

REVIEW

Open Access



Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail

Pandora Pound^{1*}  and Merel Ritskes-Hoitinga²

Abstract

Background: The pharmaceutical industry is in the midst of a productivity crisis and rates of translation from bench to bedside are dismal. Patients are being let down by the current system of drug discovery; of the several 1000 diseases that affect humans, only a minority have any approved treatments and many of these cause adverse reactions in humans. A predominant reason for the poor rate of translation from bench to bedside is generally held to be the failure of preclinical animal models to predict clinical efficacy and safety. Attempts to explain this failure have focused on problems of internal validity in preclinical animal studies (e.g. poor study design, lack of measures to control bias). However there has been less discussion of another key factor that influences translation, namely the *external* validity of preclinical animal models.

Review of problems of external validity: External validity is the extent to which research findings derived in one setting, population or species can be reliably applied to other settings, populations and species. This paper argues that the reliable translation of findings from animals to humans will only occur if preclinical animal studies are both internally *and* externally valid. We review several key aspects that impact external validity in preclinical animal research, including unrepresentative animal samples, the inability of animal models to mimic the complexity of human conditions, the poor applicability of animal models to clinical settings and animal–human species differences. We suggest that while some problems of external validity can be overcome by improving animal models, the problem of species differences can never be overcome and will always undermine external validity and the reliable translation of preclinical findings to humans.

Conclusion: We conclude that preclinical animal models can never be fully valid due to the uncertainties introduced by species differences. We suggest that even if the next several decades were spent improving the internal and external validity of animal models, the clinical relevance of those models would, in the end, only improve *to some extent*. This is because species differences would continue to make extrapolation from animals to humans unreliable. We suggest that to improve clinical translation and ultimately benefit patients, research should focus instead on human-relevant research methods and technologies.

Keywords: External validity, Preclinical animal models, Translation, Human-relevant methods

*Correspondence: pandora@safermedicines.org

¹ Safer Medicines Trust, PO Box 122, Kingsbridge TQ7 9AX, UK
Full list of author information is available at the end of the article



Introduction

Few would dispute that the pharmaceutical industry is in the midst of a productivity crisis [1–7] or that rates of translation from bench to bedside are dismal [8–15]. Similarly, few would disagree that patients are being let down by the current system of drug discovery. Evidence suggests that current levels of investment in pharmaceutical drugs are out of proportion to their impact in terms of value for money or population health [16]. There is not a drug for every disease; several thousand diseases affect humans, of which only about 500 are estimated to have any approved treatments [17]. Many of the treatments that do exist cause dangerous and undesirable reactions in humans [18]. This failure of the drug discovery system not only lets down patients but also uses an enormous amount of resources that could likely be better spent [19]. The failure of the current drug discovery model is an issue of global importance for human health. However, the first step towards resolving this issue is to identify what is going wrong.

While many factors contribute to the poor rates of translation from bench to bedside (including flawed clinical trials [20]), a predominant reason is generally held to be the failure of preclinical animal models to predict clinical efficacy [4, 6, 9, 14, 21] and safety [22, 23]. Efficacy and safety issues account for the majority of failures (52% and 24% respectively) at Phases II and III of clinical trials [24]. Attempts to explain these failures have focused on problems of internal validity in preclinical animal studies (i.e. shortcomings in study design, conduct, analysis and reporting), but there has been relatively little discussion of another key factor that influences translation, namely the *external* validity of preclinical animal models [25]. External validity is usually taken to mean the extent to which research findings derived in one setting, population or species can be reliably applied to other settings, populations and species [26]. It is a key criterion for assessing the credibility of scientific research. In the field of preclinical animal research, where the findings derived from animal studies are intended to have relevance in clinical settings, external validity is of the utmost importance. External validity is sometimes referred to as generalisability. Although the two terms are interchangeable, we use the term external validity here because generalisability, despite having a distinct methodological meaning, is often confused with translatability. In fact the two concepts are discrete; external validity/generalisability *contributes* to translatability [26] and in fact, as we shall argue, it is a prerequisite for the translation of findings from animals to humans.

In this paper we argue that the translation of findings from animals to humans can only occur reliably if preclinical animal studies are both internally *and* externally valid. We consider the relationship between internal and external

validity before going on to explore several key aspects that impact external validity in preclinical animal research. We suggest that while some problems of external validity are surmountable, the issue of human–animal species differences is not; species differences will always have an impact on external validity and the ability to translate preclinical findings to humans. We explore the implications of this conclusion.

