

# Issues with data and analyses: Errors, underlying themes, and potential solutions

Andrew W. Brown<sup>a,1</sup>, Kathryn A. Kaiser<sup>a,2</sup>, and David B. Allison<sup>a,3,4</sup>

<sup>a</sup>Office of Energetics and Nutrition Obesity Research Center, University of Alabama at Birmingham, Birmingham, AL 35294

Edited by Victoria Stodden, University of Illinois at Urbana–Champaign, Champaign, IL, and accepted by Editorial Board Member Susan T. Fiske November 27, 2017 (received for review July 5, 2017)

Some aspects of science, taken at the broadest level, are universal in empirical research. These include collecting, analyzing, and reporting data. In each of these aspects, errors can and do occur. In this work, we first discuss the importance of focusing on statistical and data errors to continually improve the practice of science. We then describe underlying themes of the types of errors and postulate contributing factors. To do so, we describe a case series of relatively severe data and statistical errors coupled with surveys of some types of errors to better characterize the magnitude, frequency, and trends. Having examined these errors, we then discuss the consequences of specific errors or classes of errors. Finally, given the extracted themes, we discuss methodological, cultural, and system-level approaches to reducing the frequency of commonly observed errors. These approaches will plausibly contribute to the self-critical, self-correcting, ever-evolving practice of science, and ultimately to furthering knowledge.

rigor | statistical errors | quality control | reproducibility | data analysis

In common life, to retract an error even in the beginning, is no easy task... but in a public station, to have been in an error, and to have persisted in it, when it is detected, ruins both reputation and fortune. To this we may add, that disappointment and opposition inflame the minds of men, and attach them, still more, to their mistakes.

Alexander Hamilton (1)

## Why Focusing on Errors Is Important

Identifying and correcting errors is essential to science, giving rise to the maxim that science is self-correcting. The corollary is that if we do not identify and correct errors, science cannot claim to be self-correcting, a concept that has been a source of critical discussion (2). Errors are arguably required for scientific advancement: Staying within the boundaries of established thinking and methods limits the advancement of knowledge.

The history of science is rich with errors (3). Before Watson and Crick, Linus Pauling published his hypothesis that the structure of DNA was a triple helix (4). Lord Kelvin misestimated the age of the earth by more than an order of magnitude (5). In the early days of the discipline of genetics, Francis Galton introduced an erroneous mathematical expression for the contributions of different ancestors to an individual's inherited traits (6). Although it makes them no less erroneous, these errors represent significant insights from some of the most brilliant minds in history working at the cutting edge of the border between ignorance and knowledge—proposing, testing, and refining theories (7). These are not the kinds of errors we are concerned with herein.

We will focus on actions that, in principle, well-trained scientists working within their discipline and aware of established knowledge of their time should have or could have known were erroneous or lacked rigor. Whereas the previously mentioned errors could only have been identified in retrospect from advances in science, our focus is on errors that often could have been prospectively avoided. Demonstrations of human fallibility—rather than human brilliance—have been and will always be present in science. For example, nearly 100 y ago, Horace Secrist, a professor and author of a text on statistical methods (8), drew substantive conclusions about business performance based on patterns

that a statistical expert of the day should have realized represented regression to the mean (9). Over 80 y ago, the great statistician "Student" published a critique of a failed experiment in which the time, effort, and expense of studying the effects of milk on growth in 20,000 children did not result in solid answers because of sloppy study design and execution (10). Such issues are hardly new to science. Similar errors continue today, are sometimes severe enough to call entire studies into question (11), and may occur with nontrivial frequency (12–14).

**What Do We Mean by Errors?** By errors, we mean actions or conclusions that are demonstrably and unequivocally incorrect from a logical or epistemological point of view (e.g., logical fallacies, mathematical mistakes, statements not supported by the data, incorrect statistical procedures, or analyzing the wrong dataset). We are not referring to matters of opinion (e.g., whether one measure of anxiety might have been preferable to another) or ethics that do not directly relate to the epistemic value of a study (e.g., whether authors had a legitimate right to access data reported in a study). Finally, by labeling something an error, we declare only its lack of objective correctness, and make no implication about the intentions of those making the error. In this way, our definition of invalidating errors may include fabrication and falsification (two types of misconduct). Because they are defined by intentionality and egregiousness, we will not specifically address them herein. Furthermore, we fully recognize that categorizing errors requires a degree of subjectivity and is something that others have struggled with (15, 16).

**Types of Errors We Will Consider.** The types of errors we consider have three characteristics. First, they loosely relate to the design of studies, statistical analysis, and reporting of designs, analytic choices, and results. Second, we focus on "invalidating errors," which "involve factual mistakes or veer substantially from clearly accepted procedures in ways that, if corrected, might alter a paper's conclusions" (11). Third, we focus on errors where there is reasonable expectation that the scientist should have or could have known better. Thus, we are not considering the missteps in thinking or procedures necessary for progress in new ideas

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Reproducibility of Research: Issues and Proposed Remedies," held March 8–10, 2017, at the National Academy of Sciences in Washington, DC. The complete program and video recordings of most presentations are available on the NAS website at [www.nasonline.org/Reproducibility](http://www.nasonline.org/Reproducibility).

Author contributions: A.W.B., K.A.K., and D.B.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. V.S. is a guest editor invited by the Editorial Board.

Published under the [PNAS license](http://www.pnas.org).

<sup>1</sup>Present address: Department of Applied Health Science, School of Public Health–Bloomington, Indiana University, Bloomington, IN 47405.

<sup>2</sup>Present address: Department of Health Behavior, School of Public Health, University of Alabama at Birmingham, Birmingham, AL 35294.

<sup>3</sup>Present address: Department of Epidemiology and Biostatistics, School of Public Health–Bloomington, Indiana University, Bloomington, IN 47405.

<sup>4</sup>To whom correspondence should be addressed. Email: [allison@iu.edu](mailto:allison@iu.edu).

Published online March 12, 2018.

and theories (17). We believe the errors of Secrist and identified by Student could have been prevented by established and contemporaneous knowledge, whereas the errors of Pauling, Kelvin, and Galton preceded the knowledge required to avoid the errors.

We find it important to isolate scientific errors from violations of scientific norms. Such violations are not necessarily invalidating errors, although they may affect trust in or functioning of the scientific enterprise. Some “detrimental research practices” (15), not disclosing conflicts of interest, plagiarism (which falls under “misconduct”), and failing to obtain ethical approval do not affect the truth or veracity of the methods or data. Rather, they affect prestige (authorship), public perception (disclosures), trust among scientists (plagiarism), and public trust in science (ethical approval). Violations of these norms have the potential to bias conclusions across a field, and thus are important in their own right, but we find it important to separate discussions of social misbehavior from errors that directly affect the methods, data, and conclusions both in primary and secondary analyses.

### **Underlying Themes of Errors and Their Contributing Factors**

The provisional themes we present build on our prior publications in this area and a nonsystematic evaluation of the literature (11, 18). The identification of these themes represents our opinions from our own vantage point; there is no mathematical proof that these are the best themes, but they have proven useful to us.

**Themes of Types of Errors.** A variety of themes or taxa of errors have been proposed. We have noted errors related to measurement, study design, replication, statistical analysis, analytical choices, citation bias, publication bias, interpretation, and the misuse or neglect of simple mathematics (18). We also briefly described a theme of invalidating errors (11), which we expand on below. Others have categorized errors by stage of the research process. Bouter et al. (16), for instance, classified “research misbehaviors” in four domains: reporting, collaboration, data collection, and study design. Many items within various themes or taxa overlap: One person’s research misbehavior may be classified as another’s statistical error.

