ARRIVE 2.0 and the British Journal of Pharmacology: updated guidance for 2020

Elliott Lilley[1] *

S. Clare Stanford[2] *

David E Kendall[3]

Stephen P. H. Alexander[3]

Giuseppe Cirino[4]

James R. Docherty[5]

Christopher H. George[6]

Paul A. Insel[7]

Angelo A. Izzo[4]

Yong Ji[8]

Reynold A. Panettieri[9]

Christopher G. Sobey[10]

Barbara Stefanska[11]

Gary Stephens[12]

Mauro Teixeira[13]

Amrita Ahluwalia[14] *

[1] Royal Society for the Prevention of Cruelty to Animals UK

[2] University College London, UK

[3] University of Nottingham, UK

[4] University of Naples Federico II, Italy

[5]Royal College of Surgeons in Ireland, Dublin, Ireland

[6] Swansea University, UK

[7] University of California, San Diego, USA

[8] Nanjing Medical University, Nanjing, China

[9] Rutgers University, USA

[10] La Trobe University, Bundoora, VIC, Australia

[11] The University of British Columbia, Canada

[12] University of Reading, UK

[13] Federal University of Minas Gerais, Belo Horizonte, Brazil

[14] William Harvey Research Institute, Queen Mary University of London, London, UK

* Authors made equal contributions to this article

Conflict of interest

The authors declare no conflicts of interest.

* **Correspondence**

Amrita Ahluwalia, British Journal of Pharmacology, The Schild Plot, 16 Angel Gate, City Road, London EC1V 2PT, UK. Email: info@bps.ac.uk

BACKGROUND

BJP has been and remains an active advocate of the ARRIVE (*Animal Research: Reporting of In Vivo Experiments)* guidelines (Kilkenny, Browne et al., 2010a) that were established by the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) in 2010. The ARRIVE guidelines need no introduction and we will not rehearse the arguments in depth here, other than to restate that the dearth of key *in vivo* experimental details has been identified as a major contributing factor to the poor reproducibility of pre-clinical research. This fact was the primary driver for establishment by the NC3Rs of the first version of the ARRIVE guidelines.

ARRIVE provided a 20-point checklist, specifying all of the experimental details (procedures and fixed factors) that should be included in manuscripts for proper reporting of animal research. The guidelines were rapidly endorsed internationally by funding bodies, universities, learned societies and, importantly, Life Science journals. Currently there are 1046 journals endorsing ARRIVE including BJP, which was one of the original six influential journals that published the guidelines in full in 2010 (Kilkenny, Browne et al., 2010b). However, ARRIVE has not had the impact that was originally hoped for in terms of full transparency, despite the widespread support and endorsement. In 2016 and in 2018 assessments of adherence of articles published in endorsing journals identified a lack of engagement. As an example, one systematic review of reports of studies investigating acute lung injury revealed that, of the items expected for ARRIVE compliance, only 45% of those advised for inclusion in the Methods, and only 29% of those for inclusion in the Results section, were present (Avey, Moher et al., 2016). Moreover, formal endorsement of the ARRIVE guidelines by journals did not necessarily improve compliance (Leung, Rousseau-Blass et al., 2018). Such disappointing reports of outcome triggered a reappraisal of the guidelines led by the NC3Rs who, in 2018, established a new, international Working Group to review and update the ARRIVE guidelines to generate ARRIVE 2.0 (Percie du Sert, Hurst et al., 2018). As with the team of experts brought together to establish the first iteration of ARRIVE, the Editor in Chief of BJP is a member of the new team that was assembled. As such, BJP has been well-placed to influence the content, testing and final publication of both iterations of the guidelines.

In 2015, BJP published an editorial reporting findings from a survey of compliance with ARRIVE in articles published in two issues of the journal in 2014 (McGrath & Lilley, 2015). The results, as above, were somewhat disappointing revealing scope for improving compliance in respect of both the design of the (animal) experiments and the description of experimental procedures. To help remedy this problem, a checklist was developed that provided an *aide memoire* of the details of the animals and research procedures that should be reported in manuscripts submitted to the journal. A separate editorial focussed on experimental design and data analysis (Curtis, Alexander et al., 2015) for all types of experimental data, including those emanating from experiments with animals, was also published in the journal. A key issue with many of the studies reported in BJP prior to this related to inadequate evidence of forethought in experimental design and inappropriate statistical analysis. At the same time, the *Instructions to Authors* were revised, to include the new rubric, and the editorial scrutiny of these aspects of the peer review process was tightened up. As a further prompt, since 2016, authors are required to make a Declaration as part of the submission process, to confirm that their manuscript is ARRIVE compliant. In addition to these measures, two new 'specialist' editors were appointed in 2016: a *Design & Analysis Advisor* and a *Consulting Editor in ARRIVE Guidelines and Animal Welfare*. The brief of the latter was to monitor ARRIVE compliance in BJP publications and to assist Senior Editors with their appraisal of manuscripts. All these changes still remain in place with the addition of a Consulting Editor in Statistical Analysis to ensure that the research reports match BJP criteria and so qualify for publication.

