SAGE Reference

The SAGE Encyclopedia of Research Design

Biological and Technical Replicates

By: Jay Hegdé
Edited by: Bruce B. Frey
Book Title: The SAGE Encyclopedia of Research Design
Chapter Title: "Biological and Technical Replicates"
Pub. Date: 2022
Access Date: July 3, 2022
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks,
Print ISBN: 9781071812129
Online ISBN: 9781071812082
DOI: https://dx.doi.org/10.4135/9781071812082.n56
Print pages: 128-130

Biological and technical replicates are two different approaches to making repeated measurements of an underlying biological phenomenon in biomedical research. It is generally a good practice to include both types of replicates in each experiment.

Broadly speaking, technical replicates are independently repeated measurements of the same sample using the same procedure. As such, these replicates represent independent measures of the noise (typically random) associated with protocols or equipment: They help measure the reproducibility of an assay and not the reproducibility of the underlying biological phenomenon. Biological replicates, on the other hand, are parallel measurements of biologically distinct samples. These replicates help capture the variation (random or otherwise) of the biological phenomenon under study and help measure its reproducibility. The distinction between technical and biological replicates is a functional one, in that it depends on which type of data variability—procedural or biological—they capture and not necessarily on how the replicates are obtained.

There is no one-size-fits-all formula for designing replicates that are optimal for a given experiment. The optimal design, including the optimal mix of technical and biological replicates in a given experiment, depends on the potential sources and magnitudes of variability in a given experiment and the questions that the experiment seeks to answer.

Historical Origins of the Replicate Nomenclature

Until the 1990s, much of the reproducibility testing in biomedical research, especially in the wet laboratory experimental sciences such as molecular and cellular biology, employed what would be considered technical replicates today. The push to systematically include biological replicates in experiments originated primarily in these fields in the 2000s with the widespread realization that biomedical research findings were not sufficiently reproducible, in large part because technical replicates by themselves did not properly account for the variability of the underlying biological phenomena. Major funding agencies in these fields, such as the U.S. National Institutes of Health, spurred the widespread adoption of the current replicate nomenclature and practices of replicate design by incentivizing researchers to employ both biological and technical replicates in their research as a way of enhancing the reproducibility of the research findings.

While the practice of quantitative measurements and statistical testing was much better established in many other fields of biomedical research, such as epidemiology, psychophysics, ecology and evolutionary biology, reproducibility of results was not necessarily commensurately better in these fields, arguably also because of poor replicate design.

Reproducibility Requires Representative Replicates

Research is primarily about learning general truths about the phenomenon under study. A set of findings is useful only to the extent that the same findings are obtained when the given experiment is independently but precisely repeated. The term reproducibility typically means this type of across-experiment reproducibility of the findings or conclusions and not of the measurements on which the conclusions are based. One accepts the mathematical reality that the measurements themselves will not be exactly reproducible from one instance to the next, be it within or across experiments, even as one expects the conclusions to be reproducible.

The only way one can draw reproducible conclusions based on inherently variable measurements is to use sound practices of statistical sampling and, where necessary, statistical testing. To the extent that the empirical measurements are truly representative of the underlying phenomenon, one can have a quantifiable degree of confidence, say 95%, that the conclusions will be reproduced when the experiment is exactly repeated. Thus, the key to obtaining reproducible results is to ensure that replicates as a group adequately represent the relevant statistical properties of the phenomenon of interest. This, in a nutshell, is the goal of replicate design: to ensure that the replicates are representative and that they adequately capture the study-relevant statistical properties, including the variability, of the phenomenon.

In biology, the substrates of phenomena of interest tend to be highly variable. Therefore, ensuring that this

biological variability is properly represented in the empirical measurements of a given phenomenon is a necessary and proper way of improving the reproducibility of the findings about the phenomenon. In other words, it is usually a good idea to include biological replicates in an experiment because the underlying biological substrates are usually variable. It follows from the elementary principles of statistics that the greater the variability of the relevant biological substrates, the larger the number of replicates needed to adequately capture this variability.

