

Improving the ACGIH threshold limit value (TLV) process

Toxicology Research and Application

Volume 2: 1–23

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/2397847318801758

journals.sagepub.com/home/tor

**Carr J Smith^{1,2} and Thomas A Perfetti³**

Abstract

The intent of this manuscript is to elucidate needed improvements in American Conference of Governmental Industrial Hygienists (ACGIH) threshold limit value (TLV) deliberations. More broadly, irreproducibility and bias adversely impact the collection, interpretation, statistical analysis, presentation, and reporting of results in many fields. In 2012, Begley and Ellis reported that scientists at Amgen had attempted to confirm published findings related to research topics of possible interest to Amgen. Fifty-three papers were deemed “landmark” studies. The authors were “shocked” when scientific findings were confirmed in only 6 (11%) cases. Many studies have confirmed that the peer-reviewed literature in biomedicine is in the midst of an irreproducibility crisis. Compounding the irreproducibility crisis is the existence of a significant bias against the publication of negative results. In the toxicology setting, negative toxicity test results are infrequently published as compared with reports that a chemical possesses a particular toxicity in a given test. Despite these deficiencies, the ACGIH states that “... the TLV[®]-CS Committee preferably relies on published, peer reviewed literature available in the public domain.” The primarily academic studies published in the peer-reviewed literature upon which ACGIH relies to determine TLVs rarely report raw data not already statistically transformed that are thus incalculable. In contrast, consideration of unpublished studies funded by industry, the vast majority of which are good laboratory practice-conducted contract lab studies, is only acceptable to ACGIH if the data owner provides the raw data to third parties upon request. This asymmetry in both the source of data emphasized, and inability to independently statistically analyze findings reported in the published academic literature, introduces a strong skew toward reliance on unverifiable although published measurements in the TLV process. Since Occupational Safety and Health Administration (OSHA) recommends that workplaces rely on ACGIH TLVs and National Institute for Occupational Safety and Health recommended exposure limits rather than older OSHA permissible exposure limit values to optimize worker safety, ACGIH should adopt a more transparent and science-based process.

Keywords

American Conference of Governmental Industrial Hygienists, threshold limit value, National Institute for Occupational Safety and Health recommended exposure limit, Occupational Safety and Health Administration PEL, transparency

Date received: 1 August 2018; accepted: 26 August 2018

Widespread irreproducibility of peer-reviewed published research results

Striving for complex procedural reproducibility predates modern science as apprenticeship in the medieval trades was based on the passing down of established methods and techniques from master to apprentice,¹ with similar systems extant in all of the civilizations of antiquity characterized by sophisticated infrastructure. Within the context of the scientific method, the reproducibility of research results

¹ Albemarle Corporation, Alabama, USA

² Department of Nurse Anesthesia, Florida State University, Tallahassee, FL, USA

³ Perfetti & Perfetti, LLC, Winston-Salem, NC, USA

Corresponding author:

Carr J Smith, Albemarle Corporation, 6400 Brindlewood Court Mobile, Alabama, USA; Department of Nurse Anesthesia, Florida State University, Tallahassee, FL 32306, USA.

Email: carr.smith@albemarle.com



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

has been an issue for many years. The first modern chemist, Robert Boyle (1627–1691), is credited as the first person to formally assert that knowledge should be founded on experimentally produced facts. According to Boyle, the relevant facts were believable to the scientific community because they were reproducible.² In 1934, Karl Popper noted that “non-reproducible single occurrences are of no significance to science.”³ In 1935, the British statistician Ronald Fischer stated that “we may say that a phenomenon is experimentally demonstrable when we know how to conduct an experiment which will rarely fail to give us statistically significant results.”⁴ The historical primacy of the necessity of reproducibility illustrates its place as a foundational concept in modern science.

In 2012, Begley and Ellis⁵ reported that scientists in the hematology and oncology department at Amgen had over the period of a decade attempted to confirm the published findings related to particular research topics of possible interest to Amgen. Fifty-three papers were deemed “landmark” studies and acknowledged that some of the data might not “holdup” because papers were deliberately selected that described something new, for example, “fresh approaches to targeting cancers or alternative clinical uses for existing therapeutics.” These authors described themselves as shocked when scientific findings were confirmed in only six (11%) cases. In summary, the state-of-the-art laboratories at Amgen could only repeat 6 of the 53 reportedly landmark preclinical cancer studies.

A similar lack of reproducibility has been reported in disciplines far afield from cancer biology, for example, in experimental psychology. Concerns about low reproducibility rates, in psychology studies, led Wicherts et al.⁶ to evaluate the feasibility of obtaining raw data for the purpose of recalculation. These authors identified 141 authors who had published in American Psychology Association (APA) journals and asked them to share their data. After 6 months, only 38 (27%) authors had responded. Wicherts and Bakker⁷ recommended that researchers publish their underlying data, and these authors provided a demonstration of how to do so. Undeterred by the unenthusiastic response encountered by Wicherts et al.,⁶ Vanpaemel et al.⁸ published a follow-up study in which 246 of 394 (62%) contacted authors of papers in APA journals did not share their data. An open science collaboration among 270 researchers from around the world used “high-powered designs and original materials” to conduct replications of 100 experimental and correlational studies published in three psychology journals. The analysis showed that the reported effect size declined by half. While 97% of the original studies reported their results as statistically significant, only 37% of replications showed statistically significant effects.⁹

The replicability problem elucidated by Amgen is such a generalizable phenomenon that the July 6, 2016 edition of the prestigious journal *Nature* was devoted to “Challenges in irreproducible research.” In an uncredited introduction to

the July 6, 2016 edition,¹⁰ the editors at *Nature* note that “There is growing alarm about results that cannot be reproduced.” In support of this contention, the editors conducted a survey and found that two-third of the 1576 researchers who responded thought that there was a reproducibility problem in science.^{10,11} A number of different factors have been hypothesized to underlie the low reproducibility rates including pressure to publish, selective reporting of research results, improper selection and application of statistical tests, fragility of protocol designs, inadequate description of methods, and failure to show all of the relevant data.¹¹ Despite these deficiencies, the American Conference of Governmental Industrial Hygienists (ACGIH) states that “... the TLV®-CS Committee preferably relies on published, peer reviewed literature that is available in the public domain.”¹² The primarily academic studies published in the peer-reviewed literature upon which ACGIH relies to determine threshold limit values (TLVs) rarely report raw data not already statistically transformed that are thus incalculable.

Bias against publication of negative results

The existence of a bias against the publication of negative results has been reported in many studies (Table 1). There are several factors incentivizing this bias. “When a dog bites a man that is not news, but when a man bites a dog that is news,”⁶⁶ encapsulates the natural tendency for the appreciation of novelty. Negative results are less novel than positive results. Second, given the underappreciation of negative results, researchers are less likely to invest the significant time required to write up and submit negative results for consideration for publication. Third, reviewers representing the best interest of peer-reviewed journals are cognizant of the desire to maintain a high citation impact factor⁶⁷ and negative results are less likely to be cited by other researchers in the field. Fourth, especially in the field of toxicology, important industrial chemicals are usually tested against a number of different biological activity endpoints.^{68,69} Negative results in these many different types of toxicological assays are not noteworthy, while positive results might either be relevant to the development of the toxicological profile of the chemical,^{68,69} construction of the safety data sheet,⁷⁰ or any number of potential regulatory actions. The existence of a difficult to quantitate but significant bias against the publication of negative results compounds the severity of the irreproducibility crisis.

Transparency in science

In 1992, our research group published the first manuscript in what was to become a long-standing collaboration with the late, great Corwin Hansch^{71–78} and following his passing at age 92 in 2011⁷⁹ with members of his research group at Pomona College.^{80–82} Dr Hansch was known as the “father of computer-assisted molecule design” for his