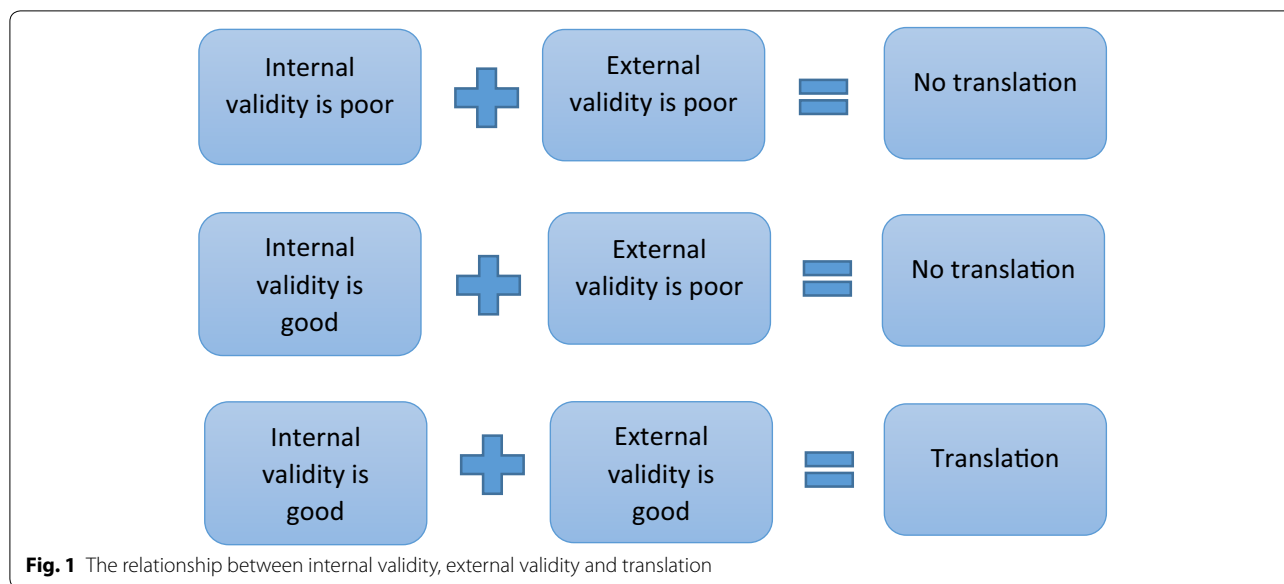
The relationship between internal and external validity

External validity is distinct from *internal* validity, which refers to the scientific robustness of a study's design, conduct, analysis and reporting. Systematic review evidence has revealed that preclinical animal studies suffer from serious problems of internal validity, in particular a failure to take measures to prevent bias, such as random allocation to groups and blinded assessment of outcomes [27–29]. If a study does not take such measures then its internal validity is poor and its findings cannot be relied upon. Studies that lack internal validity will always lack external validity [30]. For example, in the field of stroke, animal studies that were 'unblinded' overestimated the effect of the intervention by 13% compared with studies that included blinding [31]. In other words, the lack of blinding led to the benefits of the animal studies being overstated. Because the benefits were not real, they could not be applied to other populations and settings. As such, lack of internal validity led to a lack of external validity.

Some believe that if issues of internal validity were resolved (i.e. if researchers took measures to prevent bias and conducted studies according to agreed scientific standards) then the clinical translation of animal studies would be more successful [32]. However the available evidence does not support this view. In 1999 the Stroke Treatment Academic Industry Roundtable introduced a series of recommendations and standards intended to improve the quality of animal studies in stroke. Yet by 2012 translation rates had not improved [33] and the situation is no better today. Well over a thousand drugs have been tested in animal studies of stroke [15] but of these only one has translated into clinical use and the benefits of that one are controversial [34]. This may be partly because internal validity remains poor but it is also due to the fact that preclinical animal studies need to be both internally *and* externally valid if they are to translate into benefits for humans (see Fig. 1) [10].

Aspects of external validity to consider in preclinical animal research

There are several aspects of external validity that raise problems within preclinical animal research. Some of these are potentially surmountable but others are more



intractable. We begin this section by reflecting on the surmountable problems of external validity before going on to consider a problem of external validity that we regard as *insurmountable*, namely species differences. We argue that even if the following surmountable issues of external validity were resolved, the issue of species differences would continue to undermine external validity and therefore clinical translation.