**Errors producing “bad data.”** We define bad data as those acquired through erroneous or sufficiently low-quality collection methods, study designs, or sampling techniques, such that their use to address a particular scientific question is scientifically unjustifiable. In one example, self-reported energy intake has been used to estimate actual energy intake. This method involves asking people to recall their dietary intake in one or more ways, and then deriving an estimate of metabolizable energy intake from these reports. The method, compared with objective measurements of actual energy intake, turns out to be invalid (19), not just “limited” or “imperfect.” The measurement errors are sufficiently large and nonrandom that they have led to consistent and statistically significant correlations in the opposite direction from the true correlation for some relationships. Moreover, the relations between the errors and other factors are sufficiently numerous and complex that they defy simple corrections. Concerns about this method were raised decades ago (20), and yet its use is continued. We have called for its use to be discontinued (19).

Other common examples of bad data include confounding batch effects with study variables of interest (21) and cell-line misidentification or contamination (22). For confounding or contamination, the data are bad from failed design and are often unrecoverable.

Bad data represent one of the most egregious of themes of errors because there is typically no correct way to analyze bad data, and often no scientifically justifiable conclusions can be reached about the original questions of interest. It also can be one of the more difficult errors to classify, because it may depend on information like the context in which the data are being used and whether they are fit for a particular purpose.

**Errors of data management.** Errors of data management tend to be more idiosyncratic than systematic. Errors we have seen (and

sometimes made) are the result not of repeating others’ errors, but of constructing bespoke methods of handling, storing, or otherwise managing data. In one case, a group accidentally used reverse-coded variables, making their conclusions the opposite of what the data supported (23). In another case, authors received an incomplete dataset because entire categories of data were missed; when corrected, the qualitative conclusions did not change, but the quantitative conclusions changed by a factor of  $>7$  (24). Such idiosyncratic data management errors can occur in any project, and, like statistical analysis errors, might be corrected by reanalysis of the data. In some cases, idiosyncratic errors may be able to be prevented by adhering to checklists (as proposed in ref. 25).

Errors in long-term data storage and sharing can render findings nonconfirmable because data are not available to be reanalyzed. Many metaanalysts, including us, have attempted to obtain additional information about a study, but have been unable to because the authors gave no response, could not find data, or were unsure how they calculated their original results. We asked authors once to share data from a publication with implausible baseline imbalances and other potential statistical anomalies; they were unable to produce the data, and the journal retracted the paper (26). We have struggled on occasion to find our own raw data from older studies and welcome advances in data management, data repositories, and data transparency.

**Errors of statistical analysis.** Errors of statistical analysis involve methods that do not reliably lend support to the conclusions. These can occur if the underlying assumptions of the analyses are not met, the wrong values are used in calculations, statistical code is misspecified, incorrect statistical methods are chosen, or a statistical test result is misinterpreted, regardless of the quality of the underlying data. We have written about three such errors (11). First, misanalysis of cluster-randomized trials (27) may inappropriately and implicitly assume independence of observations. Worse still, when there is only one cluster per group, clusters are completely confounded with treatment, resulting in zero degrees of freedom to test for group effects. This, too, has resulted in retraction (28). Second, effect sizes for metaanalyses may inappropriately handle multiple treatment groups (e.g., assuming independence despite sharing a control group) or fail to use the correct variance component in calculations. In turn, the metaanalytic estimates from these effect-size calculations may be incorrect, and have sometimes required correction (29). Third, it is inappropriate to compare the nominal significance of two independent statistical tests as a means of drawing a conclusion about differential effects (30). This “differences in nominal significance” [DINS (31)] error is sometimes committed in studies with more than one group, in which final measurements are compared with baseline separately for each group; if one is significant and one is not, an author may erroneously conclude that the two groups are different. We have noted, and attempted to correct, DINS errors (e.g., refs. 32 and 33).

The effects of these errors on conclusions can be severe. However, when treatment effects are misanalyzed, we often cannot immediately say the conclusions are false, but rather, we can say that the analyses are unreliable for statistical inference and conclusions. Authors and editors must be contacted to resolve the issue (e.g., ref. 28). In other cases, conclusions may be obviously wrong. If a DINS error was committed in a study and the point estimates of each group are identical, it is clear that the appropriate between-group test would not be statistically significant. Fortunately, the nature of statistical errors is such that, if authors and journals are willing, and the underlying data are not bad, then errors of analysis can be corrected. Unfortunately, correction of errors often requires an arduous process that highlights limitations of the self-correcting nature of science (11).

**Errors in logic.** Albeit not an error of data or analyses, research filtered through the lens of poor logic can distort findings, resulting in conclusions that do not follow from the data, analysis, or fundamental premises.

Classical logical fallacies appear in literature. “*Cum hoc, ergo propter hoc*” (with this, therefore because of this; common from cross-sectional data) and “*post hoc, ergo propter hoc*” (after this, therefore because of this; common with longitudinal data) are two examples of errors in logic that assume observed associations are sufficient evidence for causation. Assuming causation from observational evidence is common (34, 35). In some cases, papers are careful to appropriately describe associations rather than statements of causation—like, “Dietary factors were estimated to be associated with a substantial proportion of deaths from heart disease, stroke, and type 2 diabetes,” (36). However, subsequent media hype or communications from the authors may succumb to these fallacies [e.g., “Our nation’s nutrition crisis: nearly 1,000 cardiovascular & diabetes deaths each day (!) due to poor diet” (37)].

Arguments based on authority, reputation, and *ad hominem* reasoning are also common. These arguments might focus on characteristics of the authors, the caliber of a journal, or the prestige of authors’ institutions to bolster the strength of or refute a study. In one example of *ad hominem* reasoning, an author was disparagingly identified only as a “chemical industry consultant with a competing interest” to passively dismiss arguments, while they also reasoned from authority and reputation by negatively contrasting the arguments of the other authors with “independent scientific entities” (38). Authority and reputation may serve as useful heuristics for making daily decisions; using them to support or refute the quality of the evidence in published papers is tangential to science.

Other logical fallacies are evident in the literature, but one that ties the others together is arguing that conclusions drawn from erroneous research are false—the “fallacy fallacy.” Identification of an error in a paper or reasoning cannot be used to say the conclusions are wrong; rather, we can only say the conclusions are unreliable until further analysis.

**Errors of communication.** Errors of communication do not necessarily affect data and methods, but are flaws in the logic used to connect the results to conclusions. In the simplest case, communication may be overzealous—extrapolating beyond what a study can tell us. Authors discussing benefits and limitations of animal testing in predicting human cancer risk noted, “The problem with animal testing is that animal test results are often improperly extrapolated to humans” (39). They recount studies in which dosages provided to animals were degrees of magnitude more than expected for humans. One study dosed animals with daminozide (a plant growth regulator) that would require humans to consume “28,000 pounds of apples daily for 10 years” to obtain—extrapolation errors both in species and dosage.

Other forms of erroneous extrapolation are evident. A study on responses to small, 1-d exposures may be inappropriate to extrapolate to chronic exposures (as demonstrated by ref. 40). Effects on outcomes like energy intake are linearly extrapolated to weight change (41), despite energy balance being a dynamic, nonlinear system (42). Associations across epidemiological studies are extrapolated to public health action (43). In cases of extrapolation, the study may be perfectly executed within its constraints, but just not support stated conclusions. These errors are identifiable through more thorough review of the data and the methods, which can admittedly be burdensome and challenging.

Publication, reporting, and citation biases are other forms of errors of communication that may lead to a form of bad data when considering a collection of scientific reports as data themselves. If scientists fail to publish some results for whatever reason, then the totality of data used in summarizing our scientific knowledge (e.g., metaanalysis) is incomplete.