Basic Principles of Replicate Design

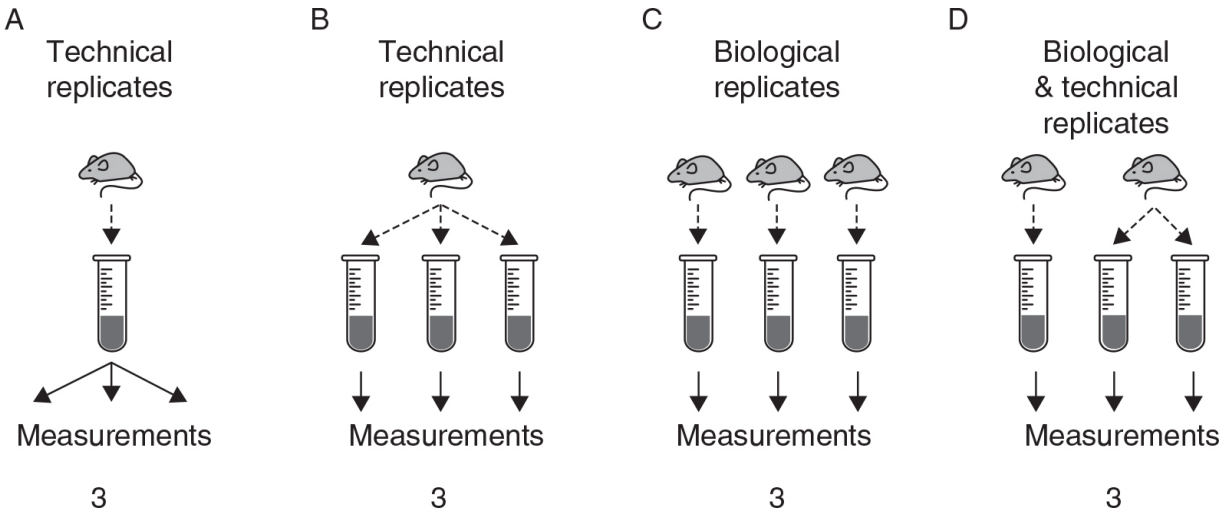
Consider a simple hypothetical experiment to determine the levels of a blood component called albumin in adult Spraque Dawley laboratory rats 24 hours after skin injury. We induce injury in a designated spot, say a hind thigh, using standard procedures in one rat. We draw a vial of blood from this rat 24 hours after the procedure and measure the albumin levels using standard procedures. We repeat this measurement twice more independently using the same vial of blood. These constitute three technical replicates because they measure the reproducibility of the albumin assay (Figure 1A). From the three replicates, we determine the mean and the standard deviation of albumin levels.

Note, however, that these findings apply only to the particular vial of blood. We have no way of evaluating whether the results are likely to be reproducible across additional blood draws from the same mouse because we have not tested any additional blood draws. Obviously, this is not a useful outcome and reflects poor replicate design. To make the results more generalizable, we make three mutually independent blood draws from this rat and measure the albumin levels in each (Figure 1B). Although the sample size remains the same at three, these technical replicates are better designed because the albumin level estimate is likely to be better reproducible for this mouse.

It is desirable to have our findings apply to all Spraque Dawley rats and not just to the one rat we tested. We therefore repeat the experiment using three different rats, making one measurement each (Figure 1C). These three biological replicates allow us to draw conclusions about the three rats in question. To the extent that these three rats are representative of all Spraque Dawley rats, the results should be reproducible across all rats of this strain when the experiment is exactly repeated.

In general, it is a good practice to include both technical and biological replicates (Figure 1D), since the two types of replicates measure different types of variability in the measurements, as noted earlier.

Figure 1 A Hypothetical Example That Illustrates the Distinction Between Technical and Biological Replicates



Note: In each case, rat blood is drawn and the concentration of albumin in the blood is measured. In each panel, dotted arrows at the top denote blood draws, and solid arrows at the bottom denote albumin measurements. (A) Repeated measurements from the same blood draw nominally represent technical replicates but reflect poor replicate design. (B) A better design for technical replicates is to make multiple independent blood draws and measure albumin level in each draw. (C) Biological replicates are those in which each blood draw is made from a different selected rat. (D) It is generally a good design practice to include both technical and biological replicates in an experiment and to adjust the relative proportions of the two types of replicates so that the replicates are representative of the phenomenon under study. Of the four replicate designs shown, the one in panel D is likely to yield the most reproducible results, even though the sample size (n = 3), and therefore the nominal statistical power, is the same in all cases. Note that statistically desirable values of n tend to be higher than 3 in actual experiments.

Additional Observations About Replicate Design

There are a few additional things we always wanted to know about replicate design but our mothers never told us. First, it ultimately does not matter whether a given replicate is designated a technical replicate or a biological replicate, as long as the sources of the replicates are faithfully kept track of. For instance, if we arbitrarily shuffle the replicate designations of one or more replicates in Figure 1, it will not in any way affect the experimental findings, our observations as to which aspects of the experiments the findings apply to, or the reproducibility of the findings. Indeed, the technical replicates in the earlier example meet some of the criteria of biological replicates, in that they capture some of the procedural variability with a biological basis, such as the variability in the induced injury across rats. In some cases, it can be difficult to decide whether a given replicate qualifies as either type of replicate or both. For instance, what constitutes a technical replicate in a study that seeks to determine whether home value appraisers estimate lower values for homes owned by African Americans than those owned by Caucasians? Moreover, there are vast areas of research where the phenomena of interest are not biological at all. For instance, what would constitute biological replicates in a study about the effect of ethanol on catalytic converters in cars? Yet, no one would dispute the importance of these research questions or of ensuring the reproducibility of the findings by designing the replicates properly. In thinking about these issues, it helps to keep in mind the aforementioned historical origins of the replicate nomenclature and that while reproducibility is desirable in all research, not all research involves biology or even experimentation.

