scientific reports

Check for updates

EDITORIAL

Improving reproducibility in animal research

Florian Frommlet

There have been increasingly lively discussions about many published scientific results failing validation by independent studies. This so-called reproducibility crisis has led to particularly strong criticism of methodological weaknesses in animal research. Inappropriate statistical methods, poor experimental design, and extreme standardization in trial design are some contributing factors to the problem. The purpose of this Collection is to present original methodologies to improve the status quo and additionally to report meta-research about the reproducibility of published animal research.

Within the last decade, it has become increasingly clear that many published scientific results, even those in prestigious journals, do not withstand the test of reproduction by independent research. Statisticians have been aware of the underlying problem for a long time, see for example Steven Goodman[1]. In his most famous paper[2], John Ioannidis quite accurately predicted the frequency of false positive results one can expect in the scientific literature. Nevertheless, it came as quite a shock—to many researchers in different fields—when they had to admit a large proportion of their published findings were wrong. This culminated in the so-called reproducibility crisis.

Preclinical research has come under particular scrutiny because the stakes are so high, both in terms of research funding and societal impact[3]. Some of the problems involved are quite well known. For instance, poor statistical training of researchers within this area leads to poor experimental design, and the use of inappropriate statistical methods. More recently, researchers have become aware that extreme standardization in animal trials may also contribute to the problem[4], with slight deviations in test conditions yielding different results. This Collection gathers together methods which aim to improve the status quo, and meta-research on the reproducibility of published animal research. The question as to the extent that the results from animal research can be translated into practical medicine is also touched on by some articles.

The group of John Ioannidis contributed a meta-analytic paper, for studies on domestic dogs with spontaneously occurring cancers. Such studies are of particular interest from a translational point of view, because dogs as model organisms are closer to humans than murine models. However, the results from Tan et al.[5] clearly show that most of the studies performed in this area suffer severely from a lack of statistical power and proper reporting, both of methods and results.

Two contributions investigated the extent to which study designs that employ heterogenization can improve the reproducibility of animal trials. The group of Helene Richter shows that the simple step of performing experiments at different times of the day introduces sufficient systematic variation to obtain more replicable results[6]. Natasha Karp and her colleagues[7] consider a multi-batch approach, where the study is planned for several small independent mini-experiments within the same lab. The idea is very similar to the approach to experimental replication that I have recently suggested myself[8], except that in addition to studying the question of statistical analysis in this context, Karp et al.[7] also demonstrate the actual performance in two case studies.

The group of Hanno Würbel presents some slightly contrasting results in their paper[9]. They systematically studied the influence of weaning time and housing conditions on a variety of behavioral and physiological outcome variables, where they observed smaller effects than expected. They conclude that more research is needed to understand which variables across labs are actually responsible for different outcomes in replication studies.

Stan Lazic contributes two papers, covering statistical aspects of the reproducibility problem. The first article[10] presents a Bayesian alternative to the analysis of covariance, in the context of analyzing organ weight changes. The second article[11] introduces a Bayesian prediction model to deal with pseudoreplications. Treating correlated data points as independent is among the most prevalent sources of statistical malpractice in animal trials. This should be rectified, either by using mixed models, or by a Bayesian approach as suggested here.

Two more contributions discuss cage effects. The first one[12] studies cage-clustered-data, in the context of research on diets and gut microbiomes. The more theoretical paper by Michael Festing[13] stresses the importance of taking into account cage effects at the level of study design. He gives the clear recommendation of using

Center for Medical Statistics, Informatics and Intelligent Systems, Section for Medical Statistics, Medical University Vienna, Vienna, Austria. email: Florian.Frommlet@meduniwien.ac.at

randomized block designs and corresponding methods for statistical analysis. However, a brief survey of preclinical studies shows that this important concept is currently hardly ever implemented in practice.

