The reproducibility crisis in preclinical research – lessons to learn from clinical research

Laia Pedro-Roig[1] and
Christoph H. Emmerich[2]
1 Trilogy Writing & Consulting GmbH, Frankfurt am Main, Germany
2 PAASP GmbH, Heidelberg, Germany

Correspondence to:

Laia Pedro-Roig
Trilogy Writing & Consulting GmbH
Falkensteiner Strasse 77
60322 Frankfurt am Main, Germany
laia.pedro.roig@trilogywriting.com
+49 69 138 2528-44

Christoph Emmerich
PAASP GmbH
Am Aukopf 14-1
69118 Heidelberg, Germany
christoph.emmerich@paasp.net
+49 696 777 8485

Abstract

In recent years, the robustness and reproducibility of preclinical data have been a topic for discussion. Quality standards and good practices are often not well defined for different in vitro methods and in vivo models, and not harmonised amongst preclinical research laboratories. This results in poorly reliable literature, has a negative impact on the bench-to-bedside time for new drugs, and increases the resources needed for clinical development. Clinical research, on the other hand, is tightly regulated and has high quality standards in place. Although improvements are slowly introduced, preclinical development (especially in its confirmatory phases) would benefit from taking a closer look and adapting more of the internationally accepted principles used in clinical research.

Reproducibility issues are gaining awareness amongst preclinical scientists: in a *Nature* survey, 52% of researchers state that there is a significant crisis.[1] The published literature is a common source for potential new drug targets used by the pharmaceutical industry, and publication results are routinely validated in-house to ensure reproducibility. According to Prinz et al., almost two-thirds of the validation projects conducted at Bayer from 2007 to 2010 showed inconsistencies in results (including some from prestigious journals).[2] Begley and Ellis reported that researchers at Amgen could only confirm the scientific findings of six out of 53 (11%) landmark studies.[3] In addition, more than 70% of the researchers who participated in the Nature survey have failed to reproduce another scientist's results, and more than 50% admitted to having failed to reproduce their own.[1]

The reproducibility crisis is a quantifiable economic problem. Venture capital firms consider that, when repeated by an independent laboratory, the experiments in at least 50% of published studies do not provide the same results.[4] In the US alone, US $28 billion per year are spent on preclinical research that is not reproducible.[5] More importantly, the lack of reproducibility has a negative impact on the bench-to-bedside time for new medicines, and increases drug development costs, as each study replication conducted by the pharmaceutical industry to validate academic research findings requires 3 to 24 months of work and US $500,000 to $2 million.[5]

Data robustness becomes even more important at later stages of preclinical research, when results determine "go/no-go" decisions for drug candidates to enter clinical testing. A meta-analysis identified higher effect sizes in animal models of stroke in studies with low quality standards.[6] This implies how low quality research standards can make drug candidates look more promising than they actually are.

What is behind the reproducibility crisis?

Participants in the Nature survey consider that the main reasons are pressure to publish and selective reporting.[1] For academics, publishing is a career essential (e.g., for research funding, job promotion, or tenure). Journal editors, referees, and grant reviewers look for the perfect story: simple, clear, and complete.[3] These demands tempt investigators to cherry pick experiments for publishing, develop hypotheses to fit the data, or keep collecting data until the desired significance level is reached (p-hacking). Competition among laboratories and pressure to publish among scientists may result in negligent controlling or reporting of experimental conditions.[2] Another issue is the bias towards publishing positive results and the difficulties in publishing results that contradict data in high-impact journals or currently established opinion (publication bias).[2] This leads to strengthening certain hypotheses, even if there is a body of unpublished evidence against them.

Published preclinical research often lacks proper quality standards in study design (e.g., blinding and randomisation) and validation of research tools that ensure the data obtained is meaningful and unbiased. Begley and Ellis observed that authors of reproducible preclinical cancer studies had paid close attention to controls, reagents, and description of the complete dataset, while in studies that could not be reproduced, data were not routinely analysed by blinded investigators and often results from only one experiment were presented.[3] According to Freedman et al., errors leading to irreproducibility of preclinical data can be due to study design, biological reagents and reference materials, laboratory protocols, and data analysis and reporting.[5]

Low quality standards can make drug candidates look more promising than they actually are.

