

# Novel biomedical research must not be a work of fiction

Mitchell A. Lazar

Institute for Diabetes, Obesity, and Metabolism, and Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA.

AAP members, colleagues, and guests, I am truly honored to serve as President of the Association of American Physicians. It has been a privilege to work with a superb council and to coordinate plans for this meeting with our sister societies, the ASCI and the APSA. I especially thank AAP past President Mary Klotman, Secretary and President-Elect Beth McNally, and Executive Director Lori Ennis. We met many times virtually, and the constraints of the pandemic forced us to pivot on several occasions. I hope that our membership is getting the most out of the virtual format that we reluctantly embraced as a way to continue our important tradition of meeting annually, after the pandemic forced us to cancel the meeting in 2020. It is wonderful to see the range and quality of biomedical science being led by our members, even at this challenging time. It is particularly gratifying to induct 140 regular and 3 international members, 71 newly elected in 2021 and 72 from last year when we were forced to cancel the meeting. Welcoming these new members to our historic and honorific society for physician-scientists is a major highlight of this meeting that will follow my presidential address.

I'm sure that many of you know that the AAP is one of North America's oldest and most prestigious professional organizations of physicians, founded by seven leading academic physicians of the era, including William Osler. With that, I am going to begin my address with a quiz. What year was the AAP founded? If this address were live and in person I'd ask audience members to raise their hands if they know the answer or even shout it out. In this virtual format you can use the chat box. Many of you may not know the

answer off the top of your heads, but are good multitaskers and can try to find the answer as I continue my virtual talk. One way would be to crowdsource via social media. However, that is notoriously unreliable. For example, Gupta and colleagues analyzed Twitter content in the wake of the Boston Marathon bombing and found that 29% of the most viral content on Twitter was rumors and fake news and 51% was mainly opinions and comments (1). Only 20% of the information was true. So you probably can't trust social media for this. Instead, you could search AAP on your browser. Very likely, the first thing you will find is that the AAP was founded in 1929. That's a long time ago, but not the right answer, because that refers to the American Academy of Pediatrics. Wrong AAP! Next on the search list will be the Association of American Publishers (1970), the American Academy of Periodontology (1914, changed to current name in 1919), the Association of American Physiatrists (1967), the Association for Academic Psychiatry (1970), and Advance Auto Parts (1932). All of these are babies relative to our AAP, which was founded in 1885, as we proudly proclaim on our website.

That our AAP was founded in 1885 is a fact. As long as they are recorded accurately, historical dates meet the definition of a fact, which according to the *Oxford English Dictionary*, is a thing that is known or proved to be true. My point is that facts matter and our AAP should stand for that. Yet it is getting harder and harder to be certain of what is an actual fact. In recent years, this has reached crisis proportions as some of our nation's most trusted reporters of facts, such as the *Washington Post* and *New York Times*, are being labeled as

fake news, especially by certain groups. In turn, many of these groups tweet and publish their own fake news. This is not a new problem. John Adams, the third president of the United States, was particularly unhappy with a newspaper called the *Philadelphia Aurora* and proclaimed, "There has been more new errors propagated by the press in the last ten years than in the hundred years before 1798." In the 19th century, the proponents of fake news included William Randolph Hearst and Joseph Pulitzer. Although Pulitzer's name is now eponymous with journalistic excellence, his newspaper, the *New York World*, was shut down for three days after it published forged documents that were claimed to have been written by Abraham Lincoln (2).

