COMMENT

Genuine replication and pseudoreplication

Stanley E. Lazic

A large sample size, or *N*, increases the sensitivity of an experiment to detect differences between treatment groups. However, the biological entity that *N* refers to may not be obvious. Defining the wrong entity can inflate the sample size and increase both false-positive and false-negative results.

The sample size is an important design consideration when planning an experiment, and journals often require the number to be justified and reported[1–3]. Researchers are asked to describe what was replicated, and a distinction between biological and technical replicates is often made. Biological replication is meant to reflect the sample size, whereas technical replication does not. However, these two terms are used inconsistently, do not capture the important characteristics of an experiment and are therefore insufficient[4].

Three biological entities or units have been proposed to better define where replication should occur[4,5]. The first is the scientific or biological unit (such as a person, animal or cell), which is the target of inference; the purpose of an experiment is to draw a conclusion about these units. The second is the experimental unit (also known as the unit of allocation or the unit of randomization), which is the entity that is randomly and independently assigned to an experimental condition. Finally, the observational unit (also known as the measurement unit) is the entity on which measurements are taken.

The design and analysis of an experiment is simple when the biological, experimental and observational units correspond to the same biological entity, but when the units refer to different entities, knowing what to replicate can be difficult. For example, an experimenter randomizes pregnant female rodents (experimental units) to a control or drug condition, and they are interested in the drug's effect on the offspring after they are born (biological units), specifically, the number of dendritic spines on neurons (observational units). The key idea is that the sample size is the number of experimental units — the number of pregnant dams — not the number of offspring or the number of neurons. The sample size will be greatly inflated if *N* is taken as the number of offspring or neurons, as these constitute pseudoreplication.

Replicating the biological unit

An experiment cannot test the hypothesis of interest unless the scientific or biological unit is replicated. For example, suppose that a difference between inbred and outbred strains of mice on a cognitive task is expected, and therefore the hypothesis is about strains and not individual mice, which are the observational units. Hence, taking ten inbred (for example, C57BL/6) and ten outbred (for example, ICR) mice will not provide a valid test of the hypothesis as there are only two strains ($N=2$). Having multiple mice from each strain is beneficial because variation between mice within a strain may be high, but with this design it is impossible to determine whether differences are between these two strains or a property of these strains (inbred versus outbred status). Hence, multiple inbred and outbred strains are required.

Requirements for genuine replication

Even if the biological entity of interest is replicated, it may be insufficient for valid statistical inference. Genuine replication requires that three conditions are met. First, the units must be independently and randomly assigned to treatment groups. This requirement defines the experimental unit and forms the basis of a statistical analysis. At times, it is acceptable to use nature's randomization, for example when comparing males versus females or F1 animals with different genotypes.

Second, the treatments should be independently applied to each experimental unit after randomization. For example, cells may have been randomly placed into the wells of a microtitre plate by the pipetting process, but then treatments (such as compounds) are applied to all cells simultaneously within each well, resulting in correlated treatment error between cells in the same well[4]. Hence, cells are unsuitable as experimental units.

Finally, the experimental units should not influence each other. Cells in the same well may influence each other via cell-to-cell connections or by competing for the same nutrients in the media. Cells undergoing necrotic death may release damage-associated molecular patterns that can affect neighboring cells. Even if cells are randomly assigned to wells by the process of pipetting, it is unrealistic to assume that they provide independent information about a treatment effect and it is impossible to prove that cells in a well are not influencing each other. It is therefore necessary to use the well as the experimental unit even if measurements are taken on individual

Prioris.ai Inc., Ottawa, Ontario, Canada.

e-mail: stan.lazic@cantab.net

https://doi.org/10.1038/s43586-022-00114-w

cells. Similarly, animals sharing a cage may influence each other on the relevant outcome measures, in which case cages are the appropriate experimental units.

Experimental and biological units can be related in several ways, which makes the experimental unit harder to identify. An experimental unit may correspond to a biological unit of interest, groups of biological units, parts of a biological unit or a sequence of observations on a biological unit. Distinguishing between genuine replication and pseudoreplication can be challenging, and a longer discussion, detailed examples and diagrams to visualize these relationships can be found elsewhere[4,5].

Summary

When planning an experiment, clearly define the biological, experimental and observational units. Ensure that the biological unit is replicated for valid inferences. Ensure that the experimental units are independently assigned to treatment groups, treatments are applied independently and that the experimental units do not influence each other. Once the appropriate experimental unit is determined, ensure that there are enough of them for the study to have adequate statistical power or otherwise meet the experimental objectives.

1. Landis, S. C. et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* **490**, 187–191 (2012).
2. Percie du Sert, N. et al. The ARRIVE guidelines 2.0: updated guidelines for reporting animal research. *PLoS Biol.* **18**, e3000410 (2020).
3. Percie du Sert, N. et al. Reporting animal research: explanation and elaboration for the ARRIVE guidelines 2.0. *PLoS Biol.* **18**, e3000411 (2020).
4. Lazic, S. E. *Experimental Design for Laboratory Biologists: Maximising Information and Improving Reproducibility* (Cambridge University Press, 2016).
5. Lazic, S. E., Clarke-Williams, C. J. & Munafò, M. R. What exactly is '*N*' in cell culture and animal experiments? *PLoS Biol.* **16**, e2005282 (2018).

Competing interests
The author declares no competing interests.