. & rev. edns. New York: Wiley; Huntington: Krieger. xix + 631 p.

Kirk R.E., 1982. Experimental design, 2nd ed. Brooks/Cole Publishing Company, Pacific Grove, California. xi + 911 p.

Kozlov M.V., 2003a. Pseudoreplication in ecological research; the problem overlooked by Russian scientists// Zhurnal Obstchei Biologii [Journal of Fundamental Biology]. V. 64. P. 292-307 (in Russian, with English summary).

Kozlov M. V. , 2003b. Pseudoreplication in Russian ecological research // Bulletin of the Ecological Society of America. V. 84. P. 45-47.

Krebs C.J., 1999. Ecological methodology, 2d edn. New York: Addison-Wesley Longman. x + 620 p.

Kroodsma D.E., 1989. Suggested experimental designs for song playbacks // Animal Behavior. V. 37. P. 600-609.

Kroodsma D.E., Byers B.E., Goodale E., Johnson S., Liu W.-C., 2001. Pseudoreplication in playback experiments, revisited a decade later // Animal Behaviour. V. 61. P. 1029-1033.

Lombardi C.M., Hurlbert S.H., 1996. Sunfish cognition and pseudoreplication // Animal Behaviour. V. 52. P. 419-422

Machlis L., Dodd P.W.D., Fentress J.C., 1985. The pooling fallacy: problems arising when individuals contribute more than one observation to the data set // Z. Tierpsychol. V. 68. P. 201-214.

Mead R., 1988. The design of experiments. Cambridge: Cambridge Univ. Press. xiv + 620 p.

Mead R., 2003. [review of *Experimental design and data analysis for biologists* by G. Quinn and M. Keough] // Biometrics. V. 59. P. 738-739.

Mead R., Curnow R.N., 1983. Statistical methods in agriculture and experimental biology. New York: Chapman and Hall. xi + 335 p.

Millar R.B., Anderson M.J., 2004. Remedies for pseudoreplication // Fisheries Research. V. 70. P. 397-407.

Milliken G.A., Johnson D.E., 1989. Analysis of messy data, vol. 2: Nonreplicated experiments. New York: Van Nostrand Reinhold. viii + 199 p.

Morrison D.A., Morris E.C., 2000. Pseudoreplication in experimental designs for the manipulation of seed germination treatments // Austral. Ecol. V. 25. P. 292-296.

Ramirez C.C., Fuentes C.E., Rodriguez L.C., Niemeyer H.M., 2000. Pseudoreplication and its frequency in olfactometric laboratory studies // J. Chem. Ecol. V. 26. P. 1423-1431.

Riley J., Edwards P., 1998. Statistical aspects of aquaculture research: Pond variability and pseudoreplication // Aquaculture Res. V. 29. P. 281-288.

Rudneva I.I., Zherko N.V., 2000. Effect of polychlorinated biphenyls on the antioxidant system and lipid peroxidation in gonads of the Black Sea scorpionfish *Scorpaena porcus* L. // Ekologija [Russian Journal of Ecology]. V. 31. P. 70-73 (in Russian, with English summary).

Ruxton G.D., Colegrave N., 2003. Experimental design for the life sciences. Oxford: Oxford Univ. Press. xviii + 114 p.

Schindler D.W., 1998. Replication versus realism: the need for ecosystem-scale experiments // Ecosystems. V. 1. P. 323-334.

Schindler D.W., Mills K.H., Malley D.F., Findlay D.L., Shearer J.A., Davies I.J., Turner M.A., Linsey G.A., Cruikhank D.R., 1985. Long-term ecosystem stress: the effects of years of experimental acidification on a small lake // Science. V. 228. P. 1395-1401.

Smirnov K.A., 2001. Impact of moose on regrowth in spruce forests of southern taiga // Lesovedenie [Forestry]. V. 0(2). P. 46-52 (in Russian).

Sokal R.R., Rohlf F.J., 1969, 1981, 1995. Biometry: The principles and practice of statistics in biological research, 1st, 2d, 3rd edns. New York: W. H. Freeman and Co. xxi + 776, xviii + 859, xix + 887 p.

Steel R.G.D., Torrie J.H., 1980. Principles and procedures of statistics, 2nd edn. New York: McGraw-Hill. xxi + 633 p.

Tatarnikov D.V., 2005. On methodological aspects of ecological experiments (comments on M.V. Kozlov publication) // Zhurnal Obstchei Biologii [Journal of Fundamental Biology]. V. 66. P. 90-93 (in Russian, with English summary).

Thomson K.S., 1984. The literature of science // Amer. Sci. V. 72. P. 185-187.

Underwood A.J., 1981. Techniques of analysis of variance in experimental marine biology and ecology // Ann. Rev. Oceanogr. Marine Biol. V. 19. P. 513-605.

Underwood A.J., 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances // Ecol. Appl. V. 4. P. 3-15.

Underwood A. J., 1997. Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge: Cambridge Univ. Press. xviii + 504 p.

Urquhart N.S., 1981. The anatomy of a study // HortScience. V. 16. P. 621-627.

Whiting-O'Keefe Q.E., Henke C., Simborg D.W., 1984. Choosing the correct unit of analysis in medical care experiments // Medical Care. V. 22. P. 1101-1114.

Wise D.H., 1993. Spiders in ecological webs. Cambridge: Cambridge Univ. Press. ix + 328 p.