Surmountable problems of external validity

The unrepresentativeness of animal samples is a problem in preclinical research. In general the standardisation of laboratory animal populations produces homogenous samples that do not extrapolate to heterogeneous human populations [26, 35]. In addition, laboratory animals may be housed in conditions that complicate the extrapolation of findings to humans. For example the biology of laboratory mice may be affected by their being housed in same-sex groups, by lack of opportunities for physical exercise, and by temperature [36] and diet [37]. Furthermore, the animals used in preclinical research tend to be young and healthy whereas many human diseases manifest in older age. For example, animal studies of osteoarthritis (OA) tend to use young animals of normal weight, whereas clinical trials focus mainly on older people with obesity [38]. Animals used in stroke studies have tended to be young whereas human stroke is largely a disease of the elderly [39]. It is not hard to see that in such cases the findings from animal studies are unlikely to be applicable to human patient populations, i.e. they will lack external validity.

Furthermore, many animal models lack the complexity required to accurately mimic human conditions.

Although there has been some success with diseases based on single gene defects that can be reproduced in animal models [16], most human diseases tend to evolve over time as part of the human life course. For example, it may be possible to grow a breast tumour on a mouse model but this does not actually represent the human experience because most human breast cancer occurs post-menopausally. While some animal species may be better models of specific diseases than others (e.g. horses have similar cartilage degenerative processes as humans [40]), in general the animal models currently used do not mimic the slow, progressive and degenerative nature of many human chronic diseases [10], nor do they involve the complexity of comorbidity or polypharmacy (human patients often take more than one type of medication). To take the example of stroke again, many people with this disease have hypertension but preclinical animal studies of stroke have generally used healthy animals without comorbidities, which results in the effects of interventions being overestimated [31]. In fact many experimental treatments for stroke are less effective in humans with hypertension [39]. Furthermore, recovery from a severe stroke can take years for humans but animals can recover from experimental stroke within days or weeks [41]. Additionally, while human stroke is highly heterogeneous the four most commonly used stroke animal models are all of ischaemic stroke [41].

Finally, animal models developed in the laboratory lack applicability to ‘real life’ clinical settings. To return to the example of OA, animals are usually given drugs for OA prophylactically, or in the early stages of OA, whereas in clinical trials humans are usually given drugs in the late stages of their disease [38]. Similarly, experimental drugs

for multiple sclerosis (MS) are most commonly administered to animals some days before neurological impairment. As these drugs may work by blocking the induction of the disease they are not relevant to the human condition because human patients cannot be identified prior to the onset of their MS. Animal models of MS can only have clinical relevance if treatment is successfully started *after* the onset of symptoms [42]. Animal models of Parkinson's disease pose a similar problem [43], as do those of inflammatory bowel disease [44]. Again, in the case of stroke, Tirilazad was able to successfully treat animals if given within 10 min of stroke induction but humans are highly unlikely to be able to access treatment for stroke within 10 min. In clinical trials humans were given Tirilazad within a more realistic 5 h and the trials were unsuccessful [39]. The choice of animal models across a range of fields appears to lack rationality in terms of evidence-base or appropriateness in relation to the relevant human condition [40, 44].

These failures of animal models to accurately represent human diseases and clinical contexts are sometimes described as failures of construct validity, which is generally understood to be a subset of external validity [10, 45]. As noted above, these failures, which may be in part due to academic pressures to produce quick, high impact papers [46] can potentially be resolved, although some of the solutions raise serious ethical issues (e.g. animal models that involve ageing and comorbidity are likely to be more externally valid but because the harms to animals are protracted they are more likely to be considered severe). Yet even if all the surmountable problems of external validity described above were resolved, one intractable problem would remain. Species differences, i.e. the differences between animals and humans in terms of their underlying biology, would continue to undermine external validity.

The insurmountable problem of external validity: species differences

Perlman [36, 47], an evolutionary biologist, points out that mice (the most frequently used animals in research) and humans have a high level of genetic homology as well as many biochemical and physiological similarities. He notes however, that the lineages that led to modern rodents and primates diverged around 85 million years ago and that since then, the species in these lineages have become adapted to very different environments. As a result, mice and humans now have very different life histories, they eat different diets, have different levels of physical activity, are exposed to different environmental toxins and pathogens and have different microbiomes. Because they harbour different sets of pathogens and microbiomes, host–pathogen and host–microbiome

coevolution has led to evolved differences between the human and mouse immune systems [36, 47]. Furthermore, Perlman [36] suggests that due to different network architectures between mice and humans and different genotype–phenotype relationships, the relationships between genotype and disease are likely to differ in these two species. Critically Perlman notes that while mice and other animals may be useful for understanding processes that arose early in evolution and that humans share with other species, they are less likely to be useful for understanding chronic non-communicable diseases because the pathogenesis of these diseases is enmeshed in our unique, evolved life histories [36, 47, 48].

