Photo: Fotolia/wimammoth9

Drug target controversy

Academic Truth or Biotech Bullshit

Are unreliable academic results a poor investment for the future? Researchers in biotechnology and the pharmaceutical industry claim the quality of academic research has become so bad that the majority of its published results cannot be reproduced. Jeremy Garwood reports on some startling revelations.

The pharmaceutical industry has been finding it increasingly hard and costly to produce new drugs. Various reasons for this apparent failure have been presented, not the least of which is that biology is incredibly complex and never ceases to throw up new surprises.

However, in the last year, a new argument has come to the fore that poses some fundamental questions about the foundations of biomedical science. Researchers in commercial, for-profit companies accuse much publicly-funded academic research of being unreliable. Too many scientific publications present positive results of pharmaceutical interest that simply cannot be reproduced by other laboratories!

But they say this is not due to conscious fraud. Instead, the industrial researchers point to research bias and academic self-interest that encourages academic researchers to publish exaggerated, unrepresentative or insufficiently proven data as positive findings.

"Academic bias & biotech failures"

For the venture capitalist, Bruce Booth, this was apparently just another topic for his weekly blog at *Life Sci VC*. But response to his comments suggests it's a widely-held view in the biotechnology industry: "Most academic research results are not reproducible!" (*lifescivc.com*, 28/03/11).

Booth from Atlas Venture in Boston helps "start and fund emerging life science companies – therapeutics, diagnostics, tools, and devices". On his site he describes himself as a "Recovering scientist turned early stage venture capitalist" (he has a PhD in immunology from Oxford University). He is also "a biotech optimist fighting gravity", reflecting the current difficulty that Booth and other venture capitalists are having, trying to convince investors that biotechnology start-up companies really can produce products and make money, despite abundant evidence to the contrary.

His blog begins by describing the failure of yet another company created around an academic lab's discoveries; in this case, a "fascinating new approach to drugging hot receptor targets". Apparently, the company failed because nobody outside the founder's lab could reproduce the key scientific results.

"The company spent $5 million or so trying to validate a platform that didn't exist. When they tried to directly repeat the academic founder's data, it never worked. Upon re-examination of the lab notebooks, it was clear the founder's lab had, at the very least, massaged the data and shaped it to fit their hypothesis. Essentially, they systematically ignored every piece of negative data."

Several times in the past decade, Booth says his own company has been affected by this "failure to repeat" problem.

The unspoken rule of academic investment

In fact, he says, it's such a frequent problem that he now assumes the following "unspoken rule" before investing:

"At least 50% of the studies published even in top tier academic journals – *Science*, *Nature*, *Cell*, *PNAS*, etc… – can't be repeated with the same conclusions by an industrial lab. In particular, key animal models often don't reproduce. This 50% failure rate isn't a data-free assertion: it's backed up by dozens of experienced R&D professionals who've participated in the (re)testing of academic findings. This is a huge problem for translational research and one that won't go away until we address it head on."

At least 50%! But why is this happening?

The reality is "we live in a tournament model world of academic research: winners get the spoils, losers get nothing. Publish or peril. Grants are really competitive, and careers are on the line. Only positive findings are typically published, not negative ones. This pressure creates a huge conflict of interest for academics, and a strong bias to write papers that support the hypotheses included in grant applications and prior publications."

Booth says there is a major public misconception when it comes to comparing academic and industrial research – this holds that "objectivity" is only to be found in academic research, while industrial research is dominated by "pervasive bias". He says this is "complete nonsense" but this view still prevails because there is a "rich" literature detailing "Pharma bias" in scientific publications. Over the previous 15 month period, *PubMed* listed 63 articles featuring "pharma conflicts of interest with academics, clini-

cal trial reporting" and general "pharma industry bias" but Booth said he could not find any that addressed "academic bias" or the lack of repeatability of academic findings.

