

Reply to ‘It is time for an empirically informed paradigm shift in animal research’

Hanno Würbel¹, Bernhard Voelkl², Naomi S. Altman³, Anders Forsman⁴, Wolfgang Forstmeier⁵, Jessica Gurevitch⁶, Ivana Jaric⁷, Natasha A. Karp⁸, Martien J. Kas⁹, Holger Schielzeth¹⁰ and Tom Van de Castele¹¹

In their correspondence about our recent Perspective article (Reproducibility of animal research in light of biological variation. *Nat. Rev. Neurosci.* **21**, 384–393 (2020))¹, Richter and von Kortzfleisch support our recommendations for a paradigm shift from rigorous standardization to systematic heterogenization in animal research (It is time for an empirically informed paradigm shift in animal research. *Nat Rev Neurosci.* <https://doi.org/10.1038/s41583-020-0369-0> (2020))². However, they argue that empirical studies are needed to demonstrate that heterogenization improves reproducibility, and how it does so, and that heterogenization should be based on controlled variation.

Although we welcome their support for the proposed paradigm shift and their call for more empirical research, we would like to emphasize that there is already good empirical evidence demonstrating that heterogenization improves reproducibility. Simulations using diverse sets of empirical data collected across multiple independent laboratories demonstrate that effective heterogenization improves reproducibility substantially, for a wide variety of outcome variables³. Further empirical research is now needed to investigate practicable solutions, as we indicated in the Perspective¹. We know of several such studies that are currently under way: at the University of Bern, some of us are investigating how heterogenization of study populations by breeder (that is, using subpopulations of mice from multiple breeders) affects reproducibility; the European consortium EQIPD is using large multi-centre studies to identify factors in study design that influence reproducibility in pre-clinical neuroscience and safety studies; and the German consortium DECIDE is monitoring a call by the Federal Ministry of Education and Research for multi-centre studies to establish best practice guidelines to improve the robustness and reproducibility of confirmatory preclinical studies. In addition, meta-analyses of large data sets of research consortia (for example, EQIPD and see REF⁴) allow identification of biological variables that explain between-centre

variation and may thus become effective heterogenization factors. Nevertheless, we need to consider the possibility that effective heterogenization may be context dependent, and specific solutions may have to be sought for specific research questions, animal models or outcome variables.

Richter and von Kortzfleisch further argue “that the concept of heterogenization relies on the introduction of systematic and hence controlled variation”². We deliberately refrained from limiting heterogenization to specific factors and procedures. Empirical multi-centre simulations have shown that uncontrolled heterogenization by centre, introducing significant variability in genotype, husbandry and study protocols, improved reproducibility substantially³. Ideally, we would be able to mimic multi-centre studies by systematically varying one or two factors that account for most of the observed between-centre variation. Richter and von Kortzfleisch refer to ‘experimenter’ as an example of such ‘umbrella factors’. We are sceptical that experimenter is a good example though and question whether such umbrella factors exist — especially ones that generalize across animal models and outcome variables. Although experimenter can have strong effects on study results^{5,6} and can be varied systematically, the variation introduced by experimenter is uncontrolled (similar to ‘laboratory’), as we can neither predict how different experimenters will affect the results nor analyse what differences between experimenters caused variation in the results. In most cases, between-study variation will probably be multifactorial and the assumption that one or two factors exist that account for most of it may be unrealistic. More realistically, we should aim to prevent unwarranted overgeneralizations by extending the inference space of animal studies using biologically meaningful heterogenization factors. Richter and von Kortzfleisch are concerned that “using outbred strains might bear the risk of inflating sample sizes” but recent evidence strongly suggests otherwise⁷. Furthermore, whereas fixed factors limit

generalization to the specific factor levels used (for example, the two genotypes in the case of two inbred strains), random factors allow generalization of results to the range of variation covered by the random factor (for example, the variety of genotypes represented by an outbred strain). However, instead of using outbred strains, genetic reference panels such as the BXD family of recombinant inbred strains⁸ or the Collaborative Cross⁹ offer more powerful ways of heterogenizing genotype. Similarly, for more powerful ways of heterogenizing environment, we might establish ‘environmental reference panels’ — a variety of ‘envirotypes’ based on a set of biologically relevant exogenous factors that are known to affect the organisms’ phenotypes^{10,11}.

In conclusion, we maintain that the study of the principle of heterogenization and its effects on reproducibility is clearly beyond the conceptual level. However, we agree with Richter and von Kortzfleisch that more research is needed to explore and validate effective and practicable study designs for specific purposes.

Hanno Würbel¹, Bernhard Voelkl², Naomi S. Altman³, Anders Forsman⁴, Wolfgang Forstmeier⁵, Jessica Gurevitch⁶, Ivana Jaric⁷, Natasha A. Karp⁸, Martien J. Kas⁹, Holger Schielzeth¹⁰ and Tom Van de Castele¹¹

¹Animal Welfare Division, Vetsuisse, University of Bern, Bern, Switzerland.

²Department of Statistics, The Pennsylvania State University, University Park, PA, USA.

³Department of Biology and Environmental Science, Linnaeus University, Kalmar, Sweden.

⁴Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, Seewiesen, Germany.

⁵Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA.

⁶Data Sciences & Quantitative Biology, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK.

⁷Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, Netherlands.

⁸Institute of Ecology and Evolution, Friedrich Schiller University Jena, Jena, Germany.

⁹Statistics and Decision Sciences, Janssen R&D, Johnson & Johnson, Beerse, Belgium.

¹⁰e-mail: hanno.wuerbel@vetsuisse.unibe.ch

<https://doi.org/10.1038/s41583-020-0370-7>

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. & von Kortzfleisch, V. It is time for an empirically informed paradigm shift in animal research. *Nat. Rev. Neurosci.* <https://doi.org/10.1038/s41583-020-0369-0> (2020).
3. 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).
4. Corrigan, J. K. et al. A big-data approach to understanding metabolic rate and response to obesity in laboratory mice. *eLife* **9**, e53560 (2020).
5. 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).
6. Sorge, R. E. et al. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat. Methods* **11**, 629–632 (2014).

7. Tuttle, A. H. et al. Comparing phenotypic variation between inbred and outbred mice. *Nat. Methods* **15**, 994–996 (2018).
8. Peirce, J. L., Lu, L., Gu, J., Silver, L. M. & Williams, R. W. A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC Genet.* **5**, 7 (2004).
9. Threadgill, D. W., Hunter, K. & Williams, R. W. Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. *Mamm. Genome* **13**, 175–178 (2002).
10. Beckers, J., Wurst, W. & de Angelis, M. Towards better mouse models: enhanced genotypes, systemic phenotyping and envirotype modelling. *Nat. Rev. Genet.* **10**, 371–380 (2009).
11. Xu, Y. Envirotyping for deciphering environmental impacts on crop plants. *Theor. Appl. Genet.* **129**, 653–673 (2016).

Competing interests

The authors declare no competing interests.

RELATED LINKS

DECIDE: <https://www.bihealth.org/en/research/quest-center/mission-approaches/quality/decide/>

EQIPD: <https://quality-preclinical-data.eu/>