P-hacking and related practices (44) [e.g., researcher degrees of freedom (45) and p-fiddling (46), among other names] represent a form of selective reporting and may also be considered errors of statistical analysis. In most cases, there is not a single, universally agreed-upon method to analyze a particular dataset, so trying multiple analyses may be considered scientifically prudent to test the robustness of findings. However, p-hacking uses

the *P* value from an analysis as the rule by which a particular analysis is chosen, rather than the appropriateness of the analysis itself, often without fully disclosing how that *P* value was chosen. Conclusions are questionable because “undisclosed flexibility in data collection and analysis allows presenting anything as significant” (45). A striking example is the publication of apparently highly statistically significant results in the “Bible Code” that were later debunked as a variant of p-hacking (9).

**Themes of Contributing Factors.** Scientists are humans; we make mistakes and ill-informed guesses, sometimes with the best of intentions. Scientific processes are intended to constrain these human foibles, but humans still report findings derived from erroneous methods, data, or interpretations. Sometimes, errors only become apparent through time and improvements in technology. Understanding and identifying what contributes to errors that cloud scientific processes may be key to improving the robustness of scientific findings.

**Ignorance.** An obvious contributing theme is simple ignorance, whether of an individual, the research team, a peer reviewer, editors, or others. Although we and others have cataloged and publicized the existence of errors, this only establishes that the errors are known to us and the scientific community broadly, but not necessarily each individual. In other words, these errors are “known unknowns”: errors known to science, but not a particular scientist. In our communications with research teams who we think have made statistical errors, the response is frequently one of surprise because they were unaware of the errors or the consequences of analytical or study design choices.

Bad examples in the literature may, themselves, perpetuate ignorance. Exposure to any errors we presented above without appropriate and repeated correction may result in a scientist presuming that the paper, methods, and logic were correct; after all, it went through peer review and remains uncorrected. Effective postpublication peer review may be particularly useful to mitigate ignorance by using such errors to serve as instructive examples of what not to do. It is also important to recognize that some errors have yet to be made, identified, or corrected, and thus the errors are presently unknown unknowns. Time may be the most critical component to reveal these yet-unidentified errors.

**Poor study inception.** A poorly conceived study presents foundational problems for the remainder of the process of conducting, analyzing, and reporting research. Study inception can bifurcate into hypothesis generation and hypothesis testing, although the two branches certainly contribute to each other. If a study is started with discovery in mind, but with no clear scientific plan, choices made along the way follow the data. This is not a problem per se, as long as the final results are communicated as a wandering exploration. Conversely, a poorly planned test of a hypothesis may allow for researchers to choose variations in methods or analyses not based on a rigorous question or theory, but on interests and expectations. A regularly used example is the experience of C. Glenn Begley, who, after failing to replicate results of another research group, was told by one of the original authors that an experiment had been tried multiple times, but they only published the results that “made the best story” (47). Generating hypotheses after the results are already known [so-called HARKing (48) or post hoc storytelling] provides the façade of a carefully conducted study, but in fact, the path from hypothesis through data collection to rigorous conclusions is short-circuited by looking at the results and applying a story that fits the data. In some respects, Gregor Mendel’s classic pea genetics studies are consistent with this latter model, with data likely too perfect to have arisen naturally (49).

**Expectations of publication.** Publications serve as academic currency, and thus academics may be under pressure to publish something—sometimes anything—to increase that currency, obtain tenure, or maintain funding. This is the so-called “publish-or-perish” paradigm. Given the expansion of the number of journals, there are fewer barriers to publishing, and a more modern expectation may include the desire to publish in higher-ranking

journals; garner more publicity; or report positive, novel, or exciting results.

There may also be personal expectations: After months of experimentation or years of data collection, something “useful” is desired of a project. Not everything is worth publishing if it does not add knowledge. If the data are bad, methods flawed, or conclusions invalid, the publication will not contribute to knowledge, but rather may detract from knowledge. Publication-based, goal-directed pressure may drive behavior away from rigorous science. In 1975, Paul Feyerabend expressed concerns over the increase in publications without a concomitant increase in knowledge by remarking, “Most scientists today are devoid of ideas, full of fear, intent on producing some paltry result so that they can contribute to the flood of inane papers that now constitutes ‘scientific progress’ in many areas” (50).

**Excitement.** Many scientists began their lines of investigation because of innate interest: a deep curiosity, a desire for discovery, or a personal connection to a problem in the world. Conducting experiments, analyzing data, and observing the world are not just aspects of science, but also represent personal interests and passions. Thus, when results provide something interesting—whether simply intellectually stimulating or of profound practical importance—passion and excitement risk overriding the fact that science is designed to be “the great antidote to the poison of enthusiasm and superstition” (51). During the Sackler Colloquium upon which this issue of PNAS is built, Sir Philip Campbell noted this dichotomy of excitement vs. rigor. Commenting on his training, he remarked, “The culture in the lab was that if I made any interesting claim about what I was discovering, my supervisor assumed it was a collaboration between Mother Nature and my equipment to tell a lie. And I really had to work to convince him if ever I really thought I got something interesting” (52).

**Resources.** Whether it be time, personnel, education, or money, rigorous science requires resources. Insufficient resources may foster errors. If time is short, adequate checks for rigor may be foregone; if there are too few personnel, a team may be insufficient to complete a project; if there is too little education, appropriate expertise may be lacking; and if there is inadequate funding, rigorous methodology may be inaccessible. Practical compromises must be made, sometimes at the cost of rigor.

**Conflicting priorities.** Insufficient checking of methods, results, or conclusions because of conflicting priorities can also contribute to the introduction or ignoring of errors. A researcher may consciously know better than to commit certain errors or shortcuts, but priorities may compete for resources, attention, or willpower. The result may be sloppy science, neglectful behavior, or a distortion of observations. In fact, there may be some disparity among scientists with respect to attending to such conflicts, with higher creativity being associated with lower levels of conscientiousness compared with those with lower creativity, according to one metaanalysis (53). It is often impossible to determine if the authors succumbed to these conflicting priorities, they intentionally deviated from scientific rigor, or they made honest errors. The most common discourse about priorities is around disclosure of potential financial conflicts, but there are many other sources of conflict. When individuals believe in an idea so fully or have built an entire career and image upon an idea, publishing something to the contrary would be to conflict with an entrenched ideology (54, 55). In other cases, the ideology an author may champion is considered righteous. In pediatric obesity, for instance, many putative causal factors are dichotomized as bad (screen time and sedentary behavior) or good (breastfeeding and family time), depending on the prevailing zeitgeist. In the interest of protecting children from obesity, researchers may succumb to White Hat Bias, which involves “distortion of information in the service of what may be perceived to be righteous ends” (56). In turn, future research may parrot the righteous stance regardless of true effects, such as when the potential for publication bias was ignored in a World Health Organization report of breastfeeding and obesity (57).

We clarify that, although intentionality is important for separating sloppiness from deliberate misconduct and possibly for addressing errors or reprimanding bad actors, both intentional and unintentional deviations from best practices cause erroneous contributions to the scientific literature.

## The Prevalence and Consequences of Errors

**Prevalence of Errors and the Ability to Detect Them.** Individual scientists have long noted and criticized errors in the literature, causing heated interchanges to this day. While we are aware of no formal, long-standing catalog of errors, either in frequency or category, efforts have critiqued common errors in focused areas of the literature (31, 58), aiming to educate the particular community where these are observed. Other groups used statistical approaches to detect data errors (59, 60).

Some individuals have made scientific critiques a personal mission. In a 1990 book focusing on methodological errors in medical research (61), Andersen states that his goal is to improve the quality of research by educating its consumers, who are, in many cases, also the source. Andersen goes on to enumerate many examples of errors since the 1950s. Systematic sampling of 149 studies from popular medical journals concluded that only 28% of the sample was considered “acceptable” (62). More than 20 y later, an analysis of 196 drug trials to treat rheumatoid arthritis concluded that 76% of the conclusions or abstracts contained “doubtful or invalid statements” (63).

