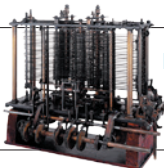


# COMMENT

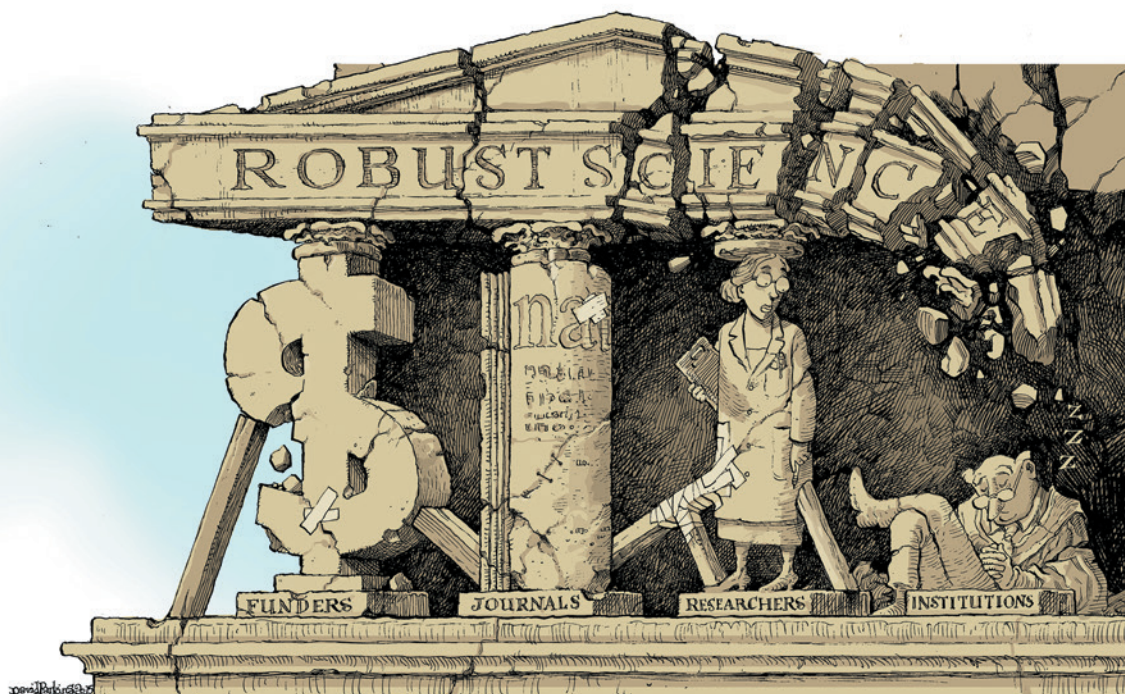
**ENERGY** Subsidize renewables to break fossil-fuel stronghold and spur carbon pricing **p.27**

**HISTORY** Richard Holmes on the scientific and cultural legacy of Ada Lovelace **p.30**



**EQUALITY** Memoir of a salmon run to nowhere **p.31**

**CONSERVATION** Offsets are often the only way to meet commitments **p.33**



## Institutions must do their part for reproducibility

Tie funding to verified good institutional practice, and robust science will shoot up the agenda, say **C. Glenn Begley, Alastair M. Buchan and Ulrich Dirnagl**.

**I**reproducible research poses an enormous burden: it delays treatments, wastes patients' and scientists' time, and squanders billions of research dollars. It is also widespread. An unpublished 2015 survey by the American Society for Cell Biology found that more than two-thirds of respondents had on at least one occasion been unable to reproduce published results. Biomedical researchers from drug companies have reported that one-quarter or fewer of high-profile papers are reproducible<sup>1,2</sup>.

Many parties are addressing the problem. Funding bodies such as the US National Institutes of Health (NIH) have announced

training initiatives<sup>3</sup> and explicitly instructed grant reviewers to consider whether experimental plans ensure rigour. New methods of data analysis and peer review have been proposed to deflate bias.

