

# Protocol for a systematic review of guidelines for rigour in the design, conduct and analysis of biomedical experiments involving laboratory animals

Jan Vollert,<sup>1</sup> Esther Schenker,<sup>2</sup> Malcolm Macleod,<sup>3</sup> Anton Bernalov,<sup>4,5</sup> Hanno Wuerbel,<sup>6</sup> Martin Christian Michel,<sup>4,7</sup> Ulrich Dirnagl,<sup>8</sup> Heidrun Potschka,<sup>9</sup> Kimberley E Wever,<sup>10</sup> Thomas Steckler,<sup>11</sup> Bruce Altevogt,<sup>12</sup> Andrew S C Rice,<sup>1</sup> on behalf of the EQIPD WP3 study group

**To cite:** Vollert J, Schenker E, Macleod M, *et al.* Protocol for a systematic review of guidelines for rigour in the design, conduct and analysis of biomedical experiments involving laboratory animals. *BMJ Open Science* 2018;2:e000004. doi:10.1136/bmjos-2018-000004

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjos-2018-000004>).

Received 15 February 2018  
Accepted 6 June 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Jan Vollert;  
j.vollert@imperial.ac.uk

## ABSTRACT

**Objective** Within the last years, there has been growing awareness of the negative repercussions of unstandardized planning, conduct and reporting of preclinical and biomedical research. Several initiatives have set the aim of increasing validity and reliability in reporting of studies and publications, and publishers have formed similar groups. Additionally, several groups of experts across the biomedical spectrum have published experience and opinion-based guidelines and guidance on potential standardized reporting. While all these guidelines cover reporting of experiments, an important step prior to this should be rigours planning and conduction of studies. The aim of this systematic review is to identify and harmonize existing experimental design, conduct and analysis guidelines relating to internal validity and reproducibility of preclinical animal research. The review will also identify literature describing risks of bias pertaining to the design, conduct and analysis of preclinical biomedical research.

**Search strategy** PubMed, Embase and Web of Science will be searched systematically to identify guidelines published in English language in peer-reviewed journals before January 2018 (box 1). All articles or systematic reviews in English language that describe or review guidelines on the internal validity and reproducibility of animal studies will be included. Google search for guidelines published on the websites of major funders and professional organisations can be found in (Box 2).

**Screening and annotation** Unique references will be screened in two phases: screening for eligibility based on title and abstract, followed by screening for definitive inclusion based on full text. Screening will be performed in SyRF (<http://syrf.org.uk>). Each reference will be randomly presented to two independent reviewers. Disagreements between reviewers will be resolved by additional screening of the reference by a third, senior researcher.

**Data management and reporting** All data, including extracted text and guidelines, will be stored in the SyRF platform. Elements of the included guidelines will be identified using a standardized extraction form.

Reporting will follow the PRISMA guidelines as far as applicable.

## INTRODUCTION

Within the last years, there has been growing awareness of the negative repercussions of unstandardised planning, conduct and reporting of preclinical and biomedical research.<sup>1 2</sup> Several initiatives have set the aim of increasing validity and reliability in reporting of (not only preclinical) studies and publications, such as CAMARADES (Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies),<sup>3</sup> NC3Rs (The National Centre for the 3Rs),<sup>4</sup> SYRCLE (SYstematic Review Center for Laboratory animal Experimentation)<sup>5</sup> and the EQUATOR (Enhancing the QUALity and Transparency Of health Research) network.<sup>6</sup> Publishers have formed similar groups (eg, *The Lancet's* REWARD (REduce research Waste And Reward Diligence) initiative<sup>7</sup>). Additionally, several experts or groups of experts across the biomedical spectrum, both clinical and preclinical, have published experience and opinion based guidelines and guidance on potential standardised reporting.<sup>8–10</sup> While many of the points raised are identical or similar between these various guidelines (in fact many experts on the field are part of more than one initiative), they differ in details, rigour and show especially distinct variance in generalisability or specific challenges for a single field. While all these guidelines cover reporting of experiments, an important step prior to this should be rigours planning and conduction of studies, which faces a similar situation.<sup>11</sup> Consequently,

