

Mind the gap – towards complete and transparent reporting of animal research

Sandra Tillmann

Translational Neuropsychiatry Unit,
Department of Clinical Medicine,
Aarhus University, Aarhus, Denmark

Correspondence to:

Sandra Tillmann
Translational Neuropsychiatry Unit
Department of Clinical Medicine
Aarhus University
Skovagervej 2, 8240 Risskov, Denmark
sti@clin.au.dk
+45 7847 1122

Abstract

Several initiatives have been taken to standardise the reporting of animal studies in peer-reviewed scientific journals, such as the ARRIVE (Animal Research: Reporting of In Vivo Experiments) and GSPC (Gold Standard Publication Checklist) guidelines. Surprisingly, many publications still lack key methodological details. As a result, animal studies are often criticised for poor scientific quality and low translatability to the clinic. To promote adherence to available guidelines, this article covers the rationales for including key parameters that are often overlooked, such as strain nomenclature, housing conditions, and behavioural test settings.

Using rodents to understand human disease

On the journey from laboratory to clinic, animal testing provides the first opportunity to characterise the safety and efficacy of a drug candidate in a living organism. Depending on the disease target, the choice of species ranges from apes to zebrafish, with mice and rats making up about 95% of all research animals. Rodents are commonly used as disease models and are therefore inherently expected to be at least somewhat predictive for a human response to a drug. For model organisms to have this potential translational value, animal studies must be



designed, conducted, analysed, and reported with the highest scientific rigour. Regrettably, they still lag behind the standards for reporting human clinical data, although both share the common aim of generating unbiased data.

Catching up with human standards

Reporting standards of human randomised controlled trials (RCTs) were improved by the introduction of the CONSORT (Consolidated Standards of Reporting Trials) statement in 1996.¹ Today, fundamental principles such as randomisation and blinding are regarded as the minimum requirements for performing and communicating science. Surprisingly, not all preclinical (and clinical) publications in peer-reviewed scientific journals seem to meet these very basic reporting standards. A 2009 survey by NC3Rs (National Centre for the Replacement, Refinement, and Reduction of Animals in Research) on the quality of reporting of publicly funded animal research in the UK and US revealed that over 85% of included publications lacked reporting of randomisation or blinding, and 41% lacked key information on

hypothesis and number/characteristics of animals.² In response to these survey results, NC3Rs developed the ARRIVE (Animal Research: Reporting of In Vivo Experiments) guidelines in 2010,³ freely available at www.nc3rs.org.uk/arrive-guidelines. They follow the example of the CONSORT statement and not only aim to improve the reporting of existing studies, but also the design of new animal experiments. The ARRIVE guidelines include a checklist of 20 items with descriptions of how to report a study comprehensively and transparently, e.g., by providing animal characteristics and statistical approach. Several other initiatives are dedicated to reducing the risk of bias in animal studies, such as CAMARADES (Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies) and SYRCLE (Systematic Review Centre for Laboratory Animal Experimentation). The latter group also published SYRCLE's Risk of Bias Tool, an adapted version of the RCT-targeted Cochrane Risk of Bias Tool for animal studies,⁴ and a Gold Standard Publication Checklist (GSPC),⁵ which partially overlaps with the ARRIVE guidelines.

Despite such efforts to enhance the reproducibility of animal research, many publications still fail to provide even basic details on experimental design and analysis. As a result, they are rated as low-quality reports and are excluded from systematic reviews and meta-analyses. Such summaries of primary research articles represent the highest level of medical evidence and are used to guide clinical decision-making. Incomplete reports, even though the study may have been conducted perfectly, are therefore a source of bias and may lead to faulty conclusions about the safety and efficacy of a drug. Most professional medical writers and editors have extensive knowledge of the publication process and already have a good understanding of why details on blinding, randomisation, sample size, study objectives, or statistical analysis matter. For this reason, this article focuses on often under- or misreported methodological sections specific to animals, and the rationales for their importance.

Despite efforts to enhance the reproducibility of animal research, many publications still fail to provide even basic details on experimental design and analysis.