Several journals, including *Nature* and *Science*, have updated their guidelines and introduced checklists. These ask scientists whether they followed practices such as randomizing, blinding and calculating appropriate sample size. *Science* has also added statisticians to its panel of reviewing editors. Philanthropic and non-profit organizations have sponsored projects to improve robustness.

Funders' policies, journal guidelines and

widespread soul-searching are necessary. But they are not sufficient.

Conspicuous by their absence from these efforts are the places in which science is done: universities, hospitals, government-supported labs and independent research institutes. This has to change. Institutions must support and reward researchers who do solid — not just flashy — science and hold to account those whose methods are questionable.

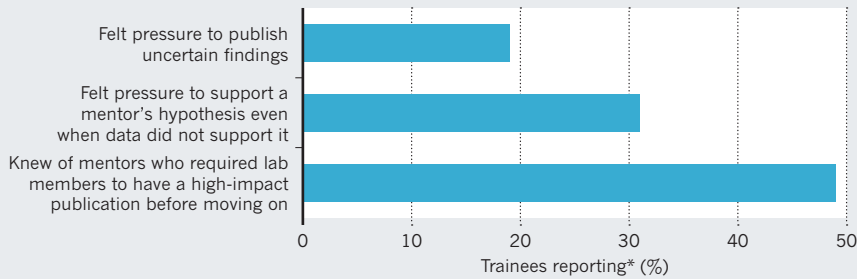
### SPOT THE SHIRKERS

Although researchers want to produce work of long-term value, multiple pressures and prejudices discourage good scientific

ILLUSTRATION BY DAVID PARKINS

## PRESSURED FINDINGS

A survey of US biomedical trainees suggests that the push to publish spurs unreliable results.



\*Online survey of ~140 trainees at the MD Anderson Cancer Center in Houston, Texas.

► practices. In many laboratories, the incentives to be first can be stronger than the incentives to be right.

Discussions of conflicts of interest typically centre on relationships with industry, but academic scientists face more pernicious, even existential temptations. Monetary rewards are often less important than the 'currency' with which scientists advance their careers: high-level publications lead to funding opportunities, promotions, awards and other forms of recognition. These markers of scientific achievement become proxies for assessment of the work itself, and further encourage spectacular, but less than substantiated, research.

Amplifying these pressures is a human prejudice in favour of our own ideas. There is a very real temptation to ignore a result that does not conform to our preconceptions, or to recast it so that it does. Data-dredging is used to find statistically significant results that justify a publication. Sound practices such as blinding, multiple repeats, validated reagents and appropriate controls<sup>4</sup> are dismissed as luxuries or nuisances.

Research institutions contribute to and benefit from these perverse incentives. They bathe in the reflected glory of their faculty; they trumpet breakthroughs published in top-tier journals, lauding achievements to the media and donors. Some even pay investigators for publications. Many require that investigators generate their salary from research grants.

An anonymous survey of around 140 trainees at the MD Anderson Cancer Center in Houston, Texas, found that nearly one-third had felt pressure to prove a mentor's hypothesis even though their experimental results did not support it, and nearly one-fifth had themselves published results they considered less than robust<sup>5</sup>. Nearly half knew of mentors who required lab members to publish a high-impact paper to complete training in their labs (see 'Pressured findings').

Although important, the checklists introduced by journals do nothing to shift the focus from results to the legitimacy of the process by which the results are produced. Researchers encounter these lists after they

have drawn conclusions and are ready to announce them — not when planning their research. There is no mechanism to verify that listed practices were actually employed.

The core instinct of scientists — scepticism — is punished by the current system. Institutions have a duty to reform it. They must shoulder their responsibility for training graduate students and postdoctoral fellows, for supporting the scientific behaviour of their faculty members and for the knowledge that emanates from their endeavours.

## GOOD INSTITUTIONAL PRACTICE

Although there are some protections against outright fraud, few institutions have strong, transparent processes in place to discourage poor-quality science or to foster objectivity. We propose that research institutions that receive public funding should apply the same kind of oversight and support to ensure research integrity as is routinely applied for animal husbandry, biosafety and clinical work.

