

Reproducibility and Repeatability in Ecology

Authors: CASSEY, PHILLIP, and BLACKBURN, TIM M.

Source: BioScience, 56(12) : 958-959

Published By: American Institute of Biological Sciences

URL: [https://doi.org/10.1641/0006-3568\(2006\)56\[958:RARIE\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2006)56[958:RARIE]2.0.CO;2)

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Reproducibility and Repeatability in Ecology

PHILLIP CASSEY AND TIM M. BLACKBURN

The quantitative synthesis of research results is of fundamental importance in seeking to develop ecological generalities and construct refutable theories. It is thus critical that published studies contain sufficient detail to allow their methods to be replicated and their results compared. In response to this need, growing attention is being paid to the publication and presentation of analytical results in ecology. Our recent experience has been that journal referees (and editors) increasingly express the opinion that results need to be accompanied by general access to the primary data on which they are based. Here we argue that the legitimate aims of formal scientific inquiry (including the publication and validation of results) do not need to infringe the intellectual property rights of publishing ecologists.

Repeatability of methods and analysis

It is widely agreed that modern scientific inference relies on the vulnerability to refutation of its general theories, which have the characteristic quality of being both testable and falsifiable (Popper 1968). In a scientific discipline such as ecology, the search for general rules and laws is greatly hampered by a high degree of historically based, context-specific contingency. Progress toward such principles will thus be best served by the ability to repeat potentially important discoveries across different ecological systems (Gurevitch et al. 2001, Koricheva 2003).

It follows from this that a repeatable study must satisfy two basic criteria. First, from the information presented, a third party must be able to perform a study using identical methodological protocols and analyze the resulting data in an identical manner. Note that it is not

necessary for the third party to obtain the same result as the original study: It is this that gives the first clues about generality. Second, the value of ecological synthesis lies in the comparability of the results of individual studies. This is not the case for published results that provide limited or no information regarding their statistical summaries, such as estimates of the size of an effect, required for comparing analytical outcomes. A published result must be presented in a manner that allows for a quantitative comparison in a later study, or it cannot be classified as repeatable. Repeatability is clearly important for the development of any field of research, and we believe it is the basic requirement for the advancement of ecological research.

Reproducibility of results

It should be apparent from the discussion above that we do not consider the publication of raw data to affect the repeatability of a study. However, the inclusion of raw data does make a study reproducible. Following Schwab and colleagues (2000), we consider a study reproducible if, from the information presented in the study, a third party could replicate the reported results identically. It is important to note that a study that is reproducible is not necessarily repeatable: Raw data and analytical methods may be presented without adequate indication of how those data were obtained.

It has been argued that a scientific study, to be acceptable for publication, should be reproducible (NRC 2003), and indeed our recent experiences have led us to believe that reproducibility is increasingly being requested by journals. This raises the question of why reproducibility might be thought desirable. Three reasons come to mind. First, it may be useful to be able to replicate exactly the

results of any given paper in extending or attempting to falsify those results. Second, reproducibility might be desirable because it protects against data loss and human error. Third, reproducibility would go some way toward protecting against deliberate fraud. Nevertheless, it is no guarantee, as anyone unscrupulous enough to fabricate analytical results is also likely to be unscrupulous enough to fabricate data.

Reproducibility versus repeatability

It is clear that there are reasons why reproducibility may be a desirable feature of scientific research. It is also clear that repeatability and reproducibility are distinct concepts. Our recent experience, however, is that the distinction between them is being increasingly blurred (or simply confused) by journal editors and referees—for example, “We are concerned that not making the original data publicly available detracts from the strength of the paper because others cannot accurately assess the support for your conclusions and because of the inherent value to the community of such a dataset” (anonymous manuscript review, 2004). We believe that this growing attitude of journal editors and referees has significant implications for all scientists who attempt to publish their research, and for their careers. Our own view is that reproducibility is less important for the advancement of ecology than is repeatability.

