Failed trials for central nervous system disorders do not necessarily invalidate preclinical models and drug targets

Anton Beshpalov, Thomas Steckler, Bruce Altevogt, Elena Koustova, Phil Skolnick, Daniel Deaver, Mark J. Millan, Jesper F. Bastlund, Dario Doller, Jeffrey Witkin, Paul Moser, Patricio O'Donnell, Ulrich Ebert, Mark A. Geyer, Eric Prinssen, Theresa Ballard and Malcolm Macleod

A recent article identified five key technical determinants that make substantial contributions to the outcome of drug R&D projects (Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* 13, 419–431 (2014))¹. Careful consideration of such determinants might be particularly valuable in the fields of neurology and psychiatry, in which successful drug development has declined precipitously over the past decade. This decline has largely been fuelled by a high failure rate in the translation of preclinical efficacy findings, caused by multiple factors (see [Supplementary information S1 \(table\)](#)), including limited training and poor protocol design, inadequate animal models, insufficiently validated therapeutic targets and problems with data handling and reporting.

Here, we focus on three factors that can be addressed immediately in order to re-evaluate the therapeutic potential of older drugs and targets and to increase the probability of success for future preclinical-to-clinical translation: data robustness, data generalizability and target engagement data, a factor that was also highlighted in the recent article¹. We argue that the many failed clinical trials in neuropsychiatry do not necessarily invalidate the potential of a drug target or an animal model. Rather, these failures indicate a need for improved experimental designs and a robust translational strategy to better inform compound and dose selection for clinical trials. We conclude that many of the drugs and targets in neuropsychiatry that have been discarded because of negative clinical trial outcomes may deserve re-evaluation using contemporary knowledge, methodology and tools.

Robustness

The problem of robustness in preclinical data is best illustrated by an example from research in amyotrophic lateral sclerosis (ALS), a severe progressive neurodegenerative disease. There is currently one approved medication for ALS, riluzole, which has only

modest effects on survival. Numerous other drug candidates have reported efficacy in a superoxide dismutase 1 (SOD1) mouse model of ALS (one of the common animal models for this disease), but none of these candidates produced an efficacy signal in clinical trials. The ALS Therapy Development Institute later rigorously retested more than 100 of those molecules in the SOD1 model (using adequate statistical power, treatment groups matched for litter and gender, blinding, uniform end point criteria, tracking of non-ALS deaths and quantitative analysis of transgene copy number prior to assigning mice to a study), and they were unable to replicate any of the previously reported preclinical efficacy findings². In this context, the lack of clinical efficacy is not surprising.

Similar problems related to deficiencies in experimental design (such as inadequate blinding and randomization) had previously been observed in studies with animal models of stroke and multiple sclerosis³. The potential impact of such experimental design problems can be assessed retrospectively for drugs that have already been approved or abandoned, and steps can be taken to improve the robustness of experiments for drugs in development (see [Supplementary information S2 \(table\)](#)).

A hallmark of the scientific method is the replication of findings both within and between laboratories. However, such replication is limited by cost, human resources, time and bioethical considerations. Additionally, “there is an almost irresistible pressure to stop when the result is about what one expects it to be,” according to Terry Quinn⁴. However, the more novel the findings of an experiment appear, the less likely they are to be true⁵, especially in the context of poorly designed and underpowered studies. This problem reflects the pressure to publish novel findings in high-impact journals before being scooped by a rival laboratory or funding runs out. The conventional value attached to such publications for career advancement and future

funding often outweighs the efforts required to rigorously challenge the novel findings; thus, verification in an independent laboratory is unlikely.

As the costs of clinical studies are so much higher than those of preclinical development, one might assume that pharmaceutical companies would conduct robust replications of key findings. This is indeed the case during lead optimization, candidate selection, testing different administration routes and the use of primary preclinical disease models. However, this is far less common for studies in the more complex disease models that are used in late stages of development and are potentially more relevant for predicting clinical efficacy.

In general, the rigorousness with which preclinical data is obtained — and the resulting robustness of the data — is quite low; few studies report randomization, blinding, sample size calculations or attrition.

Generalizability

Every laboratory has a unique combination of protocols, suppliers of tools and reagents, source of animals, and animal husbandry characteristics. As drugs are used in highly heterogeneous patient populations, efficacy observed in a single lab is more likely to be successfully translated when similar findings can also be obtained under different conditions in other laboratories. There is empirical evidence to support this assumption: the broader the range of circumstances and laboratory environments in which preclinical efficacy can be demonstrated, the higher the likelihood of detecting efficacy signals in clinical studies⁶. A recent European Union-funded initiative (the Multicentre Preclinical Animal Research Team ([MultiPART](#))) established web-based platforms for multicentre animal studies, and the National Institute of Ageing supports an [Interventions Testing Program](#) that seeks to validate the efficacy of treatments for ageing across several test sites with adequately powered, rigorous experiments using genetically heterogeneous mice of both sexes.