## Box 1 Search string

*Web of Science*

(guideline OR recommendation OR recommendations) AND (“preclinical model” OR “preclinical models” OR “disease model” OR “disease models” OR “animal model” OR “animal models” OR “experimental model” OR “experimental models” OR “preclinical study” OR “preclinical studies” OR “animal study” OR “animal studies” OR “experimental study” OR “experimental studies”)

*PubMed*

((Consensus[mh] OR Consensus development conferences as topic[mh] OR Guidelines as topic [Mesh] OR Practice guidelines as topic[mh] OR guideline[pt] OR practice guideline[pt] OR consensus development conference[pt] OR position statement\*[tiab] OR policy statement\*[tiab] OR practice parameter\*[tiab] OR best practice\*[tiab] OR standards[ti] OR guideline[ti] OR guidelines[ti] OR recommendation[ti] OR recommendations[ti]) AND (“Animal Experimentation”[Mesh] OR “Models, Animal”[Mesh] OR Preclinical model[tiab] OR Pre-clinical model[tiab] OR Preclinical models[tiab] OR Pre-clinical models[tiab] OR disease model[tiab] OR disease models[tiab] OR animal model[tiab] OR animal models[tiab] OR experimental model[tiab] OR experimental models[tiab] OR preclinical study[tiab] OR pre-clinical study[tiab] OR preclinical studies[tiab] OR pre-clinical studies[tiab] OR animal study[tiab] OR animal studies[tiab] OR animal experiment\*[tiab] OR experimental study[tiab] OR experimental studies[tiab])) OR ((Consensus[mh] OR Consensus development conferences as topic[mh] OR Guidelines as topic [Mesh] OR Practice guidelines as topic[mh] OR guideline[pt] OR practice guideline[pt] OR consensus development conference[pt] OR position statement\*[tiab] OR policy statement\*[tiab] OR practice parameter\*[tiab] OR best practice\*[tiab] OR standards[ti] OR guideline[ti] OR guidelines[ti] OR recommendation[ti] OR recommendations[ti]) AND ((Preclinical[tiab] OR Pre-clinical[tiab] OR Experimental[tiab] OR animal[tiab]) AND (Study[tiab] OR Studies[tiab] OR Model[tiab] OR Models[tiab]) AND animals[Mesh:noexp])) OR (((“Methods”[Mesh] OR “methods”[Subheading]) AND (tool[ti] OR protocol[ti])) AND (“Animal Experimentation”[Mesh] OR “Models, Animal”[Mesh] OR ((Preclinical[tiab] OR Pre-clinical[tiab] OR Experimental[tiab] OR animal[tiab]) AND (Study[tiab] OR Studies[tiab] OR Model[tiab] OR Models[tiab])) AND animals[Mesh:noexp])) OR ((position statement\*[tiab] OR policy statement\*[tiab] OR practice parameter\*[tiab] OR best practice\*[tiab] OR standards[ti] OR guideline[ti] OR guidelines[ti] OR recommendation[ti] OR recommendations[ti]) AND ((Preclinical[tiab] OR Pre-clinical[tiab] OR Experimental[tiab] OR animal[tiab]) AND (Study[tiab] OR Studies[tiab] OR Model[tiab] OR Models[tiab])) NOT medline[sb])

*Embase*

(Consensus/ or consensus development/ or practice guideline/ or position statement\*.ti,ab,kw. OR policy statement\*.ti,ab,kw. OR practice parameter\*.ti,ab,kw. or best practice\*.ti,ab,kw. OR standards.ti. OR guideline.ti. OR guidelines.ti. OR recommendation.ti. OR recommendations.ti.) AND (exp animal experiment/ or exp animal model/ or Preclinical model.ti,ab,kw. OR Pre-clinical model.ti,ab,kw. OR Preclinical models.ti,ab,kw. OR Pre-clinical models.ti,ab,kw. OR disease model.ti,ab,kw. OR disease models.ti,ab,kw. OR animal model.ti,ab,kw. OR animal models.ti,ab,kw. OR experimental model.ti,ab,kw. OR experimental models.ti,ab,kw. OR preclinical study.ti,ab,kw. OR pre-clinical study.ti,ab,kw. OR preclinical studies.ti,ab,kw. OR pre-clinical studies.ti,ab,kw. OR animal study.ti,ab,kw. OR animal studies.ti,ab,kw. OR animal experiment\*.ti,ab,kw. OR experimental study.ti,ab,kw. OR experimental studies.ti,ab,kw.) OR ((Consensus/ or