Most of the fake news we hear about pertains to political and social issues. But there is plenty of fake information about biomedical science, leading to unfounded fear of vaccines and misplaced blame for the SARS-Covid-2 pandemic, among other things. Disturbingly, there is also fake news from biomedical scientists. In the spring of 2020, the US government was hyping hydroxychloroquine for patients with Covid-19 infections despite the fact that this treatment was not endorsed by Tony Fauci, Director of the National Institute of Allergy and Infectious Diseases. Like all of you, I am so proud of Dr. Fauci, a longtime AAP member, 2000 president of our society, and 2007 Kober medalist, who advocated for science-based decisions throughout the pandemic in the face of hostile politicians and death threats to himself and his family. In May 2020, the prestigious medical journal *The Lancet* published a paper which purported to show that this treatment was associated with increased arrhythmias (3), providing a major argument against the use of hydroxychloroquine. However, within weeks of its publication, the scientific community raised questions about the study,

**Reference information:** *J Clin Invest.* 2021;131(18):e150827. <https://doi.org/10.1172/JCI150827>.

This article is adapted from a presentation at the 2016 AAP/ASCI/APSA Joint Meeting, April 15, 2016, in Chicago, Illinois, USA.

**Copyright:** © 2021, American Society for Clinical Investigation.

and after additional scrutiny, *The Lancet* retracted this paper in June 2020 (3).

It has been suggested that this is an excellent example of science policing itself (4). However, it is naive to believe that retraction of a high-profile publication is like the paper was never published in the first place. Papers do not disappear after retraction. A recent study by Candal-Pedreira and colleagues actually found an increase in postretraction citations compared with citations received preretraction (5). Moreover, the initial finding is often featured on page one of the newspaper, touted on the evening news, and tweeted by the publisher, the investigators, and their institutions. This leads to viral worldwide dissemination of the false finding, whereas the eventual retraction usually receives much less attention. The collateral damage to the reputation of scientific research is enormous. The news media and the public have a large appetite for information relating to health and rely on the peer-review system of the scientific community for accuracy in their reporting. If confidence in reputable peer-reviewed scientific publications is eroded, the public at large will lose their trust in science, above and beyond the antiscience fringe elements and conspiracy theorists. We must not allow this to happen.

The number of published peer-reviewed papers that are retracted is increasing at an exponential rate (6). But wait, as they say, there's more! Thousands of additional papers appear online as preprints without any peer review, creating pressure on our most admired and impactful biomedical journals to compete for these papers at the risk of reducing their standards of scientific rigor. So what can we do? I will now discuss four principles which I suggest that physician-scientists should remember in their roles as teachers, authors, and reviewers. I will refer to these as (a) bigger is better; (b) the three little pigs; (c) Bayes watch; and (d) look out for bullshit. Let me explain.

### Bigger is better

The biomedical community has settled on a statistical standard that, in most cases, accepts a change that has less than a 5% probability of occurring at random as being "significant." However, the danger in describing changes only in terms

of their statistical significance is that the biological relevance of the change is not taken into account. The magnitude of the change needs to be considered as well. As an example, let's use two situations in which changes in blood glucose levels are statistically significant. A tripling of blood glucose levels would be of major concern and impact over an extended period. On the other hand, a 10% change would not be immediately worrisome and of lesser long-term consequence. Put another way, it would be of less biological significance. Yet the title and abstract of many papers do not specify the magnitude of statistically significant changes, and readers scanning the papers cannot easily discern the large from the small effects. Of course, it would be ideal if everyone read all papers carefully, but this is increasingly difficult with the sheer volume of journals, let alone preprints. This is further confounded as bioinformatic approaches, including natural language processing, are used to interrogate a large body of literature, increasing the likelihood that effects of vastly different magnitude would be lumped together as causes of increased blood glucose. Furthermore, press offices and the popular press are likely to report on the phenomenon of increased blood glucose without considering the magnitude of the effect, leading to undue alarm in the case of small, albeit statistically significant, changes. It is also my impression that while these concerns about statistical versus practical significance clearly pertain to clinical studies, this issue is becoming much more frequent in preclinical basic science papers. Peer reviewers should make an assessment of biological significance in their area of expertise and consider that strongly when they formulate their recommendation. And editors should consider this in their decision.