Generalizability of preclinical data is not only an issue concerned with laboratory conditions and animal strains, age and sex. For example, there is a remarkable paucity of studies employing chronic or subchronic drug administration. Given that even a second dose of most drugs can alter the biological milieu (for example, tolerance, sensitization or receptor regulation), chronic dosing studies are important for generating the best predictions of effects in patients. It is largely unknown whether a compound's efficacy has been confirmed in properly designed studies

with chronic administration in animals before initiation of a clinical trial⁷. This information is particularly important for those indications for which preclinical models do not require repeated administration to detect drug efficacy, while the treatment duration in clinical trials can range from weeks to several months, depending on the indication.

Target engagement

In the context of hypothesis-driven drug research, any observation of clinical efficacy is serendipitous if the molecule does not engage its biological targets at the dose tested. Various modelling tools can be used to assess target engagement, and direct target occupancy assays including positron emission tomography (PET) are increasingly available. The aim is to demonstrate that, at relevant doses, the drug is present in the same compartment as the target and in appropriate free concentrations to bind to it. PET studies demonstrate receptor occupancy but not targeted downstream effects; such approaches may not be appropriate for novel drugs working through allosteric or non-competitive molecular mechanisms, and for some targets PET tracers are not yet available.

Scientists at Pfizer considered evidence of exposure at the site of action, target binding and expression of functional pharmacological activity for 44 Phase II programmes across several therapeutic areas⁸. In 43% of cases, it was reported that the target mechanism had not been adequately tested owing to the lack of evidence of target engagement. Similarly, AstraZeneca reported that 40% of efficacy failures in Phase II projects could be attributed to a lack of clear target linkage to a disease or validated animal model, and 29% could be attributed to a lack of data establishing tissue exposure¹.

Even the use of cerebrospinal fluid concentrations to guide dosing may be misleading for intracellular targets or for compounds that are actively transported across the blood–brain barrier. Similar considerations may apply to antibody therapeutics. Key parameters required for brain penetration have been discussed⁹, and the failure of AZD8529 — a positive allosteric modulator of metabotropic glutamate receptor 2 (mGluR2) — in a Phase II study in schizophrenia has been attributed, in part, to unreliable target engagement¹.

We have used the [Thomson Reuters Cortellis database](#) to identify drug development projects in schizophrenia between 1994 and 2014; we could not find evidence of biomarker-driven dose selection for 80% of 72 novel drugs subjected to Phase II clinical proof-of-concept studies (FIG. 1).

Figure 1 | Analysis of the use of biomarkers in the development of novel treatments for schizophrenia. The Thomson Reuters Cortellis database (searched on January 21, 2014) was used to identify drug development projects in schizophrenia between 1994 and 2014. We could not find evidence of biomarker-driven (for example, using positron emission tomography (PET), MRI or electroencephalography (EEG)) dose selection for 80% of 72 novel drugs that were evaluated in Phase II clinical proof-of-concept studies in this time period. 5-HT_{2A}, 5-hydroxytryptamine receptor 2A; 5-HT₆, 5-hydroxytryptamine receptor 6; α₇, α₇ nicotinic receptor; σ, σ receptor; DA D₁, dopamine receptor 1; GABA, γ-aminobutyric acid; GlyT₁, sodium- and chloride-dependent glycine transporter 1; H₃, histamine receptor 3; M₁, muscarinic acetylcholine receptor 1; mGlu, metabotropic glutamate receptor; NK₃, neurokinin receptor 3; NMDAR, N-methyl-D-aspartate receptor; PDE10, phosphodiesterase 10.

Path forward

Efforts to improve the robustness of preclinical data will lead to better study designs. Strengthening of publication policies and open access to data (including negative data) is an additional key way to improve data reliability and transparency. The Preclinical Data Forum was established with support from the European College of Neuropsychopharmacology (ECNP)¹⁰ and is developing an online platform to enable scientists to exchange unpublished data in a pre-competitive manner and to share knowledge on the use of tool compounds. This platform should facilitate disclosure of large amounts of pre-competitive information and should be paralleled by the development of consensus approaches to data robustness and demonstrating generalizability.

There may be compounds that have failed in clinical proof-of-concept studies owing to poor target engagement or that might be more appropriate for other diseases sharing similar mechanisms. The [ECNP medicines chest](#) provides a list of pharmacological tools no longer under development and can be used to obtain further clinical information on a particular target. Such target revalidation efforts should ideally occur in a pre-competitive space and may involve the development of new business models.

Finally, improved training on appropriate study designs will ensure that preclinical research is conducted to the highest standards,

with appropriate protocol design informed by statistical and power analyses similar to those processes now standard for clinical trials.

Summary and conclusions

Herein, we challenge the widely held view that the high failure rate in neuropsychiatry trials invalidates both the drug targets chosen and the preclinical models used. We argue instead that the scientific community should attach greater importance to issues of data robustness, data generalizability and target engagement when designing preclinical studies. We do not seek to detract attention from the fundamental need to strengthen our scientific understanding of disease mechanisms, improve clinical testing strategies, and develop better disease models. Our premise is that an increase in robustness, generalizability and evidence of target engagement will increase the probability of successful translation of preclinical findings into Phase II efficacy. We are optimistic that the adoption of these approaches will enhance our ability to bring improved and much needed medicines to patients.