An enduring challenge in drug development is the erroneous use and misinterpretation of preclinical data from cell lines and animal models. In vitro cell culture systems are crucial research tools for analysing complex mechanisms regulating cell biology. However, over 480 misidentified cell lines (as of November 2017) routinely used in published studies are

contaminated, very frequently with HeLa cells (list available from the International Cell Line Authentication Committee [ICLAC]).[7]

The causes behind the reproducibility crisis are not limited to a specific field (in vitro or in vivo) of preclinical research or therapeutic area. The limitations of preclinical cancer models include (i) the use of a small number of poorly characterised tumour cell lines that inadequately recapitulate human disease, (ii) the inability to capture the human tumour environment, (iii) the lack of consideration for pharmacodynamics and pharmacokinetics, (iv) the use of problematic endpoints and testing strategies, and (v) the regular exclusion of predictive biomarkers for efficacy.[3] In the amyotrophic lateral sclerosis (ALS) field, Steve Perrin and his team re-examined 100 compounds that had been identified as candidates for therapy in an ALS mouse model.[8] Most of these compounds failed to slow the disease in animals (including eight drugs that had previously looked promising, proceeded to clinical trials, and ultimately failed). These discrepancies are likely due to the low quality standards of the original publications, as most did not include statistical models to minimise experimental noise or implement blinding and randomisation procedures.

What is the current situation and what could be done?

There are no commonly accepted and followed guidelines and quality standards for preclinical research outside those intended for studies that directly support drug marketing authorisations.[9-12] Indeed, none of the existing GxP standards (Good Laboratory Practice [GLP], Good Clinical Practice [GCP], Good Manufacturing Practice [GMP], etc.) can be used to ensure high quality preclinical research outcomes. Whether conducted in an academic or industrial laboratory, this non-regulated research is, however, essential to identify and validate novel drug targets and to build the basis for successful translation of preclinical data into clinically meaningful efficacy. Thus, there is a need for new specialised Good Research Practice (GRP) guidelines that focus on study design, unbiased conduct, statistical analysis, and transparent reporting.

> The reproducibility crisis is a quantifiable economic problem.

Clinical research, on the other hand, is highly regulated and adherence to quality standards is routinely monitored. Human experimentation has strong ethical restrictions that require researchers to comply with higher research standards to avoid submitting study participants to unnecessary risks. There are several lessons that preclinical research could learn from clinical research regarding quality standards.

Lessons to learn from clinical research

Clinical research is not perfect: A recent analysis of more than 5,000 papers in eight leading medical journals showed that roughly 2% of randomised controlled clinical trials may include fabricated data or lack adequate ethical approval.[13] However, clinical research is supported by strong standards and well-established procedures, as the following examples demonstrate, which could be used in preclinical research.

Declaration of Helsinki

As the cornerstone document of clinical research ethics, the Declaration of Helsinki helps ensure that the risk to trial subjects is proportionate to the benefit expected to society. Similar codes of practice would help preclinical researchers to realise that there is an implicit responsibility in all their activities. Currently, a similar concept exists only for animal research: the 3Rs (Replacement, Reduction, and Refinement) are considered in the US Guide for the Care and Use of Laboratory Animals and the European Directive 2010/63/EU.[14,15] These guidelines encourage finding alternatives to the use of animals, using the right number of animals, refining breeding, accommodation and care, and minimising distress. Of note, "reduction" means using the minimum number of animals required

to obtain statistically significant results based on power calculations (and not less than those); the same principle is applied for sample size calculations in clinical trials.

There are several lessons that preclinical research could learn from clinical research regarding quality standards.