### The three little pigs

In 2017, AAP member and 2018 Nobel laureate Bill Kaelin exhorted scientists to "publish houses of brick, not mansions of straw" (7). This is clearly an allusion to the fable of the three little pigs. You know the story — three little pigs are trying to escape the big bad wolf. One quickly builds a house of straw, but the wolf blows it down. The second builds a house of sticks, and it's equally inadequate. But the third pig

took the time to build a house of bricks, which resisted the wolf and stood the test of time. The third pig is a metaphor for what scientists should strive to be, publishing work that is reproducible and stands the test of time. Kaelin points out that scientific work should be judged by "whether its conclusions are likely to be correct, not whether it would be important if it were true." Unfortunately, current trends seem to be in conflict with this prescription. This is largely because the most prestigious journals make publishing decisions largely based on whether the findings are novel. The more novel the work is, the more titillating it is likely to be. Ironically, when used as a noun, the word novel refers to a work of fiction, which is just the opposite of what a work of science is supposed to represent. Brembs, in making a similar point, has called for "reliable novelty" and cautions that in science "new should not trump true" (8). To the extent that journal prestige matters to scientists, it is reasonable for higher ranked journals to consider what the general interest of a new scientific story may be. However, if the results are not convincing, the work should not be published anywhere, regardless of the level of novelty.

### Bayes watch

So novelty is a good thing provided that the findings are strong, robust, reliable, and reproducible. Is more novelty always better than less novelty? Paraphrasing the eponymous theorem of Reverend Thomas Bayes, the probability of a novel conclusion being correct is related to the weight of new evidence as well as the prior probability of its being true (9). This concept has been popularized in recent years by Nate Silver (10). But I don't think that journals and reviewers take this into account as much as they should. We've all heard the expression, "If it's too good to be true it probably is." For sure, unexpected truths do emerge, but there needs to be weighty evidence before highly surprising conclusions become the prevailing scientific view. On the contrary, solid findings that advance the field may be considered *Too True to Be Good*, which is the title of a George Bernard Shaw comedy related to this point (11). Maybe he wrote it after one of his papers was rejected for lack of novelty! It does seem increasingly common that the most "selective" journals

are willing to overlook scientific flaws if the conclusions of a paper will generate a buzz that resonates with the lay press and the general public.

Yet just the opposite should be the case. Since current scientific thinking is based on a long record of mainly reproducible observations, the more incredible a new result is, the higher the standard of proof to which it should be held. I have heard some colleagues suggest that it's fine to allow really surprising findings to be published without the appropriate level of proof. They say, "Let the scientific community judge for itself and we'll see whether the conclusions will hold up." I strongly disagree with this notion. This causes unnecessary distraction for the scientific community at large and major damage when the news media and public that trust science to be grounded in facts discover that they have been misled. Furthermore, once "out there," it is difficult to put the genie back in the bottle and ignore the fact that the surprising result has been discredited. This is especially true with the permanent scientific record being easily queried by PubMed and other search engines. Indeed, on numerous occasions, I've had trainees excitedly tell me about a paper they came across that makes an incredible point without simultaneously finding the trail of publications refuting that work. Indeed, the word "incredible" is derived from the Latin word meaning not worthy of belief. So you should maintain a good deal of skepticism and not believe a surprising result without sizable and robust supporting evidence. Always be wary of colleagues who describe their findings as "incredible" or "unbelievable." As a rule, those are not such good things in science. The goal of science is to make discoveries that are reproducible and believable.

## Look out for bullshit

Science is the search for truth about the natural world. It is hubris to think that what we know is truth, but it is critical that as science builds its house of bricks, it does so based on reproducible observations. The interpretation of these observations leads to hypotheses that are tested by experimental scientists. The results of these experiments lead to revised hypotheses, and the more surprising the result, the

more the hypothesis needs to be revised. While truth is the goal, falsehood is the poison to this operation. When it occurs, it slows the progress of science and erodes confidence in the scientific enterprise.