Surveys of the literature have also cataloged invalidating errors we mentioned earlier. For cluster randomized trials in occupational therapy interventions, 7 of 10 identified studies included clustering in the analysis (64), while 19 of 83 clustered cross-over studies were unclear as to whether both clustering and cross-over effects were included in the analysis (65). Data extraction for effect-size calculation was similarly problematic, with errors found in 17 of 27 metaanalyses (66). Bakker and Wicherts surveyed reporting of statistical analyses in psychology journals and noted that 55% of articles had errors, with 18% having gross errors (67). The ability of the authors to detect and categorize errors depended both on whether statistics were reported and if they were completely and exactly reported (e.g., reporting exact test statistics with degrees of freedom vs.  $P < 0.05$ ).

New technologies and standards have increased our ability to detect some errors or research behaviors. The “statcheck” package for R can automatically test whether there are inconsistencies between  $P$  values and test statistics, and was used to discover that one in eight psychology articles had errors that could affect conclusions (14). Other software is available to detect an error of granularity, called GRIM errors, which “evaluates whether the reported means of integer data... are consistent with the given sample size and number of items” (60); e.g., if two integers are averaged, the first decimal place must be a 5 or a 0.

Clinical trial registration and reporting mandates allow the comparison of published articles against the preregistered intentions and primary outcomes across time. The proportion of large, National Heart, Lung, and Blood Institute studies reporting null findings over time has increased, which is suggestive of decreases in publication bias (68). On the other hand, the COMParte Trials Project ([compare-trials.org/](http://compare-trials.org/)) has noted that many pre-registered outcomes in trials were not reported, and many others were added to publications with no mention that the endpoints were not preregistered.

Any conversation about errors in the literature would be incomplete without discussing the peer review process. Regardless of the scientific discipline, a challenging part of peer review is what is another example of an unknown unknown. Peer reviewers read what is contained in the text of a manuscript, but are incapable of evaluating some errors if manuscripts do not present complete information. Reporting guidelines have been developed to improve reporting of a variety of study types and subdisciplines as cataloged by the Enhancing the Quality and Transparency of Health Research network; these guidelines are also useful for reviewers to identify missing information. Even

with the existence of guidelines and journal mandates, authors do not always report information specified in guidelines, nor do peer reviewers demand that the information be reported (69). Furthermore, universal operational definitions and standards of peer review remain elusive (70), although new models of open review are evolving with some journals and publishers.

**Consequences of Errors.** The systematic use of erroneous methods or bad data can affect the entirety of our knowledge of a phenomenon. This may cause harm. Poor reasoning and data collection resulted in radiation treatment of children to prevent sudden infant death syndrome in the early 1900s, resulting in >10,000 babies dying of thyroid cancer (71).

Even when the errors do not result in our collective misunderstanding of natural phenomena or a tragic loss of life, the consequences can affect the scientific enterprise in other ways. If scientists cannot produce reliable, reproducible experiments, why should the public trust the scientific enterprise? An imbalance in trust between the purveyor and consumer of knowledge has been termed a “lemon market” (72), which is an idiomatic expression used to describe a decrease in quality that occurs from information asymmetry (73). In our fields of nutrition and obesity, the constant vacillation of headlines and studies purporting that a food is good and then perhaps bad from week to week has contributed to a decreased trust in nutrition science, with many unscientific ideas being advanced through higher exposure (34, 74–77). Although lay media have contributed to sensationalism and misrepresentation of science, many of these mistaken messages originate from the scientific community itself (78). If we are conducting erroneous research, then time, resources, and money have been wasted. One estimate calculated this waste to be half of all preclinical research, amounting to an estimated \$28 billion going to irreproducible research in the United States alone in 2015 (79).

## How to Improve Conditions and Quality

Other papers in this issue offer suggestions and considerations for rigor, reproducibility, and transparency that are also relevant to the concerns we raise. We focus here on mechanisms to address data and statistically oriented errors that appear to have occurred in published papers. Experience indicates that the handling of such errors (both purported and confirmed) is haphazard, unduly slow, inconsistent, and often markedly socially inappropriate (11, 80, 81). Thus, we both offer suggestions on how to and how not to handle such errors.

## Some Suggested Principles and Practices.

**Comment on studies, data, methods, and logic, not authors.** The recent case of the criticisms inveighed against a prominent researcher’s work (82) offers some stark examples of individuals going beyond commenting on the work itself to criticizing the person in extreme terms (e.g., ref. 83). As we have said elsewhere (84), in science, three things matter: the data, the methods used to collect the data (which give them their probative value), and the logic connecting the data and methods to conclusions. Everything else is a distraction. However, in trying to counter the points of some authors or studies, some individuals resort to *ad hominem* arguments, often trying to undermine the credibility of arguments by attacking a person based on perceived expertise (85) or presumed motives, focusing especially on funding sources (86). These attacks (38) are not new (87), and remain distractions from the science itself. In our opinions, and in the opinions of some scientific societies, such attacks on fellow scientists on nonscientific grounds are unethical (*Examples of Societies and Associations Denouncing Ad Hominem Attacks*). Scientists are often protected by academic freedom, and in the United States, individuals are afforded First Amendment rights for free speech. However, freedoms are not immune to legal or social recourse, as in the case where a biotech chief executive officer was convicted of wire fraud for a misleading press release about a

## Examples of Societies and Associations Denouncing Ad Hominem Attacks

“Harassment includes speech or behavior that is not welcome or is personally offensive, whether it is based on... any other reason not related to scientific merit” (89).

“Attempting to discredit scientific opinions or individuals solely on the basis of collaborative relationships and/or funding sources has no place in the scientific process” (90).

“In a professional setting, it’s best to avoid ad hominem arguments and personal attacks, especially if they amount to slander, libel, and/or sexual harassment” (91).

“Criticism of another’s language, ideas, or logic is a legitimate part of scholarly research, but ethical researchers avoid ad hominem attacks” (92).

“Differences of opinion and disagreements... do not in and of themselves necessarily constitute harassment; involved individuals should nonetheless endeavor to be respectful and refrain from ad hominem remarks” (93).

product (88). Individuals engaging in ad hominem attacks in scientific discourse should be subject to censure.

**Respectfully raise potential concerns about invalidating errors (or plausible misconduct) and allow for due process.** If an invalidating error or misconduct has occurred, we believe the best way to proceed is to report the concern privately to some combination of the author, the journal editor, or the author’s institution. Scientists should participate in the private and due process of adjudication and, if appropriate, correct the purported error quickly. Even if subsequently found to be unsubstantiated, merely the allegation of a severe invalidating error or, worse yet, misconduct, can permanently taint individuals or important works (94, 95). “Trial by blog” is no way to adjudicate scientific knowledge or the reputations and careers of individual scientists. We do not suggest that public discourse about science, and particularly potential errors or points of clarification, should be stifled. Postpublication discussion platforms such as PubPeer, PubMed Commons, and journal comment sections have led to useful conversations that deepen readers’ understanding of papers by bringing to the fore important disagreements in the field. Informal, public platforms have unfortunately led to public ridicule [e.g., a post now removed (96)], and even legal battles by those who were the subject of public discussion (97). Professional decorum and due process are minimum requirements for a functional peer review system, and thus it seems only fair that those norms should function in postpublication peer review, too. As discussed in this issue and elsewhere (81), potential errors should be appropriately identified, verified, and corrected, while protecting both those raising the errors in good faith and those who are being accused of making honest errors. The focus must remain on the science.

**Develop and utilize uniform procedures for addressing purported invalidating errors in a timely fashion.** Our call for professional decorum and due process is, admittedly, somewhat idealistic. As we reported elsewhere (11), the process of getting errors corrected, even when going through proper channels with journals, is often some combination of absurdly slow, inept, confusing, costly, time-intensive, and unsatisfying. Papers declared by editors to contain patently incorrect conclusions are allowed to stand unretracted if an author declines to retract (98). Papers retracted because of errors in one journal are republished with the same errors in other journals (99). Journals may take more than a year to resolve an issue, failing to keep the concerned individuals apprised of progress or to provide specific timelines. Editors may abrogate their responsibility for resolving claims of invalidating errors (100), leaving it to teams of authors to make cases in opposing letters (e.g., refs. 101 and 102) and likely leaving readers confused. It seems essential that the scientific community come together to promulgate better procedures for handling

concerns about invalidating errors. The proposed Research Integrity Advisory Board (103), combined with the Committee on Publication Ethics and the International Committee of Medical Journal Editors may be guiding bodies through which this could be accomplished. Until such procedures are in place and working expeditiously, we think some scientists may still feel compelled to address their concerns publicly, and those who are accused of misdeeds may seek guidance on how to respond to accusations (104).

**Potential Solutions.** Bringing about improvements in the scientific enterprise will require varied approaches and constant vigilance from multiple stakeholders. Suggested solutions have frequently been to raise awareness or increase education. Although admitting we have a problem is the first step to fixing it, science is a human endeavor, and behavioral scientists have demonstrated how hard it is to effect change in habitual behavior and attitudes. Much like science itself, the solutions will continue to evolve and will require the involvement and coordination of various stakeholders. Fortunately, better tools are evolving.

**Considerations for education.** Educational approaches are frequently recommended as ways to fix problems in science. Indeed, our comment that many errors we see may be related to ignorance seems to suggest we believe education is a good solution. Clearly, if we do not increase awareness of problems and teach their solutions, there is little hope to address the issues. However, there are substantial challenges to implementing educational solutions. Foremost, research rigor and related topics are parts of a discipline in their own right, so simply adding them to curricula is impractical for many situations. Curricula at universities struggle to accommodate everything that may be required to be taught, with various subdisciplines pushing for more representation in already-bloated programs. Adding additional courses on study design, logical inference, data management, statistical analysis, and other topics that are important to rigor and reproducibility may require difficult curricular or time-commitment trade-offs.

One approach to the dueling concerns of time and required education is to incorporate components into a synergistic curriculum, where topics could be better integrated with existing courses. This has been attempted, for instance, by incorporating writing into laboratory courses; perhaps incorporating logic, statistical analysis, data integrity, or study design into other courses could also work. Alternatively, better preparing students to operate in a truly interdisciplinary team may alleviate the need for deep knowledge of everything. If laboratory scientists were trained to be familiar with, rather than functionally proficient in, statistical analyses, then they could perhaps better collaborate with a statistician. This divergence of expertise was recounted, if even just apocryphally, on an American Statistical Association discussion board:

A neurosurgeon phones the statistical consulting department to inform them, “I’m doing a study, and I’d rather just do my own statistics. So I don’t need your help; I just wonder if you can suggest a good statistics text.” The consulting statistician says, “I’m so glad you called! I’ve always wanted to do brain surgery; can you suggest a good text on that?” (105)

In addition, if education is not paired with other structural changes, competing priorities may overshadow the knowledge gained. Many statistical errors are already covered in required courses, and yet they persist.

**Considerations for “gatekeeper” functions.** Gatekeeper functions create circumstances in which people have no choice but to “do the right thing.” Such solutions have already been implemented in several domains, such as requirements for registration of trials. Requirements for depositing of raw data and publication of statistical code have been implemented by some journals. Some funders and contracts require the posting of results, such as for studies registered in [ClinicalTrials.gov](https://clinicaltrials.gov). One thing these functions have in common is increasing the amount of information reported—“increased transparency.” After all, it is difficult to identify errors

if insufficient information is provided to be able to evaluate the science.

These gatekeeper functions are important for forcing some actions. However, without the intrinsic buy-in from cultural shifts or extrinsic incentives, researchers may only comply within the letter of the requirements, rather than the intended spirit of rigor. Tightening requirements too far may risk the creation of a system that will fail to be flexible enough to accommodate a variety of scientific areas. In addition, some investigators have lamented that they spend, on average, almost half of their time on administrative tasks (106). Gatekeeping functions may increase this burden, and have been criticized as the bureaucratization of science (107). Burdens can be alleviated by additional resources, such as new job roles tailored to requirements within institutions, much like an interdisciplinary approach alleviates the need for a single scientist to be a polymath.

**Considerations for incentive systems.** Incentives and disincentives go beyond permitting one to pass through gatekeeper functions. Being allowed to publish after complying with gatekeeper restrictions is hardly a reward. Incentives involve rewards, ratings, and rankings that provide acknowledgment for work well done. Receiving badges for open science practices (e.g., through Badge Alliance) is one approach to extrinsic motivation via recognition. Such rewards may need additional reinforcement beyond passive community recognition, like inclusion in tenure and promotion decisions.

Disincentive systems may also be employed. For example, the National Institutes of Health can withhold funding if studies associated with their funding are not compliant with public access—this could also be considered a gatekeeper function; however, investigators being fined for failing to submit results to [ClinicalTrials.gov](https://clinicaltrials.gov) could be considered a disincentive.

Incentives and disincentives may result in “gaming the system.” Such recently questioned incentives like recognition for publishing in high-impact-factor journals resulted in journals artificially inflating their impact factor through various means of inflating self-citations (108). In any large enterprise, behavior can at best be improved incrementally. Processes need to be resilient to manipulation and should not be a substitute for critical evaluation of research outputs.

**Considerations for increasing resources.** The need to increase resources to improve rigor and reproducibility is also a common refrain. If the education, gatekeeper, and incentive solutions are to be accomplished, they will need proper funding, personnel, and buy-in from stakeholders. However, increasing resources for rigor means resources may be taken away from other endeavors unless society increases resources in toto (such as through taxes) or creative solutions are enacted.

Reapportioning resources to reinforce rigorous research could pay for itself. Rather than many small, underpowered, non-randomized studies, or collecting cross-sectional survey data with a variety of discordant, nonvalidated questionnaires, we could pool resources for consortia to provide more useful, complete, and reliable knowledge, especially for probative, hypothesis-testing research. This is not to undervalue exploratory work, but too frequently, exploratory work or pilot and feasibility studies are presented as hypothesis testing, rather than generating, research. Such a culture shift could relieve the fiscal burden of reinforcing rigor and improve gains in knowledge from enriching our corpus of research with higher-quality evidence.

Over time, many proposed solutions should gain efficiencies. Indeed, various best-practice reporting guidelines (e.g., the Consolidated Standards of Reporting Trials guidelines for human trials or the Animal Research: Reporting of In Vivo Experiments guidelines for animal studies) have been streamlined for journals to use, requiring less effort to implement.

**Considerations for shifts in scientific culture.** Increasing the intrinsic motivation to conduct rigorous science is the cornerstone for our proposed considerations. The other considerations we presented depend on individuals having the intrinsic motivation to pursue truth and see science as their vocation or passion, rather than

merely “a job.” Without a dedication to finding one’s own mistakes, scientists may not seek education, may circumvent gatekeeper functions, may game incentives, or might squander resources in favor of other priorities.

Normalizing error correction is necessary to advance the enterprise. Some have suggested retiring or replacing the all-encompassing word “retraction” (which conjures “misconduct”) with more meaningful descriptions of correction (109). “Retraction and republication” could be used to both maintain the historical record and correct the scientific literature (110). If science is to be self-correcting, encouraging authors to be active participants in the correction process is essential, and stigma should be minimized in cases of honest error.

Similarly, recognizing scientists for their contribution to the scientific quality-control process of peer review may be an important cultural shift. Journals have long listed reviewer names in annual “thanks” statements, but peer review is sometimes viewed as a burden—an added service expectation without pay—and may be relegated to a footnote on some curricula vitae. A movement toward valuing peer review led to the creation of Publons.

1. Hamilton A (1961) *The Papers of Alexander Hamilton* (Columbia Univ Press, New York).
2. Ioannidis JPA (2012) Why science is not necessarily self-correcting. *Perspect Psychol Sci* 7:645–654.
3. Salsburg D (2017) *Errors, Blunders, and Lies* (CRC, Boca Raton, FL).
4. Perkel J (2012) Should Linus Pauling’s erroneous 1953 model of DNA be retracted? *Retraction Watch*. Available at [retractionwatch.com/2012/06/27/should-linus-paulings-erroneous-1953-model-of-dna-be-retracted/](http://retractionwatch.com/2012/06/27/should-linus-paulings-erroneous-1953-model-of-dna-be-retracted/). Accessed September 21, 2017.
5. Burchfield JD (1990) *Lord Kelvin and the Age of the Earth* (Univ of Chicago Press, Chicago).
6. Galton F (1897) The average contribution of each several ancestor to the total heritage of the offspring. *Proc R Soc Lond* 61:401–413.
7. Livio M (2013) *Brilliant Blunders: From Darwin to Einstein—Colossal Mistakes by Great Scientists that Changed our Understanding of Life and the Universe* (Simon & Schuster, New York), 1st Simon & Schuster hardcover ed, p 341.
8. Sechrist R (1925) *An Introduction to Statistical Methods* (Macmillan, New York).
9. Ellenberg J (2014) *How Not to Be Wrong: The Power of Mathematical Thinking* (Penguin, New York).
10. “Student” (1931) The Lanarkshire milk experiment. *Biometrika* 23:398–406.
11. Allison DB, Brown AW, George BJ, Kaiser KA (2016) Reproducibility: A tragedy of errors. *Nature* 530:27–29.
12. Carlisle JB (2017) Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals. *Anaesthesia* 72:944–952.
13. Ercan I, et al. (2017) Examination of published articles with respect to statistical errors in veterinary sciences. *Acta Veterinaria* 67:33–42.
14. Nuijten MB, Hartgerink CH, van Assen MA, Epskamp S, Wicherts JM (2016) The prevalence of statistical reporting errors in psychology (1985–2013). *Behav Res Methods* 48:1205–1226.
15. National Academies of Sciences Engineering, & Medicine (2017) *Fostering Integrity in Research* (The National Academies Press, Washington, DC).
16. Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G (2016) Ranking major and minor research misbehaviors: Results from a survey among participants of four World Conferences on Research Integrity. *Res Integr Peer Rev* 1:17.
17. Firestein S (2016) *Failure: Why Science Is So Successful* (Oxford Univ Press, Oxford).
18. Allison DB, et al. (2015) Goals in nutrition science 2015–2020. *Front Nutr* 2:26.
19. Dhurandhar NV, et al.; Energy Balance Measurement Working Group (2015) Energy balance measurement: When something is not better than nothing. *Int J Obes* 39: 1109–1113.
20. Beaudoin R, Mayer J (1953) Food intakes of obese and non-obese women. *J Am Diet Assoc* 29:29–33.
21. Sebastiani P, et al. (2010) Genetic signatures of exceptional longevity in humans. *Science*, 10.1126/science.1190532, and retraction (2011) 333:404.
22. Hughes P, Marshall D, Reid Y, Parkes H, Gelber C (2007) The costs of using unauthenticated, over-passaged cell lines: How much more data do we need? *Bio-techniques* 43:575, 577–586, and erratum (2008) 44:47.
23. Verhulst B, Eaves L, Haterini PK (2012) Correlation not causation: The relationship between personality traits and political ideologies. *Am J Pol Sci* 56:34–51, and erratum (2016) 60:E3–E4.
24. Bent S, Tiedt TN, Odden MC, Shlipak MG (2003) The relative safety of ephedra compared with other herbal products. *Ann Intern Med* 138:468–471, and correction (2003) 138:1012.
25. Gawande A (2010) *The Checklist Manifesto: How to Get Things Right* (Metropolitan Books, New York).
26. George BJ, Brown AW, Allison DB (2015) Errors in statistical analysis and questionable randomization lead to unreliable conclusions. *J Paramed Sci* 6:153–154.
27. Li P, et al. (2015) Concerning Sichieri R, Cunha DB: Obes facts 2014; 7:221–232. The assertion that controlling for baseline (pre-randomization) covariates in randomized controlled trials leads to bias is false. *Obes Facts* 8:127–129.
28. Anonymous (2015) Retraction statement: LA sprouts randomized controlled nutrition, cooking and gardening program reduces obesity and metabolic risk in Latino youth. *Obesity (Silver Spring)* 23:2522.
29. Zalewski BM, et al. (2015) Correction of data errors and reanalysis of “The effect of glucosamine on body weight in overweight or obese children and adults: A systematic review of randomized controlled trials”. *Nutrition* 31:1056–1057.
30. Gelman A, Stern H (2006) The difference between “significant” and “not significant” is not itself statistically significant. *Am Stat* 60:328–331.
31. George BJ, et al. (2016) Common scientific and statistical errors in obesity research. *Obesity (Silver Spring)* 24:781–790.
32. Serlie MJ, Ter Horst KW, Brown AW (2016) Addendum: Hypercaloric diets with high meal frequency, but not increased meal size, increase intrahepatic triglycerides: A randomized controlled trial. *Hepatology* 64:1814–1816.
33. Allison DB, Williams MS, Hand GA, Jakicic JM, Fontaine KR (2015) Conclusion of “Nordic walking for geriatric rehabilitation: A randomized pilot trial” is based on faulty statistical analysis and is inaccurate. *Disabil Rehabil* 37:1692–1693.
34. Brown AW, Bohan Brown MM, Allison DB (2013) Belief beyond the evidence: Using the proposed effect of breakfast on obesity to show 2 practices that distort scientific evidence. *Am J Clin Nutr* 98:1298–1308.
35. Cofield SS, Corona RV, Allison DB (2010) Use of causal language in observational studies of obesity and nutrition. *Obes Facts* 3:353–356.
36. Micha R, et al. (2017) Association between dietary factors and mortality from heart disease, stroke, and type 2 diabetes in the United States. *JAMA* 317:912–924.
37. Mozaffarian D (2017) “Our nation’s nutrition crisis: nearly 1,000 cardiovascular & diabetes deaths each day (!) due to poor diet. <https://t.co/oECZQRJQQA>.” Available at <https://twitter.com/Dmozaffarian/status/839227095565815808>. Accessed March 7, 2017.
38. Trasande L, Lind PM, Lampa E, Lind L (2017) Dismissing manufactured uncertainties, limitations and competing interpretations about chemical exposures and diabetes. *J Epidemiol Community Health* 71:942.
39. American Council on Science and Health (1991) *From Mice to Men: The Benefits and Limitations of Animal Testing in Predicting Human Cancer Risk* (American Council on Science and Health, New York), 3rd Ed.
40. Louis-Sylvestre J, et al. (1989) Learned caloric adjustment of human intake. *Appetite* 12:95–103.
41. Bohan Brown MM, Brown AW, Allison DB (2015) Linear extrapolation results in erroneous overestimation of plausible stressor-related yearly weight changes. *Biol Psychiatry* 78:e10–e11.
42. Thomas DM, et al. (2013) Can a weight loss of one pound a week be achieved with a 3500-kcal deficit? Commentary on a commonly accepted rule. *Int J Obes* 37: 1611–1613.
43. Schwingshakel L, et al. (2015) Fruit and vegetable consumption and changes in anthropometric variables in adult populations: A systematic review and meta-analysis of prospective cohort studies. *PLoS One* 10:e0140846.
44. Gelman A, Loken E (2013) The garden of forking paths. Available at [www.stat.columbia.edu/~gelman/research/unpublished/\\_p\\_hacking.pdf](http://www.stat.columbia.edu/~gelman/research/unpublished/_p_hacking.pdf). Accessed July 5, 2017.
45. Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci* 22:1359–1366.
46. Gadbury GL, Allison DB (2012) Inappropriate fiddling with statistical analyses to obtain a desirable p-value: Tests to detect its presence in published literature. *PLoS One* 7:e46363.
47. Begley S (2012) In cancer science, many “discoveries” don’t hold up. *Reuters Science News*. Available at [www.reuters.com/article/us-science-cancer-idUSBRE82R12P20120328](http://www.reuters.com/article/us-science-cancer-idUSBRE82R12P20120328). Accessed July 5, 2017.
48. Kerr NL (1998) HARKing: Hypothesizing after the results are known. *Pers Soc Psychol Rev* 2:196–217.
49. Pires AM, Branco JA (2010) A statistical model to explain the Mendel–Fisher controversy. *Stat Sci* 25:545–565.
50. Feyerabend P (1975) How to defend society against science. *Radical Philos* 2:5–14.