In the absence of published studies, Booth presents his own thoughts, "I'm sure there are cases where it's truly fabrication or falsification of data, but as an optimist I believe that must be a tiny percentage: most of the time I think it's just the influence of bias."

Three manifestations

Booth suggests three ways academic bias could manifest itself:

First, the academic investigator can directly or indirectly apply pressure on their labs to publish sensational "best of all experimental" results rather than the average or typical study;

Then there's that "special sauce" that only seems to exist in the author's lab, those not-so-evident details of how the experiment was done, or what serum was used, or "what specific cells were played with, etc." All of which lead to a "local optimum of activity in the paper that can't be replicated elsewhere".

Finally, Booth points to a tendency for academics to systemically ignore contradictory data in order to support the lab's favoured hypothesis, often leading them to dismiss any conflicting findings as technical or reagent failures.

Importantly, how are venture capitalists that invest in biotech supposed to engage on cool new data when the repeatability is so low? Frankly, Booth says most VCs now avoid investing in early academic spin-outs. This is, in part, due to the "insidious impact of the sector's high failure rate with academic reproducibility (also known as 'bias')".

But for those, like himself, who are still prepared to take the risk, he offers the following advice:

First, investors should understand that "findings from a single academic lab are suspect. If other labs haven't validated it in peer reviewed literature, it's very high risk. It's probably bleeding edge rather than cutting edge. If it's only a single lab, it's likely that only a single post-doc or grad student has actually done the work. Given the idiosyncrasies of lab practices, that's a concentrated risk profile. Wait for more labs to repeat the work, or conduct a full lab notebook audit."

Secondly, "repeating the findings in an independent lab should be gating before investing". Don't dive in with financing before externally validating the data with some real "wet diligence". Sign an option agreement with a material transfer agreement (MTA), then repeat the study in a contract research lab or totally independent academic lab.

His final remark is that technology transfer offices (TTOs) in universities and public research organisations could themselves improve matters by validating the findings of an investigator's work in a reputable contract research lab that industrial partners and VCs would trust. If a TTO could show third party data supporting a lab's striking findings, the prospects for funding would increase significantly.

Booth's blog elicited a range of comments. For example, from Art Krieg, who as a medical researcher discovered the immune stimulatory CpG DNA motif (*Nature* 374: 546-9), then moved into industry (e.g. chief scientific officer at the Oligonucleotide Therapeutics Unit at Pfizer).

He also decries the common misperception that research published by academics is somehow "cleaner" than that from scientists working in industry. "In my own experience, just the opposite is true. In fact, I think your estimate that 50% of high profile academic research is irreproducible is optimistic – I think the truth is even worse, especially in very hot areas, where the pressure for an academic scientists to publish before being scooped is especially intense, and where the rewards for being seen as a leader in the field may be more immediate than the cost of publishing something that turns out to be wrong. Having worked on both sides, there is no question that scientists working in industry just aren't under these publish or perish questions, and they often apply a significantly higher level of rigor before publishing."

Anecdotal evidence

Furthermore, Krieg says, "One of the most important reasons to go to scientific congresses is to find out from friends in these fields what has been reproducible, and what hasn't been. In two cases where I've asked PIs (principal investigators) if there was any difficulty in generating data I knew to be non-reproducible, I was told the post-doc involved had to be pressured hard to get the 'right' data! The majority of these non-reproducible papers are never retracted, but those in the field all know pretty soon."

Working at a contract research organisation (CRO), commentator "Johnnyboy" agrees, "This is a very common problem. Companies (big and small) ask us to apply a model or technique in a published article, which on closer inspection turns out to have been deeply flawed, or described so poorly that it is impossible to reproduce. Trying to explain that to the client is understandably difficult, as the usual attitude out there is: 'Well, it's published, so it's true. Why can't you just do the same?'"