Phillip Cassey (e-mail: p.cassey@bham.ac.uk)
and Tim M. Blackburn (e-mail:
t.blackburn@bham.ac.uk) work in the School of
Biosciences, University of Birmingham,
Edgbaston, Birmingham B15 2TT,
United Kingdom.

We strongly advocate the honest publication of repeatable methods and comparable summary statistics, but we do not believe that this means that the publication of raw data is a scientific requisite for ecological studies. In the simplest terms, the development of general ecological theories from quantitative synthesis of results relies on the publication of analyses from independent data sets. Because reiterating a result from analyses that have already been published does not aid in this synthesis, the principal novel purpose that the publication of raw data serves is to allow other unaffiliated researchers to use these data for novel research. While that may be of considerable use, we suggest that, unlike repeatability, it is not a legitimate criterion for the publication of ecological research.

We do not think these views on repeatability versus reproducibility are specific to any one branch of ecology, or indeed biology: They apply whether results derive from field experiments, laboratory experiments, natural experiments, or comparative data. No single type of analysis produces results that are inherently more likely to require extension or falsification, to be subject to fraud, or to be lost. We can see why comparative data might be considered to be of more general use to other studies in ecology, but perceived utility to other researchers is not a valid scientific reason to demand the publication of raw data in some cases but not in others.

The future

Should reproducibility be required of ecological studies? If that is the route that ecology is to travel, then we think that at least three points need to be noted.

First, reproducibility cannot be a piecemeal requirement. We need to move on

from the current situation, in which reproducibility is requested of some papers but not of others.

Second, we do not think that journals are ready for the task of general custodianship of all the data that consistent reproducibility would force them to accumulate. What is needed is a universal protocol or framework for storing and checking data (Arzberger et al. 2004). This protocol must be developed with the rights of authors, universities or other employers, and funding bodies in mind, along with the recognition that information from papers informs the work of subsequent generations of scientists. The system should allow the free flow of this information within a framework that rewards individuals and institutions for their efforts in generating it. Examples of data protocols that address some of these issues already exist. Similarly, within a variety of disciplines, Internet storage facilities are increasingly available with the specific goals of enabling the collection, access, and sharing of historical and real-time data. We suggest that more forums are required to continue building ideas about the custodianship of data into research standards that take into account the varying interests of all parties.

Finally, and of paramount importance, reproducibility should not be at the expense of the rights of the authors. When reproducibility is made an explicit requirement of publication by a journal, with increasing competition for journal space used as leverage, authors must in effect waive all rights of ownership over the data they are required to publish. In our view, being required to give away those hard-won data for no return is not justifiable and has the potential to significantly hinder scientists'

careers, especially at their outset. In contrast, the ability to retain possession of information has the potential to spawn collaborations between the owner and other scientists on topics that the data owner may or may not have addressed otherwise. This provides those scientists who have invested their time and resources in collating important data sets with rewards within the current publish-or-perish system, by which the success of modern scientific careers is to a large degree judged. If reproducibility is to be widely incorporated into future ecological research, then it must be done consistently. Most important, it cannot be imposed by journals in a manner that infringes on the intellectual rights of publishing researchers.

References cited

- Arzberger P, Schroeder P, Beaulieu A, Bowker G, Casey K, Laaksonen L, Moorman D, Uhler P, Wouters P. 2004. An international framework to promote access to data. *Science* 303: 1777–1778.
- Gurevitch J, Curtis PS, Jones MH. 2001. Meta-analysis in ecology. *Advances in Ecological Research* 32: 199–247.
- Koricheva J. 2003. Non-significant results in ecology: A burden or a blessing in disguise? *Oikos* 102: 397–401.
- [NRC] National Research Council. 2003. *Sharing Publication-related Data and Materials: Responsibilities of Authorship in the Life Sciences*. Washington (DC): National Academies Press, Committee on Responsibilities of Authorship in the Life Sciences.
- Popper KR. 1968. *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Harper Torch Books.
- Schwab M, Karrenbach N, Claerbout J. 2000. Making scientific computations reproducible. *Computing in Science and Engineering* 2: 61–67.