If we can reinforce that scientific quality should be sought foremost in a scientific endeavor—rather than paychecks or publications, grants or grandeur—then we believe improvements will follow.

## Conclusions

Science is the process by which we come to have objective knowledge of the world. It has enriched our lives with wonder and discovery and enhanced our world with useful technology. However, throughout, it is an incremental process conducted by imperfect humans. It has always been subject to error and always will be. Still, those same flawed humans who conduct and care about science ever refine the scientific processes to reduce errors and increase rigor. Much good has been done on this path, and much good remains to be done ahead.

**ACKNOWLEDGMENTS.** This work was supported in part by National Institutes of Health Grants R25DK099080, R25HL124208, R25GM116167, P30DK056336, and P30AG050886. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NIH or any other organization.

51. Smith A (1809) *An Inquiry into the Nature and Causes of the Wealth of Nations* (J. Maynard, London).

52. National Academy of Sciences (2017) Panel Discussion, Reproducibility of Research: Issues and Proposed Remedies, Washington, DC, March 8–10, 2017. Available at <https://www.youtube.com/watch?v=7ustnetADPY>. Accessed September 21, 2017.

53. Feist GJ (1998) A meta-analysis of personality in scientific and artistic creativity. *Pers Soc Psychol Rev* 2:290–309.

54. Viswanathan M, et al. (2013) AHRQ methods for effective health care. *Identifying and Managing Nonfinancial Conflicts of Interest for Systematic Reviews* [Agency for Healthcare Research and Quality (US)], Rockville, MD.

55. Dragioti E, Dimoliatis I, Fountoulakis KN, Evangelou E (2015) A systematic appraisal of allegiance effect in randomized controlled trials of psychotherapy. *Ann Gen Psychiatry* 14:25.

56. Cope MB, Allison DB (2010) White hat bias: Examples of its presence in obesity research and a call for renewed commitment to faithfulness in research reporting. *Int J Obes* 34:84–88, discussion 83.

57. Cope MB, Allison DB (2008) Critical review of the World Health Organization's (WHO) 2007 report on 'evidence of the long-term effects of breastfeeding: Systematic reviews and meta-analysis' with respect to obesity. *Obes Rev* 9:594–605.

58. Brown AW, et al. (2015) Best (but oft-forgotten) practices: Designing, analyzing, and reporting cluster randomized controlled trials. *Am J Clin Nutr* 102:241–248.

59. Benford F (1938) The law of anomalous numbers. *Proc Am Philos Soc* 78:551–572.

60. Brown NJL, Heathers JA (2017) The GRIM test: A simple technique detects numerous anomalies in the reporting of results in psychology. *Soc Psychol Personal Sci* 8:363–369.

61. Andersen B (1990) *Methodological Errors in Medical Research* (Blackwell Scientific, New York).

62. Schor S, Karten I (1966) Statistical evaluation of medical journal manuscripts. *JAMA* 195:1123–1128.

63. Gotzsche PC (1989) Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 10:31–56.

64. Tokolahi E, Hocking C, Kersten P, Vandal AC (2016) Quality and reporting of cluster randomized controlled trials evaluating occupational therapy interventions: A systematic review. *OTJR (Thorofare, NJ)* 36:14–24.

65. Arnup SJ, Forbes AB, Kahan BC, Morgan KE, McKenzie JE (2016) The quality of reporting in cluster randomised crossover trials: Proposal for reporting items and an assessment of reporting quality. *Trials* 17:575.

66. Gotzsche PC, Hróbjartsson A, Maric K, Tendal B (2007) Data extraction errors in meta-analyses that use standardized mean differences. *JAMA* 298:430–437.

67. Bakker M, Wicherts JM (2011) The (mis)reporting of statistical results in psychology journals. *Behav Res Methods* 43:666–678.

68. Kaplan RM, Irvin VL (2015) Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLoS One* 10:e0132382.

69. Hopewell S, et al. (2014) Impact of peer review on reports of randomised trials published in open peer review journals: Retrospective before and after study. *BMJ* 349:g4145.

70. Smith R (2006) Peer review: A flawed process at the heart of science and journals. *J R Soc Med* 99:178–182.

71. Ritterman JB (2017) To err is human: Can American medicine learn from past mistakes? *Perm J* 10:7812/TPP16-181.

72. Cottrell RC (2013) Scientific integrity and the market for lemons. *Res Ethics Rev* 10:17–28.

73. Akerlof GA (1970) The market for "lemons": Quality, uncertainty and the market mechanism. *Q J Econ* 84:488–500.

74. Brown AW, Ioannidis JP, Cope MB, Bier DM, Allison DB (2014) Unscientific beliefs about scientific topics in nutrition. *Adv Nutr* 5:563–565.

75. Casazza K, et al. (2013) Myths, presumptions, and facts about obesity. *N Engl J Med* 368:446–454.

76. Casazza K, et al. (2015) Weighing the evidence of common beliefs in obesity research. *Crit Rev Food Sci Nutr* 55:2014–2053.

77. Bohan Brown MM, Brown AW, Allison DB (2013) Nutritional epidemiology in practice: Learning from data or promulgating beliefs? *Am J Clin Nutr* 97:5–6.

78. Yavchitz A, et al. (2012) Misrepresentation of randomized controlled trials in press releases and news coverage: A cohort study. *PLoS Med* 9:e1001308.

79. Freedman LP, Cockburn IM, Simcoe TS (2015) The economics of reproducibility in preclinical research. *PLoS Biol* 13:e1002165.

80. Barbour V, Bloom T, Lin J, Moylan E (2017) Amending published articles: Time to rethink retractions and corrections? *bioRxiv*:10.1101/118356.

81. Fiske ST (2016) A call to change science's culture of shaming. *APS Observer*. Available at [www.psychologicalscience.org/observer/a-call-to-change-sciences-culture-of-shaming](http://www.psychologicalscience.org/observer/a-call-to-change-sciences-culture-of-shaming). Accessed June 21, 2017.