THE CURRENT STANDARD OF REPORTING IN BJP

ARRIVE 2.0 together with an 'Explanation and Elaboration' document was loaded onto the NC3Rs website in July 2019 as a preprint (i.e., before peer review (Percie du Sert, Hurst *et al*., 2019)), and is now published in full in PLOS Biology (Percie-du Sert et al, 2020a), with simultaneous publication in several international journals including in this issue of BJP (ref). In preparation for the publication of the new guidelines we conducted surveys assessing compliance to ARRIVE in BJP over 4 years, the results of which can be seen in Table 1. These data together with the July 2019 preprint of ARRIVE 2.0 were discussed extensively by the Senior Editorial Board of BJP in December 2019. This editorial explains the ensuing changes in the journal's editorial policy as a result of those discussions and how they relate to ARRIVE 2.0.

The principle of full disclosure lies at the heart of what we expect from authors wishing to publish in BJP; and in support of this principle is the absence of a word restriction for the Methods section. Judging from articles surveyed, between July 2014 and October 2019, most authors are conscientious about compliance with our *'Instructions to Authors'* in respect of confirming ethical approval of the research, specifying the source, species and strain of animals and the inclusion of a statement on the translational relevance of the research study (Table 1, Figure 1). However, a general observation is that, although many manuscripts include lengthy details of the molecular biological or biochemical techniques that were used in the study, when experiments involved the use of animals, both the design of experiments and the description of the experimental procedures are often inadequately detailed making reproducibility challenging.

Reporting of the method for euthanasia has improved over the years but is still inconsistent, despite the need for manuscripts to be clear about when and how animals were killed. Information on housing and husbandry is often provided but, again, the level of detail is highly variable. Important elements, such as stocking density, configuration of group-housed animals (in respect of littermates, genotypes, or randomised mixed-caging, for example) provision of food (including, ideally, the composition of the laboratory diet) and water and environmental enrichment are not always disclosed, despite increasing recognition that these factors can affect the research findings (Finney, Proschogo *et al*., 2020; Reardon, 2016). The poor reporting of the provision of analgesia in the surgical context is particularly worrying. Of course, a lack of any mention of analgesia in the manuscript does not mean that none was provided; there could also be a strong scientific justification for withholding analgesia in some experiments. However, that did not apply to any of the manuscripts that were reviewed: in all those cases, analgesia could and should have been provided.

THE BJP RESPONSE TO ARRIVE 2.0

Two options were considered by the Senior Editorial team in December 2019: (i) endorse ARRIVE 2.0 guidelines and adopt them *verbatim* as official BJP policy or (ii) endorse ARRIVE 2.0 but devise our own reporting policy, which would be particularly relevant for pharmacologists. In considering the results of our surveys and the content of the ARRIVE 2.0 preprint, the team decided on the latter approach.

Authors should be reassured that nothing major has changed in practice. This journal has always taken inspiration from ARRIVE, while adopting a 'bespoke' approach to the reporting of the types of animal research carried out by pharmacologists. Authors should also note that requirements in respect of experimental design and analysis in articles published in BJP have not changed at all and can be found at: https://bpspubs.onlinelibrary.wiley.com/doi/full/10.1111/bph.14153 (Curtis, Alexander *et al*., 2018). That said, we strongly advise authors to use the NC3Rs Experimental Design Assistant (https://www.nc3rs.org.uk/experimental-design-assistant-eda) (Percie du Sert, Bamsey et al., 2017) when planning the experimental design and the data analysis.

THE BJP POLICY FOR THE FUTURE

The most striking change in the updated ARRIVE guidelines is the subdivision of the original checklist into two tables / checklists: the *'Essential 10'* and the *'Recommended set'*. The former focuses mainly on aspects of the experimental design that are essential for reducing bias (subjective or systematic), whereas the latter deals with points that pertain to the factors that authors are advised to disclose in order to enable others to replicate the experiment. Our current guidelines for experimental design and statistical analysis address many of the *'Essential 10'*, and for our modified approach we concentrate on many of the items in the *'Recommended set'.*

Our updated *'BJP Declaration of Transparency and Scientific Rigour: Checklist for Animal Experimentation'* is shown in Table 2, along with indicators of how each element maps onto ARRIVE 2.0. These amendments do not challenge the importance of either version of ARRIVE. On the contrary, we strongly recommend that all authors read ARRIVE 2.0 (Percie du Sert et al, 2020a) and familiarise themselves with the complementary *'Explanation and Elaboration'* document (Percie du Sert et al, 2020b).