Species, strains, and why C57/BL6 mice do not exist

Most authors report the species and strain of the included animals, since these factors have long been known to affect behavioural and pharmacological responses. What is perhaps less known is that substrains of a strain, and even the same strain obtained from different vendors, may exhibit distinct phenotypes. For example, Wistar rats from Harlan Laboratories and Charles River vary in their response to the same acute myocardial infarction model, in that the former have a higher survival rate despite being more sensitive to cardiomyocyte damage.⁶ Not only the vendor, but also the location of the vendor matters. Wistar rats from Harlan in the US (Hsd:WI) are behaviourally distinct from Wistar rats from Harlan in Europe (RccHanTM:WI).⁷ Although both are of the Wistar strain, they originated from different institutes (Wistar Institute, Philadelphia, PA, US vs. Zentralinstitut für Versuchstierzucht, Hannover, Germany), emphasising the need for proper documentation of strain, substrain, vendor, and vendor location. This is further demonstrated by the widespread erroneous assumption that the correct nomenclature for the popular “Black 6” mouse strain is “C57/BL6”. As the original breeder, The

Jackson Laboratory, put it in its blog post: “There is no such thing as a C57/BL6 mouse!” Instead, there are two distinct breeding colonies denoted as C57BL/6J (“J” for Jackson Laboratory) and C57BL/6N (“N” for National Institutes of Health), from which many substrains have emerged, such as C57BL/6NCrl (Charles River) and C57BL/6Jcl (Clea Japan).

Genetic and phenotypic differences both within and among populations are well established, so authors should always include the complete substrain designation indicating the laboratory maintaining the colony. For genetically modified animals (e.g. knockout or transgenic), additional information is needed. Guidelines and a checklist for reports on mutant studies have been provided by Crusio et al.⁸

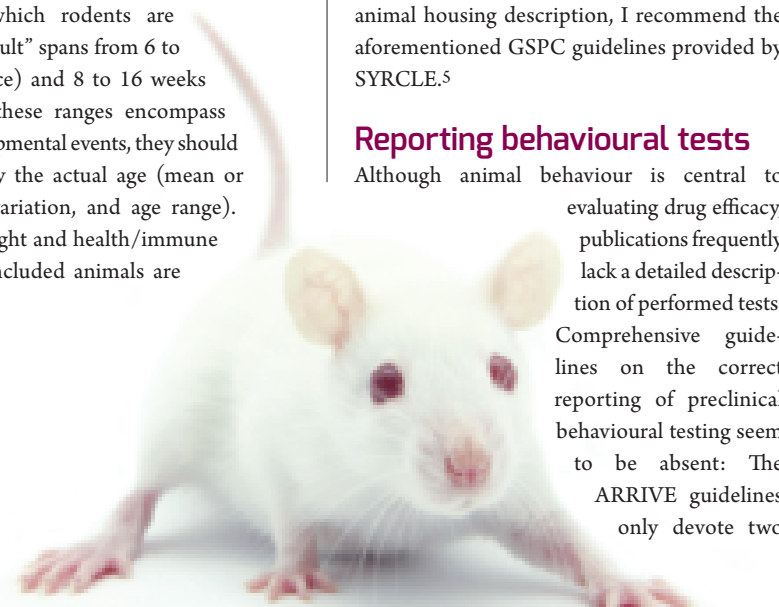
Failure to mention sex and age may result in skewed data interpretation, as these factors are known to affect pharmacokinetics and pharmacodynamics in humans and laboratory animals. Fortunately, reporting percentages have markedly increased in the last two decades; nevertheless, a study found that still only 50% of the included articles published in 2014 reported both sex and age of their mice.⁹ There is hardly an excuse for not including these variables, since they are available to all researchers and do not take up much space. Simply providing the weight or the developmental stage of the animals (e.g. “juvenile” or “adult”) instead of their age does not suffice, since these vary greatly across laboratories. According to a survey by Jackson et al.,¹⁰ the age at which rodents are considered “adult” spans from 6 to 20 weeks (mice) and 8 to 16 weeks (rats). Since these ranges encompass distinct developmental events, they should be replaced by the actual age (mean or median age, variation, and age range). Moreover, weight and health/immune status of all included animals are needed.

To enrich or not to enrich?

Laboratory housing conditions significantly influence behaviour and pharmacological response, and should be reported in detail. Since mice and rats are very social animals, isolating them may induce stress and interfere with the effects of a drug. Distinguishing between single- and group-housed animals is therefore important. Perhaps less known, mice and rats are sensitive to environmental complexity. In the 1940s, the influential psychologist and neuroscientist Donald Hebb took some laboratory rats home for his children to play with. Surprisingly to him, these animals subsequently performed better on cognitive tests than animals housed only in the laboratory. In the following decades, it was found that animals in enriched cages (containing sensory stimulation such as nesting material, tunnels, cardboard boxes, or chewing toys) but not control cages had more synapses, more and longer dendrites, and were protected against several types of brain injuries. Given these marked structural brain changes, care should be taken to adequately report the presence or lack of any cage enrichment for laboratory animals. Authors should refrain from referring to “standard cages” without further clarification, since the standard in one lab may be a cage with bedding only, while in another it may also contain shelter and toys. Publication writers should therefore clearly specify the number of animals (and sex thereof) per cage, model and size of cage and lid, presence and type of any enrichment, and frequency of cage change. Moreover, they need to describe the temperature and lighting in the room as well as the nutrition type and feeding regimen. For a comprehensive guide on the animal housing description, I recommend the aforementioned GSPC guidelines provided by SYRCLE.⁵