Nevertheless there is a strong assumption among the biomedical community that gene functions and developmental systems are conserved between animals and humans [49]. Moreover, there appears to be little interest within the biomedical community in verifying this assumption, or in the evidence emerging from evolutionary developmental biology indicating that gene functions and gene networks diverge through evolution [50]. Commentators have observed that the animal model paradigm tends to discourage any critical appraisal of species differences, encouraging instead a view that animal based findings are generally applicable to humans [50, 51] and emphasising the commonalities rather than the differences [21]. Preuss [50] suggests that if species differences are acknowledged they tend to be 'soft-peddled' or treated as 'noise,' again noting that researchers focus on 'commonalities' and 'basic uniformity' instead. But as Perlman [36] notes, biology is characterized by diversity as well as unity; evolution is 'descent with modification' [52].

Unfortunately, focusing on the commonalities without acknowledging difference is problematic. Sjoberg [53] argues that crude inferences are made about the properties of one group (humans) based on observations from another group (animals), simply because both groups have some other property in common (genetic similarity). Sjoberg uses the example of Jack and Jill: if Jack is clumsy then it might be inferred that his sibling Jill is also clumsy. However there is no evidence that Jill is clumsy and the argument is based solely on the observation that Jack and Jill have genetic properties in common. This reasoning, which relies purely on an *assumption* of similarity (rather than its empirical demonstration), underpins the use of animals as models of human disease. As Wall and Shani [21] note, the assumption is that if two systems are homologous then they are likely to function similarly. However this is incorrect; while some molecular pathways may appear identical between humans and animals, there may be differences, for example in specific receptors and enzymes, that will cause them to behave

very differently [54]. Non-human primates are often cited as having great genetic similarity with humans, but this belies the fact that in complex living systems even minor differences can result in significant differences in biological processes and outcomes [55]. The case of TGN1412, which was tested in non-human primates precisely because of their close relation with humans, amply demonstrates this. After just a few minutes of being infused with a dose 500 times smaller than that found safe in animal studies, all six human volunteers started suffering severe cytokine release syndrome leading to severe inflammation and multiple organ failure [56]. Wall and Shani [21] suggest that in some cases animal models can serve as a good analogue to study general principles, *but not specific details*. Details matter when it comes to developing safe and effective drugs for humans. As they write, 'On average, the extrapolated results from studies using tens of millions of animals fail to accurately predict human responses.' Consequently, they conclude that it is probably inadvisable to use animal models for extrapolation.

What can be done about the problem of species differences?

Transgenic mouse models were intended to enhance the external validity of animal models but as Geerts [9] suggests, if translation rates are anything to go by they have failed. This is because the paradigm suffers from the same problems; the SOD1 transgenic mouse, for example, appears to mimic humans in terms of some of the characteristics of motor neurone disease, but this is no guarantee that the same mechanisms are involved [10, 57]. Lynch [49] suggests that a way forward might be to empirically demonstrate (rather than assume) the similarity between animal and human genes with regard to the function being studied. Likewise, Seok et al. [12] suggest that researchers should specify in advance the extent to which their animal model mimics the molecular behaviour of the key genes and key pathways thought to be important for the human disease under investigation. While this would appear to provide an answer however, it potentially leads us towards another problem of reasoning, namely the 'extrapolator's circle'. In other words, if we want to determine whether a mechanism in animals is sufficiently similar to the mechanism in humans to justify extrapolation, we must know how the relevant mechanism in humans operates. But if we already know about the mechanism in humans then the initial animal study is likely to have been redundant [58] (depending upon the purpose of that animal study [59]).

Consequently we suggest that animal–human species differences constitute a problem of external validity that cannot be overcome. Imagine that the next several

decades are spent resolving the myriad problems of internal validity and the surmountable aspects of external validity (i.e. the representativeness of animal samples and the clinical relevance of animal models). While vast resources would be expended and colossal numbers of animals used and killed in this endeavour the end result would be only modest; the robustness of animal studies and the clinical relevance of animal models would likely improve *to some extent*. This unremarkable result would be due to the fact that despite improvements in animal models, the intractable issue of species differences would remain and would continue to make extrapolation from animals to humans unreliable. Along the way there might be some serendipitous findings but serendipity is not a reliable scientific method. Thus decades from now pre-clinical animal studies would still fail to reliably and consistently predict human responses and the findings from preclinical animal models would still fail to translate into benefits for humans. This is essentially the uninspiring scenario proposed by those who insist that the answer to the problem of translation lies in improving animal studies and animal models. This scenario is particularly unexciting given that current attempts to improve matters have so far failed [33, 60, 61].