In terms of details that should be included in every manuscript submitted to the BJP, the new checklist for reporting animal research will consist of the following six mandatory elements (see also, Table 2):

Animal demographics: Details of the source, species, strain, sex, age and/or weight range of the animals used in the study must be given.

Experimental procedures: A comprehensive account of the experimental procedure with particular attention to the use of any pharmacologically active agents at any stage of the study, including: anaesthesia, analgesia, antibiotics or any veterinary treatment administered for welfare purposes. Authors should also confirm that, for all surgical procedures, precautions to ensure aseptic conditions were applied throughout. Details of post-surgical analgesia and care of the animals should be provided, as should any procedures for monitoring the animals during post-operative recovery: e.g., the frequency and duration of the observations, temperature control *etc*. If analgesia cannot be given, this needs to be justified scientifically. A statement to confirm that the animals were killed by an individual trained to carry out euthanasia, together with a description of how and when animals are killed, must also be included. In this context, authors should bear in mind that only certain methods for euthanasia are permitted in the UK (and other countries governed by Directive 2010/63/EU) and these are stringently regulated. As a consequence, it might not be possible to consider a manuscript for publication in this journal if euthanasia has been carried out using a procedure that is regarded as unacceptable in the UK.

One notable change to our policy is that whenever welfare assessments, or other precautions, were necessary during any stage of the research, these must now be reported, together with a clear definition of humane end-points. We acknowledge that these aspects of experimental conduct may well be implemented routinely in individual laboratories but, hitherto, they have not been perceived as an essential aspect of the research report. We want authors to include this information, as a matter of course, not least because other researchers in the field would find it important and helpful.

Justification for validity of the choice of animal species or model: Full details of the specific animal model that was used and the scientific justification for the choice of that model in the context of achieving the research objectives should be provided. This aspect of the report can be challenging, but we hope to encourage authors to be conservative and realistic when claiming to have used an animal 'model' of a complex human disorder (see next section).

Ethical review: A statement to confirm that the research reported in the manuscript was awarded local ethical approval is essential. BJP seeks to publish research on animals that conforms to standards upheld in the UK.

Housing and husbandry: Some elements should always be included, for example: type of housing, type of cage (open or individually ventilated) stocking density, food and water provision, bedding material, environmental enrichment, lighting regimen. For fish, details of the tank and the number of fish in each one should be included. Information on other elements should be included if they are crucial for the study or if they could affect the experimental results: e.g., the time of day the study took place, nature of handling/capture. The configuration of animals within cages should be specified, also.. For instance, mixed or separate housing of genetically-altered / wildtypes or disease-susceptible / normal animals

Animal welfare: Any welfare-related assessments, measurements and interventions that were carried out before, during, or after the study, especially if these were intended to ameliorate or limit the harms to the animals (e.g., humane end-points) must be reported.

Another change is that both ARRIVE 2.0 and our revised BJP Declaration have removed the obligation to report any advance in the 3Rs (*Replacement, Reduction* and *Refinement*) that has emerged from the research. This is partly because compliance with this rubric has been negligible. Nevertheless, the BJP Editorial Board recognises the importance of the 3Rs as part of the ethical framework for all research that uses animals and so we encourage authors to include this information whenever their findings have made a useful contribution to any of the 3Rs.

ANIMAL MODELS AND SCIENTIFIC VALIDITY

The ARRIVE 2.0 guidelines are certainly an important step in helping to address the problem of poor reproducibility and translation of research findings in the biomedical sciences. However, successful translation also depends on a sound rationale for the research. So far, the attempts to improve reproducibility have somewhat overshadowed the growing scepticism about the validity of 'animal models' of some human disorders.

In many cases, the creation of an animal model in order to carry out experiments, which would not be permitted in humans, is an entirely reasonable objective. That would be the case for human disorders that have an established link with a single genetic abnormality, such as cystic

fibrosis, Down syndrome and hypertrophic cardiomyopathy. Even in cases where a specific genetic mutation is a common, but not invariable, cause of the disorder (e.g., Fragile X syndrome, amyotrophic lateral sclerosis), it is still reasonable to back-translate the mutation to create an animal model to investigate the underlying biological abnormalities and potential treatments. Few people would argue that those research findings are not instructive or that the description of the animal as a 'model' of the human disorder is scientifically unjustified.

