I

n my previous editorial, I pointed out that the reproducibility of scientific papers would improve dramatically if the focus of the academic enterprise shifted from the short term to the long term: such a shift would discourage publication of sloppy or fraudulent work, because neither sloppiness nor fraud is reproducible and, therefore, neither can stand the test of time. In this editorial, I wish to elaborate on the overarching importance of the test of time in science and medicine.

For those of us who are involved in scientific research, time is the ultimate arbiter, the supreme judge. Its power is absolute and its verdict final—it cannot be appealed or overruled. Time has unfettered jurisdiction over both the validity and the value of research—that is, it determines whether research findings are true and whether they are important, respectively. With regard to validity, only time can tell whether a study can be reproduced and confirmed by others to the point
that its results are widely accepted as correct. Similarly, with regard to value, only time can tell whether a discovery has a significant impact on the field. (Parenthetically, this concept of time as the supreme judge applies not only to research but to most human activities.)

Unfortunately, the academic enterprise does not typically use the test of time to gauge the merit of published manuscripts; instead, it often relies on such inappropriate metrics as the number of times a paper is cited in the first 2 to 3 years after publication or the impact factor and prestige of the journal in which the paper appears. These metrics are profoundly misleading. Neither the validity nor the value of a paper can and should be measured by how big a splash it makes in the short term (eg, the number of times it is cited in the first 2–3 years after publication). Many papers that appear to be “hot” are highly cited soon after publication, only to be forgotten later on because they become irrelevant or are found to be flawed. Nor can the validity and value of a paper be properly assessed by the prestige or impact factor of the journal that publishes it. When a prestigious journal accepts a manuscript, the decision to publish that paper reflects only the opinion of a very small group of individuals (frequently less than 5), that is, the reviewers and editors. Does the opinion of 3 to 4 persons necessarily guarantee that a manuscript will be valid and important? As for the impact factor of the journal, it is a common (and deleterious) mistake to assume that a high numerical value of this metric implies a high scientific value of an individual paper published in that journal. Apart from its innumerable limitations (which have been lamented ad nauseam in recent years and thus will not be regurgitated here), the journal’s impact factor reflects only the average citation rate of all its articles; it cannot accurately gauge the merit of an individual paper. In any journal (including the most prestigious ones), the citations of individual papers exhibit a huge variability, such that a minority of the manuscripts accounts for the majority of the citations that drive the impact factor; consequently, it is not uncommon for papers to be accepted in a high-impact journal and still receive few citations.

The validity and value of a research study are best assessed by determining how it stands the test of time: Have other investigators been able to reproduce it over the subsequent years? Has it advanced the field? How long did its impact on the field last? Was it just a fleeting meteor? Answers to these fundamental questions usually require more than 2 or 3 years. Yet, it is customary nowadays for headlines to appear in both scientific and nonscientific journals and other news media immediately after a paper appears (long before the test of time has been applied), extolling its merits or denouncing its flaws, particularly if it deals with a “sexy” or “hot” topic. Grants, promotions, prizes, recognitions, and fame are awarded or denied on the basis of the initial splash (or lack thereof), not of the final verdict of time. And when the test of time fails to corroborate the initial assessment, it is often too late for the early decisions to be changed, and few people notice it anyway; as explained in my previous editorial, the attention span of our society is too short (and getting shorter) and the pace of discovery is too feverish (and getting more feverish) for most people to ponder why a study has failed to deliver on its promises. By the time the error in the initial assessment becomes obvious, grants, promotions, prizes, recognitions, and fame have been awarded, and most people have moved on to other things; few take the time to stop and look back at what was said or done several years before. In short, in the contemporary academic world, many important decisions are made without waiting for the test of time.

As I explained previously, this combination of short attention span and short memory contributes powerfully to the irreproducibility of scientific studies because it removes rewards for publishing reproducible work and deterrents for publishing nonrepeatable experiments. The remedy to this situation would be to change the academic reward system in a manner that emphasizes reproducibility and long-term impact; however, realistically, it is unlikely that such a change will occur.

The history of medicine is replete with articles that made a huge splash soon after publication but then were forgotten because they could not be reproduced or because they lost relevance. Conversely, there are many examples of papers that initially generated little interest but later were recognized as scientific and medical milestones. In the 1980s, the idea that antioxidants, anti-inflammatory agents, and anti-neutrophil therapies limit myocardial infarct size generated tremendous excitement and academic and corporate interest (as well as innumerable citations), only to fade away under the ashes of history when the initial work could not be reproduced. I doubt many readers even remember the high-profile publications suggesting that postischemic myocardial dysfunction (“stunning”) is caused by proteolytic degradation of troponin I; again, that idea was not confirmed by others. The use of gene therapy to combat bradyarrhythmias (thereby making pacemakers obsolete) was heralded as a breakthrough and produced copious early citations and news media headlines, only to be abandoned later on. Conversely, Alexander Fleming’s report of the discovery of penicillin in 1929 went almost unnoticed for an entire decade. When the first report of thrombolytic therapy in patients with acute myocardial infarction was published in 1959, few realized that this approach would be a major breakthrough; its impact was not appreciated until the mid-1980s. In all of these cases, the passage of time made it clear that the initial perception was erroneous. One could fill an entire book with similar examples.

I wish to point out that I am not writing anything new here. The concept of time as the supreme judge may be foreign to our culture but was well recognized in antiquity. “Time, as it grows old, teaches all things”, wrote the Greek tragedian Aeschylus (525–456 BC) in Prometheus Bound. (This is probably the earliest known ancestor of the modern idiom “Time will tell”.) Five centuries later, the Roman stoic philosopher Seneca (c. 4 BC–AD 65) admonished, in his book De Ira (On Anger), that “We should always allow some time to elapse; time discloses the truth” (Dandum semper est tempus: veritatem dies aperit). Wouldn’t science (and the world in general) be much better off if these sage words were heeded?
Time is the best judge of both scientific validity and scientific importance. Studies that are scientifically valid will be reproduced for years after publication. Papers that are important will have a lasting impact on their field and will be remembered and cited for many years (not just 2–3 years) after they appear; the rest will be quickly forgotten. Time is infallible—it always sorts out hype from science and truth from error. But, alas, time is also a slow judge, taking many years to deliberate. And in our increasingly fast-paced and ADHD-like* culture, who has the patience to wait for time to render its verdict? In the age of Twitter, 24-hour news cycles, immediate feedback, and short attention span, Seneca’s wise precept is all but forgotten. There is no time for the test of time.

*ADHD: Attention deficit hyperactivity disorder

Disclosures

None.

References

Dandum semper est tempus: The Crucial Importance of (and Increasing Disregard for) the Test of Time

Roberto Bolli

Circ Res. 2015;117:755-757
doi: 10.1161/CIRCRESAHA.115.307613

Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2015 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/117/9/755

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org//subscriptions/