An alternative approach, and one taken by an increasing number of scientists, is to consider a paradigm for drug discovery that cuts out the uncertainty introduced by species differences [4, 6, 7, 62]. Within this paradigm new, human-relevant approaches and technologies are considered, such as the generation of human induced pluripotent stem cells (iPSC), which can be used to create disease- or patient-specific cell lines for testing potential drugs, micro-physiological systems known as 'organ-on-chips', and human organoids (three dimensional cell cultures that incorporate key features of organs). Many of these new techniques integrate with *in silico* approaches and with systems biology, seen by many as having potential to revolutionise medicine [63, 64] and drug discovery [2]. Given that the return on investment for developing a new drug decreased from 10% in 2010 to 3% in 2017, the pharmaceutical industry certainly perceives a need to do things differently [5] and there is some considerable optimism, both within industry [4, 6] and elsewhere, about the potential of these new approaches to increase the speed and accuracy of drug discovery. The US is making significant investments in organ-on-chip technologies [7] and the Netherlands is aiming to phase out animal use in the regulatory safety testing of medicines and chemicals by 2025 [65] regarding new technologies as able to increase research relevance and deliver more reliable risk assessments whilst maintaining existing safety levels [66]. Systematic reviews are being used to review legacy data on drug safety [67] and this evidence, alongside low-risk

approaches such as microdosing in clinical trials [68], could provide a valuable safeguard during a transition to new technologies [69]. Systematic reviews will also have a role in synthesising emerging evidence about the efficacy of new technologies. Initial findings suggest that organ-on-chips [70, 71] and *in silico* approaches [72] may have advantages over animal studies in terms of predicting adverse drug reactions. New physiologically relevant technologies also appear more capable of illuminating mechanisms of toxicity than animal studies [73–75].

Conclusion

We have argued that translation from animals to humans can only occur if preclinical animal studies are both internally and externally valid. We have also suggested that external validity consists of potentially modifiable features (e.g. representativeness of animal samples, clinical relevance of animal models) and unmodifiable features (animal–human species differences). Thus we suggest that while some aspects of animal models can be improved to a limited extent, they can never be fully externally valid because of the uncertainty introduced by animal–human species differences. If the aim is to improve clinical translation and ultimately address patients' needs for safe and effective treatments, the first step is to acknowledge where current systems are failing.

We noted that those conducting preclinical animal research appear to downplay the problem of animal–human species differences but interestingly, other researchers and commentators in the field do similarly. Although they may briefly acknowledge that species differences constitute a problem for external validity, the tendency is to focus on other, potentially modifiable, aspects of external validity [10, 26]. This is perhaps understandable, since acknowledging the issue of species differences entails confronting the possibility that the preclinical animal research paradigm no longer has a great deal to offer. That possibility is alarming, not only to scientists who conduct animal research but also to those attempting to improve it. Yet there is a way forward. Research methods and technologies that are physiologically relevant to humans obviate the need for animals and thus eliminate the problem of animal–human species differences. As a recent industry report [6] concluded, the time has come to humanise medicine. For the sake of patients and animals, we agree.

Authors' contributions

PP conceived the idea for this paper and wrote the first draft. MR-H contributed conceptually with ideas and comments. Both authors critically revised and edited subsequent drafts before approving the final version. Both authors read and approved the final manuscript.

Author details

¹ Safer Medicines Trust, PO Box 122, Kingsbridge TQ7 9AX, UK. ² SYRCLC, Department for Health Evidence, Radboud University Medical Center, PO Box 9101, Route 133, 6500 HB Nijmegen, The Netherlands.

Acknowledgements

Thank you to Kathy Archibald for commenting on drafts of this paper.

Competing interests

PP declares that she has no competing interests. MR-H is a member of the council of management of the UK registered company Laboratory Animals Ltd (LAL). LAL issues the journal *Laboratory Animals*. The position is unpaid but travel to LAL meetings is reimbursed. The journal's profits are used for charitable purposes, subsidising educational projects in laboratory animal science and welfare.

