

It is time for an empirically informed paradigm shift in animal research

S. Helene Richter and Vanessa von Kortzfleisch 

In their recent Perspective article (Voelkl et al. Reproducibility of animal research in light of biological variation. *Nat. Rev. Neurosci.* **21**, 384–393 (2020))¹, Voelkl et al. recommend the use of systematic heterogenization in animal studies. The rationale for this recommendation lies in the current best practice of rigorously standardizing (that is, homogenizing) conditions within experiments. This practice has been repeatedly criticized to limit experiments' inference to the specific experimental conditions used and hence to cause, rather than cure, poor reproducibility in animal experiments^{2,3}. Instead, it has been suggested to embrace biological variation and to use it actively as a tool for making study populations more representative and the results more meaningful and reproducible (for example, see REFS^{4,5}). We greatly appreciate the recommendations of Voelkl et al. and agree on the importance of a paradigm shift in animal research. However, we would like to draw attention to some points that, from our perspective, deserve more attention.

It is right that “direct evidence for the standardization fallacy is currently limited to simulations across replicate studies and only a few dedicated experimental studies”¹. However, we think it is even more critical to highlight the lack of studies going beyond the standardization fallacy, proving the benefits of systematic heterogenization. Hypothesis-driven comparisons of standardized and heterogenized designs are needed, demonstrating, for example, that the latter leads to the better reproducibility of treatment effects. Until now, only three empirical studies have adopted such an approach. A single-laboratory study showed that heterogenizing mouse study populations across two environmental factors improved the reproducibility of behavioural data³. However, in a multi-laboratory situation, the same approach did not yield similarly promising

results⁶. A third, ecological study investigated different strategies in microcosm experiments, showing that genetic heterogenization lowered between-laboratory variation⁷.

Another point we wish to highlight is the lack of practical solutions for how to heterogenize a study population in an effective and feasible way. Voelkl et al. wrote: “heterogenization may be based on controlled variation, for instance by systematically varying the genotype [...], the state and history of the individual [...], or the test condition [...]. Alternatively, heterogenization may be based on uncontrolled variation, for example by using outbred study populations, by splitting experiments into multiple independent batches of animals or by conducting multilaboratory studies.” Indeed, two recent simulation studies hinted towards better reproducibility in experiments that are either heterogenized across laboratories or the “time of day at which an experiment is conducted”^{8,9}. However, approaches involving heterogenizing conditions across housing conditions and age classes did not lead to the desired improvements⁶. Thus, it is probably far more difficult to address the practical issues than has been suggested. Moreover, we argue that the concept of heterogenization relies on the introduction of systematic and hence controlled variation. Introducing uncontrolled variation instead (for example, by using outbred strains) might bear the risk of inflating sample sizes, as it is hard to control for this variation in the experimental design or the statistical analysis¹⁰. We therefore strongly recommend the use of heterogenization factors that can be systematically varied and act as kind of ‘umbrella factors’ covering plenty of known and unknown background variables at the same time (for example, the experimenter^{11,12}).

Taken together, although the acceptance of the heterogenization concept has greatly

increased over the past decade (for example, see REF¹³), it is supported by only a few empirical studies. Proving the concept empirically and not just theoretically, however, appears particularly important to overcome existing lab traditions and to “establish systematic heterogenization of study populations as a new standard”¹. We therefore appeal to the scientific community not to remain at the conceptual level but, instead, to explore and validate novel strategies for putting this concept into practice.

There is a reply to this letter by Würbel, H. et al. *Nat. Rev. Neurosci.* <https://doi.org/10.1038/s41583-020-0370-7> (2020)

S. Helene Richter  and Vanessa von Kortzfleisch 
Department of Behavioural Biology, University of Münster, Münster, Germany.
[✉]e-mail: richterh@uni-muenster.de
<https://doi.org/10.1038/s41583-020-0369-0>

1. Voelkl, B. et al. Reproducibility of animal research in light of biological variation. *Nat. Rev. Neurosci.* **21**, 384–393 (2020).
2. Richter, S. H., Garner, J. P. & Würbel, H. Environmental standardization: cure or cause of poor reproducibility in animal experiments? *Nat. Methods* **6**, 257–261 (2009).
3. Richter, S. H., Garner, J. P., Auer, C., Kunert, J. & Würbel, H. Systematic variation improves reproducibility of animal experiments. *Nat. Methods* **7**, 167–168 (2010).
4. Richter, S. H. Systematic heterogenization for better reproducibility in animal experimentation. *Lab Anim.* **46**, 343–349 (2017).
5. Voelkl, B. & Würbel, H. Reproducibility crisis: are we ignoring reaction norms? *Trends Pharmacol. Sci.* **237**, 509–510 (2016).
6. Richter, S. H. et al. Effect of population heterogenization on the reproducibility of mouse behavior: a multi-laboratory study. *PLoS ONE* **6**, 0016461 (2011).
7. Milcu, A. et al. Genotypic variability enhances the reproducibility of an ecological study. *Nat. Ecol. Evol.* **2**, 279–287 (2018).
8. Bodden, C. et al. Heterogenising study samples across testing time improves reproducibility of behavioural data. *Sci. Rep.* **9**, 8247 (2019).
9. Voelkl, B., Vogt, L., Sena, E. S. & Würbel, H. Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biol.* **16**, e2003693 (2018).
10. Festing, M. F. & Altman, D. G. Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR J.* **43**, 244–258 (2002).
11. Richter, S. H. Automated home-cage testing as a tool to improve reproducibility of behavioral research? *Front. Neurosci.* **14**, 383 (2020).
12. Chesler, E. J., Wilson, S. G., Lariviere, W. R., Rodriguez-Zas, S. L. & Mogil, J. S. Influences of laboratory environment on behavior. *Nat. Neurosci.* **5**, 1101–1102 (2002).
13. National Academies of Sciences, Engineering, and Medicine. Reproducibility and Replicability in Science (The National Academies Press, 2019).

Competing interests

The authors declare no competing interests.