Continued

## Box 1 Continued

consensus development/ or practice guideline/ or position statement\*.ti,ab,kw. OR policy statement\*.ti,ab,kw. OR practice parameter\*.ti,ab,kw. or best practice\*.ti,ab,kw. OR standards.ti. OR guideline.ti. OR guidelines.ti. OR recommendation.ti. OR recommendations.ti.) AND ((Preclinical.ti,ab,kw. OR Pre-clinical.ti,ab,kw. OR Experimental.ti,ab,kw. OR animal.ti,ab,kw.) adj2 (Study.ti,ab,kw. OR Studies.ti,ab,kw. OR Model.ti,ab,kw. OR Models.ti,ab,kw.)) AND animal.mp.) OR ((methodology/ or experimental design/ or study design/) and (tool.ti. or protocol.ti.) and (exp animal experiment/ or exp animal model/ or ((Preclinical.ti,ab,kw. OR Pre-clinical.ti,ab,kw. OR Experimental.ti,ab,kw. OR animal.ti,ab,kw.) adj2 (Study.ti,ab,kw. OR Studies.ti,ab,kw. OR Model.ti,ab,kw. OR Models.ti,ab,kw.))) AND animal.mp.)

it is hard for researchers to decide which guidelines to follow, especially at the stage of planning future studies.

The aim of this systematic review is to identify and harmonise existing experimental design, conduct and analysis guidelines relating to preclinical animal research. The review will also identify literature describing (either through primary research or systematic review) risks of bias pertaining to the design, conduct and analysis of preclinical biomedical research. Reporting standards will be considered if they refer to topics that are important to consider at planning or experimental, not only at reporting, stage. This review will focus on internal validity and reproducibility of experimental design, conduct and analysis. While we realise that factors like animal housing and welfare are

## Box 2 List of funders and organisations

*Professional neuroscientific organisations*

Society for Neuroscience (USA)  
Cognitive Neuroscience Society (USA)  
American College for Neuropsychopharmacology (USA)  
Federation of European Neuroscience Societies (EU)  
European Brain and Behaviour Society (EU)  
European College of Neuropsychopharmacology (EU)  
British Neuroscience Association (UK)

*Major funders*

National Institute of Health & Howard Hughes Medical Institute (USA)  
Chinese Academy of Sciences & National Natural Sciences Foundation of China (China)  
Japan Society for the Promotion of Science & Japan Neuroscience Society (Japan)  
European Research Council & Horizon 2020 & Innovative Medicines Initiative (EU)  
Wellcome Trust & Medical Research Council (UK)  
Deutsche Forschungsgemeinschaft (Germany)  
L'agence Nationale de la Recherche & Pasteur Foundation (France)  
Dirección General de Investigación Científica y Técnica & Instituto de Salud Carlos III (Spain)  
Ministry of Instruction, Universities, and Research (Italy)  
Ministry of Education and Science, Russian Science Foundation and Russian Foundation for Fundamental Research (Russia)  
Ministry of Science and Higher Education (Poland)  
Swiss National Science Foundation (Switzerland)  
ZonMw (Netherlands)