The Collection also includes papers directly concerned with the practical aspects of animal trials. One contribution, from Andelius et al.[14], studies how insertion trauma of intracerebral devices affects variability of outcome measurements. They conclude that the time to stabilization differs depending on the device. This is important information for obtaining valid baseline values, and may have an effect on the reproducibility of studies using microdialysis. Another paper introduces PiDose, a system that automatically weighs the individual animals of a study, and based on that, provides the daily dosage of drug solution through each animal's drinking water[15]. This system allows a reduction in interaction with experimenters, and thus, potentially helps decrease study bias. Finally, the Cardioprotection Large Animal Platform (CIBER-CLAP) is presented by Rossello et al.[16], which was established in Spain to ascertain whether results from cardiovascular animal trials are reproducible, before the next step is taken, and greater complexity is introduced in a clinical setting.

Although the Collection has now been published, submissions are still welcome on a rolling basis. Papers accepted will then be added to the Collection as and when they are published. I would like to thank all the authors who have contributed to this Special Collection and I sincerely hope that their work presented here will help to improve the reproducibility of preclinical animal trials.

References

1. Goodman, S. N. A comment on replication, p-values and evidence. *Stat. Med.* **11**, 875–879. https://doi.org/10.1002/sim.4780110705 (2005).
2. Ioannidis, J. P. Why most published research findings are false. *PLoS Med.* **2**, e124. https://doi.org/10.1371/journal.pmed.0020124 (2005).
3. Harris, R. *Rigor Mortis: How Sloppy Science Creates Worthless Cures, Crushes Hope, and Wastes Billions* (Basic Books, New York, 2017).
4. Richter, S. H., Garner, J. P. & Würbel, H. Environmental standardization: cure or cause of poor reproducibility in animal experiments?. *Nat. Methods* **6**, 257 (2009).
5. Tan, Y. J., Crowley, R. J. & Ioannidis, J. P. A. An empirical assessment of research practices across 163 clinical trials of tumor-bearing companion dogs. *Sci. Rep.* **9**, 11877. https://doi.org/10.1038/s41598-019-48425-5 (2019).
6. Bodden, C. *et al.* Heterogenising study samples across testing time improves reproducibility of behavioural data. *Sci. Rep.* **9**, 8247. https://doi.org/10.1038/s41598-019-44705-2 (2019).
7. Karp, N. A. *et al.* A multi-batch design to deliver robust estimates of efficacy and reduce animal use—a syngeneic tumour case study. *Sci. Rep.* **10**, 6178. https://doi.org/10.1038/s41598-020-62509-7 (2020).
8. Frommlet, F. & Heinze, G. Experimental replications in animal trials. *Lab. Anim.* https://doi.org/10.1177/0023677220907617 (2020).
9. Bailoo, J. D. *et al.* Effects of weaning age and housing conditions on phenotypic differences in mice. *Sci. Rep.* **10**, 11684 (2020).
10. Lazic, S. E., Semenova, E. & Williams, D. P. Determining organ weight toxicity with Bayesian causal models: improving on the analysis of relative organ weights. *Sci. Rep.* **10**, 6625. https://doi.org/10.1038/s41598-020-63465-y (2020).
11. Lazic, S. E. *et al.* A Bayesian predictive approach for dealing with pseudoreplication. *Sci. Rep.* **10**, 2366. https://doi.org/10.1038/s41598-020-59384-7 (2020).
12. Basson, A. R. *et al.* Artificial microbiome heterogeneity spurs six practical action themes and examples to increase study power-driven reproducibility. *Sci. Rep.* **10**, 1–19 (2020).
13. Festing, M. The "Completely randomised" and the "Randomised block" are the only experimental designs suitable for widespread use in pre-clinical research. *Sci. Rep.* **10**, 17577. https://doi.org/10.1038/s41598-020-74538-3 (2020).
14. Andelius, T. C. K. *et al.* Consequence of insertion trauma – effect on early measurements when using intracerebral devices. *Sci. Rep.* **9**, 10652. https://doi.org/10.1038/s41598-019-47052-4 (2019).
15. Woodard, C. L. *et al.* PiDose: an open-source system for accurate and automated oral drug administration to group-housed mice. *Sci. Rep.* **10**, 11584. https://doi.org/10.1038/s41598-020-68477-2 (2020).
16. Rossello, X. *et al.* CIBER-CLAP (CIBERCV Cardioprotection Large Animal Platform): a multicenter preclinical network for testing reproducibility in cardiovascular interventions. *Sci. Rep.* **9**, 20290. https://doi.org/10.1038/s41598-019-56613-6 (2019).

Competing interests