Reporting behavioural tests

Although animal behaviour is central to evaluating drug efficacy, publications frequently lack a detailed description of performed tests. Comprehensive guidelines on the correct reporting of preclinical behavioural testing seem to be absent: The ARRIVE guidelines only devote two



lines to this effect, in which they prompt the author to clearly define the experimental outcome (i.e. behavioural change). GSPC guidelines also lack a complete list, although they at least regard the frequency of handling as a relevant variable. Given the multitude and complexity of available behavioural tests, a concise checklist is no easy task.

However, some general considerations apply to most behavioural tests and should be reported in all behavioural publications to facilitate scientific transparency and unbiased data interpretation. The most obvious consideration is the precise description of the behavioural test. For example, the Novel Object Recognition test, a popular cognition test in rodents, consists of multiple steps: habituation, training, and testing. The total test duration is 1 to 4 days, depending on the length of the habituation of the animal to the arena. Several lengths have been reported across laboratories, spanning from 2 minutes to as much as 2 hours. As long as the durations of each phase and the intervals between them are reported, this is not a major source of concern. However, it becomes a problem when authors deviate from the original protocol without mentioning their modifications; or worse, when they only provide the name of the test. In all tests comprising multiple phases, all durations and intervals in between should be given, along with any time allowed for acclimatisation to the behavioural rooms immediately prior to testing. The general rule of thumb here is to provide as many details as possible: If water is involved (e.g. Morris Water Maze or Forced Swim Test), provide the temperature and frequency of changing; if an apparatus is used, always report the size and material, as these can affect the response to a drug agent.

From a human perspective, it is easy to imagine that we probably perform better at a cognition test during the day than in the middle of the night. The same has been observed in (nocturnal) rodents – the time of testing (active vs. inactive phase) significantly affects their behaviour. Testing

Failure to mention sex and age may result in skewed data interpretation, as these factors are known to affect pharmacokinetics and pharmacodynamics in humans and laboratory animals.

should therefore preferably be conducted in the dark (active) phase, unless it is known that a particular measure is not impacted by circadian rhythm. This does not only apply to behaviour, but also to the collection of any tissue that is sensitive to the photoperiod. In addition to information on the light cycle in the animal room, publications should

therefore include the time of the day in which behavioural data or tissue were collected, as well as the light intensity (in lux) where applicable. Authors should also state if all tests were conducted by the same experimenter and how often the rats were handled prior to testing. Moreover, if more than one test was used, the number of days in between tests and the order of the tests should be reported, especially if one of them is stressful for the animals. Timelines of experimental events are a good way to graphically represent complex designs or testing batteries.

Drug intervention and tissue collection

Any drug treatment must be clearly described, including the dose per weight, volume of injection, route of administration, frequency (including the time of the day it was given), exact vehicle, and method of preparation (e.g. sonication, multiple dilutions, etc.). Moreover, the euthanasia process and tissue collection must be documented, such as euthanasia method, time of day, whether all animals were euthanised on the same day, if randomisation was applied, and how tissue was collected. Instead of simply writing “liver tissue was obtained”, it should be mentioned which lobe the tissue was

taken from and how it was stored during and after the collection process. This might be obvious to many, but papers frequently lack basic information on tissue processing, e.g. dissection method, centrifugation speed, or freezer temperature. If decapitation is used, include measures of how rodents were protected from the smell of blood to prevent any confounding hormonal effects (e.g., they might have been housed in an adjacent room and brought into the decapitation room by an experimenter free of blood scent).

How can we improve reporting standards?

By including or omitting methodological details, authors tremendously influence the quality of the article and the ability to draw meaningful conclusions from the results. With the omission of important details, readers are left to assume the worst-case scenario – that they have not been considered or performed. As a consequence, the overall relevance and quality of the data might be assessed as poor whether it is or is not. Unless research is adequately reported, the time, effort, and resources invested are wasted. In the case of preclinical studies, wasted resources may mean unnecessary loss of animal lives, which should be prevented at all costs. It is therefore the responsibility of the author to adhere to current guidelines such as ARRIVE or GSPC to reduce the risk of bias and maintain a high scientific standard in preclinical research. Journals need to implement adherence to these guidelines (e.g., by requesting a filled-out checklist at submission) rather than just endorsing them passively. They also need to provide authors with sufficient space to include all relevant details, either in the manuscript body or in supplementary files. New medical writers and researchers could greatly benefit from education and training opportunities that address the issues mentioned in this

article in greater detail. Transparency may also be increased by animal registries such as www.preclinicaltrials.eu, in which researchers pre-register details of their experiment in an online database like www.clinicaltrials.gov for human trials.

