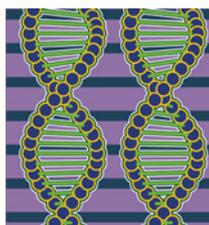




Setback for glucose reabsorption inhibitors p645



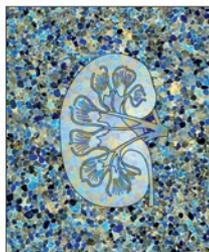
DNA patent decision p650



Head of Patient-Centered Outcomes Research Institute discusses its agenda p652



Identifying R&D outliers p653



New drug to prevent transplant rejection p655



## Reliability of ‘new drug target’ claims called into question

Bayer halts nearly two-thirds of its target-validation projects because in-house experimental findings fail to match up with published literature claims, finds a first-of-a-kind analysis on data irreproducibility.

*Asher Mullard*

An unspoken industry rule alleges that at least 50% of published studies from academic laboratories cannot be repeated in an industrial setting, wrote venture capitalist Bruce Booth in a recent [blog post](#). A first-of-a-kind analysis of Bayer’s internal efforts to validate ‘new drug target’ claims now not only supports this view but suggests that 50% may be an underestimate; the company’s in-house experimental data do not match literature claims in 65% of target-validation projects, leading to project discontinuation.

“People take for granted what they see published,” says John Ioannidis, an expert on data reproducibility at Stanford University School of Medicine in California, USA. “But this and other studies are raising

deep questions about whether we can really believe the literature, or whether we have to go back and do everything on our own.”

For the non-peer-reviewed analysis, Khusru Asadullah, Head of Target Discovery at Bayer, and his colleagues looked back at 67 target-validation projects, covering the majority of Bayer’s work in oncology, women’s health and cardiovascular medicine over the past 4 years. Of these, results from internal experiments matched up with the published findings in only 14 projects, but were highly inconsistent in 43 (in a further 10 projects, claims were rated as mostly reproducible, partially reproducible or not applicable; see article online at [go.nature.com/utqTzl](http://go.nature.com/utqTzl)). “We came up with some shocking examples of discrepancies between published data

and our own data,” says Asadullah. These included inability to reproduce: over-expression of certain genes in specific tumour types; and decreased cell proliferation via functional inhibition of a target using RNA interference.

Irreproducibility was high both when Bayer scientists applied the same experimental procedures as the original researchers and when they adapted their approaches to internal needs (for example, by using different cell lines). High-impact journals did not seem to publish more robust claims, and, surprisingly, the confirmation of any given finding by another academic group did not improve data reliability. “We didn’t see that a target is more likely to be validated if it was reported in ten publications or in two publications,” says Asadullah.

Although the analysis is limited by a small sample size, and cannot itself be checked because of company confidentiality concerns, other studies point to similarly sobering conclusions. In one study researchers tried, and largely failed, to repeat the findings of published microarray gene expression analyses by working directly from the data sets the original conclusions were drawn from (*Nature Genet.* **41**, 149–155; 2009). In another study, when an identical sample of proteins was sent to different proteomics laboratories, the vast majority failed to independently, and therefore reproducibly, identify all of the component proteins (*Nature Methods* **6**, 423–430; 2009).

Although data to date primarily focus on the reliability of academic published findings, industry-based research may have the same issues. “I don’t want to make the point that academia is bad and industry is good,” says Asadullah. “Results from pharma and biotech companies probably wouldn’t fare much better in terms of reproducibility.” Indeed, the high failure rates observed for Phase III trials in recent years (*Nature Rev. Drug Discov.* **10**, 87; 2011) — which in part reflect an inability to reproduce positive Phase II findings — suggest that the problem spreads into all stages of the pipeline.

### Why don’t the data hold up?

The underlying causes of irreproducibility are not all equally problematic or damning. In some cases, irreproducible results may represent scientific food for thought: perhaps an assumed constant is actually variable from laboratory to laboratory, leading to the possibility of further discovery and improved scientific understanding. In other cases, irreproducibility — and the fact that biomedical research often ultimately advances through trial, error and revision — is simply a cost of science.

