The role of systematic reviews in identifying the limitations of preclinical animal research, 2000–2022: part 2

Merel Ritskes-Hoitinga[1,2] and Pandora Pound[3]
[1]Radboud MC, 6525 GA Nijmegen, Netherlands
[2]Aarhus University, 8000 Aarhus C, Denmark
[3]Safer Medicines, Kingsbridge, TQ7 9AX, UK
Corresponding author: Merel Ritskes-Hoitinga. Email: j.ritskes-hoitinga@uu.nl

In Part 1 of this publication we provided a partial history of preclinical systematic reviews in both the UK and the Netherlands during the years 2000–2022.[1] Here in Part 2, we conclude by considering some of the evidence generated by these reviews, and its implications for preclinical research and for medicine.

Evidence from systematic reviews

In 2008, Michael Bracken, at Yale University, summarised some of the developments in the field, highlighting the worryingly poor quality of animal studies and making the case for more preclinical systematic reviews.[2] Between 2005 and 2010, Korevaar et al.[3] estimated that 163 preclinical systematic reviews had been published, while 246 were identified between 2009 and 2013 by Mueller et al.[4] As more and more animal studies were scrutinised as part of the systematic review process, it gradually became apparent that much animal research was conducted to a low standard and was therefore unable to generate robust, reliable data. This made uncomfortable reading for animal researchers, who were found to report low rates of random allocation, allocation concealment and blinded outcome assessment.[5–7] Studies that take these accepted precautions to reduce biases are less likely to suggest differential effects than studies that do not observe these precautions. It soon became evident that large bodies of animal research had overstated the benefits of their experimental interventions. Tsilidis et al.[8] demonstrated this clearly in the field of preclinical neurological research, as did Crossley et al.[9] in the field of preclinical stroke research. The accumulating preclinical systematic reviews also revealed that animal samples are typically small, leading to underpowered and therefore unreliable studies, as Emily Sena, convenor

of CAMARADES, showed in her 2014 overview.[10] In short, systematic reviews provided overwhelming evidence that animal studies suffer from poor experimental design and a lack of scientific rigour, raising doubts about the robustness of their findings and consequently, their clinical relevance.

Selective analysis and biased outcome reporting – the practice of reporting only the most positive outcomes and analyses from among the many performed and studied – was also revealed to be a problem in animal research.[8] Again, this leads to an overestimate of beneficial treatment effects, ultimately creating a body of evidence with an inflated proportion of studies with positive results. Incomplete reporting was revealed to be another limiting factor. Even basic information, such as the number of animals used in experiments, was found to be missing, as was reporting on attrition.[11] This – the loss of animals through death or exclusion – can dramatically alter the results of a study and, again, have the effect of making animal studies appear more positive than they actually are. Publication bias (the phenomenon whereby studies are more likely to be published if they present 'positive' findings) was found to be a significant problem,[3,4] leading once more to the benefits of animal studies being overstated.[12] And citation bias, first reported in the clinical field,[13] was found to be an issue in animal research. A German study of 109 investigator brochures, the documents presented to ethics review boards by those applying to conduct Phase I and II trials in humans, revealed that only 6% of the preclinical animal studies referenced in the brochures reported an outcome demonstrating no effect; the vast majority – 82% – were described as reporting positive findings.[14]

Unsurprisingly then, when scientists from Astra Zeneca reviewed 255 protocols for forthcoming animal experiments, they found that over half needed amending to ensure proper experimental

design, appropriate sample sizes and measures to control bias.[15] And when pharmaceutical companies conducted in-house validation of data coming from academia, they found that much of it was irreproducible[16]; in other words, the experiments did not produce the same results when independently repeated.[17] This problem – which has come to be known as the reproducibility crisis – is due to poor experimental design and poor scientific conduct and is compounded by incomplete reporting. A number of key papers were written on this topic, including by Ioannidis,[18] Leist and Hartung[19] and Begley and Ioannidis.[20] Although it is beyond the scope of this article to describe them, many initiatives were developed to improve the quality of animal study design and reporting, and to address problems such as publication bias, one of the most famous of which is the ARRIVE guideline.[21] Another initiative is European Quality in Preclinical Data (EQIPD),[22] an EU consortium that has assembled preclinical researchers from both academia and industry to identify how the quality of preclinical science could be improved. One of its outputs, for example, is a systematic review of existing guidelines for preclinical animal studies, resulting in 58 recommendations.[23]