ICH E6 (GCP)

The International Conference for Harmonisation (ICH) guideline E6 covers ethical and scientific quality standards for designing, conducting, recording, and reporting clinical trials, and enhances data credibility. Amongst other, ICH E6 includes the following concepts:

- The Independent Ethics Committee (IEC) or Institutional Review Board (IRB) are independent bodies constituted of medical, scientific, and non-scientific members who ensure protection of the rights, safety, and well-being of the participants of a clinical trial by reviewing and approving essential trial aspects such as the protocol and its amendments, or the suitability of investigators and facilities. The Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC)[16] and the International Council for Laboratory Animal Science (ICLAS)[17] carry out a similar function in animal experimentation: promoting proper treatment and ethical use of animals in science. Other areas of preclinical research, such as cell line or in vitro work, still lack a mechanism to obtain feedback on quality and relevance.
- Adequate and accurate source documents and trial records must be maintained. Source data should be Attributable, Legible, Contemporaneous, Original, and Accurate (ALCOA principles to ensure data integrity), as well as complete. Furthermore, any changes to source data should be traceable and not obscure the original entry (i.e., audit trail should be maintained). In preclinical research, there is still no clear consensus on which is the full set of essential parameters to be recorded in a specific experiment. Furthermore, the use of lab notebooks as tools to record research results is not standardised, and practices such as data witnessing are often not implemented.
- GCP states that qualified individuals should be involved in the conduct of a trial. Frequently, errors in preclinical research result from the incorrect analysis of data, suggesting biostatisticians to have a more relevant position in the experimental planning and analysis of preclinical data. In addition, medical writers as specialists in guidelines and good reporting practices have become essential in clinical research, and could equally contribute to enhance quality standards in preclinical research reporting.
- Audits serve to evaluate trial conduct and compliance with the protocol, standard operating procedures (SOPs), GCP, and regulatory requirements. In non-regulated preclinical research, monitoring of compliance with quality requirements that ensure unbiased conduct of research is increasingly becoming the focus of discussion, and has already been performed at contract research organisations (CROs) offering preclinical services. Furthermore, these routine audits could be interesting for agencies funding preclinical research.

As mentioned, preclinical research studies intended to support drug marketing applications are governed by strict regulations set by GxP.[9-12] However, these standards are not suitable for non-regulated, preclinical biomedical research and there is a need for the specialised set of GRP guidelines already discussed. Regarding in vitro cell culture, Good Cell Culture Practice (GCCP) (principles for standardisation, rationalisation, and international harmonisation of cell and tissue culture laboratory practices) has already been defined.[18] Nevertheless, consensus procedures for unambiguous authentication and identification of cell lines are still missing, and cell line misidentification, contamination, and genotypic and phenotypic instability remain issues.

ICH E8 and ICH E9

The ICH E8 guideline ("General Considerations for Clinical Trials") provides recommendations for the design, methodology, and analysis of clinical trials, and ICH E9 ("Statistical Principles for Clinical Trials") attempts to harmonise the principles of statistical methodology applied to them. Recently, international research consortia started to conduct so-called preclinical Phase III trials (i.e., multicentre, randomised, blinded animal studies) to test drug efficacy. These preclinical trials allow larger sample sizes and reduce bias, thus improving robustness and translational predictability. They also address the reaction norm issue (whether response of an organism to an experimental treatment can be affected by environmental factors such as food and housing conditions).[19] Trials combining data from different centres with slightly different environmental conditions are well suited to analyse the robustness of effects and the reproducibility of in vivo experiments.

Transparency

The EMA Policy 70 is an attempt to enhance transparency by publishing clinical data for medicines once the decision making process on an application for an EU-wide marketing authorisation is complete.[20] This implies having open access to full datasets from those trials. In similar ways, some journals publishing biomedical preclinical research have now implemented "open data" policies: publications need to include full datasets, biological properties of all samples, and complete methodology. The Transparency and Openness Promotion (TOP) guidelines advise journals and funding agencies on how to incentivise transparency in planning and reporting preclinical research.[21]

Registration of clinical trials

Clinical trials need to be registered, as this avoids reporting bias, a common problem in preclinical research.[22] Notably, an increasing number of journals in preclinical research now offer the "Registered Reports" publishing format, in which peer review is conducted prior to data collection, based on the importance of the research question and the quality of the methodology. Article acceptance for publication is ensured unless quality assurance or unresolvable reporting problems arise.