Science does police itself, in the form of paper retractions, and these mainly occur following overt scientific misconduct, where the scientist intentionally dupes colleagues and reviewers to make false claims (12). This form of scientific fraud is hardly new and indeed has been highlighted by the late Phil Majerus, an AAP member and 1991 Kober lecturer, in his 1982 ASCI presidential address (13) as well as the late Robert Petersdorf (14), an AAP member and 1996 Kober medalist. Scientific misconduct is likely motivated by ambition as well as, in some cases, a conviction by the scientist that they know what the result should be, even if it cannot be demonstrated by controlled experiments. In any case, scientific misconduct is a form of lying. The perpetrators know that they are reporting false data, and the explanation for its nonreproducibility is clear.

Despite the increasing number of retracted papers, many more published findings are nonreproducible than can be explained by misconduct leading to official retraction in the literature. For example, researchers at pharmaceutical companies could replicate only 11% to 25% of published findings that they pursued in their own laboratories (15, 16). The cause of this depressing failure to reproduce published results is probably multifactorial. In many cases, it is likely that the nonreproducibility results from subtle but important variations in protocols and approaches used by different investigators at multiple sites. Another contributor to nonreproducibility may be unconscious bias on the part of the scientists towards proving their hypothesis, which can result in selective use and interpretation of data. This is more insidious than volitional misconduct in that the scientists do not believe that they are reporting false data and the explanation for its nonreproducibility is often murky and in many cases never definitively established.

In his 1988 ASCI presidential address (17), Bob Lefkowitz, the 2001 AAP president, 2011 Kober medalist, and 2012 Nobel laureate, pointed out that this cause of nonreproducibility is a form of bullshit,

paying homage to Harry Frankfurt's illuminating and aptly titled book *On Bullshit* (18). This concept was recently furthered by Bergstrom and West in their book *Calling Bullshit* (19). Frankfurt suggested that the difference between a lie and bullshit is that while the liar, or perpetrator of scientific misconduct, knows that they are lying, the bullshitter does not. Because lies can be recognized as such, he posited that "bullshit is a greater enemy of the truth than lies are." Similarly, results that are nonreproducible for unknown reasons tend to rear their ugly heads and linger in a more sinister manner than data that have been proven to be falsified and retracted.

Clearly these are not new issues, but they are on the rise and a more palpable threat to the biomedical science enterprise because of the proliferation of journals and preprint servers as well as both conventional and social media hyping the most surprising new findings. We cannot rely on every paper being read carefully by everyone potentially interested in the findings, so the onus is increasingly on the front end, that is the release of new putative scientific facts into the scientific world. At a time when facts are being challenged, it is more important than ever for science and scientists to be seen as an unbiased and reliable source of valid information about the natural world.

One step toward a solution, as has been suggested by others, including current AAP council member John Ioannidis (20), is to align career success with reproducibility. Current metrics for success in obtaining grants, academic promotions, and scientific honors lean towards rewarding novelty and perceived impact. These are important, but the system is becoming unbalanced and we should be equally rewarding reproducibility. This can be a metric that is empirically assessed by documenting subsequent papers that directly reproduce the original findings. Even more credit should be given when later work adds one or more bricks to the solid house of scientific progress. A shining example is the basic science done by AAP member Drew Weissman and his colleagues, whose discovery of RNA modifications that prevent their immunogenicity (21) paved the way to RNA vaccines with the potential to lead us out of the SARS-Covid-2 pandemic. We should all be grateful and thankful that these key discoveries,

made more than 15 years ago, have stood the test of time and serve as the foundation for a scientific solution to the ravages of the current pandemic.

Scientists must strive to maintain their authority and influence progress in the world. As noted by AAP's 2022 Kober medalist Linda Fried in her 2017 AAP presidential address (22), scientific knowledge itself is a public good. That must not be based on science fiction, but on scientific facts. I implore members of the AAP as well as ASCI and the future physician-scientists of APSA to stand for the highest standards of rigor, robustness, and reproducibility in science. Our integrity is on the line along with the future of biomedical science and its promise to lead to better health for all.