Availability of data and materials

Not applicable.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

The authors received no specific funding for this work.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 5 September 2018 Accepted: 31 October 2018

Published online: 07 November 2018

References

1. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. 2010;9(3):203.
2. Loscalzo J. Personalized cardiovascular medicine and drug development: time for a new paradigm. *Circulation*. 2012;125(4):638–45.
3. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. *Nat Biotechnol*. 2014;32(1):40.
4. Innovate UK. A non-animal technologies roadmap for the UK: advancing predictive biology. 2015. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/474558/Roadmap_NonAnimalTech_final_09Nov2015.pdf. Accessed 22 May 2018.
5. Deloitte. A new future for R&D? Measuring the return from pharmaceutical innovation 2017. 2017. <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/life-sciences-health-care/deloitte-uk-measuring-roi-pharma.pdf>. Accessed 10 July 2018.
6. BioIndustry Association and Medicines Discovery Catapult. State of the Discovery Nation 2018 and the role of the Medicines Discovery Catapult. 2018. <https://md.catapult.org.uk/FlipBuilder/mobile/index.html>. Accessed 25 May 2018.
7. Marshall LJ, Austin CP, Casey W, Fitzpatrick SC, Willett C. Recommendations toward a human pathway-based approach to disease research. *Drug Discov Today*. 2018. <https://doi.org/10.1016/j.drudis.2018.05.038>.
8. Contopoulos-loannidis DG, Ntzani EE, Ioannidis JPA. Translation of highly promising basic science research into clinical applications. *Am J Med*. 2003;114:477–84.
9. Geerts H. Of mice and men. Bridging the translational disconnect in CNS drug discovery. *CNS Drugs*. 2009;23(1):915–26.
10. Van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, Macleod MR. Can animal models of disease reliably inform human studies? *PLoS Med*. 2010;7(3):e1000245.