Anton Beshpalov was previously at Neuroscience Research, AbbVie, 6706 Ludwigshafen, Germany. Present address: Partnership for Assessment and Accreditation of Scientific Practice, Am Aukopf 14/1, D-69118 Heidelberg, Germany and Institute of Pharmacology, Pavlov Medical University, 197022 St Petersburg, Russia.

Thomas Steckler is at Janssen Research and Development, B-2340 Beerse, Belgium.

Bruce Altevogt is at Global Policy and International Public Affairs, Pfizer, New York, 10017 New York, USA.

Elena Koustova and Phil Skolnick are at the National Institute on Drug Abuse, National Institutes of Health, Bethesda, Maryland 20892, USA.

Daniel Deaver is at Non-Clinical Research and Development, Alkermes, Waltham, Massachusetts 02451, USA.

Mark J. Millan is at the Institut de Recherche Servier, 78290 Croissy sur Seine, France.

Jesper F. Bastlund is at Neuroscience Research, H. Lundbeck A/S, Copenhagen, 2500 Valby, Denmark.

Dario Doller was previously at Discovery Chemistry & DMPK; Lundbeck Research USA, Paramus, New Jersey 07652, USA. Present address:

Concert Pharmaceuticals, Inc., 99 Hayden Avenue, Lexington, Massachusetts 02421, USA.

Jeffrey Witkin is at Neuroscience Discovery Research, Lilly Research Labs, Eli Lilly and Company, Indianapolis, Indiana 46285, USA.

Paul Moser was previously at the Pierre Fabre Research Institute, 81106 Castres, France.

Present address: BIAL-Portela & Ca S.A., Avenida da Siderurgia Nacional, 4745-457 São Mamede do Coronado, Portugal.

Patricio O'Donnell is at the Neuroscience and Pain Research Unit, Pfizer, Cambridge, Massachusetts 02139, USA.

Ulrich Ebert is at Boehringer Ingelheim Pharma, 55218 Ingelheim am Rhein, Germany.

Mark A. Geyer is at the University of California San Diego, La Jolla, California 92093, USA.

Eric Prinssen and Theresa Ballard are at Roche Pharma Research and Early Development, Neuroscience, Ophthalmology and Rare Diseases, Roche Innovation Center Basel, CH-4070 Basel, Switzerland.

Malcolm Macleod is at Edinburgh University, Old College, South Bridge, Edinburgh EH8 9YL, UK.

Correspondence to A.B. anton.bespalov@paasp.net

doi:10.1038/nrd.2016.88

Published online 17 Jun 2016

1. Cook, D. *et al.* Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* **13**, 419–431 (2014).
2. Scott, S. *et al.* Design, power, and interpretation of studies in the standard murine model of ALS. *Amyotroph. Lateral Scler.* **9**, 4–15 (2008).
3. van der Worp, H. B. *et al.* Can animal models of disease reliably inform human studies? *PLoS Med.* **7**, e1000245 (2010).
4. Quinn, T. Don't stop the quest to measure Big G. *Nature* **505**, 455 (2014).
5. Ioannidis, J. P. Why most published research findings are false. *PLoS Med.* **2**, e124 (2005).
6. Richter, S. H. *et al.* Systematic variation improves reproducibility of animal experiments. *Nat. Methods* **7**, 167–168 (2010).

7. Bespalov, A. *et al.* Drug tolerance: a known unknown in translational neuroscience. *Trends Pharmacol. Sci.* <http://dx.doi.org/10.1016/j.tips.2016.01.008> (2016).
8. Morgan, P. *et al.* Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. *Drug Discov. Today* **17**, 419–424 (2012).
9. Di, L. *et al.* Demystifying brain penetration in central nervous system drug discovery. *J. Med. Chem.* **56**, 2–12 (2013).
10. Steckler, T. *et al.* The preclinical data forum network: A new ECNP initiative to improve data quality and robustness for (preclinical) neuroscience. *Eur. Neuropsychopharmacol.* **25**, 1803–1807 (2015).

Acknowledgements

The authors thank Martien Kas (Utrecht University), Michael Decker (AbbVie), Lynn Butler-David (Exciva), Martin Weber (Genentech), Jurgen Gottowik (Roche) and Katja Brose (Cell Press) for stimulating discussion and helpful comments.

Competing interests statement

The authors declare [competing interests](#): see Web version for details.

FURTHER INFORMATION

ECNP medicines chest: <https://www.ecnp.eu/projects-initiatives/ECNP-medicines-chest.aspx>

Interventions Testing Program: <https://www.nia.nih.gov/research/dab/interventions-testing-program-itp>

MultiPART: <http://www.dcn.ed.ac.uk/multipart/>

Thomson Reuters Cortellis database: <https://cortellis.thomsonreuterslifesciences.com/>

SUPPLEMENTARY INFORMATION

See online article: [S1](#) (table) | [S2](#) (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF