
SAM OGDEN

Publish houses of brick, not mansions of straw

Papers need to include fewer claims and more proof to make the scientific literature more reliable, warns **William G. Kaelin Jr**.

I worry about sloppiness in biomedical research: too many published results are true only under narrow conditions, or cannot be reproduced at all. The causes are diverse, but what I see as the biggest culprit is hardly discussed. Like the proverbial boiled frog that failed to leap from a slowly warming pot of water, biomedical researchers are stuck in a system in which the amount of data and number of claims in individual papers has gradually risen over decades. Moreover, the goal of a paper seems to have shifted from validating specific conclusions to making the broadest possible assertions. The danger is that papers are increasingly like grand mansions of straw, rather than sturdy houses of brick.

The papers leading to my 2016 Lasker prize (with Gregg Semenza and Peter Ratcliffe, for discovering how cells sense oxygen) were published more than a decade ago. Most would be considered quaint, preliminary and barely publishable today. One — showing that a tumour-suppressor protein was required for oxygen signalling — would today be criticized for failing to include a clear mechanism and animal experiments (O. Iliopoulos *et al. Proc. Natl Acad. Sci. USA* **93,** 10595–10599; 1996). Another, showing that the protein's main target underwent an oxygen-dependent modification, was almost rejected because we hadn't identified the enzyme responsible (M. Ivan *et al. Science* **292,** 464–468; 2001). Fortunately, an experienced editor intervened, arguing that publication would open the search for the enzyme to other groups; such reprieves seem less common today.

What is driving today's 'claims inflation'? One factor is the emphasis that funding agencies place on impact and translation. Another is that technological advances have made it easier to generate data, which can be accommodated in online supplements. Both factors encourage reviewers and editors to demand extra experiments that are derivative, tangential to the main conclusion or aimed at increasing impact. And it has always taken more courage to accept a paper than to reject it with suggestions for more experiments. That can create perverse incentives by linking acceptance to a preordained result. I fear that reviewers are especially inclined to ask for more when funding is tight, as it is now.

In years past, an interesting observation described in Figure 1 would lead to a series of experiments focused on its robustness. When I was a postdoctoral fellow, an entire paper could consist of the detection of two proteins that bound to one another and the follow-up experiments to establish that binding occurred in living cells. Today, data supporting such an assertion would consist of one or two experiments described in Figure 1 (or worse, Supplemental Figure 1). The rest of the paper would describe work spanning diverse scientific disciplines that elevate the claims and culminate in a figure with a patina of clinical relevance.

Unfortunately, this breadth often compromises depth. Multiple corroborating lines of evidence are essential to make inferences from experimental data, because any individual approach has pitfalls and limitations. I fear the literature has devolved from papers making a single major claim that is proved in multiple ways to papers having multiple claims, each with a single reed of support. The final figures of papers today often seem a bridge too far.

Overly broad claims push the peer-review system past its limit. Although I am a seasoned reviewer, I find it difficult to wade through the increasing amount of data in papers, and often encounter material where I am not an expert. If this trend continues, it will be necessary to take mini-sabbaticals to review papers. Editors might successfully gather reviewers with complementary backgrounds to examine such broad papers, but they do so at the expense of having multiple experts scrutinize the same experiments. And I worry that the supplemental section, which reviewers tend to inspect less thoroughly, can be used to bury weak data.

OVERLY BROAD CLAIMS PUSH THE PEER-REVIEW SYSTEM PAST ITS LIMIT.

Other unintended consequences are delays in communicating new knowledge and prolongation of training periods because professional advancement becomes yoked to producing a magnum opus that takes years to complete. Unanswered questions and unexplained results are often perceived as weaknesses that jeopardize publication. This can encourage bad behaviours, such as cherry picking data so that nothing seems incomplete, inconsistent or unexplained. We should appreciate that papers strengthen science when they candidly acknowledge limitations and puzzling results.

Lack of knowledge is the true bottleneck to clinical translation. We need to stop telling basic scientists, especially trainees, that their work's value lies in its translatability. We must return to more careful examination of research papers for originality, experimental design and data quality, and adopt more humility about predicting impact, which can truly be known only in retrospect (transformative discoveries such as restriction enzymes, yeast cell-cycle mutants and CRISPR–Cas9 were once considered simply oddities of nature). We should also place more emphasis on the quality of a body of work and whether it has enabled subsequent discoveries, and focus less on where individual papers are published.

The main question when reviewing a paper should be whether its conclusions are likely to be correct, not whether it would be important if it were true. Real advances are built with bricks, not straw. ■

William G. Kaelin Jr *is a professor of medicine at the Dana–Farber Cancer Institute in Boston, Massachusetts.*
e-mail: william_kaelin@dfci.harvard.edu