"Jay", a postdoc, wrote, "In academia, people are well aware of the problem, but as stake holders they are very unlikely to do anything about it, true to the motto: 'Just don't rock the boat!' But has anybody ever considered the actual costs of this issue to society? The money spent on trying to reproduce other people's results is enormous. I am a postdoc and have, on two occasions, been involved in a project that was entirely based on somebody else's data and bias. On both occasions I wasted almost one year before the project was abandoned and the public investment (my lowly salary and reagent costs) evaporated into thin air. From my point of view, the way academic research is conducted these days is deeply flawed, starting with funding allocation, career structures/prospects, publishing and peer review. It's all rather corrupt. There is no truly independent oversight." But more job security in public research might help, "People who don't have to be afraid of losing their job and 'careers' are more likely to escape the pressures that lead to flawed data, which in the end would save society, as a whole, a lot of resources."

Big Pharma and compromised drug discovery

An anonymous academic researcher "working in a high profile and competitive area", confirms that "the vast majority of papers (especially those in high profile journals) are not reproducible. In the case of genetics, most papers are real (since it's hard to fudge a mutation). However, most so-called 'functional' papers (in the past this was called cell biology) are not reproducible. But it is not in anyone's interest to publicly call them 'bullshit,' so no one does (or, very, very rarely). Since papers in high profile journals tend to beget papers in high profile journals, this type of science becomes self-perpetuating."

"Respisci", a researcher in biotech says, "I don't think that I am alone in being unable to reproduce other's results. I recall an instance during my post-doc where, in experiment after experiment, our lab consistently demonstrated a 30% effect of X on Y, while the published results had >70%. Our PI was getting more and more frustrated with his own team and questioned the skill/expertise of our group (and we were questioning ourselves). During a visit by the head of the lab, which had published the findings, when asked about this 70% effect, he causally replied 'that was a best result. Typically we see 30%'. To this day, I can still recall the shockwave that went through that room. For nearly a year our team was trying to reproduce results, which even the parent lab couldn't achieve. On top of that our results were correct, which taught me an important lesson in believing in my own team/data. I fear that academic bias is alive and well."

Photo: Fotolia/Arlwasabi

No new drugs due to unreliable data from academia – do society and patients have to swallow the bitter pill?

Well, of course, despite what Booth and his fellow bloggers and commentators might say, critics can always dismiss it as 'anecdotal evidence'. Where are the hard facts? Can you really prove that academic results are as bad as you claim?

This is when Big Pharma decided to take an interest. Germany's Bayer AG is a huge company (>110,000 employees, total revenue 36.5 billion euros in 2011), famous for discovering aspirin and heroin; Bayer HealthCare, its pharmaceutical and medical products subgroup, accounts for half its revenue.

In their article: "Believe it or not: how much can we rely on published data on potential drug targets?" three of Bayer HealthCare's drug discovery researchers decided to look at their own experiences and came up with some "quantitative data" (*Nature Reviews Drug Discovery,* 10: 712).

Florian Prinz and Thomas Schlange work on Target Research in Wuppertal and Berlin, while Khusru Asadullah is Head of Target Discovery in Berlin. They agreed with Bruce Booth that there seems to be a general impression among scientists, both in academia and industry, that many published results are hard to reproduce. However, "To our knowledge, 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."

Substantial investments

Exactly this is their speciality: Target Identification and Validation at the earliest stages of drug discovery (see text box for "The Drug Discovery Pipeline"). At Bayer Target Discovery, they find candidate drug targets from public sourcing, "in particular based on reports published in the literature and presented at conferences" as well as in-house target identification campaigns and in-licensing.

However, when they transfer 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." This is why, 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."

The three aims of their target validation work are:

- to increase confidence in the biology of the targets with an unbiased approach;
- to provide assays that need to be reliable during later stages, such as compound optimization;
- 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." However, they can't say the same for the academic data: "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."

So, to get a measure of how big a problem this is, Prinz & co. conducted an analysis covering the previous four year's R&D on their early stage in-house projects (i.e. target identification and target validation) for Bayer's three main strategic research fields: oncology, women's health and cardiovascular diseases.

