

Believe it or not: how much can we rely on published data on potential drug targets?

Florian Prinz, Thomas Schlange and Khusru Asadullah

A recent report by Arrowsmith noted that the success rates for new development projects in Phase II trials have fallen from 28% to 18% in recent years, with insufficient efficacy being the most frequent reason for failure (Phase II failures: 2008–2010. *Nature Rev. Drug Discov.* **10**, 328–329 (2011))¹. This indicates the limitations of the predictivity of disease models and also that the validity of the targets being investigated is frequently questionable, which is a crucial issue to address if success rates in clinical trials are to be improved.

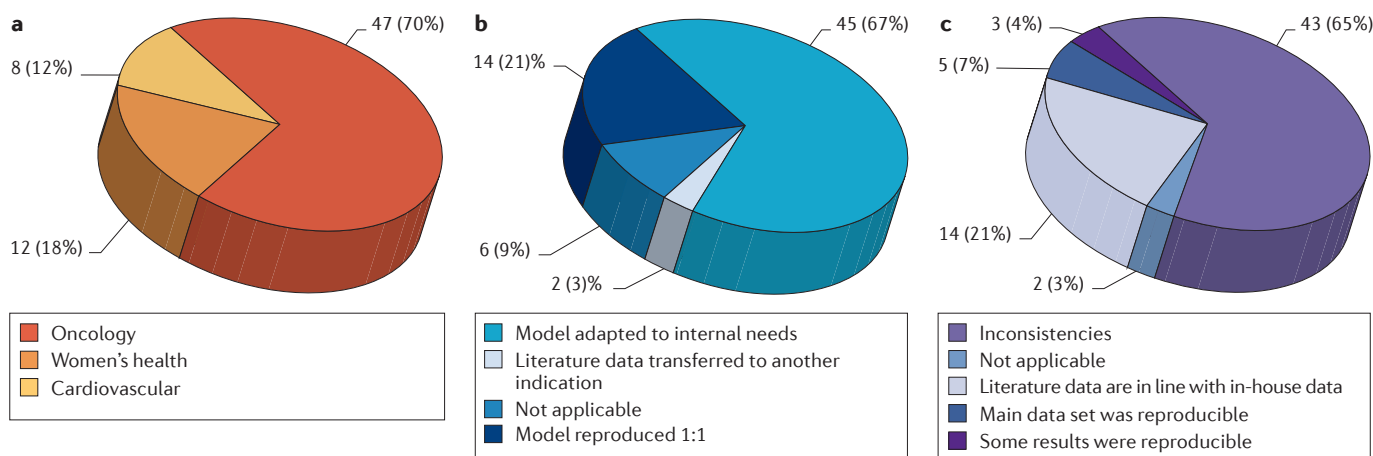
Candidate drug targets in industry are derived from various sources, including in-house target identification campaigns, in-licensing and public sourcing, in particular based on reports published in the literature and presented at conferences. During the transfer of projects from an academic to a company setting, the focus changes from ‘interesting’

to ‘feasible/marketable’, and the financial costs of pursuing a full-blown drug discovery and development programme for a particular target could ultimately be hundreds of millions of Euros. Even in the earlier stages, investments in activities such as high-throughput screening programmes are substantial, and thus the validity of published data on potential targets is crucial for companies when deciding to start novel projects.

To mitigate some of the risks of such investments ultimately being wasted, most pharmaceutical companies run in-house target validation programmes. However, validation projects that were started in our company based on exciting published data have often resulted in disillusionment when key data could not be reproduced. Talking to scientists, both in academia and in industry, there seems to be a general impression that many

results that are published are hard to reproduce. However, there is an imbalance between this apparently widespread impression and its public recognition (for example, see REFS 2,3), and the surprisingly few scientific publications dealing with this topic. Indeed, to our knowledge, so far there has been no published in-depth, systematic analysis that compares reproduced results with published results for wet-lab experiments related to target identification and validation.