Table 1. Bias reported in scientific publications.^a

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³			
Field of study	Estimated magnitude of bias ^b	Source of bias	Citation
Biomedical clinical research	Moderate	Studies with significant or positive results were more likely to be published than those with non-significant or negative results, thereby confirming findings from a previous HTA (Health Technology Assessment) report. There was convincing evidence that outcome reporting bias exists and has an impact on the pooled summary in systematic reviews.	Song et al. ¹⁴
Psychology	Moderately high	The extreme view of the “file drawer problem” is that journals are filled with only about the 5% of the studies that show type-I errors, while the file drawers are filled with the 95% of the studies that show non-significant results. Quantitative procedures for computing the tolerance for filed and future null results are reported and illustrated, and the implications are discussed.	Rosenthal ¹⁵
Human genetics	Moderately high	Here, we have evaluated by meta-analysis 370 studies addressing 36 genetic associations for various outcomes of disease. We show that significant between-study heterogeneity (diversity) is frequent, and that the results of the first study correlate only modestly with subsequent research on the same association. The first study often suggests a stronger genetic effect than is found by subsequent studies. Both bias and genuine population diversity might explain why early association studies tend to overestimate the disease protection or predisposition conferred by a genetic polymorphism. We conclude that a systematic meta-analytic approach may assist in estimating population-wide effects of genetic risk factors in human disease.	Ioannidis et al. ¹⁶
Human genetics	Moderate	Maximal between-study variances were more likely to be recorded early in the 44 eligible meta-analyses of genetic associations than in the 37 meta-analyses of health-care interventions ($p = 0.013$). At the time of the first heterogeneity assessment, the most favorable-ever result in support of a specific association was more likely to appear than the least favorable-ever result (22 vs. 10, $p = 0.017$); the opposite was seen at the second heterogeneity assessment (15 vs. 5, $p = 0.031$). Such a sequence of extreme opposite results was not seen in the clinical trials meta-analyses. The estimated between-study variance decreased over time in genetic association studies ($p = 0.010$), but not in clinical trials ($p = 0.30$). In contrast to prospective trials, a rapid early sequence of extreme, opposite results is frequent in retrospective hypothesis-generating molecular research.	Ioannidis and Trikalinos ¹⁷
Psychology	Minimal	Some scientists attribute the “decline effect” to statistical self-correction of initially exaggerated outcomes, also known as “regression to the mean.” No one can be sure of this interpretation, or even test it, because we do not generally have access to ‘negative results’: experimental outcomes that were not noteworthy or consistent enough to pass peer review and be published.	Schooler ¹⁸
Genetic epidemiology	Moderate	Newly discovered true (non-null) associations often have inflated effects compared with the true effect sizes. The main reasons for this inflation are as follows: First, theoretical considerations prove that when true discovery is claimed based on crossing a threshold of statistical significance and the discovery study is underpowered, the observed effects are expected to be inflated. Second, flexible analyses coupled with selective reporting may inflate the published discovered effects. Third, effects may be inflated at the stage of interpretation due to diverse conflicts of interest. Fourth, discovered effects are not always inflated, and under some circumstances, may be deflated, for example, in the setting of late discovery of associations in sequentially accumulated overpowered evidence, in some types of misclassification from measurement error, and in conflicts causing reverse biases.	Ioannidis ¹⁹

(continued)

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³		
Field of study	Estimated magnitude of bias ^b	Source of bias
Biomedical clinical research	Moderately high	One hundred two trials with 122 published journal articles and 3736 outcomes were identified. Overall, 50% of efficacy and 65% of harm outcomes per trial were incompletely reported. Statistically significant outcomes had a higher odd of being fully reported compared with non-significant outcomes for both efficacy (pooled odds ratio, 2.4; 95% confidence interval [CI], 1.4–4.0) and harm (pooled odds ratio, 4.7; 95% CI, 1.8–12.0) data. In comparing published articles with protocols, 62% of trials had at least 1 primary outcome that was changed, introduced, or omitted. Eighty-six percent of survey responders (42/49) denied the existence of unreported outcomes despite clear evidence to the contrary. The reporting of trial outcomes is not only frequently incomplete but also biased and inconsistent with protocols. Published articles, as well as reviews that incorporate them, may therefore be unreliable and overestimate the benefits of an intervention. To ensure transparency, planned trials should be registered, and protocols should be made publicly available prior to trial completion.
Biomedical clinical research	Moderately high	Forty-eight trials were identified with 68 publications and 1402 outcomes. The median number of participants per trial was 299, and 44% of the trials were published in general medical journals. A median of 31% (10th–90th percentile range 5–67%) of outcomes measured to assess the efficacy of an intervention (efficacy outcomes) and 59% (0–100%) of those measured to assess the harm of an intervention (harm outcomes) per trial were incompletely reported. Statistically significant efficacy outcomes had a higher odd than non-significant efficacy outcomes of being fully reported (odds ratio 2.7; 95% confidence interval 1.5–5.0). Primary outcomes differed between protocols and publications for 40% of the trials. Selective reporting of outcomes frequently occurs in publications of high-quality government-funded trials.
Biomedical clinical research	Moderately high	Results of 519 trials with 553 publications and 10,557 outcomes were identified. Survey responders (response rate 69%) provided information on unreported outcomes but were often unreliable—for 32% of those who denied the existence of such outcomes there was evidence to the contrary in their publications. On average, over 20% of the outcomes measured in a parallel group trial were incompletely reported. Within a trial, such outcomes had a higher odds of being statistically non-significant compared with fully reported outcomes (odds ratio 2.0 (95% confidence interval 1.6 to 2.7) for efficacy outcomes; 1.9 (1.1 to 3.5) for harm outcomes). The most commonly reported reasons for omitting efficacy outcomes. Incomplete reporting of outcomes within published articles of randomized trials is common and is associated with statistical non-significance. The medical literature therefore represents a selective and biased subset of study outcomes, and trial protocols should be made publicly available.
Medical/pharmacological researchers	Moderate	The frequency with which scientists fabricate and falsify data or commit other forms of scientific misconduct is a matter of controversy. Many surveys have asked scientists directly whether they have committed or know of a colleague who committed research misconduct, but their results appeared difficult to compare and synthesize. This is the first meta-analysis of these surveys. To standardize outcomes, the number of respondents who recalled at least one incident of misconduct was calculated for each question, and the analysis was limited to behaviors that distort scientific knowledge: fabrication, falsification, “cooking” of data, and so on. Meta-regression showed self-reports surveys, surveys using the words “falsification” or “fabrication,” and mailed surveys yielded lower percentages of misconduct. When these factors were controlled for, misconduct was reported more frequently by medical/pharmacological researchers than others. Considering that these surveys ask sensitive questions and have other limitations, it appears likely that this is a conservative estimate of the true prevalence of scientific misconduct.

(continued)

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³		
Field of study	Estimated magnitude of bias ^b	Source of bias
Psychology	Moderately high	Cases of clear scientific misconduct have received significant media attention recently, but less flagrantly questionable research practices may be more prevalent and, ultimately, more damaging to the academic enterprise. Using an anonymous elicitation format supplemented by incentives for honest reporting, we surveyed over 2000 psychologists about their involvement in questionable research practices. The impact of truth-telling incentives on self-admissions of questionable research practices was positive, and this impact was greater for practices that respondents judged to be less defensible. Combining three different estimation methods, we found that the percentage of respondents who have engaged in questionable practices was surprisingly high. This finding suggests that some questionable practices may constitute the prevailing research norm.
Biomedical and life-science research	Moderate	A detailed review of all 2047 biomedical and life-science research articles indexed by PubMed as retracted on May 3, 2012 revealed that only 21.3% of retractions were attributable to error. In contrast, 67.4% of retractions were attributable to misconduct, including fraud or suspected fraud (43.4%), duplicate publication (14.2%), and plagiarism (9.8%). Incomplete, uninformative or misleading retraction announcements have led to a previous underestimation of the role of fraud in the ongoing retraction epidemic. The percentage of scientific articles retracted because of fraud has increased almost 10-fold since 1975. Retractions exhibit distinctive temporal and geographic patterns that may reveal underlying causes
Biomedical clinical research	Moderate	There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias.
Psychology	Moderate	Humans make decisions with uncertainty. Evaluations (scientific and non-scientific) are subject to unconscious heuristics and biases. The author shows that our decisions about how to design experiments and how to interpret the results are subject to bias, sometimes with serious results. For the person who believes that professional scientists can somehow easily “control” bias away, Rosenthal presents an important counterargument. The problems documented by the author have been taught previously in many disciplines; however, these teachings are not studied today. These teachings may have gone out of fashion, but most certainly the problems not gone away.
Animal behavior research	Minimal	Authors reviewed several hundred published articles from 1970 to 2010 in five leading animal behavior journals and found that these two methods for minimizing or eliminating bias were rarely reported (<10% of articles reviewed). In contrast, a journal focusing on human infant behavior research was far more rigorous in incorporating methods to avoid bias (>80% of articles reviewed). The lack of reporting attempts to minimize bias in animal behavior studies suggests that, at best, many researchers view blind analyses of data or inter-rater reliability as unimportant components of research or, if carried out, unnecessary to report in a manuscript. At worst, the lack of reporting attempts to minimize bias suggests that some published behavioral research may be unreliable. We are aware of constraints imposed by fieldwork and data collecting issues that make blind data comparisons or inter-rater reliability assessments sometimes difficult or unfeasible. However, given that research ethicists often emphasize the fundamental importance of trust and transparency in science, we urge authors, reviewers, and editors of manuscripts to ensure that at least one of these two methods of reducing and reporting observer bias occurs.