11. Howells DW, Sena ES, O'collins V, Macleod MR. Improving the efficiency of the development of drugs for stroke. *Int J Stroke*. 2012;7(5):371–7.
12. Seok J, Warren S, Cuenca A, Mindrinos M, Baker H, Xu W, et al. Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc Natl Acad Sci*. 2013;110(9):3507–12.
13. Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimer's Res Ther*. 2014;6(4):37. <https://doi.org/10.1186/alzrt269>.
14. Perrin S. Preclinical research: make mouse studies work. *Nature*. 2014;507:423–5.
15. O'Collins VE, Macleod MR, Donnan GA, Horky LL, van der Worp BH, Howells DW. 1,026 experimental treatments in acute stroke. *Ann Neurol*. 2006;59(3):467–77.
16. Jones R, Wilsdon J. The biomedical bubble. 2018. Nesta. <https://www.nesta.org.uk/report/biomedical-bubble/>. Accessed 13 Aug 2018.
17. NCATS 2017. Transforming translational science. Fall 2017. <https://ncats.nih.gov/files/NCATS-factsheet.pdf>. Accessed 9 Aug 2018.
18. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, Farrar K, Park BK, Breckenridge AM. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ*. 2004;329(7456):15–9.
19. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374(9683):86–9.
20. Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. *Trials*. 2017;18:122. <https://doi.org/10.1186/s13063-017-1870-2>.
21. Wall RJ, Shani M. Are animal models as good as we think? *Theriogenology*. 2008;69:2–9.
22. Waring MJ, Arrowsmith J, Leach AR, Leeson PD, Mandrell S, Owen RM, et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov*. 2015;14(7):475–86.
23. Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of investigational drugs in late-stage clinical development and publication of trial results. *JAMA Intern Med*. 2016;176(12):1826–33.
24. Harrison RK. Phase II and phase III failures: 2013–2015. *Nat Rev Drug Discov*. 2016;15:817–8.
25. Henderson VC, Kimmelman J, Fergusson D, Grimshaw JM, Hackam DG. 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(7):e1001489.
26. Bailoo JD, Reichlin TS, Würbel H. Refinement of experimental design and conduct in laboratory animal research. *ILAR J*. 2014;55(3):383–91.
27. Hooijmans CR, Ritskes-Hoitinga M. Progress in using systematic reviews of animal studies to improve translational research. *PLoS Med*. 2013;10(7):e1001482.
28. Hirst J, Howick J, Aronson J, Roberts N, Perera R, Koshiaris C, et al. The need for randomization in animal trials: an overview of systematic reviews. *PLoS ONE*. 2014;9(6):e98856.
29. Henderson VC, Demko N, Hakala A, MacKinnon N, Federico CA, Fergusson D, et al. A meta-analysis of threats to valid clinical inference in preclinical research of sunitinib. *Elife*. 2015;4:e08351.
30. Consort Statement. Section 21: generalisability. 2010. <http://www.consort-statement.org/checklists/view/32-consort-2010/120-generalisability>. Accessed 21 May 2018.
31. Crossley NA, Sena E, Goehler J, Horn J, van der Worp B, Bath PM, Macleod M, Dirnagl U. Empirical evidence of bias in the design of experimental stroke studies: a metaepidemiologic approach. *Stroke*. 2008;39(3):929–34.
32. Dirnagl U, Endres M. Found in translation. *Stroke*. 2014;45:1510–8.
33. Sutherland BA, Minnerup J, Balami JS, Arba F, Buchan AM, Kleinschnitz C. Neuroprotection for ischaemic stroke: translation from the bench to the bedside. *Int J Stroke*. 2012;7(5):407–18.
34. Sandercock PA, Ricci S. Controversies in thrombolysis. *Curr Neurol Neurosci Rep*. 2017;17(8):60.
35. Voelkl B, Vogt L, Sena ES, Würbel H. Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biol*. 2018;16(2):e2003693. <https://doi.org/10.1371/journal.pbio.2003693>.
36. Perlman RL. Mouse models of human disease. An evolutionary perspective. *Evol Med Public Health*. 2016;2016(1):170–6.
37. Martin B, Ji S, Maudsley S, Mattson MP. 'Control' laboratory rodents are metabolically morbid: why it matters. *Proc Natl Acad Sci*. 2010;107(14):6127–33.
38. Malfait AM, Little CB. On the predictive utility of animal models of osteoarthritis. *Arthritis Res Ther*. 2015;17(1):225.
39. Howells D, Macleod M. Evidence-based translational medicine. *Stroke*. 2013;44:1466–71.
40. de Vries RB, Buma P, Leenaars M, Ritskes-Hoitinga M, Gordijn B. Reducing the number of laboratory animals used in tissue engineering research by restricting the variety of animal models. *Articular cartilage tissue engineering as a case study. Tissue Eng Part B Rev*. 2012;18(6):427–35.
41. Corbett D, Carmichael ST, Murphy TH, Jones TA, Schwab ME, Jolkonen J, et al. Enhancing the alignment of the preclinical and clinical stroke recovery research pipeline: consensus-based core recommendations from the stroke recovery and rehabilitation roundtable translational working group. *Int J Stroke*. 2017;12(5):462–71.
42. Vesterinen HM, Sena E, French-Constant C, Williams A, Chandran S, Macleod M. Improving the translational hit of experimental treatments in multiple sclerosis. *Mult Scler J*. 2010;16(9):1044–55.
43. Zeiss CJ, Allore HG, Beck AP. Established patterns of animal study design undermine translation of disease-modifying therapies for Parkinson's disease. *PLoS ONE*. 2017;12(2):e0171790.
44. Zeeff SB, Kunne C, Bouma G, de Vries RB, te Velde AA. Actual usage and quality of experimental colitis models in preclinical efficacy testing: a scoping review. *Inflamm Bowel Dis*. 2016;22(6):1296–305.
45. Vervliet B, Raes F. Criteria of validity in experimental psychopathology: application to models of anxiety and depression. *Psychol Med*. 2013;43(11):2241–4.
46. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, Salman RA, Chan AW, Glasziou P. Biomedical research: increasing value, reducing waste. *Lancet*. 2014;383(9912):101–4.
47. Perlman RL. Response to: is animal research sufficiently evidence based to be a cornerstone of biomedical research? *BMJ*. 2014;348:g3387.
48. Perlman RL. *Evolution and medicine*. Oxford: Oxford University Press; 2013.
49. Lynch VJ. Use with caution: developmental systems divergence and potential pitfalls of animal models. *Yale J Biol Med*. 2009;82(2):53.
50. Preuss TM. Who's afraid of *Homo sapiens*? *J Biomed Discov Collab*. 2006;1(1):17.
51. Langley G. Considering a new paradigm for Alzheimer's disease research. *Drug Discov Today*. 2014;19(8):114–124.
52. Darwin C. *On the origin of species by means of natural selection*. London: John Murray; 1859.
53. Sjoberg E. Logical fallacies in animal model research. *Behav Brain Funct*. 2017;13(1):3.
54. Mestas J, Hughes CC. Of mice and not men: differences between mouse and human immunology. *J Immunol*. 2004;172(5):2731–8.
55. Bailey J, Taylor K. Non-human primates in neuroscience research: the case against its scientific necessity. *ATLA*. 2016;43(1):43–69.
56. Attarwala H. TGN1412: from discovery to disaster. *J Young Pharm*. 2010;2(3):332–6.
57. Greek R, Hansen L. Questions regarding the predictive value of one evolved complex adaptive system for a second: exemplified by the SOD1 mouse. *Prog Biophys Mol Biol*. 2013;113:231–53.
58. Howick J, Glasziou P, Aronson J. Problems with using mechanisms to solve the problem of extrapolation. *Theor Med Bioeth*. 2013;34(4):275–91.
59. Kimmelman J, Mogil JS, Dirnagl U. Distinguishing between exploratory and confirmatory preclinical research will improve translation. *PLoS Biol*. 2014;12(5):e1001863.
60. Leung V, Rousseau-Blass F, Beauchamp G, Pang DSJ. 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*. 2018;13(5):e0197882.
61. Enserink M. Sloppy reporting on animal studies proves hard to change. *Science*. 2017;357(6358):1337–8.
62. Ronaldson-Bouchard K, Vunjak-Novakovic G. Organs-on-a-chip: a fast track for engineered human tissues in drug development. *Cell Stem Cell*. 2018;22(3):310–24.