To do this, they simply distributed a questionnaire to all the relevant scientists at Bayer. This asked them for details of the published academic data used in their respective projects and how well their own in-house data matched up with the published results. Had their own in-house results significantly affected the project outcomes?

Replies came back from 23 heads of laboratory, providing details of 67 projects. Oncology research accounted for 70% of these.

They found that the relevant published data was "in line with our in-house findings" for only 14 out of the 67 projects (21%). Of these, just one project perfectly reproduced the reported data, although 12 others could be "adapted to internal needs" while the last one was not applicable.

However, in two-thirds of the projects (43 out of 67), the inconsistencies between published results and their in-house data was big enough for Bayer to completely terminate the projects!

These inconsistencies "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".

Observed lack of reproducibility

Why was this happening? The Bayer researchers wondered whether heterogeneous experimental conditions could be an explanation, for example, that there were differences in the cell lines or assay formats used between the academic and industrial labs, but this "was not crucial" for the detected discrepancies.

Instead, they found a clear distinction: either the published results were reproducible and showed transferability in other models (which is what they wanted), or there were inconsistencies between published and in-house data, even when they could reproduce the published experimental procedures "1:1".

When they looked in more detail at the original papers reporting the irreproducible data, they were surprised to discover that "even publications in prestigious journals or from several independent groups did not ensure reproducibility". In fact, their 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".

The Bayer scientists proposed several reasons for their observed lack of reproducibility:

- Bad statistics: Incorrect or inappropriate statistical analysis of results or insufficient sample sizes, which result in potentially high numbers of irreproducible or even false results.
- Publication pressures: "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)."
- Positive bias: 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.

Photo: www.danisfoundation.org

The Drug Discovery 'Pipeline'

- **Target identification** – the 'target' is usually the naturally-existing cellular or molecular structure involved in the pathology of interest. The drug-in-development is meant to act on this target.
- **Target validation** – this involves gathering functional information about the target. A better scientific understanding and publication history exists for 'established' targets, describing how the target functions both in normal physiology and human pathology.
- **Target to hit** – find compounds that are active against ('hit') the selected target, e.g. by high-throughput screening.
- **Hit to lead** – confirm the 'hit' and expand information about the identified molecule.
- **Lead optimisation** – synthesise lead compounds, new analogues with improved potency, reduced off-target activities and physiochemical/metabolic properties suggestive of reasonable *in vivo* pharmacokinetics.
- **Pre-clinical** – determine a product's preliminary safety profile, including 'absorption, distribution, metabolism, excretion and toxicity' (ADMET) testing on animals. Estimate a safe starting dose of the drug for clinical trials in humans.
- **Clinical trials:**

Phase I: pharmacovigilance and dose-ranging using increasing sub-therapeutic doses (5 - 100 people; determines whether drug is safe to check for efficacy)
Phase II: testing of drug using therapeutic doses (ca. 100-300 people; determines whether drug can have any efficacy)
Phase III: testing of drug for intended use as therapy (300-3000 patients; determines a drug's therapeutic effect)
Phase IV: post-marketing surveillance, monitoring drug use in public; observe drug's long term effects on anyone seeking treatment from their physician.

▸ Ineffective peer review: "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 undetected. Adding to this problem, many initially rejected papers will subsequently be published in other journals without substantial changes or improvements."

Despite this, the Bayer researchers insist, "We are not reporting fraud, but a lack of reproducibility. We do not want to make the point that our experimental data are correct, whereas data from other groups are 'false'."

Nevertheless, "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" (they noted that the industry's overall success rates at Phase II trials had fallen from 28% to 18% between 2006 and 2010).

Big Pharma and poor cancer therapy

The revelations from Bayer prompted another Big Pharma researcher to analyse his own company files. In "Raise standards for preclinical cancer research" (*Nature* 483: 531-3) Glenn Begley, head of Hematology and Oncology Research at Amgen (total revenue >$15 billion for 2011) explained that the quality of published preclinical data was a significant contributor to failure in oncology trials.