82. McCook A (2017) Cornell finds mistakes—not misconduct—in papers by high-profile nutrition researcher. *Retraction Watch*. Available at <https://retractionwatch.com/2017/04/06/cornell-finds-mistakes-not-misconduct-papers-high-profile-nutrition-researcher/>. Accessed June 21, 2017.

83. Anaya J (2017) The Donald Trump of food research. *Medium.com*. Available at <https://medium.com/@OmnesRes/the-donald-trump-of-food-research-49e2bc7daa41>. Accessed September 21, 2017.

84. Allison DB (2017) Letter re Hearing on Pros and Cons of Restricting SNAP Purchases. Committee on Agriculture, House of Representatives, 115th Congress, 1st Session (U.S. Government Publishing Office, Washington, DC).

85. Klement RJ, et al. (2017) Need for new review of article on ketogenic dietary regimes for cancer patients. *Med Oncol* 34:108.

86. Zaraska M (2016) What big meat wants you to think: Research is frequently funded by the industry. *New York Daily News*. Available at [www.nydailynews.com/opinion/marta-zaraska-big-meat-article-1.2669374](http://www.nydailynews.com/opinion/marta-zaraska-big-meat-article-1.2669374). Accessed July 5, 2017.

87. Rothman KJ (1993) Conflict of interest. The new McCarthyism in science. *JAMA* 269:2782–2784.

88. Memorandum Appeal from the United States District Court for the Northern District of California, United States v. W. SCOTT HARKONEN, M.D. (2012). Available at [cdn.ca9.uscourts.gov/datastore/memoranda/2013/03/04/11-10209.pdf](http://cdn.ca9.uscourts.gov/datastore/memoranda/2013/03/04/11-10209.pdf). Accessed July 5, 2017.

89. American Association for the Advancement of Science (2015) AAAS Annual Meeting Code of Conduct. Available at [meetings.aaas.org/policies](http://meetings.aaas.org/policies). Accessed September 21, 2017.

90. The Obesity Society (2014) Acceptance of Financial Support from Industry for Research, Education & Consulting. Available at [www.obesity.org/publications/position-and-policies/research-education-consulting](http://www.obesity.org/publications/position-and-policies/research-education-consulting). Accessed September 21, 2017.

91. American Philosophical Association (2016) Code of Conduct. Available at [www.apaonline.org/page/codeofconduct](http://www.apaonline.org/page/codeofconduct). Accessed September 21, 2017.

92. National Communication Association (1999) A Code of Professional Ethics for the Communication Scholar/Teacher. Available at [https://www.natcom.org/sites/default/files/pages/1999\\_Public\\_Statements\\_A\\_Code\\_of\\_Professional\\_Ethics\\_for\\_%20the\\_Communication\\_Scholar\\_Teacher\\_November.pdf](https://www.natcom.org/sites/default/files/pages/1999_Public_Statements_A_Code_of_Professional_Ethics_for_%20the_Communication_Scholar_Teacher_November.pdf). Accessed September 21, 2017.

93. Human Biology Association (2016) Code of Ethics of the Human Biology Association (HBA) to be circulated to membership prior to a vote by the HBA Membership at the annual business meeting. Available at <https://humbio.org/sites/default/files/files/HBA%20Code%20of%20Ethics%20Final.pdf>. Accessed September 21, 2017.

94. Chawla DS (2016) "We should err on the side of protecting people's reputation:" Management journal changes policy to avoid fraud—interview with Patrick Wright. *Retraction Watch*. Available at [retractionwatch.com/2016/07/13/management-journal-changes-its-policy-to-avoid-fraud-qa-with-editor-patrick-wright/](https://retractionwatch.com/2016/07/13/management-journal-changes-its-policy-to-avoid-fraud-qa-with-editor-patrick-wright/). Accessed June 22, 2016.

95. McCook A (2017) It's not just whistleblowers who deserve protection during misconduct investigations, say researchers—interview with Sven Hendrix. *Retraction Watch*. Available at [retractionwatch.com/2017/05/29/hot-just-whistleblowers-deserve-protection-misconduct-investigations-say-researchers/](https://retractionwatch.com/2017/05/29/hot-just-whistleblowers-deserve-protection-misconduct-investigations-say-researchers/). Accessed June 22, 2017.

96. Public Library of Science (2015) Post removed by PLOS—the fight over transparency: Round two. *PLOS Biologue*. Available at [blogs.plos.org/biologue/2015/08/13/the-fight-over-transparency-round-two/](http://blogs.plos.org/biologue/2015/08/13/the-fight-over-transparency-round-two/). Accessed July 5, 2017.

97. Vence T (2014) Pathologist sues PubPeer users. Available at [www.the-scientist.com/?articles.view/articleNo/41322/title/Pathologist-Sues-PubPeer-Users/](http://www.the-scientist.com/?articles.view/articleNo/41322/title/Pathologist-Sues-PubPeer-Users/). Accessed July 5, 2017.

98. Oransky I (2015) When should a paper be retracted? A tale from the obesity literature. *Retraction Watch*. Available at [retractionwatch.com/2015/04/24/when-should-a-paper-be-retracted-a-tale-from-the-obesity-literature/](https://retractionwatch.com/2015/04/24/when-should-a-paper-be-retracted-a-tale-from-the-obesity-literature/). Accessed July 5, 2017.

99. Scudellari M (2017) "Strange. Very strange." Retracted nutrition study reappears in new journal. *Retraction Watch*. Available at [retractionwatch.com/2017/03/28/strange-strange-retracted-nutrition-study-reappears-new-journal/](https://retractionwatch.com/2017/03/28/strange-strange-retracted-nutrition-study-reappears-new-journal/). Accessed September 22, 2017.

100. Hauner H (2015) Quality management in scientific publishing—The importance to critically scrutinize scientific work. *Obes Facts* 8:125–126.

101. Kaiser KA, Parman MA, Kim E, George BJ, Allison DB (2016) Potential errors and omissions related to the analysis and conclusions reported in Cuspidi C, et al., AJH 2014; 27(2):146-156. *Am J Hypertens* 29:780–781.

102. Cuspidi C, Rescaldani M, Tadic M, Sala C, Grassi G (2016) Response to "potential errors and omissions related to the analysis and conclusions reported in Cuspidi C, et al., AJH 2014; 27(2):146-156". *Am J Hypertens* 29:782–783.

103. Mervis J (2017) U.S. report calls for research integrity board. *Science* Available at: [www.sciencemag.org/news/2017/04/us-report-calls-research-integrity-board](http://www.sciencemag.org/news/2017/04/us-report-calls-research-integrity-board). Accessed July 5, 2017.

104. Hendrix S (2014) What to do when you are falsely accused of scientific fraud? *smartsiencecareer*. Available at [www.smartsiencecareer.com/falsely-accused/](http://www.smartsiencecareer.com/falsely-accused/). Accessed September 21, 2017.

105. Norris DC (2017) ASA Connect, 17. RE: Common statistical errors in biomedical research? Available at [#bm15. Accessed July 5, 2017.](http://community.amstat.org/communities/community-home/digestviewer/viewthread?MessageKey=6af5728d-e8a4-42e3-b3a5-ed7c71c4195b&CommunityKey=6b2d607a-e31f-4f19-8357-020a8631b999&tab=digestviewer)

106. Rockwell S (2009) The FDP faculty burden survey. *Res Manag Rev* 16:29–44.

107. Anonymous (2010) The registration of observational studies—When metaphors go bad. *Epidemiology* 21:607–609.

108. Chorus C, Waltman L (2016) A large-scale analysis of impact factor biased journal self-citations. *PLoS One* 11:e0161021.

109. Enserink M (2017) Rethinking the dreaded r-word. *Science* 356:998.

110. The Lancet (2015) Correcting the scientific literature: Retraction and republication. *Lancet* 385:394.