But negative, potentially addressable, systemic causes of irreproducibility remain pervasive and troublesome. In an essay provocatively titled “Why Most Published Research Findings Are False”, Ioannidis points out key culprits, including investigator prejudice, incorrect statistical methods, competition in hot fields and publishing bias, whereby journals are more likely to publish positive and novel findings than negative or incremental results (*PLoS Med.* **2**, e124; 2005).

A failure to work to industry standards may also contribute to the problem. To this end, Philip Cohen, who runs the Division of Signal Transduction Therapy at the University of Dundee, UK, and works closely with multiple pharmaceutical partners, has enacted preventive measures to safeguard the validity

Results from pharma and biotech companies probably wouldn’t fare much better

of his team’s results. “Many researchers borrow clones from other labs without ever checking them out properly,” he explains. “We stopped borrowing materials back in 1996, because of problems we inherited from the clones and samples sent in by other labs.”

Booth, at Atlas Venture in Boston, USA, adds that a lack of industry perspective may be a factor as well. For instance, the quintessential biomedical claim that a potential compound is ‘safe and well tolerated’ in animals may be included in a published paper even when the academic investigators did not look at the relevant industrial toxicity readouts. “It’s not that the academic scientists have made a fraudulent statement; they just define ‘safe and well tolerated’ differently from the industrial ones.” As academia and industry continue to forge closer ties — through the formation of partnerships and potentially via the US National Institute of Health’s new National Center for Advancing Translational Sciences (NCATS) programme — the issues of industry standards and perspective may fade away.

Asadullah’s unexpected discovery that even findings reported by more than one team tend to be irreproducible in industry hands highlights another problem: a potential lack of true independence between academic groups. In addition to sharing materials, “people may be running the same experiments with the same biases, or maybe they feel committed to finding the same results”, says Ioannidis. “These are not really truly independent replications,” he adds.

And at the far end of the negative spectrum, unfortunately, also lies outright fraud. Everyone hopes that the prevalence of such behaviour is rare, but it nonetheless happens. In one high-profile case, for example, Duke University had to suspend cancer clinical trials that used biomarkers to match patients to therapy after learning that the underlying science may have been faked (*Nature* **469** 139–140; 2011).

### Avoiding irreproducibility

In light of its internal analysis, Bayer has already changed its approach to target validation, says Asadullah. “We’ve become much more cautious when working with published targets.” Whereas the company sometimes used to set up high-throughput screens against published targets prior to internal validation, they now

always check out at least some of the data themselves at the bench first.

Booth, as an investor, relies on contract research organizations (CROs) to confirm published data before moving forward. “We’ll often set up a very small amount of capital — call it a few hundred thousand dollars — and use it to validate academic claims.” Another approach he takes is to cultivate ties with researchers whose work has repeatedly proved to be reproducible. “There are investigators who we have worked with repeatedly, who use the right controls, the right level of rigor and have the right level of scepticism about their own data.”

But against the backdrop of falling research and development (R&D) budgets, increased financial risk-aversion and high attrition rates, broader strategies to improve data robustness are needed as well. One possibility, proposes Asadullah, would be to create a precompetitive consortium that could assess the reproducibility of reported findings. The advantages gained by not having to pursue dead-end claims would offset any lost competitive edge. Another, he adds, would be to encourage the scientific community — including academic and industry players, as well as journal editors — to publish more negative data.

Ioannidis further argues that researchers need to be rewarded for actions that improve data robustness. “We could have some kind of bonus — funding or recognition — for people who publically deposit their samples, data and protocols, and who show that findings are reproducible,” he says. Where such a framework — or funding to implement it — would come from, however, is unclear.

Another possibility, suggests Booth, could be to tap into the expertise of university technology transfer offices (TTOs). Whereas many TTOs have set up microseed funds to spin companies out of academic findings, he argues that the TTOs would better serve both the universities and the broader community by funding either CROs or other research teams to independently validate claims. “Their data packages would then become so much stronger and better able to attract early-stage investors,” he says. “A lot of biotechs are currently formed sooner than they should be,” he adds.

Ultimately, however, more data may be needed before the field can move forward effectively. This first analysis from Bayer provides an important glimpse of the prevalence of irreproducibility, says Booth, but still only represents a small sample set. “It would be great if we could get more companies to pool their data on irreproducibility so that we can unravel the drivers of the translatability of academic findings,” he adds.