Poorly conducted, unreliable research has consequences, for humans, animals and society. It can be dangerous. Corticosteroids, for example, were found to benefit animals with brain injury, and tirilazad was beneficial for animals with acute stroke, but both drugs increased the risk of people dying when they proceeded to clinical trials.[6] Systematic reviews have also revealed a great deal of redundancy and waste in animal research. In 2010, Sena et al. demonstrated in a cumulative meta-analysis that the beneficial effects of tissue plasminogen activator for stroke had been well documented in animal models by 2001, but research using several thousand animals continued for several years afterwards.[24] And of course, if the results of preclinical studies are unreliable, then that research is also a waste of time, resources and animals' lives. In 2014, *The Lancet* held a conference on research waste in both clinical and preclinical science, highlighting that this could be avoided at every stage of the research process, i.e. funding, conduct and regulation.[25] It was a clear call to action and led, among other things, to the founding of the Ensuring Value in Research (EVIR) Funder Forum, an international group of funders committed to avoiding waste and increasing the value of funded research.[26] Ritskes-Hoitinga made contact with the forum, with the upshot that a preclinical working group was established, with Ritskes-Hoitinga as co-chair.

Also, in 2014, Pound returned to the field after an absence of 10 years and teamed up with Michael Bracken to review developments.[27] Their paper, again published in the *BMJ,* provided an overview of the evidence accruing from preclinical systematic reviews. They noted that shortcomings in almost every aspect of the scientific design, conduct and reporting of animal studies were contributing to an inability to translate into benefits for humans. This time their paper was warmly received, indicating that the scientific climate had changed considerably. No longer was it considered heretical to discuss the limitations or challenge the validity of animal research.

In 2018, a *BMJ* investigation concluded that an Oxford University research group had been selective in the reporting of their animal study results to gain funding and approval for human trials of a TB booster vaccination.[28] The group had gained funding for the human trials, but the trials had ultimately failed. An earlier systematic review of the animal data concluded that insufficient evidence had existed to support claims about the efficacy of the vaccine booster and that the claims had been overstated.[29] Highlighting the problem of selective outcome reporting, the *BMJ* commented on the group's 'pick and mix' approach, claiming that some of the animal studies showing adverse effects had been omitted from the preclinical evidence. In an accompanying editorial, Ritskes-Hoitinga and colleague Kim Wever outlined steps that needed to be taken to improve the conduct and quality of preclinical research.[30]

More recent developments

In January 2017, SYRCLE moved out of the animal facility and into the Department of Health Evidence at Radboud University. The animal facility's users and research directors had begun to withdraw support for Ritskes-Hoitinga following an interview she gave to the Dutch newspaper *Trouw* in 2013, in which – having considered the evidence on the poor quality and reporting of animal studies – she stated that animal testing could be reduced by 80%. This had shocked the Dutch animal science community, and two colleagues – Professors Frauke Ohl and Coenraad Hendriksen from Utrecht University – had disagreed openly in a letter to the newspaper. Ritskes-Hoitinga was advised that, as the manager of an animal facility, her role was to provide a service to users, not comment on the science. In her new department she had greater freedom to investigate the evidence and embarked on a series of studies, including a collaborative project between Utrecht and Radboud Universities that attempted to identify factors contributing to translational success. A scoping review performed as part of this project

found that rates of translation from animal to human studies ranged from 0 to 100 and appeared to be random, with no indication of factors that might increase its likelihood.[31] At a symposium in 2019 to mark the end of the project, epidemiologist John Ioannidis stated his view that animal testing could be reduced by 90%.[32]

Around this time Ritskes-Hoitinga and Pound began to collaborate. They published on the problem of external validity of animal studies, arguing that even if all the problems of internal validity in animal research were resolved, species differences would continue to make translation to humans unreliable.[33] In doing so, they were drawing attention to the problem that Ibn Sina had highlighted a thousand years previously. With colleague Christine Nicol, Pound also conducted a retrospective harm-benefit analysis by reanalysing the animal data from Perel et al.'s 2007 study.[6] Using Bateson's Cube[34] to weigh the harms to animals used in the research against the benefits to humans that resulted, and taking into account the importance and quality of the research studies, they concluded that fewer than 7% of the 212 animal studies scrutinised were permissible.[35] In a later paper, Pound and Ritskes-Hoitinga highlighted that, while prospective preclinical systematic reviews (i.e. those conducted prior to human trials) allow valuable scrutiny of the preclinical animal data, they are not necessarily able to reliably predict the safety and efficacy of an intervention, or safeguard clinical trial participants.[36] A systematic review is only as good as the studies it includes and if the primary animal studies cannot reliably predict safety and efficacy in humans, the systematic review findings will reflect this.