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³		
Field of study	Estimated magnitude of bias ^b	Source of bias
Physical sciences, biological sciences, social sciences	Moderate (depending on the science tested)	<p>Fanelli analyzed 2434 articles published in all disciplines and that declared to have tested a hypothesis. It was determined how many articles reported a "positive" (full or partial) or "negative" support for the tested hypothesis. If the hierarchy hypothesis is correct, then researchers in "softer" sciences should have fewer constraints to their conscious and unconscious biases, and therefore report more positive outcomes. Results confirmed the predictions at all levels considered: discipline, domain and methodology broadly defined. Controlling for observed differences between pure and applied disciplines, and between articles testing one or several hypotheses, the odds of reporting a positive result were around five times higher among articles in the disciplines of psychology and psychiatry and economics and business compared to space science, 2.3 times higher in the domain of social sciences compared to the physical sciences, and 3.4 times higher in studies applying behavioral and social methodologies on people compared to physical and chemical studies on nonbiological material. In all comparisons, biological studies had intermediate values. These results suggest that the nature of hypotheses tested and the logical and methodological rigor employed to test them vary systematically across disciplines and fields, depending on the complexity of the subject matter and possibly other factors (e.g. a field's level of historical and/or intellectual development). On the other hand, these results support the scientific status of the social sciences against claims that they are completely subjective, by showing that, when they adopt a scientific approach to discovery, they differ from the natural sciences only by a matter of degree.</p> <p>There has been a number of high-profile academic fraud cases in China recently underscoring the problems of an academic-evaluation system that places disproportionate emphasis on publications. Chinese universities often award cash prizes, housing benefits or other perks on the basis of high-profile publications, and the pressure to publish seems to be growing. The journal Acta Crystallographica Section E has retracted 70 published crystal structures that they allege are fabrications by researchers at Jiangangshan University in Jiangxi province. Further retractions, the editors say, are likely.</p>
Crystallography	Minimal	Qiu ³⁰
Physical sciences, biological sciences, social sciences	Minimal	<p>How does publication pressure in modern-day universities affect the intrinsic and extrinsic rewards in science? Using a worldwide survey among demographers in developed and developing countries, the authors show that the large majority perceive the publication pressure as high, but more so in Anglo-Saxon countries and to a lesser extent in Western Europe. However, scholars see both the pros (upward mobility) and cons (excessive publication and uncitedness, neglect of policy issues, etc.) of the so-called publish-or-perish culture. By measuring behavior in terms of reading and publishing, and perceived extrinsic rewards and stated intrinsic rewards of practicing science, it turns out that publication pressure negatively affects the orientation of demographers towards policy and knowledge sharing. There are no signs that the pressure affects reading and publishing outside the core discipline.</p> <p>There is increasing concern that the genetic literature may be distorted by various biases, such as publication bias, which may lead to a misleading impression of the strength of evidence for a putative gene-disease association. Meta-analysis is one means by which a more accurate estimate of the strength of evidence for such association may be obtained, as well as offering a means by which potential biases may be identified. The authors present evidence that the location where a study is conducted is associated with the degree to which it represents an over-estimate of the true effect size, as subsequently estimated using meta-analytical techniques. The results indicate that studies published in North America may represent a relative over-estimate of the true effect size, compared to those published in Europe or elsewhere.</p>
Genetics	Moderate	van Dalen and Henkens ³¹ Munafò et al. ³²

(continued)

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³		
Field of study	Estimated magnitude of bias ^b	Source of bias
Industrial relations research	Minimal	This article develops and applies several meta-analytic techniques to investigate the presence of publication bias in industrial relations research, specifically in the union-productivity effects literature. Publication bias arises when statistically insignificant results are suppressed or when results satisfying prior expectations are given preference. Like most fields, research in industrial relations is vulnerable to publication bias. Unlike other fields such as economics, there is no evidence of publication bias in the union-productivity literature, as a whole. However, there are pockets of publication selection, as well as negative autoregression, confirming the controversial nature of this area of research. Meta-regression analysis reveals evidence of publication bias (or selection) among U.S. studies.
Physical sciences, biological sciences, social sciences	Moderate	Concerns that the growing competition for funding and citations might distort science are frequently discussed but have not been verified directly. Of the hypothesized problems, perhaps the most worrying is a worsening of positive-outcome bias. A system that disfavors negative results not only distorts the scientific literature directly but might also discourage high-risk projects and pressure scientists to fabricate and falsify their data. This study analyzed over 4600 articles published in all disciplines between 1990 and 2007, measuring the frequency of articles that, having declared to have “tested” a hypothesis, reported a positive support for it. The overall frequency of positive supports has grown by over 22% between 1990 and 2007 with significant differences between disciplines and countries. The increase was stronger in the social and some biomedical disciplines. The United States had published, over the years, significantly fewer positive results than Asian countries (and particularly Japan) but more than European countries (and in particular the United Kingdom). Methodological artifacts cannot explain away these patterns, which support the hypotheses that research is becoming less pioneering and/or that the objectivity with which results are produced and published is decreasing.
Physical sciences, biological sciences, social sciences	Moderate	The growing competition and “publish or perish” culture in academia might conflict with the objectivity and integrity of research, because it forces scientists to produce “publishable” results at all costs. Papers are less likely to be published and to be cited if they report “negative” results (results that fail to support the tested hypothesis). Therefore, if publication pressures increase scientific bias, the frequency of “positive” results in the literature should be higher in the more competitive and “productive” academic environments. This study verified this hypothesis by measuring the frequency of positive results in a large random sample of papers with a corresponding author based in the United States. Across all disciplines, papers were more likely to support a tested hypothesis if their corresponding authors were working in states that, according to NSF data, produced more academic papers per capita. The size of this effect increased when controlling for state’s per capita R&D expenditure and for study characteristics that previous research showed to correlate with the frequency of positive results, including discipline and methodology. Although the confounding effect of institutions’ prestige could not be excluded (researchers in the more productive universities could be the most clever and successful in their experiments), these results support the hypothesis that competitive academic environments increase not only scientists’ productivity but also their bias. The same phenomenon might be observed in other countries where academic competition and pressures to publish are high.

(continued)

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³		
Field of study	Estimated magnitude of bias ^b	Source of bias
Psychology	Minimal	In this article, we accomplish two things. First, we show that despite empirical psychologists' nominal endorsement of a low rate of false-positive findings (≤ 0.05), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis. Second, we suggest a simple, low-cost, and straightforwardly effective disclosure-based solution to this problem. The solution involves six concrete requirements for authors and four guidelines for reviewers, all of which impose a minimal burden on the publication process.
Psychology	Minimal	Empirical replication has long been considered the final arbiter of phenomena in science, but replication is undermined when there is evidence for publication bias. Evidence for publication bias in a set of experiments can be found when the observed number of rejections of the null hypothesis exceeds the expected number of rejections. Application of this test reveals evidence of publication bias in two prominent investigations from experimental psychology that have purported to reveal evidence of extrasensory perception and to indicate severe limitations of the scientific method. The presence of publication bias suggests that those investigations cannot be taken as proper scientific studies of such phenomena, because critical data are not available to the field. Publication bias could partly be avoided if experimental psychologists started using Bayesian data analysis techniques.
Psychology	Minimal	The article examines a sample of 91 recent meta-analyses published in American Psychological Association and Association for Psychological Science journals and the methods used in these analyses to identify and control for publication bias. Of the 91 studies analyzed, 64 (70%) made some effort to analyze publication bias, and 26 (41%) reported finding evidence of bias. Approaches to controlling publication bias were heterogeneous among studies. Of these studies, 57 (63%) attempted to find unpublished studies to control for publication bias. Nonetheless, those studies that included unpublished studies were just as likely to find evidence for publication bias as those that did not. Authors of meta-analyses themselves were overrepresented in unpublished studies acquired, as compared with published studies, suggesting that searches for unpublished studies may increase rather than decrease some sources of bias. A subset of 48 meta-analyses for which study sample sizes and effect sizes were available was further analyzed with a conservative and newly developed tandem procedure of assessing publication bias. Results indicated that publication bias was worrisome in about 25% of meta-analyses.
Psychology	Minimal	A recent set of articles in <i>Perspectives on Psychological Science</i> discussed inflated correlations between brain measures and behavioral criteria when measurement points (voxels) are deliberately selected to maximize criterion correlations (the target article was Vul et al., 2009). However, closer inspection reveals that this problem is only a special symptom of a broader methodological problem that characterizes all paradigmatic research, not just neuroscience. Researchers not only select voxels to inflate effect size, they also select stimuli, task settings, favorable boundary conditions, dependent variables and independent variables, treatment levels, moderators, mediators, and multiple parameter settings in such a way that empirical phenomena become maximally visible and stable. In general, paradigms can be understood as conventional setups for producing idealized, inflated effects. Although the feasibility of representative designs is restricted, a viable remedy lies in a reorientation of paradigmatic research from the visibility of strong effect sizes to genuine validity and scientific scrutiny.

(continued)

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³			
Field of study	Estimated magnitude of bias ^b	Source of bias	Citation
Psychology	Minimal	In order to study the prevalence, nature (direction), and causes of reporting errors in psychology, we checked the consistency of reported test statistics, degrees of freedom, and <i>p</i> values in a random sample of high- and low-impact psychology journals. In a second study, we established the generality of reporting errors in a random sample of recent psychological articles. Our results, on the basis of 281 articles, indicate that around 18% of statistical results in the psychological literature are incorrectly reported. Inconsistencies were more common in low-impact journals than in high-impact journals. Moreover, around 15% of the articles contained at least one statistical conclusion that proved, upon recalculation, to be incorrect; that is, recalculation rendered the previously significant result insignificant, or vice versa. These errors were often in line with researchers' expectations.	Bakker and Wicherts ⁴⁰
Health services research	Moderate	Health services research often involves multilevel data, with participants nested in families, clinicians, case managers, community care programs, HMOs, or countries. Longitudinal studies are also common, and studies may be both multilevel and longitudinal. Flexible and efficient analytic methods for data that are multilevel, longitudinal, or both are increasingly accessible to researchers. However, the implications of these methods for research design have not yet been well explored. As a result, researchers remain quite uninformed about crucial design decisions. The use of multilevel modeling to investigate organizational phenomena is rapidly increasing. Unfortunately, little advice is readily available for organizational researchers attempting to determine statistical power when using multilevel models or when determining sample sizes for each level that will maximize statistical power. This often leads to bias in studies. This article presents an introduction to statistical power in multilevel models. The unique factors influencing power in multilevel models and calculations for estimating power for simple fixed effects, variance components, and cross-level interactions are presented.	Raudenbush et al. ⁴¹
Physical sciences, biological sciences, social sciences	Moderate	MLM is a key method used by social and personality psychologists in their research. The rationale for MLM is important to understand, especially as applied to the scientific fields. Bias can be introduced into studies in a variety of way. MLM can assist researchers to avoid unnecessary bias by illustrating how the method can and should be applied in research.	Scherbaum and Ferrer ⁴²
Psychology	Moderate	Manuscript selection bias is the selective publication of manuscripts based on study characteristics other than quality indicators. One reason may be a perceived editorial bias against the research from less-developed world. We aimed to compare the methodological quality and statistical appeal of trials from countries with different development status and to determine their association with the journal impact factors and language of publication. Four hundred records of clinical trials were explored by country economic status from 1993 to 2003. Country income had an inverse linear association with the presence of randomization (χ^2 for trend = 5.6, <i>p</i> = 0.02) and a direct association with the use of blinding (χ^2 for trend = 6.9, <i>p</i> = 0.008); although in low-income countries the probability of blinding was increased from 36% in 1993 to 46% in 2003. In 1993, the results of 68% of high-income trials and 64.7% of other groups were statistically significant; but in 2003, they were 66% and 82%, respectively. Study sample size and income were the only significant predictors of journal impact factor. The impact of country development on manuscript selection bias is considerable and may be increasing over time. It seems that one reason may be more stringent implementation of the guidelines for improving the reporting quality of trials on developing world researchers. Another reason may be the presumptions of the researchers from developing world about the editorial bias against their nationality.	Nezlek ⁴³
Biomedical Clinical Research	Moderate		Yousefi-Nooraie et al. ⁴⁴

(continued)

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³			
Field of study	Estimated magnitude of bias ^b	Source of bias	Citation
Biomedical clinical research	Moderate	Many authors believe that there are biases in scientific publications. Editorial biases include publication bias, which refers to those situations where the results influence the editor's decision, and editorial bias refers to those situations where factors related to authors or their environment influence the decision. This article includes an analysis of the situation of editorial biases. One bias is where mainly articles with positive results are accepted, as opposed to those with negative results. Another is latent bias, where positive results are published before those with negative results. In order to examine editorial bias, this article analyses the influence of where the article originated; the country or continent, academic center of origin, belonging to cooperative groups, and the maternal language of the authors. The article analyses biases in the editorial process in the publication of funded clinical trials. Editorial biases exist. Authors, when submitting their manuscript, should analyze different journals and decide where their article will receive adequate treatment. Reviewers increasingly are asked to review manuscripts from outside their own country, but whether they are more likely to recommend acceptance of such manuscripts is not known. To assess whether US reviewers or non-US reviewers evaluate manuscripts differently, depending on whether the manuscripts are submitted from outside the United States or from the United States. A retrospective analysis of all original submissions received by Gastroenterology in 1995 and 1996. Reviewers ranked manuscripts in four decision categories: accept, provisionally accept, reject with resubmission, or reject. The percentage of non-US manuscripts placed in each decision category by United States ($n = 2355$) and non-US reviewers ($n = 1297$) was nearly identical ($p = 0.31$). However, US reviewers recommended acceptance of articles submitted by US authors more often than did non-US reviewers ($p = 0.001$). Non-US reviewers ranked US articles slightly more favorably than non-US articles ($p = 0.09$), while US reviewers ranked US articles much more favorably ($p = 0.001$). Reviewers from the United States and outside the United States evaluate non-US papers similarly and evaluate papers submitted by US authors more favorably, with US reviewers having a significant preference for US papers.	Matías-Guiu and García-Ramos ⁴⁵
Medical research	Moderate	Two hundred and nine manuscripts were reviewed. Commercial funding was not found to be associated with a positive study outcome ($p = 0.668$). Studies with a positive outcome were no more likely to be published than were those with a negative outcome ($p = 0.410$). Studies with a negative outcome were of higher quality ($p = 0.003$) and included larger sample sizes ($p = 0.05$). Commercially funded ($p = 0.027$) and US-based ($p = 0.020$) studies were more likely to be published, even though those studies were not associated with higher quality, larger sample sizes, or lower levels of evidence ($p = 0.24-0.79$). Commercially funded studies submitted for review were not more likely to conclude with a positive outcome than were nonfunded studies, and studies with a positive outcome were no more likely to be published than were studies with a negative outcome. These findings contradict those of most previous analyses of published (rather than submitted) research. Commercial funding and the country of origin predict publication following peer review beyond what would be expected on the basis of study quality. Studies with a negative outcome, although seemingly superior in quality, fared no better than studies with a positive outcome in the peer-review process; this may result in inflation of apparent treatment effects when the published literature is subjected to meta-analysis.	Link ⁴⁶
Medical research	Moderate	Two hundred and nine manuscripts were reviewed. Commercial funding was not found to be associated with a positive study outcome ($p = 0.668$). Studies with a positive outcome were no more likely to be published than were those with a negative outcome ($p = 0.410$). Studies with a negative outcome were of higher quality ($p = 0.003$) and included larger sample sizes ($p = 0.05$). Commercially funded ($p = 0.027$) and US-based ($p = 0.020$) studies were more likely to be published, even though those studies were not associated with higher quality, larger sample sizes, or lower levels of evidence ($p = 0.24-0.79$). Commercially funded studies submitted for review were not more likely to conclude with a positive outcome than were nonfunded studies, and studies with a positive outcome were no more likely to be published than were studies with a negative outcome. These findings contradict those of most previous analyses of published (rather than submitted) research. Commercial funding and the country of origin predict publication following peer review beyond what would be expected on the basis of study quality. Studies with a negative outcome, although seemingly superior in quality, fared no better than studies with a positive outcome in the peer-review process; this may result in inflation of apparent treatment effects when the published literature is subjected to meta-analysis.	Lynch ⁴⁷

(continued)

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³				
Field of study	Estimated magnitude of bias ^b	Source of bias		Citation
Research in animal behavior	Negligible	Confirmation bias is a tendency of people to interpret information in a way that confirms their expectations. A long-recognized phenomenon in human psychology, confirmation bias can distort the results of a study and thus reduce its reliability. While confirmation bias can be avoided by conducting studies blind to treatment groups, this practice is not always used. Authors conducted a meta-analysis, using studies on nestmate recognition in ants, to compare the outcomes of studies that were conducted blind with those that were not. Nestmate recognition studies typically perform intra- and inter-colony aggression assays, with the a priori expectation that there should be little or no aggression among nestmates. Aggressive interactions between ants can include subtle behaviors such as mandible flaring and recoil, which can be hard to quantify, making these types of assays prone to confirmation bias. The survey revealed that only 29% of our sample of 79 studies were conducted blind. These studies were more likely to report aggression among nestmates if they were conducted blind (73%) than if they were not (21%). Moreover, it was found that the effect size between nestmate and non-nestmate treatment means is significantly lower in experiments conducted blind than those in which colony identity is known (1.38 vs. 2.76). The implications of the impact of confirmation bias for research that attempts to obtain quantitative syntheses of data from different studies are discussed.		van Wilgenburg and Elgar ⁴⁸
Physics	Negligible	The beliefs of physicists can bias their results toward their expectations in a number of ways. The authors survey a variety of historical cases of expectation bias in observations, experiments, and calculations.		Jeng ⁴⁹
Physical sciences, biological sciences, social sciences	Minimal	The authors failed to observe any “US effect” at all among genetic research, thus contradicting observations made by an independent, smaller study in genetic epidemiology. This suggests that the prevalence of this and similar patterns needs to be assessed field by field. However, previous observations suggest that a US propensity to report positive and statistically significant results is not limited to particular methodologies and may cut across many disciplines.		Fanelli ⁵⁰
Biomedical clinical research	Moderate	Statistical tests of heterogeneity and bias, in particular publication bias, are very popular in meta-analyses. These tests use statistical approaches whose limitations are often not recognized. Moreover, it is often implied with inappropriate confidence that these tests can provide reliable answers to questions that are not of statistical nature. Statistical heterogeneity is only a correlate of clinical and pragmatic heterogeneity and the correlation may sometimes be weak. Similarly, statistical signals may hint to bias, but seen in isolation they cannot fully prove or disprove bias in general, let alone specific causes of bias, such as publication bias in particular. Both false-positive and false-negative signals of heterogeneity and bias can be common, and their prevalence may be anticipated based on some rational considerations. The author discusses the major common challenges and flaws that emerge in using and interpreting statistical tests of heterogeneity and bias in meta-analyses. I discuss misinterpretations that can occur at the level of statistical inference, clinical/pragmatic inference, and specific cause attribution.		Ioannidis ⁵¹
Physical sciences, biological sciences, social sciences	Negligible	By integrating and translating the current methodological and statistical work into a practical guide, the authors of this text provide readers with a state-of-the-art introduction to the various approaches to doing meta-analysis and provide the readers with information on various sources of bias.		Lipsey and Wilson ⁵²

