environmental microbiology

Environmental Microbiology (2011) **13**(6), 1383–1386 doi:10.1111/j.1462-2920.2011.02445.x

Opinion

Replication, lies and lesser-known truths regarding experimental design in environmental microbiology

Jay T. Lennon[1,2]*

[1]*W.K. Kellogg Biological Station and* [2]*Department of Microbiology and Molecular Genetics, Michigan State University, 3700 East Gull Lake Drive, Hickory Corners, MI 49060, USA.*

Summary

A recent analysis revealed that most environmental microbiologists neglect replication in their science (Prosser, 2010). Of all peer-reviewed papers published during 2009 in the field's leading journals, slightly more than 70% lacked replication when it came to analyzing microbial community data. The paucity of replication is viewed as an 'endemic' and 'embarrassing' problem that amounts to 'bad science', or worse yet, as the title suggests, lying (Prosser, 2010). Although replication is an important component of experimental design, it is possible to do good science without replication. There are various quantitative techniques – some old, some new – that, when used properly, will allow environmental microbiologists to make strong statistical conclusions from experimental and comparative data. Here, I provide examples where unreplicated data can be used to test hypotheses and yield novel information in a statistically robust manner.

Introduction

The quality and impact of science is contingent upon good experimental design. When designing a study, scientists must consider issues such as randomization, block effects and controls. In addition, careful decisions need to be made about how to allocate experimental units to achieve an appropriate level of replication. Replication is the fundamental way by which we quantify random and systematic variation in a study system. In many instances, it is the partitioning of this variance that allows us to make inferences about the outcomes of our experiments with some level of certainty. Therefore, when studies are conducted without replication, there is risk of drawing weak or invalid conclusions. Nevertheless, environmental microbiologists can deliberately design or fortuitously encounter unreplicated data sets. If handled with the proper statistical procedures, some unreplicated data can be used to test hypotheses and generate novel insight into microbial processes.

Received 9 December, 2010; accepted 22 January, 2011. *For correspondence. E-mail lennonja@msu.edu; Tel. (+1) 269 671 2340; Fax (+1) 269 671 2104.

Unreplicated regression

Let us revisit the hypothetical example of the undergraduate student sampling lake bacteria (Prosser, 2010). The student observed more bacteria in one lake than in a second lake. However, he only obtained a single observation from each lake. With such limited sampling, no conclusions can be drawn about bacterial abundances in these two lakes. The supervisor recommends that the student collect additional replicate samples from the two lakes. Because it is relatively easy to enumerate cell densities, let us assume the student obtains 10 replicate samples from different locations in each lake ($n = 20$). This will capture some of the spatial heterogeneity within the lakes and provide ample statistical power to test the null hypothesis that bacterial densities are the same in both lakes. The student analyses the resulting data with a Student's *t*-test and concludes that, in fact, bacterial abundance is statistically greater in the first lake than the second lake.

But now what? Let us assume the student and supervisor have access to basic water chemistry data and find that total phosphorus (TP) is much higher in the first lake than in the second lake. This would provide a convenient explanation for the observed differences between the lakes since freshwater bacteria are often limited by phosphorus availability. With these data alone, however, it is impossible to conclude what factors are responsible for the differences in bacterial abundance between the two lakes. The student could have designed his study slightly differently, and with the same number of observations (and no within-lake replication), obtained more information about controls on bacterial abundance. Using a comparative approach (see Gasol and Duarte, 2000), he

could have measured bacterial abundance one time from 20 different lakes and tested whether there was a positive statistical relationship with TP using regression analysis.