Even when there is no such clear causal link, it is valid to describe an animal as a 'model' when it expresses an abnormality that is convincingly analogous to the diagnostic biomarker(s) in humans. Examples of these would include: neoplasia, Cushing's / Addison's disease, diabetes, skin pigmentation diseases, myasthenia gravis. Even though treatment strategies that are effective in the animals often do not translate into humans, it is still reasonable to regard them as 'models' of the human condition and such research has contributed a great deal to our understanding of the illnesses.

It is far more difficult to be confident about the validity of animal models of complex, multifactorial human illnesses, which can comprise a mixture of any of several pathological, physiological and behavioural abnormalities, which can be primary or secondary features of the illness. These would include conditions such as: hypertension, heart failure, renal failure, metabolic syndrome, neurological and psychiatric disorders. In all these cases, the profile and/or severity of the underlying problems can differ substantially from patient to patient, but still meet the criteria for the same broad clinical diagnosis.

In such cases, we hope to encourage authors to be more circumspect about claims that their 'animal model' is analogous to the human condition. This is especially important for 'models' that are based on evidence from procedures that are normally used as drug screens to predict therapeutic efficacy in humans (see, for example: Stanford, 2017). In such cases, authors should justify their assumption that the abnormality that is being evaluated in the baseline condition (i.e, with no experimental intervention such as drug treatment or genetic alteration) is a valid model of the human illness or a particular aspects of that illness. On the other hand, we want to discourage non-committal descriptions such as 'cirrhosis-like', 'epilepsy-like' or 'autism-like', or vague terms, such as 'asthma' instead of 'respiratory allergy'. Instead, our aim is to encourage

authors to be precise about the extent to which aspect(s) of the human disorder are expressed in the animal model, and also to acknowledge its limitations. Where these issues have been very well rehearsed for any particular model, it is appropriate for authors to cite the relevant literature explaining why the animal model is appropriate.

We believe this is timely advice because there is burgeoning interest in research of 'endophenotypes' in which a specific aspect of an animal's normal / abnormal physiology or behaviour can be mapped onto an underlying genetic mutation or physiological system (e.g., neuronal network, or other physiological feedback loop) – and which are often not confined to a single human disorder. This approach further acknowledges that some human illnesses are better regarded as assemblies of endophenotypes, rather than unitary disorders. Obviously, the definition of an endophenotype also needs scrupulous validation, but this change of mindset will be essential for successful progress in stratified and personalised medicine.

A final point, addressed only indirectly by ARRIVE 2.0, concerns the validity of some research procedures that are used to produce the 'model' (see: Stanford, 2020). Examples include some experimental interventions to induce end-stage heart failure, or the use of environmental stressors that are intended to alter the physiology and/or behaviour of the animals. Examples include the use of environmental stressors, such as electric shocks or a series of unpredictable stressors that change from day to day, sometimes for several weeks. In most cases, the individual stressors are mild and can arguably be regarded as analogous to challenges faced by some humans and which impair their mental health, but this is not always the case. Some of these experiments involve prolonged bouts of stressors, which would be regarded as severe, especially when their cumulative harm is taken into account. In such cases, the BJP will require authors to provide assurance that the severity and duration of the aversive stimuli was the minimum required to meet the scientific objectives of the study, particularly when using a series of unpredictable stressors, for instance. Also, in order to underpin the validity of such animal 'models', authors should include a statement to clarify the ways in which the aversive stimuli used in the research are relevant to the manifestation of the disorder of interest, which is being modelled in the animals.

Concluding remarks

In response to the publication of the revised ARRIVE guidelines (ARRIVE 2.0), the senior editorial board of this journal has revised the policy on reporting of animal research. The primary principle remains the same as before, which is to encourage full disclosure of all relevant information and to expect to see evidence for high standards of animal welfare. It is important to remember that this is not simply for the benefit of the animals. High standards of experimental design, reporting and animal welfare are crucial if the research is to be reproducible and translatable. In that context, we encourage authors to read Drummond and Fowler (2013).

The journal has, for some time, taken the view that the prevailing standards expected by the regulatory framework in the UK should be reflected in the journal's publications. For example, studies of tobacco products are not permitted in the UK and the BJP will not publish them. Similarly, death as an endpoint, in non-regulatory studies would not normally be acceptable (with the exception of studies of anti-cancer and certain antimicrobial drugs, for example), but when current regulatory guidelines still mandate mortality endpoints, we expect humane endpoints to be described fully and implemented. Some experiments that involve procedures that would be regarded as 'severe' but can be scientifically justified, nonetheless. However, the justification for some such procedures might not be clear. In most of these cases, a discussion between the editor(s) and the authors would seek to confirm whether or not there was a robust justification using such procedures, or humane endpoints, regardless of any confirmation of local ethical approval for the study. If this turns out to be the case, the manuscript may be published.