(continued)

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³		
Field of study	Estimated magnitude of bias ^b	Source of bias
Physical sciences, biological sciences, social sciences	Negligible	The metafor package provides functions for conducting meta-analyses in R. The package includes functions for fitting the meta-analytic fixed- and random-effects models and allows for the inclusion of moderators variables (study-level covariates) in these models. Meta-regression analyses with continuous and categorical moderators can be conducted in this way. Functions for the Mantel-Haenszel and Peto's one-step method for meta-analyses of 2×2 table data are also available. Finally, the package provides various plot functions (for example, for forest, funnel, and radial plots) and functions for assessing the model fit, for obtaining case diagnostics, and for tests of publication bias.
Biomedical clinical research	Moderate	Clinicians commonly misinterpret systematic review abstracts: a recent study showed many arrived at incorrect conclusions, and only 62% correctly identified the direction of the main effect. Interpreting numerical results requires statistical knowledge that many clinicians lack. To ensure correct interpretation, abstracts should give the direction and size of effects both in words and numerically. Because systematic reviews are important and widely used summaries of primary research. The authors decided to examine a sample of systematic review abstracts to assess the nature and extent of any deficiencies in reporting. For 42% of abstracts of systematic reviews, the direction of the main effect either could not be determined or needed to be inferred. Statistical uncertainty was also poorly reported: 24% of abstracts reported neither a confidence interval nor a p-value. Because many readers can only, or will only, read a systematic review's abstract, clear presentation of the main results is vital. Although abstracts should present estimates of effect and confidence intervals, interpretation of the results should not require statistical knowledge. Given the high level of innumeracy among journal readers, the main results should be presented in both words and numbers. Although replication in a wider sample of journals is desirable, the apparent poor quality of systematic review abstracts deserves attention from authors, reviewers, and journal editors.
Biomedical clinical research	Moderate	ORB in randomized trials has been identified as a threat to the validity of systematic reviews. Previous work highlighting this problem is limited to considering a single primary review outcome. The aim of this study was to assess ORB across all efficacy outcomes in the Cochrane systematic reviews of cystic fibrosis. Systematic reviews of interventions for cystic fibrosis published on the Cochrane Library by the Cochrane Cystic Fibrosis and Genetic Disorders Group before 2010 were assessed for discrepancies in outcomes between review protocol and full review. ORB in eligible trials was also assessed for all efficacy review outcomes. Two authors independently classified each outcome using a nine-point classification system developed by the Outcome Reporting Bias in Trials study. These classifications were used to inform the assessment of the risk of bias for selective outcome reporting for each trial. Forty-six Cochrane cystic fibrosis systematic reviews were included. The median number of primary outcomes, number of trials, and participants per trial in the reviews were 3 (IQR 2, 3), 4 (IQR 2, 8), and 21 (IQR 14, 41), respectively. Eighteen reviews (39%, 18/46) had a discrepancy in outcomes between protocol and full review. Thirty-seven reviews were eligible to be included in the ORB assessment. When considering review primary outcomes and all review outcomes, ORB was suspected in at least one trial in 86% and 100%, respectively. Assessment of ORB within a systematic review of a single primary outcome underestimates the risk of ORB in comparison with the assessment of multiple primary and secondary outcomes. ORB in trials is highly prevalent within systematic reviews of cystic fibrosis when assessed across all outcomes. This could be reduced by the development of a core outcome set for trials and systematic reviews in cystic fibrosis.

(continued)

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³			
Field of study	Estimated magnitude of bias ^b	Source of bias	Citation
Biomedical clinical research	Moderate	<p>Outcome reporting bias (ORB) has been identified as a threat to the validity of evidence-based-medicine. Trial outcomes with statistically significant results are more likely to be published than non-significant outcomes. The ORBIT study investigated ORB in Cochrane reviews but only considered primary outcomes and included only two CFGD reviews. The prevalence and impact of ORB in CFGD reviews is unknown. Eighty-two reviews published prior to 2010 were included; 21 identified no randomized controlled trials. Sixty-one reviews were considered further containing 405 included trials and 21 trials excluded due to “no relevant outcome data.” Between protocol and review, there were three reviews that upgraded secondary outcomes to primary, 15 downgraded primary outcomes to secondary, two reviews included outcomes that were not in the protocol and two reviews excluded outcomes that were originally in the protocol. Nine reviews removed secondary outcomes between protocol and review and six reviews added secondary outcomes. Sixteen review protocols did not distinguish between primary and secondary outcomes. The assessment of ORB is ongoing at present. ORB is a problem in all areas of research. A core set of outcomes for genetic disorders will be an important step in reducing ORB and standardizing outcome measures in these clinical areas.</p>	Dwan et al. ⁵⁷
Biomedical clinical research	Minimal	<p>Cochrane abstracts are the most frequently accessed and used part of a Cochrane review. In randomized trials, evidence shows that authors don't always report the primary outcome in the abstract and are more likely to report a clinically or statistically significant outcome. This may also be the case in abstracts of systematic reviews. To assess whether reporting of outcomes is consistent between the full text and abstract of Cochrane reviews, the authors included all new reviews published in Issue 4, 2009 of The Cochrane Database of Systematic Reviews, where the primary outcome(s) were clearly stated in the full text and a meta-analysis had been conducted ($n = 64$); they excluded nonintervention reviews. The median number of primary outcomes per review was two (range 1–10). Only 44 (69%) reviews reported all primary outcomes from the text in the abstract. Twelve (19%) reported only some of the primary outcomes in the abstract, compared with the full text, and eight (13%) failed to report any primary outcomes in the abstract. Of the 56 (88%) reviews that reported one or more primary outcomes in the abstract, only four (7%) stated this was a primary outcome and only eight (14%) reported the relative and absolute effect size and 95% confidence interval or p-value. In 33 reviews (59%), there was no absolute effect size given, in 11 (20%), the result was only stated as “significant” or “not significant,” in three (5%), only a NNT was given, and in one (2%), only the relative effect size was stated with no p-value or confidence interval. The preliminary findings suggest evidence of incomplete and selective reporting in abstracts of Cochrane reviews.</p>	Hopewell and Beller ⁵⁸

(continued)

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013)¹³

Field of study	Estimated magnitude of bias ^b	Source of bias	Citation
Biomedical clinical research	Minimal	<p>The purpose of prespecifying primary outcomes in the systematic review process is to define the most clinically relevant outcomes and to protect against bias. Adding, omitting, or changing review outcomes once the protocol is published can result in bias. The objective was to investigate the discrepancies between primary outcomes listed in protocols and the subsequent reviews published in The Cochrane Library. To identify non-publication of review protocols and ill-defined primary outcomes. The authors examined new systematic reviews published between Issue 4, 2006 and Issue 2, 2007 of The Cochrane Library. For each review, discrepancies between the primary outcome(s) listed in the review protocol and the review itself were identified by a statistician and lead review authors were contacted to provide reasons for the discrepancies. A total of 297 reviews were in the study cohort. For the primary outcome measures, there were 49 (16%) reviews that disagreed with the primary outcome(s) specified in the protocol; nine including at least one new primary outcome not specified in the protocol, five excluding at least one primary outcome specified in the protocol and 35 where a protocol outcome was either upgraded or downgraded to a primary/secondary review outcome, respectively. A further 24 (8%) reviews had no protocol registered on the Cochrane Library while 14 (5%) reviews could not be assessed for outcome reporting bias due to poor primary outcome definition. Discrepancies between primary outcomes specified in the review and protocol are common. The most common reasons for these discrepancies were: (1) recommendation by editors/peer reviewers; and (2) recognition of the importance of the outcome before/after reading the results to the included trials. The seriousness of bias arising from the reasons provided by the review authors for such discrepancies, non-registration of review protocols and ill-defined outcome definition will be discussed. Adding, omitting, or changing outcomes after a systematic review protocol is published can result in bias because it increases the potential for unacknowledged or post hoc revisions of the planned analyses. The main objective of this study was to look for discrepancies between primary outcomes listed in protocols and in the subsequent completed reviews published on the Cochrane Library. A secondary objective was to quantify the risk of bias in a set of meta-analyses where discrepancies between outcome specifications in protocols and reviews were found. New reviews from three consecutive issues of the Cochrane Library were assessed. For each review, the primary outcome(s) listed in the review protocol and the review itself were identified and review authors were contacted to provide reasons for any discrepancies. Over a fifth (64/288, 22%) of protocol/review pairings were found to contain a discrepancy in at least one outcome measure, of which 48 (75%) were attributable to changes in the primary outcome measure. Where lead authors could recall a reason for the discrepancy in the primary outcome, there was found to be potential bias in nearly a third (8/28, 29%) of these reviews, with changes being made after knowledge of the results from individual trials. Only 4 (6%) of the 64 reviews with an outcome discrepancy described the reason for the change in the review, with no acknowledgment of the change in any of the eight reviews containing potentially biased discrepancies. Outcomes that were promoted in the review were more likely to be significant than if there was no discrepancy (relative risk 1.66, 95% CI (1.10, 2.49), $p = 0.02$). In a review, making changes after seeing the results for included studies can lead to biased and misleading interpretation if the importance of the outcome (primary or secondary) is changed on the basis of those results. Our assessment showed that reasons for discrepancies with the protocol are not reported in the review, demonstrating an under-recognition of the problem. Complete transparency in the reporting of changes in outcome specification is vital; systematic reviewers should ensure that any legitimate changes to outcome specification are reported with reason in the review.</p>	Kirkham et al. ⁵⁹
Biomedical clinical research	Minimal	<p>Adding, omitting, or changing outcomes after a systematic review protocol is published can result in bias because it increases the potential for unacknowledged or post hoc revisions of the planned analyses. The main objective of this study was to look for discrepancies between primary outcomes listed in protocols and in the subsequent completed reviews published on the Cochrane Library. A secondary objective was to quantify the risk of bias in a set of meta-analyses where discrepancies between outcome specifications in protocols and reviews were found. New reviews from three consecutive issues of the Cochrane Library were assessed. For each review, the primary outcome(s) listed in the review protocol and the review itself were identified and review authors were contacted to provide reasons for any discrepancies. Over a fifth (64/288, 22%) of protocol/review pairings were found to contain a discrepancy in at least one outcome measure, of which 48 (75%) were attributable to changes in the primary outcome measure. Where lead authors could recall a reason for the discrepancy in the primary outcome, there was found to be potential bias in nearly a third (8/28, 29%) of these reviews, with changes being made after knowledge of the results from individual trials. Only 4 (6%) of the 64 reviews with an outcome discrepancy described the reason for the change in the review, with no acknowledgment of the change in any of the eight reviews containing potentially biased discrepancies. Outcomes that were promoted in the review were more likely to be significant than if there was no discrepancy (relative risk 1.66, 95% CI (1.10, 2.49), $p = 0.02$). In a review, making changes after seeing the results for included studies can lead to biased and misleading interpretation if the importance of the outcome (primary or secondary) is changed on the basis of those results. Our assessment showed that reasons for discrepancies with the protocol are not reported in the review, demonstrating an under-recognition of the problem. Complete transparency in the reporting of changes in outcome specification is vital; systematic reviewers should ensure that any legitimate changes to outcome specification are reported with reason in the review.</p>	Kirkham et al. ⁶⁰