**Box 3 Extraction form**

1. Matching or balancing treatment allocation of animals.
2. Matching or balancing sex of animals across groups.
3. Standardised handling of animals.
4. Randomised allocation of animals to treatment.
5. Randomisation for analysis.
6. Randomised distribution of animals in the animal facilities.
7. Monitoring emergence of confounding characteristics in animals.
8. Specification of unit of analysis.
9. Addressing confounds associated with anaesthesia or analgesia.
10. Selection of appropriate control groups.
11. Concealed allocation of treatment.
12. Study of dose–response relationships.
13. Use of multiple time points measuring outcomes.
14. Consistency of outcome measurement.
15. Blinding of outcome assessment.
16. Establishment of primary and secondary end points.
17. Precision of effect size.
18. Management of conflicts of interest.
19. Choice of statistical methods for inferential analysis.
20. Recording of the flow of animals through the experiment.
21. A priori statements of hypothesis.
22. Choice of sample size.
23. Addressing confounds associated with treatment.
24. Characterisation of animal properties at baseline.
25. Optimisation of complex treatment parameters.
26. Faithful delivery of intended treatment.
27. Degree of characterisation and validity of outcome.
28. Treatment response along mechanistic pathway.
29. Assessment of multiple manifestations of disease phenotype.
30. Assessment of outcome at late/relevant time points.
31. Addressing treatment interactions with clinically relevant comorbidities.
32. Use of validated assay for molecular pathways assessment.
33. Definition of outcome measurement criteria.
34. Comparability of control group characteristics to those of previous studies.
35. Reporting on breeding scheme.
36. Reporting on genetic background.
37. Replication in different models of the same disease.
38. Replication in different species or strains.
39. Replication at different ages.
40. Replication at different levels of disease severity.
41. Replication using variations in treatment.
42. Independent replication.
43. Addressing confounds associated with experimental setting.
44. Addressing confounds associated with setting.
45. Preregistration of study protocol and analysis procedures.
46. Pharmacokinetics to support treatment decisions.
47. Definition of treatment.
48. Interstudy standardisation of end point choice.
49. Define programmatic purpose of research.
50. Interstudy standardisation of experimental design.
51. Research within multicentre consortia.
52. Critical appraisal of literature or systematic review during design phrase.
53. (Multiple) free text.

highly important for reproducibility of experiments, they will not be considered in this initial systematic review (SR), which focuses on internal validity. It is planned to analyse

the influence of animal care and use at a later point in a separate SR.

**INCLUSION AND EXCLUSION CRITERIA**

This study will include all articles or systematic reviews in English language that describe or review guidelines on the internal validity (‘to what extent do the study results reflect a true cause–effect of the intervention?’) and reproducibility of the design, conduct and analysis of preclinical animal studies. Articles that focus strictly on toxicity or veterinary drugs only will not be included. Literature not focussing on guidelines, but describing (either through primary research or systematic review) risks of bias pertaining to the design, conduct and analysis of preclinical biomedical research, will also be considered. Although reporting standards are not the key primary objective of this systematic review these will also be searched, screened and extracted, as they can contain useful information that should be considered not only at reporting, but already at planning or experimental stage.

**SCREENING AND ANNOTATION**

After combining the search results from all sources, potential duplicates or publication of identical guidelines by the same author group in various journals will be identified prior to screening, based on PubMed ID, digital object identifier (DOI) and title, journal and author list. Unique references will then be screened in two phases: (1) screening for eligibility based on title and abstract, followed by (2) screening for definitive inclusion based on full text. Screening will be performed in SyRF (<http://syrf.org.uk>). Each reference will be randomly presented to two independent reviewers. Reviewers are not blinded to the authors of the presented record. In the first stage, two authors will screen the title and abstract of the retrieved records for eligibility based on predefined inclusion criteria (see above). The title/abstract screening stage will focus on sensitivity (‘could the paper be of any interest?’).

Articles included after the title-abstract screening will undergo concurrent full-text screening for definitive inclusion. We will attempt to obtain full-text versions of all included articles through open access, interlibrary loan or by contacting authors directly. Articles for which no full-text version can be obtained will be excluded from the review.

In both screening stages, disagreements between reviewers will be resolved by additional screening of the reference by a third, senior researcher, who is blind to the individual judgements of the first two reviewers.

**DATA MANAGEMENT**

All references returned from the searches will be downloaded, with entries organised by DOI (if available, or weblink alternatively), publication date, and title. All data,



including extracted text and guidelines, will be stored in the SyRF platform.

## STUDY QUALITY, META-ANALYSIS AND RISK OF BIAS ASSESSMENT

These typical stages of systematic reviews are not relevant for this study, as it focusses on guidelines rather than experimental data.

However, both reviewers will rate extracted guidelines rated based on the following system (Ia being the lowest level of provenance, IIIb being the highest):

- I. Recommendations of individuals or small groups of individuals based on individual experience only
  - a. Published stand alone.
  - b. Endorsed or initiated by at least one publisher or scientific society.
- II. Recommendations by groups of individuals, including a Delphi process
  - a. Published stand alone.
  - b. Endorsed or initiated by at least one publisher or scientific society.
- III. Recommendations based on a systematic review
  - a. Published stand alone.
  - b. Endorsed or initiated by at least one publisher or scientific society.