Regression is a flexible and powerful statistical approach that can be used in comparative or experimental studies. Importantly, it does not require replication. For those who have been indoctrinated to replicate, it can seem a bit reckless, or even wrong to deliberately design a study that relies on unreplicated regression. Some people view regression as an estimation procedure, not a hypothesis-testing tool, but there is no theoretical or mathematical basis for this view (Cottingham *et al*., 2005). Unreplicated regression is sometimes perceived as risky, too. What happens if, for unexplained reasons, one of your experimental units does not 'behave' well (i.e. it is an outlier)? Won't this sink your entire experiment? Not necessarily. Unreplicated regression designs are often more powerful than replicated designs, such as analysis of variance (ANOVA). Although they share the same underlying mathematical framework and similar types of assumptions, the matrix of independent variables in ANOVA has more parameters than a regression design given the same number of observations (Cottingham *et al*., 2005). Thus, all else being equal, a researcher is more likely to reject the null hypothesis when it is false (i.e. a lower type II error rate) using unreplicated regression than with replicated ANOVA. Furthermore, regression designs provide extra information. Specifically, the parameter values (i.e. intercepts and slopes) from regression analyses can be used for making predictions about how a variable of interest (e.g. bacterial abundance) will respond to changing conditions (e.g. increased TP). These parameters also allow researchers to detect non-linearities and thresholds along environmental gradients, and can be valuable for the development of simulation models used to explore the behaviour of more complex systems (Gotelli and Ellison, 2004). Last, there are philosophical and historical views that influence how scientists design experiments. For example, in the field of ecology during the early 1980s, scientists grew less accepting of the descriptive approaches that were common at that time (Strong *et al*., 1984). As a result, more attention was devoted to the rigorous design of experiments, the vast majority of which were analysed using ANOVA (e.g. Hurlbert, 1984; Werner, 1998). Arguably, the discipline of ecology matured during this transition period, but it has also been argued that the ANOVA mindset can be restrictive and act as a 'mental straightjacket' that limits our ability to conceptualize processes and interactions (Werner, 1998).

Unreplicated time series

Often in environmental microbiology, we quantify how populations and communities change through time to better understand processes such as succession or recovery from perturbations. For a number of reasons, replication tends to be less common in studies that focus on temporal dynamics. First, because of time or monetary constraints, environmental microbiologists may design studies with a limited number of experimental units (see excuse *i* in Prosser, 2010). This results in an inherit trade-off between the degree of (spatial) replication and the resolution of temporal sampling. Second, the benefits of replication are not always obvious in certain situations. For example, in whole ecosystem experiments (e.g. phosphorus enrichment of a lake), it can be challenging to identify appropriate 'controls' (Carpenter, 1990; Schindler, 1998), and rarely are treatments of this scale replicated. Lastly, long-term observational studies can become interrupted by unplanned events (e.g. oil spill). The resulting 'natural' experiments provide unique opportunities for scientific inquiry, but logistical or ethical issues may constrain the design and analysis of the study (Miao and Carstenn, 2006).

Fortunately, there is a suite of statistical techniques that can be used to make robust inferences about unreplicated time-series data (e.g. Diggle, 1990; Pole *et al*., 1994; Bence, 1995). Occasionally, these techniques are grounded in different statistical philosophies (Bayesian versus Frequentist), but in general, they rely on approaches that go beyond what is taught to graduate students in a traditional microbiology programme. Take for example a study that examined how the variability and predictability of microbial dynamics were affected by nutrient enrichment. Lake 227 in north-western Ontario, Canada is a famous lake that has been the focus of whole-ecosystem manipulations since 1969. Cottingham and colleagues (2000) examined a time series of fossil pigments collected from sediment varves during unperturbed (1944–1965) and fertilized (1969–1990) time periods. Microbial community composition was reconstructed annually yielding a unique 40-year time series. However, because these observations were temporally autocorrelated, they could not technically be treated as replicates. The researchers overcame this hurdle using a Bayesian technique called dynamic linear modelling (DLM), which explicitly deals with the non-independence of time-series data (Pole *et al*., 1994). Results from the DLM analysis revealed that that the forecast uncertainty of microeukaryotic and bacterial phototrophs increased when Lake 227 underwent nutrient enrichment. Despite the lack of replication in the time series, this experiment uniquely demonstrated that eutrophication can decrease the predictability of community and ecosystem dynamics. Less traditional techniques are also being applied to unreplicated time-series data to understand microbial species interactions in laboratory-scale bioreactors (Trosvik *et al*., 2008).

Replication and inference space

In the preceding sections, I provided examples of quantitative procedures that allow scientists to analyse some unreplicated data sets without violating any statistical 'rules'. Nevertheless, the lack of replication seriously impinges upon the inference space of a study. Let us return to the undergraduate student who now wants to characterize bacterial composition via deep sequencing of the 16S rRNA genes using a single sample from each of the two lakes. Once again, even though there is no replication in his study, there are approaches that would permit the student to statistically compare the two samples. Randomization procedures, including Monte Carlo simulations and other permutation-based algorithms, provide researchers with the opportunity to resample their data and test whether multivariate sequences obtained from non-replicated samples came from the same statistical population (e.g. Solow, 1993; Schloss *et al.*, 2004). Using one of these techniques, the student finds that the two samples are statistically distinct. However, the inference from his analysis cannot be extended beyond the two samples that were analysed. In other words, the student must stop short of drawing conclusions about differences in the composition of bacteria between the two lakes. How can this problem be remedied? As Prosser (2010) points out, it costs about the same amount of money to generate 90 000 sequences from one sample as it costs to generate 30 000 sequences from three replicate samples. Therefore, the student would have gained more inference about the potential differences in bacterial composition between the lakes by allocating effort towards replication rather than sequencing depth.