(continued)

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³		
Field of study	Estimated magnitude of bias ^b Source of bias	Citation
Biomedical clinical research	<p>Recent studies have shown that final reports do not always correspond to what authors originally set out to study in protocols. This phenomenon, which has only been assessed in primary studies, can affect the quality of reporting and can generate selective reporting of results. The objective was to assess the agreement between the outcomes stated in protocols and those reported in systematic reviews. Original protocols and the subsequent full reports of Cochrane systematic reviews (thereafter called "Cochrane protocol review pairs" published in Issues 2, 3, and 4 of 2005 and Issue 1 of 2006 were first identified and then a random sample of 186 CPRPs was drawn from them. For each CPRP, agreement between protocols and final reports was assessed. A CPRP was classified as "in agreement" if the outcomes were present in both protocols and reviews or when the reviewers report that an outcome was not available in primary studies. Agreement between classification of primary/secondary outcomes was also assessed. A preliminary analysis based on 19 CPRPs has shown that: four (21%) did not provide the protocol, in five (33%), disagreement between the protocol and the final report was observed, and in two (13%), primary and secondary outcomes were switched. Results based on a larger sample will be provided at the Colloquium. One of the stated advantages in Cochrane systematic reviews is the commitment of authors to prepare and publish a protocol before a systematic review is carried out. Therefore, it is important in The Cochrane Collaboration to assess whether this is happening in practice and this study was set out for this.</p>	Parmelli et al. ⁶¹
Biomedical clinical research	<p>The protocol is considered an important element when a study is conducted. First, it makes a study replicable by independent researchers; second, it provides the researchers with a guide for the conduct of the study; and finally, it protects from biases due to prespecifying hypotheses of the research. Evidence suggests that there are discrepancies between protocols and final studies in terms of outcomes in primary literature. The objective was to evaluate whether discrepancies exist between outcomes and SR protocols in Cochrane SRs. This analysis is based on a sample of 60 SRs. Outcomes of AE were reported in 39 SRs (65%). The disagreement with protocols was observed in 28 SRs (47%) for effectiveness outcomes and in 9 SRs (23%) for AE outcomes. Eleven SRs (18%) included at least a new outcome and 17% excluded at least 1 outcome. In those 18 SRs (30%) in which a change in type of outcome was observed, 6 (10%) upgraded at least 1 outcome, 9 (15%) downgraded at least one outcome, and 3 (5%) did both. Similar results were observed for AE outcomes. The median number of outcomes per protocol was 9 and it was 11 in SRs (range 2–45). We are currently analyzing the association between type of change and statistical significance. The results will be presented considering a larger sample size. The preliminary results show that discrepancies between protocols and SRs are not uncommon. Most discrepancies are due to change in the typologies of outcomes: many outcomes that were primary in the protocols became secondary, or undefined, in the SRs.</p>	Parmelli et al. ⁶²

(continued)

Table 1. (continued)

Bias reported in scientific publications (Fanelli and Ioannidis, 2013) ¹³		
Field of study	Estimated magnitude of bias ^b	Source of bias
Biomedical clinical research	Minimal	Publication of research protocols minimizes bias by explicitly stating a priori hypotheses and methods without prior knowledge of results. The authors conducted a retrospective comparative study to assess the extent to which the content of published Cochrane reviews had changed compared with their previously published protocols and to assess any potential impact these changes may have had in introducing bias to the study. They identified previously published protocols for new Cochrane reviews appearing in The Cochrane Library; 2000, issue 3. The texts of published protocols and completed reviews were compared. Two raters independently identified changes to the different sections of the protocol and classified the changes as none, minor, or major. Of the 66 new Cochrane reviews, they identified a previously published protocol for 47 reviews. Of these, 43 reviews had at least 1 section that had undergone a major change compared with the most recently published protocol. The greatest variation between protocols and reviews was in the methods section, in which 68% of reviews ($n = 32$) had undergone a major change. Changes made in other sections that may have resulted in the introduction of bias included narrowing of objectives, addition of comparisons or new outcome measures, broadening of criteria for the types of study design included, and narrowing of types of participants included. Research protocols, even if published, are likely to remain, at least to some extent, iterative documents. They found that a large number of changes were made to Cochrane reviews, some of which could be prone to influence by prior knowledge of results. Even if many of the changes between protocol and review improve the overall study, the reasons for making these should be clearly identified and documented within the final review.
Genomics	Moderate	Prior to applying genomic predictors to clinical samples, the genomic data must be properly normalized to ensure that the test set data are comparable to the data upon which the predictor was trained. The most effective normalization methods depend on data from multiple patients. From a biomedical perspective, this implies that predictions for a single patient may change depending on which other patient samples they are normalized with. This test set bias will occur when any cross-sample normalization is used before clinical prediction. The authors demonstrated that results from existing gene signatures which rely on normalizing test data may be irreproducible when the patient population changes composition or size using a set of curated, publicly available breast cancer microarray experiments. As an alternative, they examined the use of gene signatures that rely on ranks from the data and show why signatures using rank-based features can avoid test set bias while maintaining highly accurate classification, even across platforms.

MLM: multilevel modeling; ORB: outcome reporting bias; CFGD: cystic fibrosis and genetic disorder; CPRP: Cochrane protocol review pair; SR: systematic review; AE: adverse events.

^aAll results under the "source of bias" column are excerpts from the original publication. Bolded type highlights sections of particular relevance. Bolded italics denote direct discussion of bias against publication of negative results.

^bEstimated magnitude of bias: Ratings are subjective; however, the criteria set were as follows: Extreme: Bias is expected to negatively affect human life; Moderately high: There is a reasonable possibility that the bias could negatively affect human life; Moderate: There is a low possibility that the bias could negatively affect human life; however, the consequences of the bias are serious to the public; Minimal: There is no realistic possibility that the bias could negatively affect human life; however, the consequences of the bias are serious to the public; Negligible: There is no possibility that the bias could affect human life, and the bias is not of a serious nature to the public.

Patil et al.⁶⁵

development of quantitative structure activity relationships.⁷⁹ This first manuscript eventually became titled, “Mutagenic activity of a series of synthetic and naturally occurring heterocyclic amines in *Salmonella*.” One of us (Smith) presented the first draft to Dr Hansch and he asked, “Where are the raw data?” in reference to the raw Ames assay *Salmonella* bacteria revertant mutation values at each dose level. Smith replied that in an attempt to save journal page space that the raw data had been intentionally omitted. Dr Hansch explained that any published paper should be able to be replicated by scientists anywhere else in the world from the data contained in the manuscript. The raw data were placed into a table and were published in the study by Smith et al.⁷³ and in every other research paper published by our group over the intervening 26 years.