Across guidelines, the elements will be ranked based on the frequency of appearance across the included guidelines.

## REPORTING

Elements of the included guidelines will be identified using the extraction form from [box 3](#). Additionally, reporting will follow the Preferred Reporting Items for Systematic Review and Meta-analysis guidelines as far as applicable.

### Author affiliations

<sup>1</sup>Pain Research, Department of Surgery and Cancer, Imperial College London, London, UK

<sup>2</sup>Institut de recherches, Servier, Croissy-sur-Seine, France

<sup>3</sup>Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

<sup>4</sup>Partnership for Assessment and Accreditation of Scientific Practice, Heidelberg, Germany

<sup>5</sup>Valdman Institute of Pharmacology, Pavlov Medical University, Saint Petersburg, Russia

<sup>6</sup>Division of Animal Welfare, VPH Institute, Vetsuisse Faculty, University of Bern, Bern, Switzerland

<sup>7</sup>Universitätsmedizin Mainz, Johannes-Gutenberg-Universität Mainz, Mainz, Germany

<sup>8</sup>Department of Experimental Neurology, Charité Universitätsmedizin Berlin, Berlin, Germany

<sup>9</sup>Institute of Pharmacology, Toxicology, and Pharmacy, Ludwig-Maximilians-University, Munich, Germany

<sup>10</sup>Department for Health Evidence, Systematic Review Centre for Laboratory Animal Experimentation, Nijmegen Institute for Health Sciences, Radboud university medical centre, Nijmegen, The Netherlands

<sup>11</sup>Janssen Pharmaceutica NV, Beerse, Belgium

<sup>12</sup>Pfizer Inc, New York City, New York, USA

**Acknowledgements** This work is part of the European Quality In Preclinical Data (EQIPD) consortium.

**Collaborators** The EQIPD WP3 study group members are: Judi Clark, Emily Sena, Bruno Boulanger, Gernot Riedel, Bettina Platt, Annesha Sil, Martien J Kas, Bernhard Voelkl, Mathias Jucker, Bettina M Wegenast-Braun, René Bernard, Esmeralda Heiden, Ann-Marie Waldron, Maarten Loos, Merel Ritskes-Hoitinga, Tom Van De Castele, Pim Drinkenburg, Juan Diego Pita Almenar, David Gallacher, Henk Van Der Linde, Anja Gilis, Greet Teuns, Karsten Wicke, Sabine Grote, Bernd Sommer, Janet Nicholson, Sanna Janhunen, Sami Virtanen, Kristin Cheng, Sylvie Ramboz, Emer Leahy, Isabel A Lefevre, Fiona Ducrey, Javier Guillen, Patri Vergara, Thomas Inghram.

**Contributor** JV drafted this protocol, all authors commented, corrected and agreed on the final version. All authors took part in the planning and design of this study.

**Funding** This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777364. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

**Competing interests** None declared.

**Provenance and peer review** Not commissioned, externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

## REFERENCES

1. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* 2011;10:712.
2. Kilkenny C, Parsons N, Kadoszewski E, *et al*. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One* 2009;4:e7824.
3. Hirst JA, Howick J, Aronson JK, *et al*. The need for randomization in animal trials: an overview of systematic reviews. *PLoS One* 2014;9:e98856.
4. du Sert NP, Bamsley I, Bate ST, *et al*. The experimental design assistant. *Nat Methods* 2017;14:1024–5.
5. Hooijmans CR, Rovers MM, de Vries RB, *et al*. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.
6. Simera I, Moher D, Hirst A, *et al*. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med* 2010;8:24.
7. Macleod MR, Michie S, Roberts I, *et al*. Biomedical research: increasing value, reducing waste. *Lancet* 2014;383:101–4.
8. 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;38:167–82.
9. Smith AJ, Clutton RE, Lilley E, *et al*. PREPARE: guidelines for planning animal research and testing. *Lab Anim* 2018;52:135–41.
10. Kilkenny C, Browne WJ, Cuthill IC, *et al*. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol* 2010;8:e1000412.
11. Henderson VC, Kimmelman J, Fergusson D, *et al*. Threats to validity in the design and conduct of preclinical efficacy studies: a systematic review of guidelines for in vivo animal experiments. *PLoS Med* 2013;10:e1001489.