Reporting guidelines

Many journals require that authors follow the Consolidated Standards of Reporting Trials (CONSORT) statement, an evidence-based, minimum set of recommendations for complete and transparent reporting of randomised controlled trials.[23] Several reporting guidelines have been developed for preclinical research, including National Institutes of Health's (NIH) Principles and Guidelines for Reporting Preclinical Research, Nature's checklist, the Animal Research: Reporting of In Vivo Experiments (ARRIVE) guidelines,[24] and the Cell Press's Structured Transparent Accessible

Reporting (STAR) Methods.[25] They are intended to prompt authors to disclose technical and statistical information and reviewers to consider relevant aspects for research reproducibility. However, the huge variability of experimental designs and analytical techniques needs to be accounted for. The community-driven approach to this situation was the definition of "minimum information" checklists. The Minimum Information About a Microarray Experiment (MIAME), developed in 2001, was the first of such guidelines, and details which information needs to be provided to ensure reproducibility and unambiguous interpretation of microarray-based data.[26] Similar guidelines for other preclinical research techniques are described at the Minimum Information about Biological and Biomedical Investigations (MIBBI) portal, although only a few methods are covered so far.[27]

A word of caution

Clinical and preclinical research are not directly comparable. In basic and preclinical research, scientists require enough freedom to use their creativity, which is key to the advancement of science and thus the development of novel drug candidates and innovative medicines. However, science progresses by building on existing knowledge, making rigorous, reproducible, high quality studies crucial.

The importance of finding a compromise between the need to trust conclusions of published research findings and the freedom for scientists to explore and innovate, has led to the concept of exploratory and confirmatory preclinical studies: at the exploratory stage, statistical testing and low quality standards should be acceptable as long as the experimental procedure is transparently described. However, for confirmatory studies (aimed to demonstrate robust and reproducible treatment effects), proper study design and implementation of the highest quality standards are essential, even if time- and resource-consuming.[28] Preclinical studies supporting decision making processes (e.g., whether to advance to animal studies or to first-in-human trials) should, therefore, be designed and treated as carefully as any clinical trial.

Conclusion

Some of the concepts from clinical research are already starting to be applied in the preclinical setting, and various approaches to enhance the robustness of preclinical data are being considered (strict adherence to quality standards, multicentre collaborations, data sharing, etc.). It seems worth noting that clinical research has gone a long way to improve its quality standards. These developments may also illuminate the path for preclinical research.

Acknowledgements

The authors would like to thank James Visanji, manager at Trilogy Writing & Consulting GmbH, and Anton Bespalov, managing partner at PAASP GmbH, for useful advice and discussion.

Disclaimers

The opinions expressed in this article are the authors' own and not necessarily shared by their employers or EMWA.

Conflicts of interest

The authors are employed by Trilogy Writing & Consulting (LPR) and PAASP GmbH (CE).

References

1. Baker M. Is there a reproducibility crisis? Nature. 2016;533(7604):452–4.
2. 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(9):712–3.

3. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012;483(7391):531–3.
4. Osherovich L. Hedging against academic risk. SciBX 2011;4.
5. Freedman LP, Cockburn IM, Simcoe TS. The economics of reproducibility in preclinical research. PLoS Biol. 2015;13(6):e1002165.
6. Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke. 2008;39(10):2824–9.
7. International Cell Line Authentication Committee (ICLAC). [cited 2017 November 13]. Available from: http://iclac.org/.
8. Perrin S. Preclinical research: Make mouse studies work. Nature 2014;507(7493): 423–5.
9. Good Laboratory Practices (GLP) for Non-Clinical Laboratory Studies, 21 C.F.R. Part 58. [cited 2017 November 13]. Available from: https://www.fda.gov/1ohrms/ dockets/98fr/980335s1.pdf.
10. Directive 2004/10/EC of the European Parliament and of the Council of 11 February 2004 on the harmonisation of laws, regulations and administrative provisions relating to the application of the principles of good laboratory practice and the verification of their applications for tests on chemical substances. Official J Eur Comm. 2004;L50/44. [cited 2017 November 13]. Available from: http://www.bfr.bund.de/cm/349/directive_2004_10_ec.pdf.
11. Directive 2004/9/EC of the European Parliament and of the Council of 11 February 2004 on the inspection and verification of good laboratory practice (GLP). Official J Eur Comm. 2004;L50/28. [cited 2017 November 13]. Available from: http://www.bfr.bund.de/cm/349/directive_2004_9_ec.pdf.
12. 89/569/EEC Council Decision of 28 July 1989 on the acceptance by the European Economic Community of an OECD decision/recommendation on compliance with principles of good laboratory practice. Official J Eur Comm. 1989;L315/32. [cited 2017 November 13]. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:1989:315:TOC.
13. Carlisle JB. Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals. Anaesthesia. 2017;72(8):944–52.
14. National Research Council (US) Committee for the Update of the Guide for the Care and Use of Laboratory Animals. Guide for the care and use of laboratory animals. 8th edition. Washington (DC): National Academies Press (US); 2011.
15. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. Official J Eur Comm. 2010;L276/33. [cited 2017 November 13]. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32010L0063.
16. Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). [cited 2017 November 13]. Available from: https://www.aaalac.org/.
17. International Council for Laboratory Animal Science (ICLAS). [cited 2017 November 13]. Available from: http://iclas.org/.
18. Coecke S, Balls M, Bowe G, Davis J, Gstraunthaler G, Hartung T, et al. Guidance on good cell culture practice. A report of the second ECVAM task force on good cell culture practice. Altern Lab Anim. 2005;33(3):261–87.
19. Voelkl B, Würbel H. Reproducibility crisis: Are we ignoring reaction norms? Trends Pharmacol Sci. 2016;37(7):509–10.
20. EMA/240810/2013, POLICY/0070 of 2 October 2014. European Medicines Agency policy on publication of clinical data for medicinal products for human use. [cited 2017 November 13]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2014/10/WC500174796.pdf.
21. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. Science. 2015;348(6242): 1422–5.
22. Pavlica S. The need for registration of preclinical studies. Med Writ. 2013;22(2):131–3.
23. Consolidated Standards of Reporting Trials (CONSORT) statement. [cited 2017 November 13]. Available from: http://www.consort-statement.org/.
24. Animal Research: Reporting of In Vivo Experiments (ARRIVE) guidelines. [cited 2017 November 13]. Available from: https://www.nc3rs.org.uk/arrive-guidelines.
25. Structured Transparent Accessible Reporting (STAR) Methods. [cited 2017 November 13]. Available from: http://www.cell.com/star-methods.
26. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME) toward standards for microarray data. Nat Genet. 2001;29(4):365–71.
27. Fairsharing.org collections. Minimum Information for Biological and Biomedical Investigations (MIBBI). [cited 2017 November 13]. Available from: https://fairsharing.org/collection/MIBBI.
28. Kimmelman J, Mogil JS, Dirnagl U. Distinguishing between exploratory and confirmatory preclinical research will improve translation. PLOS Biol. 2014;12(5):e1001863.

Author information

Laia Pedro-Roig, PhD, has 10 years of research experience in cell signalling, including preclinical development in Parkinson's disease. She is currently a medical writer at Trilogy Writing & Consulting.

Christoph Emmerich, PhD, is a trained biochemist with over 12 years of experience in preclinical drug development. He is co-founder of PAASP GmbH, a Heidelberg-based company that focuses on establishing quality standards in preclinical research and provides assessment of operational risks to facilitate decision making in early-phase drug discovery projects.