Many researchers across various fields have recognized the necessity of transparency. Concerns over the inability to reproduce scientific results are shared by scientists, funding agencies, academic journals, and the general public.⁵ Studies examining research practices in numerous fields have identified a number of inappropriate procedures and potential sources of error including the following: inadequate study designs,⁸³ insufficient reporting of methods and study results,⁸⁴ errors in statistical analyses and interpretation,^{85,86} and unclear, difficult to understand, or misleading data visualization.^{87–89} Increased transparency can ameliorate, but not eliminate, some of the less than optimal research practices in current practice. For example, data transparency is particularly helpful in identifying errors in statistical analysis and interpretation and inappropriate data visualization. In addition, transparency in protocol design can assist an outside reviewer in evaluating the potential degree of colinearity or confounding in an experimental procedure, for example, bias introduced due to increased scrutiny of the experimental group as compared with the control group. A classic example of this differential level of examination was provided by Feinstein and Esdaile⁹⁰ in their description of artifactual detection of an increased breast cancer rate due to mammograms being conducted on the experimental group, while community incidence rates were used for the control group.

The vast majority of published academic papers report their results in the form of tables, graphs, or figures. Raw data are rarely provided. It is usually very difficult or not possible to extract the original data points representing the actual experimental measurements from graphs and figures. In addition, even results reported in table form usually contain data that have been transformed via some statistical technique and has possibly already been subjected to removal of data outliers.⁹¹ In contrast with the majority of the academic literature, contract laboratory-issued toxicology test reports sponsored by industry are conducted under good laboratory practice (GLP), good clinical practice (GCP), and good manufacturing practice and contain extensive raw data tables.⁹² The United States Food and Drug Administration (FDA) requires that preclinical data to

be submitted as part of a New Drug Application be collected under GLP/GCP.^{93–101} It should be noted that while ACGIH relies primarily on the published peer-reviewed academic literature that typically does not report raw data, ACGIH does not request the raw data from these sources. In contrast, when an industrial concern submits non-published data to ACGIH, the submission is typically in the form of a contract laboratory final report conducted under GLP and thus containing the raw data.

A little appreciated possible risk factor for the introduction of statistical error into the academic literature is the wide dissemination of spreadsheet and statistical calculation software in association with the spread of the personal computer throughout the research community. Personal computers began appearing in academic research laboratories in the early 1980s¹⁰² and initially tended to be shared work stations. By the mid-1980s, many researchers were using powerful personal computers for their personal use.¹⁰³ The availability of spreadsheet software, data graphing capability, and computerized statistical analysis programs beget the era of the amateur statistician.¹⁰⁴ Both cost and convenience were factors in the wide use of these technologies, as contract statistical fees can be substantial; typical costs are from US\$100 to US\$350 per hour depending on a variety of factors¹⁰⁵ and assigned shared statistical support can require researchers to enter a queue and await their turn to have data analyzed by a professional statistician. Different geographical regions probably display varying percentages of nonprofessional statistical analysis based on cost pressures. For example, academic research groups in Japan infrequently use a professional statistician to analyze their research data (personal communication, Professor Tatsuyuki Kakuma, Kurume University, Japan). The concern regarding statistical analysis by a nonprofessional is usually not that the selected statistical test is improperly conducted, but rather, relates to the selection of the appropriate statistical test or tests based on the particular characteristics of the data set. Professional statisticians evaluate a data set as to whether parametric or nonparametric tests are appropriate,¹⁰⁶ whether the data are skewed,¹⁰⁷ require logarithmic transformation,¹⁰⁸ need corrections for potential artifacts resulting from making a large number of statistical comparisons, for example, Bonferroni correction,¹⁰⁹ and consider criteria for including or excluding data point outliers.¹¹⁰

ACGIH TLVs are both predecessor and contemporary of National Institute for Occupational Safety and Health recommended exposure limits and Occupational Safety and Health Administration permissible exposure limits

The Occupational Safety and Health Administration (OSHA) and the National Institute for Occupational Safety and Health (NIOSH) were created on December 29, 1970,

when President Richard M Nixon signed the Occupational Safety and Health Act.¹¹¹ The relative youth of OSHA and NIOSH can be contrasted with the formation of the predecessor of ACGIH in 1938.¹¹² In 1941, the TLVs for Chemical Substances (TLV[®]—CS) Committee was established and became a standing committee in 1944. In 1946, ACGIH adopted the first list of 148 exposure limits, at that time termed maximum allowable concentrations (MACs). In 1956, the term MAC was changed to TLV. Currently, ACGIH provides TLVs for over 700 chemical substances and physical agents.^{113,114} In January 2016, ACGIH[®] became a 501(c)(3) charitable scientific organization and is not a regulatory agency of the United States or any other country. Therefore, from a legal perspective, ACGIH TLVs are nonbinding recommendations and not legally binding regulations.¹¹⁵

The term “recommendation” belies the historical, regulatory, public health, and economic impact of the ACGIH TLV. The centrality of the role played by TLVs is due to the limited number of authoritative bodies in the United States issuing similar exposure limits. OSHA issues legally binding permissible exposure limits (PELs).^{116,117} However, OSHA PELs are rarely updated and are not generally relied upon. In a January 2016 interview, then OSHA administrator David Michaels stated that

Many of these PELs are dangerously out of date and do not adequately protect workers. Past efforts to update our PELs have largely been unsuccessful. Since 1971, OSHA has successfully established or updated PELs for only about 30 chemicals. We have issued only one new exposure limit since the year 2000. As a result, many workers are currently being exposed to levels of chemicals that are legal, but not safe.¹¹⁸

Due to OSHA’s lack of confidence in its own PELs, the agency developed an online resource, known as the annotated permissible exposure limits (or annotated PEL) tables. The stated purpose of the annotated PELs is “to enable employers to voluntarily adopt newer, more protective workplace exposure limits than those required by OSHA.” The annotated PEL tables display a side-by-side comparison of OSHA PELs for general industry; NIOSH recommended exposure limits; California OSHA PELs; and ACGIH TLVs.¹¹⁹ Therefore, the agency of the US Federal Government with primary responsibility for worker safety recommends that workplaces rely on the TLVs set by ACGIH.

Conclusions

ACGIH TLV levels are extremely influential. ACGIH TLVs were in widespread industrial hygiene practice for many years before any other workplace exposure limits were established.¹²⁰ OSHA requires the ACGIH TLV to be listed on safety data sheets.¹¹⁹ OSHA recommends that its own legally binding PELs not be considered, but rather

that other exposure limit recommendations including the ACGIH TLV guide workplace safety standards.¹²¹ The ACGIH TLV directly influences the following: selection and use of personal protective equipment; installation of workplace ventilation equipment; monitoring of workplace air levels; decisions by customers as to the particular chemical selected for a given industrial application or process; decisions by chemical manufacturers regarding whether to produce a particular chemical or keep producing a particular chemical; and decisions by chemical company research and development departments as to whether a chemical substitute should be sought for a chemical with a TLV suggestive of suboptimal toxicity.

Despite the serious adverse consequences of an inaccurately determined TLV, the current TLV determination process is primarily dependent on an academic peer-reviewed literature in the midst of an irreproducibility crisis not yet fully explained or addressed. In addition, a well-established bias against the publication of negative results can additionally skew evaluations of the toxicological profile of a chemical toward over-estimates of hazard.

The degree of emphasis placed by ACGIH on the academic peer-reviewed literature assumes that the reported results are reproducible, unbiased, and representative of the actual level of toxicological hazard. ACGIH has on occasion heavily relied upon peer-reviewed academic studies that have been criticized on methodological and mechanistic grounds, for example, Smith et al.¹²² The lack of transparency inherent in the ACGIH TLV process represents a glaring deficiency in the current system that is inconsistent with authoritative recommendations regarding data transparency.¹⁰

Circumstances that particularly merit independent third-party statistical review are most notably academic studies employing non-standardized measurement methods, but contributing to the ACGIH numerical calculation of the TLV. In such cases, the appealing party would be responsible for any costs incurred via an independent third-party statistical reanalysis of the raw data. At no time would the appealing party be in possession of the academic institution’s data. Refusal to submit the raw data to independent statistical analysis would remove that data from consideration in setting the TLV. This procedure is consistent with the text and spirit of the Freedom of Information Request Act for publicly funded studies¹²³ and with the reforms being implemented throughout the scientific community.⁹

