ILLUSTRATION BY GAVIN POTENZA

of a policy that promotes rapid, open access to observing data, following the protocols developed in the International Polar Year[9]. Frameworks for helping to plan and coordinate long-term observing activities across the scientific community and other sectors need to be established.

The community-based observing networks from the International Polar Year, which focus on variables related to local environmental threats or benefits, are a good start. But to be accessible to others, these data should be entered into wider networks such as those of the WMO. Similar to the practice of joint resource management[10], the scientific community, stakeholders and decision-makers all need to be included in governance from the outset to help ensure relevance and efficiency.

Opportunities remain for the private sector to contribute to such collaborative networks. Offering up commercial vessels or infrastructure as platforms for scientific observations, sharing data and engaging the research community in the planning stages of industry observing programmes would go a long way towards establishing a 'network of networks'.

Last month, I was fortunate to be out in a small boat off Toksook Bay in Alaska with ice experts and hunters from the Yup'ik people. We were surrounded by jagged, fast-moving chunks of ice that, to me, seemed hostile. To my companions, it was all in a day's work. I recalled a sentiment I had heard from a marine-mammal expert in Barrow, more than 1,000 kilometres farther north, where the ice is now unstable. He stated that the key to adapting to increasingly dynamic ice is to learn from those to the south, such as in Toksook Bay. The charge to the scientific community is to help to create a foundation for such mutual learning to occur. ■

Hajo Eicken *is professor of geophysics at the University of Alaska Fairbanks, Fairbanks, Alaska 99775, USA.*
e-mail: hajo.eicken@gi.alaska.edu

1. Markus, T., Stroeve, J. C. & Miller, J. *J. Geophys. Res. Oceans* **114,** C12024 (2009).
2. Rampal, P., Weiss, J. & Marsan, D. *J. Geophys. Res. Oceans* **114,** C05013 (2009).
3. Lovecraft, A. L., Meek, C. & Eicken, H. *Polar Geogr.* **36,** 105–125 (2013).
4. Screen, J. A. & Simmonds, I. *Nature* **464,** 1334–1337 (2010).
5. Calder, J., Eicken, H. & Overland, J. in *Understanding Earth's Polar Challenges: International Polar Year 2007–2008 — Summary by the IPY Joint Committee* (eds Krupnik, I. *et al.*) 405–410 (CCI Press, 2011).
6. National Research Council. *Seasonal to Decadal Predictions of Arctic Sea Ice: Challenges and Strategies* (National Academies Press, 2012).
7. Kauker, F. *et al. Geophys. Res. Lett.* **36,** L03707 (2009).
8. Lindsay, R. *et al. Geophys. Res. Lett.* **39,** L21502 (2012).
9. Parsons, M. A. *Nature* **458,** 830 (2009).
10. Berkes, F. *J. Environm. Mgmt.* **90,** 1692–1702 (2009).

Six red flags for suspect work

C. Glenn Begley explains how to recognize the preclinical papers in which the data won't stand up.

A few months ago, I received a desperate e-mail from a postdoctoral scientist. Researchers — including me and my colleagues — had just reported that the majority of preclinical cancer papers in top-tier journals could not be reproduced, even by the investigators themselves[1,2]. The postdoc pleaded with me to identify those papers, saying: "I could be wasting my time working on that project." This was true, but we had signed confidentiality agreements that prevented us from revealing the specific papers. Furthermore, identifying them would not address the broader, systemic issues in research and publishing that create a plethora of papers that don't stand up to scrutiny.

There were some glaring differences ▶

between the 90% of papers that we could not reproduce and the few papers that we could. In our initial exercise[2], we contacted researchers whose work we were unable to reproduce to discuss discrepancies. Occasionally, experiments were repeated by the original authors — the most dramatic results came from investigators who could not reproduce their own work, when performed in their own laboratory, using their own reagents. The only difference the second time was that they had to perform the experiments blinded.

Many of the investigators whose work could not be reproduced were, however, prepared to honestly describe their experimental approaches to us in confidence. These non-reproducible papers shared a number of features, including inappropriate use of key reagents, lack of positive and negative controls, inappropriate use of statistics and failure to repeat experiments. If repeated, data were often heavily selected to present results that the investigators 'liked'. These, we found, are common flaws of non-reproducible papers, which apply to all basic biological research. Addressing them during the writing, editing and reviewing of research could go a long way towards creating a more robust scientific enterprise.

"Many of these flaws were identified and expunged from clinical studies decades ago."

So, here are six questions that every author, editor, reviewer and reader should ask themselves when evaluating a research paper.

SIX QUESTIONS

Were experiments performed blinded? It is much easier to obtain the result that makes the best story and that best fits a hypothesis when experiments are performed by unblinded investigators. So, first, check the methods and figure legends. For instance, animal studies, *in vitro* work and reading of gels — which are used in protein or DNA separation — can and should all be done, or at the very least reviewed, by an investigator blinded to the experimental versus control groups. Even rare lone investigators can introduce some level of blinding. It is unusual to find blinded studies in basic research in top-tier journals. If experiments are performed blinded, it increases the likelihood that the work will stand the test of time.