Unreplicated results ≠ lies, just as replicated results ≠ truth

More often than not, replication is a critical ingredient in a well-executed and influential experiment. Therefore, a lack of replication in the field of environmental microbiology is alarming and may reflect, in part, inadequate training in biostatistics (Prosser, 2010). A *traditional* course in biostatistics should emphasize the importance of replication. A *progressive* course in biostatistics, however, should convey that there are some situations where replication is either not feasible or not necessary, depending on the question and system (see Ellison and Dennis, 2010). Ultimately, environmental microbiologist must appreciate that replication determines a study's inference space, but that it also provides a relatively straightforward way to accept or reject hypotheses, which is an essential tool for testing theory (Prosser *et al.*, 2007). It is also important that environmental microbiologists understand that there are quantitative approaches for analysing unreplicated data sets; when applied appropriately, these techniques offer flexibility, yield robust conclusions and may generate novel insight into the ecology of microbial systems.

Acknowledgements

I thank K.C. Bird, J.M. Palange and three anonymous reviewers for helpful comments on an earlier version of the manuscript. Long discussions with K.L. Cottingham and B.L. Brown have shaped my opinions on experimental design. This is Kellogg Biological Station contribution # 1583.

References

Bence, J.R. (1995) Analysis of short time-series: correcting for autocorrelation. *Ecology* **76:** 628–639.

Carpenter, S.R. (1990) Large scale perturbation: opportunities for innovation. *Ecology* **71:** 2038–2043.

Cottingham, K.L., Rusak, J.A., and Leavitt, P.R. (2000) Increased ecosystem variability and reduced predictability following fertilisation: evidence from palaeolimnology. *Ecol Lett* **3:** 340–348.

Cottingham, K.L., Lennon, J.T., and Brown, B.L. (2005) Knowing when to draw the line: designing more informative ecological experiments. *Front Ecol Environ* **3:** 145–152.

Diggle, P.J. (1990) *Time Series: A Biostatistical Introduction*. Oxford, UK: Clarendon Press.

Ellison, A.M., and Dennis, B. (2010) Paths to statistical fluency for ecologists. *Front Ecol Environ* **8:** 362–370.

Gasol, J.M., and Duarte, C.M. (2000) Comparative analyses in aquatic microbial ecology: how far do they go? *FEMS Microbiol Ecol* **31:** 99–106.

Gotelli, N.J., and Ellison, A.M. (2004) *A Primer of Ecological Statistics*. Sunderland, UK: Sinauer.

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

Miao, S.L., and Carstenn, S. (2006) A new direction for large-scale experimental design and analysis. *Front Ecol Environ* **4:** 227.

Pole, A., West, A.E., and Harrison, J. (1994) *Applied Bayesian Forecasting and Time Series Analysis*. New York, USA: Champan & Hall.

Prosser, J.I. (2010) Replicate or lie. *Environ Microbiol* **12:** 1806–1810.

Prosser, J.I., Bohannan, B.J.M., Curtis, T.P., Ellis, R.J., Firestone, M.K., Freckleton, R.P., *et al.* (2007) The role of ecological theory in microbial ecology. *Nat Rev Microbiol* **5:** 384–392.

Schindler, D.W. (1998) Replication versus realism: the need for ecosystem-scale experiments. *Ecosystems* **1:** 323–334.

Schloss, P.D., Larget, B.R., and Handelsman, J. (2004) Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl Environ Microbiol* **70:** 5485–5492.

Solow, A.R. (1993) A simple test for change in community structure. *J Anim Ecol* **62:** 191–193.

Strong, D.R., Simberloff, D., Abele, L.G., and Thistle, A.B. (1984) *Ecological Communities: Conceptual Issues and Evidence*. Princeton, NJ, USA: Princeton University Press.

Trosvik, P., Rudi, K., Naes, T., Kohler, A., Chan, K.S., Jakobsen, K.S., and Stenseth, N.C. (2008) Characterizing mixed microbial population dynamics using time-series analysis. *ISME J* **2:** 707–715.

Werner, E.E. (1998) Ecological experiments and a research program in community ecology. In *Experimental Ecology: Issues and Perspectives*. Resetarits, W.J., and Bernardo, J. (eds). New York, USA: Oxford Press, pp. 3–26.