This would allow medical writers to refer to the registry and select the key information relevant to the article at hand, which may improve the quality of future reports and contribute to less



bias. Ultimately, the value and clinical meaningfulness of animal studies hinges on the thorough reporting of experimental methods. Authors and medical writers involved in the publication process should therefore be aware of the importance of including even seemingly small details, as these may alter the reproducibility and generalisability of the study outcomes.

Acknowledgements

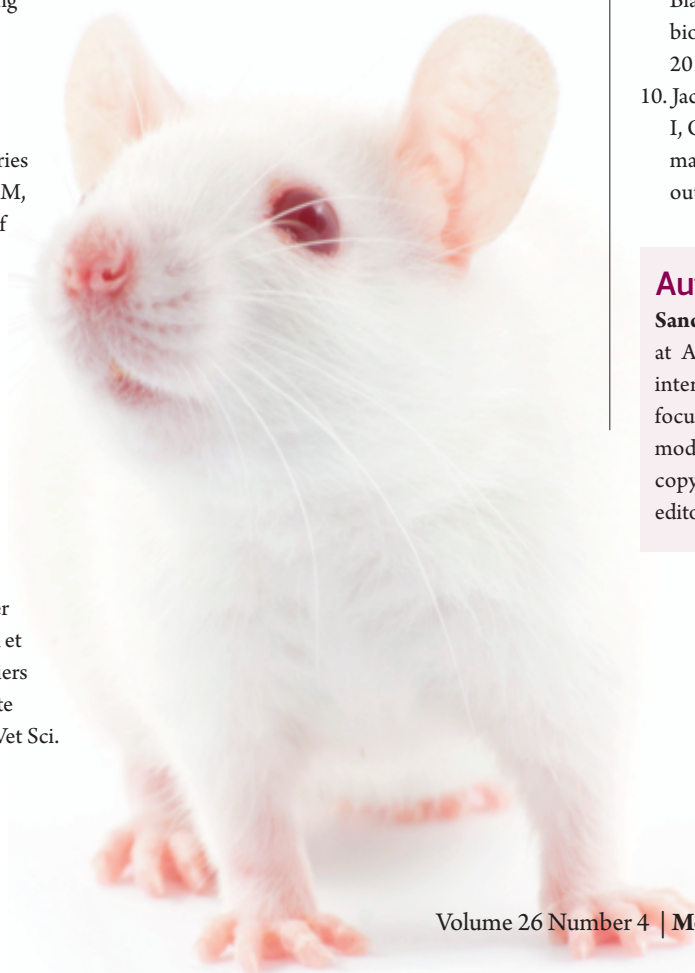
Special thanks to Denise Happ for her excellent comments on this article.

Conflicts of interest

The author declares no conflicts of interest.

References

1. Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. *BMJ*. 1996;313(7057):570–1.
2. Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One*. 2009;4(11):e7824.
3. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol*. 2010;8(6):e1000412.
4. Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol*. 2014;14:43.
5. Hooijmans CR, Leenaars M, Ritskes-Hoitinga M. A gold standard publication checklist to improve the quality of animal studies, to fully integrate the Three Rs, and to make systematic reviews more feasible. *Altern Lab Anim*. 2010;38(2):167–82.
6. Naaijken BA, van Dijk A, Meinster E, Kramer K, Kamp O, Krijnen PA et al. Wistar rats from different suppliers have a different response in an acute myocardial infarction model. *Res Vet Sci*. 2014;96(2):377–9.
7. Palm S, Hävermark Å, Meyerson BJ, Nylander I, Roman E. When is a Wistar a Wistar? Behavioural profiling of outbred Wistar rats from five different suppliers using the MCSF test. *Appl Anim Behav Sci*. 2011;135(1-2):128–37.
8. Crusio WE, Goldowitz D, Holmes A, Wolfer D. Standards for the publication of mouse mutant studies. *Genes Brain Behav*. 2009;8(1):1–4.
9. Flórez-Vargas O, Brass A, Karystianis G, Bramhall M, Stevens R, Cruickshank S et al. Bias in the reporting of sex and age in biomedical research on mouse models. *Elife*. 2016;5:e13615.
10. Jackson SJ, Andrews N, Ball D, Bellantuono I, Gray J, Hachoumi L et al. Does age matter? The impact of rodent age on study outcomes. *Lab Anim*. 2017;51(2):160–9.



Author information

Sandra Tillmann is a final-year PhD student at Aarhus University (Denmark) with an interest in medical writing. Her research focuses on the gut microbiome in animal models of depression. She is also a freelance copyeditor of scientific manuscripts and an editorial trainee at a neuropsychiatric journal.