## Acknowledgments

I thank Drs. Althier Lazar, Myles Brown, and Morrie Birnbaum, and Mr. Aaron Lazar for their helpful comments.

Address correspondence to: Mitchell A. Lazar, Institute for Diabetes, Obesity, and Metabolism, and Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA. Email: [lazar@pennmedicine.upenn.edu](mailto:lazar@pennmedicine.upenn.edu)

- Gupta A, Lamba H, Kumaraguru P. \$1.00 per RT #BostonMarathon #PrayForBoston: Analyzing fake content on Twitter. Paper presented at 2013 APWG eCrime Researchers Summit; 2013; San Francisco, California, USA. <https://doi.org/10.1109/eCRS.2013.6805772>. Accessed July 27, 2021.
- Guilford G. Fake news isn't a recent problem in the US—it almost destroyed Abraham Lincoln. *Quartz*. November 28, 2016. Accessed July 27, 2021. <https://qz.com/842816/fake-news-almost-destroyed-abraham-lincoln>.
- Mehra MR, et al. Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis [published online May 22, 2020]. *Lancet*. [https://doi.org/10.1016/S0140-6736\(20\)31180-6](https://doi.org/10.1016/S0140-6736(20)31180-6). Retracted in: *Lancet*. 2020;395(10240):1820.
- O'Brian M. Retractions of coronavirus research show process of science working as it should. *UBNow*. July 7, 2020. Accessed July 27, 2021. <http://www.buffalo.edu/ubnow/stories/2020/07/obrian-conversation-covid-retractions.html>.
- Candal-Pedreira C, et al. Does retraction after misconduct have an impact on citations? A pre-post study. *BMJ Glob Health*. 2020;5(11):e003719.
- Bhatt B. On retractions in biomedical literature. *Towards Data Science*. May 24, 2019. Accessed July 27, 2021. <https://towardsdatascience.com/on-retractions-in-biomedical-literature-1565e773559e>
- Kaelin WG. Publish houses of brick, not mansions of straw. *Nature*. 2017;545(7655):387.
- Brembs B. Reliable novelty: new should not trump true. *PLoS Biol*. 2019;17(2):e3000117.
- Bayes T. An essay towards solving a problem in the doctrine of chances. *Philos Trans*. 1763;53:370–418.
- Silver N. *The Signal and the Noise: Why so Many Predictions Fail-but Some Don't*. Penguin Press; 2012.
- Shaw GB. *Too True to Be Good*. Boston, 1932.
- Campos-Varela I, Ruano-Raviña A. Misconduct as the main cause for retraction. A descriptive study of retracted publications and their authors. *Gac Sanit*. 2019;33(4):356–360.
- Majerus PW. Fraud in medical research. Presidential address delivered before the 74th annual meeting of the American Society for Clinical Investigation, Washington, DC, 8 May 1982. *J Clin Invest*. 1982;70(1):213–217.
- Petersdorf RG. The pathogenesis of fraud in medical science. *Ann Intern Med*. 1986;104(2):252–254.
- Prinz F, et al. Believe it or not: how much can we rely on published data on potential drug targets?. *Nat Rev Drug Discov*. 2011;10(9):712.
- Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature*. 2012;483(7391):531–533.
- Lefkowitz RJ. The spirit of science. Presidential address to the American Society for Clinical Investigation, Washington, DC, 30 April 1988. *J Clin Invest*. 1988;82(2):375–378.
- Frankfurt H. *On Bullshit*. Princeton University Press; 2005.
- Bergstrom CT, West JD. *Calling Bullshit: The Art of Skepticism in a Data-Driven World*. Random House; 2020.
- Ioannidis JPA. How to make more published research true. *PLoS Med*. 2014;11(10):e1001747.
- Karikó K, et al. Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity*. 2005;23(2):165–175.
- Fried LP. The public good of science for health. *J Clin Invest*. 2017;127(10):3561–3567.