

COMMENT

TECHNOLOGY From training to therapy — applications of virtual reality surveyed **p.402**

CULTURE Biography of YouTube maps a parallel universe of viral video **p.403**

PLASTICS China's ban on imported waste could boost sustainability **p.405**



OBITUARY Calestous Juma, African science champion, remembered **p.406**

ILLUSTRATION BY DAVID PARKINS



Repeating experiments is not enough

Verifying results requires disparate lines of evidence — a technique called triangulation. **Marcus R. Munafò** and **George Davey Smith** explain.

Several studies across many fields estimate that only around 40% of published findings can be replicated reliably. Various funders and communities are promoting ways for independent teams to routinely replicate the findings of others. These efforts are laudable, but insufficient. If a study is skewed and replications

recapitulate that approach, findings will be consistently incorrect or biased. Consider a commonly used assay in which the production of a fluorescent protein is used to monitor cell activity. If the compounds used to manipulate cell activity are also fluorescent, as has happened¹, reliably repeatable results will not yield robust conclusions.

We have both spent much of our careers advocating ways to increase scientific certainty. One of us (M.R.M.) participated in work by UK funding agencies to develop strategies for reproducible science, and helped to craft a manifesto for reproducibility². But replication alone will get us only so far. In some cases, routine replication might ►

▶ actually make matters worse. Consistent findings could take on the status of confirmed truths, when they actually reflect failings in study design, methods or analytical tools.

We believe that an essential protection against flawed ideas is triangulation³. This is the strategic use of multiple approaches to address one question. Each approach has its own unrelated assumptions, strengths and weaknesses. Results that agree across different methodologies are less likely to be artefacts.

Isn't this how science is meant to operate? Perhaps so, but scientists in today's hyper-competitive environment often lose sight of the need to pursue distinct strands of evidence.

The problem was aptly described in May 2017, when cancer researcher William Kaelin lamented that the goal of the scientific paper had shifted from testing narrow conclusions in multiple ways to making a broadening series of assertions, each based on limited evidence⁴. Consequently, he said, "papers are increasingly like grand mansions of straw, rather than sturdy houses of brick".

The scientific community should address this lack of depth strategically and establish practices that facilitate triangulation. Specifically, we advocate a system to support multidisciplinary teams, each created around a common question (see "Triangulation"). This, we believe, would result in robust insights — mansions of stone.

SPECIOUS ROBUSTNESS

We rarely see projects that aim to prove a point from multiple views. Psychology, epidemiology and the clinical sciences are all geared towards producing statistically significant, definitive studies centred on an endpoint that supports a hypothesis. In parts of the biological sciences, a manuscript's acceptance often depends on a 'capstone' study showing animal efficacy, so pursuing

TRIANGULATION

A checklist.

- The different approaches address the same underlying question.
- The key sources of bias for each approach are explicitly acknowledged.
- For each approach, the expected directions of all key sources of potential bias are made explicit, where feasible.
- Ideally, some of the approaches being compared will have potential biases that are in opposite directions.
- Ideally, results from more than two approaches — which have different and unrelated key sources of potential biases — are compared. [Source: ref. 3](#)



BSIP/UG/GETTY

It took many lines of evidence to show that maternal smoking results in babies with low birth weights.

that single experiment becomes more important than carefully probing an idea from all directions. Moreover, these studies are often presented as having implications for human health without including any tests in humans.

Although many studies in the basic sciences include some element of triangulation, they rarely do enough of it.

In our field of epidemiology, there are countless examples of spurious, persistent findings. Large observational studies frequently produce precise conclusions that are precisely wrong. A correlation between *X* and *Y* might be real in that it genuinely describes an observed association between variables, but is one that does not reflect cause and effect. No amount of replication or statistical adjustment can resolve this, and one of us (G.D.S.) has devoted more than two decades to developing methods that support stronger causal inference in observational epidemiology, drawing on disciplines from the basic sciences to economics.

An illuminating example is the oft-observed J-shaped curves that chart correlation between a condition and health outcome⁵.

For instance, multiple studies show that people who consume low levels of alcohol are healthier than heavy drinkers and teetotalers, leading several researchers to conclude that moderate alcohol consumption promotes health. But other factors, such as unhealthy people being advised to give up drinking, would explain the same shape. Similarly, repeated observations that being slightly overweight is associated with the highest life expectancy might be explained by illness (including processes leading up to the manifestation of a disease, which itself can result in reduced weight); by physicians treating overweight individuals more aggressively; and by other favourable characteristics

of overweight individuals, such as lower smoking rates.

How can one tell that a consistently observed relationship between a behaviour and a health outcome is causal? One example in which triangulation has helped is in establishing that smoking during pregnancy results in babies with lower birth weights⁶. That is different from the simple observation that women who smoke are more likely to have babies who weigh less. Smokers tend to have other characteristics that are also associated with low birth weight, such as low income, less education or more drug use.