Second, proper replicate design requires that the goals of the study and the planned data analyses be precisely specified beforehand. This is especially important when the underlying phenomena are complex, multivariate, and/or subject to dynamic change. For instance, in the case of the aforementioned rat experiment, we need to decide which aspects of the underlying phenomenon we want to draw conclusions about and how broadly we want to draw them: injury to which body regions, which types of the injury (e.g., abrasions, cuts, chemicals, or burns), which ages, and so forth. Specifying the research question has the effect of specifying which statistical properties are relevant to the study and which are not, thus making replicate design more tractable. Without specifying the study parameters in this fashion, we would risk either having to obtain an unmanageably large number of replicates to try and capture all potential statistical variability of the underlying substrates or designing poor replicates that fail to capture the study-relevant variations of the phenomenon, and thereby reducing the reproducibility of the research findings. Specifying the planned tests is necessary for, among other things, planning the number of various replicates. For instance, in the aforementioned rat experiment, we planned no statistical tests because the goal of this simple experiment was to simply estimate the albumin levels, not to test any hypotheses. For a more complex experiment, in which we test the hypothesis that albumin levels in rats with skin injury are higher than in control rats that underwent a sham procedure, we would need to obtain replicates from both the treatment group and control group of rats. The numbers of the replicates do not necessarily have to be the same between the two groups. For instance, sham injury may be quite consistent from one rat to the next, so that fewer technical replicates might suffice for the control group.

Third, replicate design is closely related to, but not the same as, sample size calculation. A common, recommended practice is to first perform power analyses based on the expected strength of the effect under study (estimated based on the best available information from published results or pilot data), planned statistical tests, and the desired level of statistical significance and statistical power. This will yield the required

total number of replicates (or sample size n). The n value can then be broken down into the desired numbers of technical versus biological replicates based on the estimated variability from various sources.

A fourth, related principle is that it is the responsibility of the researcher to report the replicate design, along with the rest of the study methods in sufficient detail as to enable other researchers to replicate the findings independently. Reporting a study poorly is tantamount to designing it poorly.

Finally, there are many cases in which a statistically optimal replicate design is not possible, not desirable, or both. For instance, in invasive studies of neural activity in monkey brains, it is typical to use only two monkeys because using additional monkeys is inadvisable without a compelling reason. Instead, researchers typically study a large number of individual neurons in either monkey, which results in a statistically suboptimal nested replicate design. One can nonetheless draw the best possible reproducible conclusions from such nested data using commonly available statistical tools. In cases such as this, the imperatives of sound statistical design must be balanced against other principles of sound research, and reproducibility must be maximized using the best available alternative methods.

See also Animal Research; Nested Sampling; Power Analysis; Replication; Sample Size

Jay Hegdé
http://dx.doi.org/10.4135/9781071812082.n56
10.4135/9781071812082.n56

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & van der Sluis, S. (2014). A solution to dependency: Using multilevel analysis to accommodate nested data. Nature Neuroscience, 17(4), 491–496. doi: http://dx.doi.org/10.1038/nn.3648.
Blainey, P., Krzywinski, M., & Altman, N. (2014). Points of significance: Replication. Nature Methods, 11(9), 879–880. doi: http://dx.doi.org/10.1038/nmeth.3091.
Collins, F. S., & Tabak, L. A. (2014). Policy: NIH plans to enhance reproducibility. Nature, 505(7485), 612–613. doi: http://dx.doi.org/10.1038/505612a.
Maddox, J. (1992). Is molecular biology yet a science? Nature, 355, 201. doi:http://dx.doi.org/10.1038/355201a0.
Naegle, K., Gough, N. R., & Yaffe, M. B. (2015). Criteria for biological reproducibility: What does “n” mean? Science Signaling, 8(371), fs7. doi: http://dx.doi.org/10.1126/scisignal.aab1125.
Vaux, D. L., Fidler, F., & Cumming, G. (2012). Replicates and repeats—What is the difference and is it significant? A brief discussion of statistics and experimental design. EMBO Reports, 13(4), 291–296. doi: http://dx.doi.org/10.1038/embor.2012.36.