Despite a promising start, the number of preclinical systematic reviews remains disappointingly low, raising questions about the extent to which the evidence-based approach has been accepted within preclinical research. And although two more international symposia on systematic reviews in laboratory animal science were held, one in Edinburgh in 2013 and one in Washington in 2014, no further meetings have taken place. Then in 2021, Radboud University Medical Center suddenly decided to withdraw all funding from SYRCLE; preclinical systematic reviews were apparently no longer a priority for them.

Nevertheless, systematic reviews have been pivotal in preclinical research. In highlighting the shortcomings of animal studies, they enabled, for the first time, an open and constructive debate about the value of animal research – a debate that focused on the science, rather than the ethics of this research. For decades, scientists had sidelined any challenges to the practice of animal research as ethical rather than

scientific, referring back to the 3Rs and regulations, but the challenge presented by systematic reviews came from within the scientific community and could not be ignored. Preclinical systematic reviews have not only exposed shortcomings in the internal validity of animal research (i.e. its design, conduct and reporting); in highlighting the poor track record of animal research in translation to humans, they have also exposed its lack of external validity. In this respect, two issues are now clear. First, the inability of many, if not most animal models to replicate complex human diseases; and second, the problem of species differences.

While some of the limitations of animal research can, at least in theory, be addressed over time (i.e. internal validity and some aspects of the animal models themselves), the evolved differences between species present an altogether different problem. Evolutionary theory indicates that species differences will always make the extrapolation of animal findings to humans unreliable.[33,37] Bearing in mind that this is a fundamental flaw in the animal research paradigm, and moreover one that is insurmountable, are projects that aim to improve animal models and animal research a good use of scarce funding resources?

Many outside the field of animal research argue that the way forward is not to try to improve this research but to replace it with methodologies and technologies that are directly relevant to humans.[38] New approaches based on human biology include in vitro cell models such as organoids and organs-on-a-chip, as well as those using computer simulations and artificial intelligence. While they face the challenges that any research must deal with, in other words the need to ensure internal validity and reproducibility, they cut out the 'noise' that animal studies introduce into clinical translation, producing data that are applicable to humans and that are therefore externally valid. Yet despite these new approaches often performing better than animal studies (see for example Dirven et al.[39] on the relative performance of in vitro and in vivo methodologies for predicting drug-induced liver injury in humans), scientists appear reluctant to relinquish traditional practices. In a 2020 study, Pound and colleague Rebecca Ram reviewed scientists' opinions about the limitations of animal models of stroke.[40] They found that while many viewed species differences as a significant problem in preclinical stroke research, the vast majority were reluctant to abandon their animal models, with only 1 of 80 authors advocating a focus on human-relevant research instead.

Consequently, two streams of research are proceeding in parallel and mostly in isolation from each other. On the one hand, research based on

human biology is being advanced and developed as a direct response to the perceived limitations of animal research, and – on the other – animal research continues as usual, albeit with an eye on research improvement. In terms of the latter, CAMARADES is still going strong,[41] with several projects underway, including the development of 'living systematic reviews'.[42] Nevertheless, change seems to be coming. In the UK, a couple of high-profile animal laboratories are set to close, the reasons being 'a move to using alternative technologies'[43] and 'a changing scientific landscape',[44] while the Medical Research Council's new 'Experimental Medicine' programme[45] funds research that focuses on 'the human as the ultimate experimental animal for improving human health', noting that this is now possible due to advances in non-invasive techniques such as medical imaging, sensors and ex vivo analyses. The Dutch Parliament has a specific transition programme, 'Transitie naar Proefdiervrije Innovatie,[46] which aims to ensure that the Netherlands is a frontrunner in the transition to innovate without using laboratory animals. Meanwhile, the United States Environmental Protection Agency has committed to ending the use of mammals in chemical testing,[47] also aiming to be a frontrunner in the adoption of human-relevant technologies.[48] And in September 2021, the European Parliament voted by a stunning majority of 667 to 4 to develop a coordinated plan to replace animal experiments with innovative, non-animal methodologies.[49,50] Might this be the rumblings of a scientific revolution?