To conduct animal research, investigators must hold licences and undergo continuous education. Institutions appoint delegates to monitor compliance, and those delegates are held to account by regulators. Similar oversight is used for work with radioactivity and human embryonic stem cells.

These functions could be broadened to encompass established guidelines for research conduct, such as the ARRIVE (Animal Research: Reporting of In Vivo Experiments) and MIAME (Minimum Information About a Microarray Experiment) guidelines, and data sharing as required by the NIH and the National Science Foundation.

Standards already exist that define good laboratory practice to test chemicals for toxicity, good manufacturing practice and good clinical practice. These systems were introduced to ensure a degree of consistency, quality and integrity. Procedures

**"Most institutions will not make the necessary move unless forced."**

are in place to ensure compliance.

The scientific community should come up with a similar system for research, which we term good institutional practice (GIP). If funding depended on a certified record of compliance with GIP, robust research would get due recognition.

At a minimum, GIP should consist of the following tenets.

### Routine discussion of research methods.

Many labs already comb through data and methods as a group before submitting a paper. Such discussions should be broadened and formalized across an institution. Regular department and cross-department meetings should be established to dissect manuscripts in preparation. Methods and processes (rather than conclusions) would be debated just as a competitor's paper might be critiqued in a journal club. Primary research material would be available. This practice is roughly analogous to the 'Morbidity and Mortality' conferences routine in hospitals, in which working hours are also intense.

Regular critique sessions help scientists to learn to defend their science without feeling defensive. Investigators publicly hold each other to account, and trainees learn what to demand of their own research. Anxieties can be raised informally, highlighting institutional weaknesses and systematic errors. The practice also puts a short-term focus on what has traditionally been a long-term reward: a reputation for careful science.

**Reporting systems.** Also well-established in clinical medicine is a system to anonymously flag occurrences that did or could have jeopardized a patient's care. Such systems are often the only way workers dare to raise concerns and admit mistakes. Similarly colleagues, graduate students and postdocs should be able to discuss concerns about sloppy science without jeopardizing their careers. Designated co-mentors, a departmental ombudsman or existing university offices of research integrity could be charged with providing a forum for informal, confidential discussions. Any formal reports should be investigated in a balanced and impartial way.

**Training and standards.** Some sloppiness stems from ignorance. Many investigators determine whether trainees are ready to move on by gauging the number and impact factors of their publications; instead, supervisors should base such decisions on whether their lab members understand research methods and process. Compulsory institutional training should ensure a common understanding of rigorous experimental design, research standards and objective evaluation of data. Faculty members and trainees should demonstrate their ability to spot problems such as 'cherry picking' data to make the best



story. Compliance with research standards, including data-sharing, should be supported, audited and acknowledged.

**Records and quality management.** Laboratory notebooks and records must be available for independent review. Electronic laboratory notebooks facilitate collaboration, supervision and record keeping, and can link records to the original data. One of our institutions (U.D.'s) is now adopting these system-wide. Random audits should be conducted to guarantee that experimental data are duly recorded and that elements of good research practice are routine. Such spot-checks are commonplace in industry.

**Appropriate incentive and evaluation systems.** Institutions should find ways to deter non-compliance with guidelines, poor mentoring and scientific sloppiness. Faculty members with poor records should face loss of laboratory space and trainees, decreased funding and potential demotion. Conversely, faculty members who excel as mentors and careful experimentalists should be rewarded. Appropriate metrics should be developed so that promotions are based on robustness and high-quality mentoring, rather than simply on high-profile publications<sup>6</sup>. Surveys such as that conducted at MD Anderson exemplify one way in which administrators can gain the insight necessary to improve the research environment. Institution-level metrics could help to monitor overall performance and remind all researchers and administrators of their responsibility to the scientific community.

**Enforcement.** Institutions should investigate egregious lapses and record them in a routine, transparent way. Departments of research integrity or other centres of excellence should be funded, staffed and given enough authority to prevent, detect, investigate and penalize poor-quality research. They should also be charged with promoting an institutional culture that nurtures robustness.

#### GETTING TO GIP

The systems needed to promote reproducible research must come from institutions — scientists, funders and journals cannot build them on their own. These kinds of changes will require additional money, infrastructure, personnel and paperwork. The load on institutions and investigators will be real, but so is the burden of irreproducible research. Even if it is accompanied by an apparent decrease in productivity, the resulting increase in research quality will be well worth the costs.

Still, most institutions will not make the necessary moves unless forced. Funding bodies should make GIP a prerequisite for receiving a grant. The concept has gained some traction: last year, Science Foundation Ireland announced plans to conduct external audits on some of the labs that it supports.

There will not be one ideal solution. Faculty members, trainees and administrators will need to come together for honest, difficult discussions to restructure institutions. Neither scientists nor institutions should engage in mere box checking; new practices must restrain sloppiness while interfering only minimally with the many scientists who are behaving well.

Large-scale change is possible. In the 1970s, clinical research had little rigour or oversight. Now clinical trials routinely include concurrent control groups, double-blinding, pre-specified endpoints, power calculations to determine patient numbers and analysis plans that thwart bias. In addition, primary data are available for independent statistical analysis by regulatory authorities. At the time, these changes were controversial; many physicians believed them to be unnecessary and regressive.

Nothing an institution can do will prevent misconduct altogether. This is not the goal. Rather, it is to support the work of well-meaning scientists, to reduce the waste from biased results, and to relieve some of the pressures that encourage sloppy science. ■

**C. Glenn Begley** is chief scientific officer at TetraLogic Pharmaceuticals, Malvern, Pennsylvania, USA. **Alastair M. Buchan** is dean of medicine and head of the Medical Science Division, Oxford NIHR Biomedical Research Centre, University of Oxford, UK. **Ulrich Dirnagl** is director of the Center for Stroke Research at the Universitätsmedizin Charité Berlin, Germany. e-mail: [cgbegley@tlog.com](mailto:cgbegley@tlog.com)

1. Begley, C. G. & Ellis, L. M. *Nature* **483**, 531–533 (2012).
2. Prinz, F., Schlange, T. & Asadullah, K. *Nature Rev. Drug Discov.* **10**, 712–713 (2011).
3. Collins, F. S. & Tabak, L. A. *Nature* **505**, 612–613 (2014).
4. Begley, C. G. *Nature* **497**, 433–434 (2013).
5. Mobley, A., Linder, S. K., Braeuer, R., Ellis, L. M. & Zwilling, L. *PLoS ONE* **8**, e63221 (2013).
6. Ioannidis, J. P. & Khoury, M. J. *J. Am. Med. Assoc.* **312**, 483–484 (2014).

# Push renewables to spur carbon pricing

Make wind and solar power even cheaper by opening up access to the electricity grid and ending fossil-fuel subsidies, urge **Gernot Wagner** and colleagues.

Putting a price on carbon dioxide and other greenhouse gases to curb emissions must be the centrepiece of any comprehensive climate-change policy. We know it works: pricing carbon creates broad incentives to cut emissions. Yet the current price of carbon remains much too low relative to the hidden environmental, health and societal costs of burning a tonne of coal or a barrel of oil<sup>1</sup>. The global average price is below zero, once half a trillion dollars of fossil-fuel subsidies are factored in.

Momentum towards effective carbon pricing is building. California, joined by the Canadian province of Quebec, leads by pricing 85% of such emissions at around US\$12 per tonne. Sweden applies the highest value globally on half of its carbon dioxide emissions, at up to \$125 per tonne. The European Union has the largest system in terms of tonnes covered, pricing 45% of its greenhouse-gas emissions at about \$8 per tonne. China is experimenting with regional cap-and-trade systems. And the US Clean Power Plan encourages states

to meet emissions-reduction targets through market-based mechanisms. Yet global emissions continue to climb.

The current inadequacy of carbon pricing stems from a catch-22. Policymakers are more likely to price carbon appropriately if it is cheaper to move onto a low-carbon path. But reducing the cost of renewable energies requires investment, and thus a carbon price.

In our view, the best hope of ending this logjam rests with tuning policies to drive down the cost of renewable power ▶