Early research in the pharmaceutical industry, with a dedicated budget and scientists who mainly work on target validation to increase the confidence in a project, provides a unique opportunity to generate a broad data set on the reproducibility of published data. To substantiate our incidental observations that published reports are frequently not reproducible with quantitative data, we performed an analysis of our early (target identification and validation) in-house projects in our strategic research fields of oncology, women’s health and cardiovascular diseases that were performed over the past 4 years (FIG. 1a). We distributed a questionnaire to all involved scientists from target discovery, and queried names, main relevant published data (including citations), in-house data obtained and their relationship to the published data, the impact of the results obtained for the outcome of the projects, and the models



d

	Model reproduced 1:1	Model adapted to internal needs (cell line, assays)	Literature data transferred to another indication	Not applicable
In-house data in line with published results	1 (7%)	12 (86%)	0	1 (7%)
Inconsistencies that led to project termination	11 (26%)	26 (60%)	2 (5%)	4 (9%)

Figure 1 | Analysis of the reproducibility of published data in 67 in-house projects. **a** | This figure illustrates the distribution of projects within the oncology, women’s health and cardiovascular indications that were analysed in this study. **b** | Several approaches were used to reproduce the published data. Models were either exactly copied, adapted to internal needs (for example, using other cell lines than those published, other assays and so on) or the published data was transferred to models for another indication. ‘Not applicable’ refers to projects in which general hypotheses could not be verified. **c** | Relationship of published data to in-house data. The proportion

of each of the following outcomes is shown: data were completely in line with published data; the main set was reproducible; some results (including the most relevant hypothesis) were reproducible; or the data showed inconsistencies that led to project termination. ‘Not applicable’ refers to projects that were almost exclusively based on in-house data, such as gene expression analysis. The number of projects and the percentage of projects within this study (**a–c**) are indicated. **d** | A comparison of model usage in the reproducible and irreproducible projects is shown. The respective numbers of projects and the percentages of the groups are indicated.

that were used in the experiments and publications. The questionnaire can be obtained from the authors.

We received input from 23 scientists (heads of laboratories) and collected data from 67 projects, most of them (47) from the field of oncology. This analysis revealed that only in ~20–25% of the projects were the relevant published data completely in line with our in-house findings (FIG. 1c). In almost two-thirds of the projects, there were inconsistencies between published data and in-house data that either considerably prolonged the duration of the target validation process or, in most cases, resulted in termination of the projects because the evidence that was generated for the therapeutic hypothesis was insufficient to justify further investments into these projects.

We wondered whether heterogeneous experimental conditions could be an explanation for the frequent inconsistencies (FIG. 1b). Interestingly, a transfer of the models — for example, by changes in the cell lines or assay formats — was not crucial for the discrepancies that were detected. Rather, either the results were reproducible and showed transferability in other models, or even a 1:1 reproduction of published experimental procedures revealed inconsistencies between published and in-house data (FIG. 1d). Furthermore, despite the low numbers, there was no apparent difference between the different research fields. Surprisingly, even publications in prestigious journals or from several independent groups did not ensure reproducibility. Indeed, our analysis revealed that the reproducibility of published data did not significantly correlate with journal impact factors, the number of publications on the respective target or the number of independent groups that authored the publications.

Our findings are mirrored by ‘gut feelings’ expressed in personal communications with scientists from academia or other companies, as well as published observations. An unspoken rule among early-stage venture capital firms that “at least 50% of published studies, even those in top-tier academic journals, can’t be repeated with the same conclusions by an industrial lab” has been recently reported (see Further information) and discussed⁴. The challenge of reproducibility — even under ideal conditions — has also been highlighted, indicating that even in an optimal setting (the same laboratory, the same people, the same tools and the same assays, with experiments separated by 5 months), there were substantial variations, as the intra- and interscreen reproducibility of two genome-scale small interfering RNA screens was influenced by the methodology of the analysis and ranged from 32–99% (REF. 5).

There may be several reasons for the observed lack of reproducibility. Among these, incorrect or inappropriate statistical analysis of results or insufficient sample sizes, which result in potentially high numbers of irreproducible or even false results, have been discussed⁶. Among the more obvious yet unquantifiable reasons, there is immense competition among laboratories and a pressure to publish. It is conceivable that this may sometimes result in negligence over the control or reporting of experimental conditions (for example, a variation in cell-line stocks and suppliers, or insufficient description of materials and methods). There is also a bias towards publishing positive results, as it is easier to get positive results accepted in good journals. It remains to be studied further whether there are indeed hurdles to publishing results that contradict data from high-impact journals or the currently established scientific opinion in a given field, which could lead to the literature supporting a certain hypothesis even if there are many (unpublished) data arguing against it. One might speculate that the above mentioned issues should be eliminated by the peer review system. However, reviewers have no time and no resources to reproduce data and to dig deeply into the presented work. As a consequence, errors often remain undetected⁷. Adding to this problem, many initially rejected papers will subsequently be published in other journals without substantial changes or improvements^{8,9}.