Were basic experiments repeated? This is crucial to know in any study. Unfortunately, repetitions are seldom performed. Western blotting (a technique that uses antibodies to detect specific proteins in a mixture) and similar analyses are often performed only once, and when the desired result is obtained, that result is shown. Studies using RNA interference frequently show the results of a single experiment. Often only one or two cell lines are examined. If reports fail to state that experiments were repeated, be sceptical.

Were all the results presented? Inappropriate data selection is a crucial issue. Most western blots show only a sliver of the gel with the majority of bands cropped. Although many of these cropped bands may be extraneous, their removal falsely implies that the antibody could detect only the desired protein, which is rarely the case. In addition, size standards are often not shown. Without them, the reader cannot have any confidence that the bands identified are even remotely of the correct size. It can be valuable to compare the results of other experiments in the paper that used the same antibody: the pattern of bands should be the same across experiments.

It is always beneficial to cross-check images. Since the *Journal of Cell Biology* began routinely screening images, it has had to revoke 1% of acceptances after finding digitally manipulated image files[3]. Beware the 'typical result'; ask to see all of the results. One investigator admitted to us that he selected the one atypical result that supported his hypothesis and ignored the majority of experiments that did not.

Were there positive and negative controls? Often in the non-reproducible, high-profile papers, the crucial control experiments were excluded or mentioned as 'data not shown'. Yet it is impossible to evaluate data properly without reviewing the controls. Another common practice is to show photos of gels that are over-exposed and well outside the linear range of the film. Over-exposure of the controls makes it impossible to assess the relative amounts of total material being compared. When arguing that there is a difference between samples in the intensity of a specific signal, it is crucial to know that equivalent amounts of total sample were compared. But with over-exposed controls, that difference is obscured, and an alleged difference between samples may simply be the consequence of loading more total sample. A publication that hides the controls should be viewed with caution.

Were reagents validated? Several errors are common here. Of course, it is vital to know that the selected antibodies detect only the antigen under study. Yet, typically, the crucial western blot (showing only a single band) or other analyses that validate the reagent are not shown. Instead there is often a reference to an earlier paper, which does not show the essential data either. There are also examples of investigators using an antibody even when the manufacturer declares it unfit for that particular purpose. Experiments with small-molecule inhibitors are particularly problematic. Investigators choose to attribute the desired effect to their favourite molecule, ignoring the multiple other targets affected by the inhibitor, or consign the key experiments that allegedly demonstrate their lack of relevance to 'data not shown'.

Were statistical tests appropriate? Improper statistical analysis is commonly seen in animal studies, in which results are collected over a long time. On such a time curve, two points may be highlighted and declared to be significantly different from points on the control curve, even though the totality of the two curves is essentially the same. Check that the statistical test has been applied to the whole curve, rather than just to selected points along it (the position of the asterisk marking the statistical *P* value is an important clue)[4].

Remarkably, these six flaws are common to many papers, even those that we did not include in our original analysis. As an informal exercise, I recently thumbed through the pile of high-profile journals on my desk: the vast majority included at least one paper — and often more — that contained one or more of the basic flaws outlined here. What is also remarkable is that many of these flaws were identified and expunged from clinical studies decades ago. In such studies it is now the gold standard to blind investigators, include concurrent controls, rigorously apply statistical tests and analyse all patients — we cannot exclude patients because we do not like their outcomes.

Why do we repeatedly see these poor-quality papers in basic science? In part, it is down to the fact that there is no real consequence for investigators or journals. It is also because many busy reviewers (and disappointingly, even co-authors) do not actually read the papers, and because journals are required to fill their pages with simple, complete 'stories'. And because of the apparent failure to recognize authors' competing interests — beyond direct financial interests — that may interfere with their judgement.

Every biologist wants and often needs to get a paper into *Nature* or *Science* or *Cell*, yet the scientific community fails to recognize the perverse incentive this creates. Some of these issues could be readily addressed by publishing only blinded, replicated and appropriately controlled preclinical experiments. Isn't that what my postdoc colleague expected we were doing already? ■

C. Glenn Begley *is senior vice-president at TetraLogic Pharmaceuticals, Malvern, Pennsylvania, USA.*
e-mail: cgbegley@tetralogicpharma.com

1. Prinz, F., Schlange, T. & Asadullah, K. *Nature Rev. Drug Discov.* **10,** 712 (2011).
2. Begley, C. G. & Ellis, L. M. *Nature* **483,** 531–533 (2012).
3. Rossner, M. 'How to Guard Against Image Fraud' *The Scientist* (1 March 2006); available at http://go.nature.com/bjjbe4
4. Vaux, D. L. *Nature* **492,** 180–181 (2012).

The author declares competing financial interests: see go.nature.com/qeu6pr for details.