Triangulation means explicitly choosing analytical approaches that depend on different assumptions. For example, if a woman's partner smokes during her pregnancy, many of the same confounders apply as in maternal smoking, but the association with lower birth weight is much weaker. Birth weight can also be analysed according to levels of cigarette taxation across US states, which reduces the effects of confounders. And analyses can compare the birth weights of siblings whose mother smoked during one pregnancy but not another.

Mendelian randomization is a technique developed specifically to probe causal relationships. In cohorts grouped according to whether or not people carry a genetic variant associated with greater cigarette consumption in those who smoke, mothers who smoke and carry the variant tended to have babies who weighed less; non-smokers with the same variant did not. Taken together, these studies make it clear that maternal smoking affects birth weight directly⁶.

REPLICATION FIXATION

Replication has received considerable attention; triangulation has not. Maybe one reason replication has captured so much

interest is the often-repeated idea that falsification is at the heart of the scientific enterprise. This idea was popularized by Karl Popper's 1950s maxim that theories can never be proved, only falsified. Yet few experiments, including replication attempts, are explicitly set up to falsify a theory. In fact, we worry that an overemphasis on repeating experiments could provide an unfounded sense of certainty about findings that rely on a single approach.

Moreover, philosophers of science have moved on since Popper. Better descriptions of how scientists actually work include what epistemologist Peter Lipton called in 1991 "inference to the best explanation", or the search for the "loveliest" explanation⁷. This draws on older ideas that championed abductive over deductive reasoning — looking for likely explanations rather than deriving explanations from first principles. This spirit is also captured in the idea of consilience put forward by polymath William Whewell in the mid-nineteenth century and popularized in the 1990s by naturalist E. O. Wilson. This posits that strong theories emerge from the synthesis of multiple lines of evidence, as when Charles Darwin proposed evolution by natural selection.

Unlike consilience, triangulation suggests the deliberate use of different methods. It is the approach to inference that aligns most closely with how many philosophers feel scientists come to understand reality. But most scientists would be hard-pressed to describe it. Researchers typically receive extensive training in experimental methods and design, but little in approaches to causal inference. They are left with no framework to guide scientific pursuit.

CREDIT SHIFT

Triangulation usually requires input from multiple methodologies or disciplines. An elegant historical example is continental drift. In the early 1900s, geophysicist Alfred Wegener noticed that the shape of the west coast of Africa seems to fit that of the east coast of South America. He sought evidence to support the continental-drift theory from a wide range of sources, such as palaeontology (fossils from the same period appeared on both continents) and geology (glacier markings indicated that the continents were once close). In today's environment, scientists would need to contribute to multi-disciplinary projects, with studies providing distinct lines of evidence.

Encouraging such an approach will require fundamental changes to the way in which credit is attributed and to how peer review is conducted. In the current system, few authorship positions count much towards credit — in biomedical science, say, it typically falls just to the corresponding

and other starred authors, as well as to first authors.

To support triangulation, we recommend a shift to a contributorship model, similar to the credits that roll at the end of a film — a long list of individuals with their contributions described fully and specifically⁸. This will require academics to potentially forgo 'senior authorship' positions. It would also make it easier for early-career researchers to specify their unique contribution to a paper when applying for promotion or another position.

Peer review would change too. Instead of a few reviewers looking at the entire manuscript, several would do so, each focusing closely on a particular substudy. In this way, submissions that use multiple, diverse techniques will get appropriate scrutiny, helping to avoid the publication of papers that are like "grand mansions of straw".

Finally, funders, research institutions and journals would need to explicitly support publication of weightier articles. Or perhaps we need to develop formal ways — beyond simple citations — to explicitly link and recognize substudies that triangulate a single question.

A proposal published early last year advocated for a new category of paper that combines hypothesis-generating work with robust, pre-registered confirmatory studies conducted by qualified independent labs⁹. Papers involving triangulation in a way we propose will clearly often involve considerable work coordinating groups of researchers from different disciplines. Reviewers and tenure committees should find ways to value them appropriately. ■

Marcus R. Munafò is programme lead and **George Davey Smith** is director at the Medical Research Council's Integrative Epidemiology Unit at the University of Bristol, UK.
e-mail: marcus.munafò@bristol.ac.uk

1. Baell, J. & Walters, M. A. *Nature* **513**, 481–483 (2014).
2. Munafò, M. R. *et al.* *Nature Hum. Behav.* **1**, 0021 (2017).
3. Lawlor, D. A., Tilling, K. & Davey Smith, G. *Int. J. Epidemiol.* **45**, 1866–1886 (2016).
4. Kaelin, W. G. Jr *Nature* **545**, 387 (2017).
5. Chokshi, D. A., El-Sayed, A. M. & Stine, N. W. *J. Am. Med. Assoc.* **314**, 1339–1340 (2015).
6. Krieger, N. & Davey Smith, G. *Int. J. Epidemiol.* **45**, 1787–1808 (2016).
7. Lipton, P. *Inference to the Best Explanation* 2nd edn (Routledge, 2004).
8. Rennie, D., Yank, V. & Emanuel, L. *J. Am. Med. Assoc.* **278**, 579–585 (1997).
9. Mogil, J. S. & MacLeod, M. R. *Nature* **542**, 409–411 (2017).