We are aware that our data set — albeit quite large for wet-lab science — is still rather small and its statistical significance can be questioned. We are also aware that our own experimental results might also be irreproducible in other laboratories. However, the aim of our target validation work is: first, to increase confidence in the biology of the targets with an unbiased approach; second, to provide assays that need to be reliable during later stages such as compound optimization; and third, to transfer these assays to various laboratories in other departments in-house. With an average project duration of 6–12 months, numerous well-established cellular and *in vivo* models and several independent and often specialized laboratories that are involved in the projects with highly qualified scientists who are dedicated to target discovery, we feel confident that our data are quite reliable. It is important, however, to emphasize that we do not want to make the point that our experimental data are correct, whereas data from other groups are ‘false’. We are not reporting fraud, but a lack of reproducibility. In fact, to our knowledge, none of the studies that our internal projects were based on was retracted or suspected to be flawed.

However, with reasonable efforts (sometimes the equivalent of 3–4 full-time employees over 6–12 months), we have frequently been unable to reconfirm published data.

Our observations indicate that literature data on potential drug targets should be viewed with caution, and underline the importance of confirmatory validation studies for pharmaceutical companies and academia before larger investments are made in assay development, high-throughput screening campaigns, lead optimization and animal testing. Effective target validation, however, should not just be confirmatory, but should complement the knowledge on a particular target. An in-depth biological understanding of a target is required and should contribute to a reduction in the high attrition rates that are observed in early clinical development.

Florian Prinz is at Target Research Berlin, Bayer HealthCare, Müllerstraße 178, 13342 Berlin, Germany.

Thomas Schlange is at Target Research Wuppertal, Bayer HealthCare, Aprather Weg 18a, 42096 Wuppertal, Germany.

Khusru Asadullah is Vice President and Head of Target Discovery at Bayer HealthCare, Müllerstraße 178, 13342 Berlin, Germany.

Correspondence to K.A.

e-mail: khusru.asadullah@bayer.com

1. Arrowsmith, J. Phase II failures: 2008–2010. *Nature Rev. Drug Discov.* **10**, 328–329 (2011).
2. Lehrer, J. The truth wears off: is there something wrong with the scientific method? *The New Yorker* [online], http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer (2010).
3. Freeman, D. H. Lies, damned lies, and medical science. *The Atlantic* [online], <http://www.theatlantic.com/magazine/archive/2010/11/lies-damned-lies-and-medical-science/8269> (2010).
4. Osherovich, L. Hedging against academic risk. *SciBX* 14 Apr 2011 (doi:10.1038/scibx.2011.416).
5. Barrows, N. J., Le Sommer, C., Garcia-Blanco, M. A. & Pearson, J. L. Factors affecting reproducibility between genome-scale siRNA-based screens. *J. Biomol. Screen.* **15**, 735–747 (2010).
6. Ioannidis, J. P. Why most published research findings are false. *PLoS Med.* **2**, e124 (2005).
7. Schroter, S., et al. What errors do peer reviewers detect, and does training improve their ability to detect them? *J. R. Soc. Med.* **101**, 507–514 (2008).
8. Nemery, B. What happens to the manuscripts that have not been accepted for publication in *Occupational and Environmental Medicine*? *Occup. Environ. Med.* **58**, 604–607 (2001).
9. McDonald, R. J., Cloft, H. J. & Kallmes, D. F. Fate of submitted manuscripts rejected from the *American Journal of Neuroradiology*: outcomes and commentary. *Am. J. Neuroradiol.* **28**, 1430–1434 (2007).

Acknowledgements

We would like to thank B. Kreft and T. Zollner for their valuable contributions to this project, S. Schoepe for support in the data analysis and S. Decker for support with bioinformatics analysis of the results.

Competing interests statement

The authors are employees of Bayer Healthcare.

FURTHER INFORMATION

Life Sci VC: Academic bias & biotech failures: http://lifescivc.com/2011/03/academic-bias-biotech-failures/#_undefined_0

ALL LINKS ARE ACTIVE IN THE ONLINE PDF