It takes many years before the clinical applicability of initial preclinical observations is known. Therefore, the results of preclinical studies "must be very robust to withstand the rigours and challenges of clinical trials, stemming from the heterogeneity of both tumours and patients". This is why the Amgen replication team of "about 100 scientists" try to confirm relevant published findings before pursuing a particular line of research.

Over the previous decade, Glenn Begley had identified findings from 53 "landmark" publications (papers in top journals, from reputable labs) for his team to double-check and reproduce.

Begley said he knew some of the data might not hold up because they had deliberately selected completely new findings, such as fresh approaches to targeting cancers or alternative clinical uses for existing therapeutics. However, they discovered that only 6 out of 53 cases (11%) could be confirmed, i.e. 47 of the 53 "findings" could not be reproduced – 89% failure! "Even knowing the limitations of preclinical research, this was a shocking result!"

The shocked Begley explained, "These are the studies the pharmaceutical industry relies on to identify new targets for drug development. But if you're going to place a $1 million or $2 million or $5 million bet on an observation, you need to be sure it's true. As we tried to reproduce these papers, we became convinced you can't take anything at face value" (*Reuters.com*, 28/03/12: "In cancer science, many 'discoveries' don't hold up").

Faced with so many findings that they could not reproduce, Begley said that Amgen made an attempt to contact the original authors. They discussed what might account for their inability to confirm the results. Some authors let them borrow antibodies and other materials used in the original study. Others even allowed them to repeat experiments in their labs under the original authors' direction.

Talking to the authors

However, in his *Reuters*' interview, Begley revealed that authors also required Amgen to sign a confidentiality agreement barring them from disclosing data at odds with the original findings. This means that "the world will never know" which 47 studies, many of them highly cited, are apparently wrong.

He said the most common response from the challenged scientists was "you didn't do it right". Meanwhile, others worried that something less innocuous explained the lack of reproducibility. At a cancer conference, Begley even had breakfast with the lead scientist of one of the problematic studies. He described how they "went through the paper line-by-line, figure-by-figure. I explained that we re-did their experiment 50 times and never got their result. He said they'd done it six times and got this result once, but put it in the paper because it made the best story!"

Comparing the 'good' and 'bad' papers, Begley notes that in the reproducible studies, the authors had paid close attention to controls, reagents, investigator bias and describing the complete data set.

Meanwhile, in papers where results could not be reproduced, "investigators frequently presented the results of one experiment, such as a single Western-blot analysis. They sometimes said they presented specific experiments that supported their underlying hypothesis, but that were not reflective of the entire data set".

In fact, there was an inherent bias since the investigators were not "blinded" to the experimental versus control groups. This meant researchers knew, which cell line, or mouse, got a particular treatment or had cancer. This is a problem when data are subject to interpretation since a researcher who is "intellectually invested in a theory is more likely to interpret ambiguous evidence in its favor".

A flawed system

What reasons underlie the publication of erroneous, selective or irreproducible data? For Begley, the answer is clear, "The academic system and peer-review process tolerates and perhaps even inadvertently encourages such conduct. To obtain funding, a job, promotion or tenure, researchers need a strong publication record, often including a first-authored high-impact publication. Journal editors, reviewers and grant-review committees often look for a scientific finding that is simple, clear and complete – a 'perfect' story. It is therefore tempting for investigators to submit selected data sets for publication, or even to massage data to fit the underlying hypothesis. But there are no perfect stories in biology. Journals and grant reviewers must allow for the presentation of imperfect stories, and recognize and reward reproducible results, so that scientists feel less pressure to tell an impossibly perfect story to advance their careers."

Public investment in academic research costs billions of euros a year. However, "success" is largely measured by the number of published research papers and their journal 'impact factors', not by the quality of the science. Ironically, in many countries, public researchers are now actively encouraged to help industry. But how much is that help really worth?

Jeremy Garwood