Declarations

References

1. Ritskes-Hoitinga M and Pound P. The role of systematic reviews in identifying the limitations of preclinical animal research, 2000–2022: part 1. *J R Soc Med* 2022; 115: 186–192.
2. Bracken MB. Why animal studies are often poor predictors of human reactions to exposure. *JLL Bull: Comment Hist Treat Eval* 2008; 102: 120–122.
3. Korevaar DA, Hooft L and ter Riet G. Systematic reviews and meta-analyses of preclinical studies: publication bias in laboratory animal experiments. *Lab Anim* 2011; 45: 225–230.
4. Mueller KF, Briel M, Strech D, Meerpohl JJ, Lang B, Motschall E, et al. Dissemination bias in systematic reviews of animal research: a systematic review. *PLoS One* 2014; 9: e116016.
5. Henderson VC, Demko N, Hakala A, MacKinnon N, Federica CA, Fergusson D, et al. A meta-analysis of threats to valid clinical inference in preclinical research of sunitinib. *Elife* 2015; 4: e08351.
6. Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, et al. Comparison of treatment effects between animal experiments and clinical trials: systematic review. *BMJ* 2007; 334: 197.
7. Hirst JA, Howick J, Aronson JK, Roberts N, Perera R, Koshiaris C, et al. The need for randomization in animal trials: an overview of systematic reviews. *PLoS One* 2014; 9: e98856.
8. Tsilidis KK, Panagiotou OA, Sena ES, Aretouli E, Evangelou E, Howells DW, et al. Evaluation of excess significance bias in animal studies of neurological diseases. *PLoS Biol* 2013; 11: e1001609.
9. Crossley NA, Sena E, Goehler J, Horn J, van der Worp B, Bath PMW, et al. Empirical evidence of bias in the design of experimental stroke studies: a meta-epidemiologic approach. *Stroke* 2008; 39: 929–934.
10. Sena ES, Currie GL, McCann SK, Macleod MR and Howells DW. Systematic reviews and meta-analysis of preclinical studies: why perform them and how to appraise them critically. *J Cereb Blood Flow Metab* 2014; 34: 737–742.
11. Holman C, Piper SK, Grittner U, Diamantaras AA, Kimmelman J, Siegerink B, et al. Where have all the rodents gone? The effects of attrition in experimental research on cancer and stroke. *PLoS Biol* 2016; 14: e1002331.
12. Sena ES, van der Worp HB, Bath PM, Howells DW and Macleod MR. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol* 2010; 8: e1000344.
13. Gøtzsche PC. Reference bias in reports of drug trials. *Br Med J (Clin Res Ed)* 1987; 295: 654–656.
14. Wieschowski S, Chin WWL, Federico C, Sievers S, Kimmelman J and Strech D. Preclinical efficacy studies in investigator brochures: do they enable risk-benefit assessment? *PLoS Biol* 2018; 16: e2004879.
15. Peers IS, South MC, Ceuppens PR, Bright JD and Pilling E. Can you trust your animal study data? *Nat Rev Drug Discov* 2014; 13: 560.
16. Prinz F, Schlange T and 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.
17. Begley CG and Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature* 2012; 483: 531–533.

18. Ioannidis JP. Extrapolating from animals to humans. *Sci Transl Med* 2012; 4: 151ps15.

19. Leist M and Hartung T. Inflammatory findings on species extrapolations: humans are definitely not 70-kg mice. *Arch Toxicol* 2013; 87: 563–567.

20. Begley CG and Ioannidis JP. Reproducibility in science: improving the standard for basic and preclinical research. *Circ Res* 2015; 116: 116–126.

21. Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, et al. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLoS Biol* 2020; 18: e3000410.

22. European Quality in Preclinical Data (EQIPD). See https://quality-preclinical-data.eu/ (last checked 18 March 2022).

23. Vollert J, Schenker E and Macleod M; The EQIPD WP3 Study Group Members. Systematic review of guidelines for internal validity in the design, conduct and analysis of preclinical biomedical experiments involving laboratory animals. *BMJ Open Sci* 2020; 4: e100046.

24. Sena ES, Briscoe CL, Howells DW, Donnan GA, Sandercock PA and Macleod MR. Factors affecting the apparent efficacy and safety of tissue plasminogen activator in thrombotic occlusion models of stroke: systematic review and meta-analysis. *J Cereb Blood Flow Metab* 2010; 30: 1905–1913.