63. Hood L. Systems biology and p4 medicine: past, present, and future. *Rambam Maimonides Med J*. 2013;4(2):e0012. <https://doi.org/10.5041/RMMJ.10112>.
64. Gan W. Interview with a thought leader on systems medicine—Weiniu Gan, PhD. *Syst Med*. 2018;1(1):9–10.
65. RIVM. National Institute for Public Health and the Environment, RIVM. Roadmap for animal-free innovations in regulatory safety assessment. 2018. <https://www.rivm.nl/dsresource?objectid=571f1221-8ada-4488-99ef-2424acd50d30>. Accessed 26 June 2018.
66. NCAD. Netherlands National Committee for the protection of animals used for scientific purposes. *Transition to non-animal research: on opportunities for the phasing out of animal procedures and the stimulation of innovation without laboratory animals*. 2016. <https://www.ncadierproevenbeleid.nl/documenten/rapport/2016/12/15/ncad-opinion-transition-to-non-animal-research>. Accessed 26 Apr 2018.
67. Birnbaum LS, Thayer KA, Bucher JR, Wolfe MS. Implementing systematic review at the National Toxicology Program: status and next steps. *Environ Health Perspect*. 2013;121(4):a108.
68. Burt T, Yoshida K, Lappin G, Vuong L, John C, Wildt SN, et al. Microdosing and other phase 0 clinical trials: facilitating translation in drug development. *Clin Transl Sci*. 2016;9(2):74–88.
69. Casati S. Integrated approaches to testing and assessment. *Basic Clin Pharmacol Toxicol*. 2018;123:51–5.
70. Baker M. Tissue models: a living system on a chip. *Nature*. 2011;471(7340):661.
71. Barrile R, van der Meer AD, Park H, Fraser JP, Simic D, Teng F, et al. Organ on chip recapitulates thrombosis induced by an anti-CD154 monoclonal antibody: translational potential of advanced microengineered systems. *Clin Pharmacol Ther*. 2018; <https://doi.org/10.1002/cpt.1054>.
72. Passini E, Britton OJ, Lu HR, Rohrbacher J, Hermans AN, Gallacher DJ, et al. Human in silico drug trials demonstrate higher accuracy than animal models in predicting clinical pro-arrhythmic cardiotoxicity. *Front Physiol*. 2017;8:668.
73. Bavli D, Prill S, Ezra E, Levy G, Cohen M, Vinken M, et al. Real-time monitoring of metabolic function in liver-on-chip microdevices tracks the dynamics of mitochondrial dysfunction. *Proc Natl Acad Sci*. 2016;113(16):E2231–40.
74. Prill S, Bavli D, Levy G, Ezra E, Schmälzlin E, Jaeger MS, et al. Real-time monitoring of oxygen uptake in hepatic bioreactor shows CYP450-independent mitochondrial toxicity of acetaminophen and amiodarone. *Arch Toxicol*. 2016;90(5):1181–91.
75. Van Esbroeck AC, Janssen AP, Cognetta AB, Ogasawara D, Shpak G, van der Kroeg M, et al. Activity-based protein profiling reveals off-target proteins of the FAAH inhibitor BIA 10-2474. *Science*. 2017;356(6342):1084–7.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