Finally, raising the standard of research reporting will certainly help to improve reproducibility but, to improve translation will also require a more realistic, evidence-based appraisal of the validity of the experimental procedures and research 'models'. To that end, we want to encourage authors to consider the extent to which the models they have investigated really do offer insight into the causes, pathology and treatment of multifactorial human disorders. In cases where there is any doubt, authors should be assured that more modest objectives and clarity about the specific features of the disorder that have been 'modelled' would attract even more confidence and scientific merit.

References

Avey MA-O, Moher D, Sullivan KJ, Fergusson D, Griffin G, Grimshaw JM, *et al.* (2016). The Devil Is in the Details: Incomplete Reporting in Preclinical Animal Research. PLOS ONE 11**:** e0166733.

Curtis MJ, Alexander S, Cirino G, Docherty JR, George CH, Giembycz MA, *et al.* (2018). Experimental design and analysis and their reporting II: updated and simplified guidance for authors and peer reviewers. British Journal of Pharmacology 175**:** 987-993.

Drummond GB, Vowler Sl (2013). Do as you would be done by: write as you would wish to read. British Journal of Pharmacology 168 1043–1047.

Finney CA, Proschogo NW, Snepvangers A, Holmes NM, Westbrook RF, & Clemens KJ (2020). The effect of standard laboratory diets on estrogen signaling and spatial memory in male and female rats. 215.

Kilkenny C, Browne WJ, Cuthill IC, Emerson M, & Altman DG (2010a). Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLoS Biol 8**:** e1000412.

Kilkenny C, Browne W Fau - Cuthill IC, Cuthill Ic Fau - Emerson M, Emerson M Fau - Altman DG, & Altman DG (2010b). Animal research: reporting in vivo experiments: the ARRIVE guidelines. British Journal of Pharmacology 160**:** 1577-1579.

Leung V, Rousseau-Blass F, Beauchamp G, & Pang DSJ (2018). ARRIVE has not ARRIVEd: Support for the ARRIVE (Animal Research: Reporting of in vivo Experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. PLOS ONE 13**:** e0197882.

McGrath JC, Lilley E (2015). Implementing Guidelines on Reporting Research Using Animals (ARRIVE Etc.): New Requirements for Publication in BJP. British Journal of J Pharmacology 172(13):3189-93.

Percie du Sert N, Bamsey I, Bate ST, Berdoy M, Clark RA, Cuthill I, *et al.* (2017). The Experimental Design Assistant. PLOS Biology 15**:** e2003779.

Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Altman DG, Avey MT, *et al.* (2018). Revision of the ARRIVE guidelines: rationale and scope. BMJ Open Science 2**:** e000002.

Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, *et al.* (2019). The ARRIVE guidelines 2019: updated guidelines for reporting animal research. bioRxiv**:** 703181.

Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, Clark A, Cuthill IC, Dirnagl U, Emerson M, Garner P, Holgate ST, Howells DW, Karp NA, Lazic SE, Lidster K, MacCallum CJ, Macleod M, Pearl EJ, Petersen O, Rawle F, Reynolds P, Rooney K, Sena ES, Silberberg SD, Steckler T, & Würbel H (2020a). The ARRIVE guidelines 2.0: updated guidelines for reporting animal research. PLoS Biol. doi: 10.1371/journal.pbio.3000410

Percie du Sert N, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, Clark A, Cuthill IC, Dirnagl U, Emerson M, Garner P, Holgate ST, Howells DW, Hurst V, Karp NA, Lazic SE, Lidster K, MacCallum CJ, Macleod M, Pearl EJ, Petersen O, Rawle F, Reynolds P, Rooney K, Sena ES, Silberberg SD, Steckler T, & Würbel H (2020b). Reporting animal research: Explanation and Elaboration for the ARRIVE guidelines 2.0. PLoS Biol. doi: 10.1371/journal.pbio.3000411

Reardon S (2016). A mouse's house may ruin experiments. Nature 530.

Stanford SC (2017) Confusing preclinical (predictive) drug screens with animal 'models' of psychiatric disorders, or 'disorder-like' behaviour, is undermining confidence in behavioural neuroscience. J Psychopharmacol. 31(6):641-643.

Stanford SC (2020) Some Reasons Why Preclinical Studies of Psychiatric Disorders Fail to Translate: What Can Be Rescued from the Misunderstanding and Misuse of Animal 'Models'? Altern Lab Anim. (in press)