25. Lancet Series. Increasing value, reducing waste. See www.thelancet.com/series/research (last checked 18 March 2022).

26. Ensuring Value in Research (EVIR). See https://evir.org (last checked 18 March 2022).

27. Pound P and Bracken MB. Is animal research sufficiently evidence based to be a cornerstone of biomedical research? *BMJ* 2014; 348: g3387.

28. Cohen D. Oxford vaccine study highlights pick and mix approach to preclinical research. *BMJ* 2018; 360: j5845.

29. Kashangura R, Sena ES, Young T and Garner P. Effects of MVA85A vaccine on tuberculosis challenge in animals: systematic review. *Int J Epidemiol* 2015; 44: 1970–1981.

30. Ritskes-Hoitinga M and Wever K. Improving the conduct, reporting, and appraisal of animal research. *BMJ* 2018; 360: j4935.

31. Leenaars CHC, Kouwenaar C, Stafleu FR, Bleich A, Ritskes-Hoitinga M, De Vries RBM, et al. Animal to human translation: a systematic scoping review of reported concordance rates. *J Transl Med* 2019; 17: 223.

32. Ritskes-Hoitinga M, Leenaars C, Beumer W, Coenen-de Roo T, Stafleu F and Meijboom FLB. Improving translation by identifying evidence for more human-relevant preclinical strategies. *Animals* 2020; 10: 1170.

33. Pound P and Ritskes-Hoitinga M. Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail. *J Transl Med* 2018; 16: 304.

34. Bateson P. When to experiment on animals. *New Sci* 1986; 109: 30–32.

35. Pound P and Nicol CJ. Retrospective harm benefit analysis of pre-clinical animal research for six treatment interventions. *PLoS One* 2018; 13: e0193758.

36. Pound P and Ritskes-Hoitinga M. Can prospective systematic reviews of animal studies improve clinical translation? *J Transl Med* 2020; 18: 15.

37. Perlman RL. Mouse models of human disease: an evolutionary perspective. *Evol Med Public Health* 2016; 1: 170–176.

38. Herrmann K, Pistollato F and Stephens ML. Beyond the 3Rs: expanding the use of human-relevant replacement methods in biomedical research. *Altex* 2019; 36: 343–352.

39. Dirven H, Vist GE, Bandhakavi S, Mehta J, Fitch SE, Pound P, et al. Performance of preclinical models in predicting drug-induced liver injury in humans: a systematic review. *Sci Rep* 2021; 11: 6403.

40. Pound P and Ram R. Are researchers moving away from animal models as a result of poor clinical translation in the field of stroke? An analysis of opinion papers. *BMJ Open Sci* 2020; 4: e100041.

41. CAMARADES (Collaborative Approach to Meta-analysis and Review of Animal Experimental Studies). See www.ed.ac.uk/clinical-brain-sciences/research/camarades/about-camarades (last checked 18 March 2022).

42. CAMARADES projects. See www.ed.ac.uk/clinical-brain-sciences/research/camarades/research-portfolio/current-projects (last checked 18 March 2022).

43. Else H. Genomics institute to close world-leading animal facility. *Nature* 2019; 569: 612.

44. Else H. Proposal to close UK mouse-research centre is 'major threat'. *Nature* 2019, 26 June. doi: https://doi.org/10.1038/d41586-019-02002-y

45. MRC (Medical Research Council). See www.ukri.org/what-we-offer/browse-our-areas-of-investment-and-support/experimental-medicine/ (last checked 18 March 2022).

46. TPI (Transitie Proefdiervrije Innovaties in the Netherlands). See www.animalfreeinnovationtpi.nl/ (last checked 18 March 2022).

47. EPA (Environmental Protection Agency). See www.epa.gov/newsreleases/administrator-wheeler-signs-memo-reduce-animal-testing-awards-425-million-advance (last checked 18 March 2022).

48. Wheeler AR. Memorandum 10 September 2019: directive to prioritize efforts to reduce animal testing. United States Environmental Protection Agency. See www.epa.gov/sites/default/files/2019-09/documents/image2019-09-09-231249.pdf (last checked 18 March 2022).

49. European Parliament. See www.europarl.europa.eu/news/en/press-room/20210910IPR11926/meps-demand-eu-action-plan-to-end-the-use-of-animals-in-research-and-testing (last checked 18 March 2022).

50. Haahr T. MEPs demand EU action plan to end the use of animals in research and testing (Press Release) European Parliament. See www.europarl.europa.eu/news/en/press-room/20210910IPR11926/meps-demand-eu-action-plan-to-end-the-use-of-animals-in-research-and-testing. (last checked 18 March 2022).