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Wallis P. Apprenticeship and training in premodern England, Working Papers on the nature of evidence: how well do 'facts' travel? No. 22/07, Editor Jon Adams 1–46, <http://www.lse.ac.uk/collection/economichistory/> (2007, accessed 29 July 2018).
- Shapin S and Schaffer S. *Leviathan and the air-pump*. Princeton: Princeton University Press, 1985.
- Popper K. *The logic of scientific discovery*. Abingdon-on-Thames: Routledge, 1959, pp. 1–66. ISBN 0-41527843-0. Translated from the 1934 German original, titled Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft, which literally translates as, Logic of Research: On the Epistemology of Modern Natural Science.
- Fisher RA. *The design of experiments*, (1971), 9th ed. New York: Macmillan, 1935, pp. 1–14. ISBN 0-02-844690-9.
- Begley CG and Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature* 2012; **483**: 531–533.
- Wicherts JM, Borsboom D, Kats J, et al. The poor availability of psychological research data for reanalysis. *Am Psychol* 2006; **61**(7): 726–728. DOI: 10.1037/0003-066X.61.7.726. PMID 17032082.
- Wicherts JM and Bakker M. Publish (your data) or (let the data) perish! Why not publish your data too?" *Intelligence* 2012; **40**(2): 73–76. DOI: 10.1016/j.intell.2012.01.004.
- Vanpaemel W, Vermorgen M, Deriemaeker L, et al. Are we wasting a good crisis? The availability of psychological research data after the storm. *Collabra* 2015; **1**(1): 1–5. DOI:10.1525/collabra.13.
- Open Science Collaboration (OSC). Estimating the reproducibility of psychological science. *Science* 2015; **349**(6251): aac4716. DOI: 10.1126/science.aac4716.
- Editors at Nature. Reality check on reproducibility. *Nature* 2016; **533**(7604): 437. DOI: 10.1038/533437a.
- Baker M. 1500 Scientists lift the lid on reproducibility: survey sheds light on the 'crisis' rocking research. *Nature* 2016; **533**(7604): 452–454. DOI: 10.1038/533452a.
- ACGIH, Statement of position regarding the TLVs® and BEIs®. This Statement was adopted by the ACGIH® Board of Directors. 1 March 2002, <https://www.acgih.org/tlv-bei-guidelines/policies-procedures-presentations/tlv-bei-position-statement> (2002, accessed 29 July 2018).
- Fanelli D and Ioannidis JPA. US studies may overestimate effect sizes in softer research. *Proceedings of the National Academy of Sciences of the United States of America* 2013; **110**(37): 15031–15036. DOI:10.1073/pnas.1302997110.
- Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010; **14**(8): iii, ix–xi, 1–193.
- Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull* 1979; **86**(3): 638–641.
- Ioannidis JPA, Ntzani EE, Trikalinos TA, et al. Replication validity of genetic association studies. *Nat Genet* 2001; **29**(3): 306–309.
- Ioannidis JPA and Trikalinos TA. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol* 2005; **58**(6): 543–549.
- Schooler J. Unpublished results hide the decline effect. *Nature* 2011; **470**(7335): 437.
- Ioannidis JPA. Why most discovered true associations are inflated. *Epidemiology* 2008; **19**(5): 640–648.
- Chan AW, Hróbjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004a; **291**(20): 2457–2465.
- Chan AW, Kroleza-Jerčić K, Schmid I, et al. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004b; **171**(7): 735–740.
- Chan AW and Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005; **330**(7494): 753–756.
- Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One* 2009; **4**(5): e5738.
- John LK, Loewenstein G and Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol Sci* 2012; **23**(5): 524–532.
- Fang FC, Steen RG and Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci USA* 2012; **109**(42): 17028–17033.
- Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005; **2**(8): e124.
- Rosenthal R. *Experimenter effects in behavioural research*, Enlarged ed. New York: Irvington, 1976.
- Burghardt GM, Bartmess LeVasseur JN, Browning SA, et al. Perspectives—minimizing observer bias in behavioral studies: a review and recommendations. *Ethology* 2012; **118**(6): 511–517.
- Fanelli D. "Positive" results increase down the hierarchy of the sciences. *PLoS One* 2010; **5**(4): e10068.
- Qiu J. Publish or perish in China. *Nature* 2010; **463**(7278): 142–143.
- van Dalen HP and Henkens K. Intended and unintended consequences of a publish-or-perish culture: a worldwide survey. *J Am Soc Inf Sci Technol* 2012; **63**(7): 1282–1293.
- Munafò MR, Attwood AS and Flint J. Bias in genetic association studies: effects of research location and resources. *Psychol Med* 2008; **38**(8): 1213–1214.
- Doucouliaagos H, Laroche P and Stanley TD. Publication bias in union-productivity research? *Relat Industr/Industr Relat* 2005; **60**(2): 320–347.
- Fanelli D. Negative results are disappearing from most disciplines and countries. *Scientometrics* 2012; **90**(3): 891–904.
- Fanelli D. Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS One* 2010; **5**(4): e10271.

36. Simmons JP, Nelson LD and Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci* 2011; **22**(11): 1359–1366.
37. Francis G. Too good to be true: publication bias in two prominent studies from experimental psychology. *Psychon Bull Rev* 2012; **19**(2): 151–156.
38. Ferguson CJ and Brannick MT. Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychol Methods* 2012; **17**(1): 120–128.
39. Fiedler K. Voodoo correlations are everywhere—not only in neuroscience. *Perspect Psychol Sci* 2011; **6**(2): 163–171.
40. Bakker M and Wicherts JM. The (mis)reporting of statistical results in psychology journals. *Behav Res Method* 2011; **43**(3): 666–678.
41. Raudenbush SW, Spybrook J, Liu X, et al. *Optimal design for longitudinal and multilevel research*. Chicago: University of Chicago Press, 2005. Version 1.55.
42. Scherbaum CA and Ferreter JM. Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organ Res Methods* 2009; **12**(2): 347–367.
43. Nezlek JB. *Multilevel modeling for social and personality psychology*. London: Sage, 2011.
44. Yousefi-Nooraie R, Shakiba B and Mortaz-Hejri S. Country development and manuscript selection bias: a review of published studies. *BMC Med Res Methodol* 2006; **6**: 37.
45. Matías-Guiú J and García-Ramos R. Editorial bias in scientific publications. *Neurologia* 2011; **26**(1): 1–5.
46. Link AM. US and non-US submissions: an analysis of reviewer bias. *JAMA* 1998; **280**(3): 246–247.
47. Lynch JR, Cunningham MRA, Warne WJ, et al. Commercially funded and United States-based research is more likely to be published; good-quality studies with negative outcomes are not. *J Bone Joint Surg Am* 2007; **89**(5): 1010–1018.
48. van Wilgenburg E and Elgar MA. Confirmation bias in studies of nestmate recognition: a cautionary note for research into the behaviour of animals. *PLoS One* 2013; **8**(1): e53548.
49. Jeng M. A selected history of expectation bias in physics. *Am J Phys* 2006; **74**(7): 578–583.
50. Fanelli D. When east meets west . . . does bias increase? A preliminary study on South Korea, United States and other countries. In: *Proceedings of the 8th international conference on webometrics, informetrics and scientometrics and 13th COLLNET meeting* (eds HN Choi, et al.), Seoul, Korea, 23–26 October 2012, pp. 218–223. Seoul: Korea Institute of Science and Technology Information.
51. Ioannidis JPA. Interpretation of tests of heterogeneity and bias in meta-analysis. *J Eval Clin Pract* 2008; **14**(5): 951–957.
52. Lipsey MW and Wilson DB. *Practical meta-analysis*. Thousand Oaks: Sage, 2001.
53. Viechtbauer W. Conducting meta-analyses in {R} with the {Metafor} package. *J Stat Softw* 2010; **36**(3): 1–48.
54. Beller EM, Glasziou PP, Altman DG, et al. PRISMA for abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Med* 2013; **10**(4): e1001419.
55. Beller EM, Glasziou PP, Hopewell S, et al. Reporting of effect direction and size in abstracts of systematic reviews. *JAMA* 2011b; **306**(18): 1981–1982.
56. Dwan K, Kirkham JJ, Williamson PR, et al. Selective reporting of outcomes in randomised controlled trials in systematic reviews of cystic fibrosis. *BMJ Open* 2013; **3**(6): e002709.
57. Dwan K, Wiliamson P, Gamble C, et al. Investigating outcome reporting bias in Cochrane Cystic Fibrosis and Genetic Disorder (CFGD) reviews. In: Bero L, Montgomery P, Robinson K, et al (eds) *Bringing Evidence-Based Decision-Making to New Heights. Abstracts of the 2010 Joint Colloquium of The Cochrane and Campbell Collaborations*, Keystone, USA, 18–22 October 2010. Hoboken: John Wiley & Sons.
58. Hopewell S and Beller E. Is there any evidence of selective reporting of outcomes in abstracts of Cochrane reviews?. In: Bero L, Montgomery P, Robinson K, et al (eds) *Bringing Evidence-Based Decision-Making to New Heights. Abstracts of the 2010 Joint Colloquium of The Cochrane and Campbell Collaborations*, Keystone, USA, 18–22 Oct 2010. Hoboken: John Wiley & Sons.
59. Kirkham J, Altman D and Williamson P. ORBIT study: outcome reporting bias in trials - primary outcomes in Cochrane systematic reviews. In: *Abstracts of the 17th Cochrane Colloquium*, Singapore, 11–14 October 2009. Hoboken: John Wiley & Sons.
60. Kirkham JJ, Altman DG and Williamson PR. Bias due to changes in specified outcomes during the systematic review process. *PLoS One* 2010; **5**(3): e9810.
61. Parmelli E, D’Amico R, Minozzi S, et al. Were the outcomes reported in systematic reviews stated in protocols? A systematic review. In: *Come to the craic. Abstracts of the 14th Cochrane Colloquium*, Dublin, UK, 23–26 October 2006. Hoboken: John Wiley & Sons.
62. Parmelli E, Liberati A and D’Amico R. Reporting of outcomes in systematic reviews: comparison of protocols and published systematic reviews. In: *Evidence-based health care for all. Abstracts of the 15th cochrane colloquium*, Sao Paulo, Brazil, 23–27 October 2007. Hoboken: John Wiley & Sons.
63. Silagy CA, Middleton P and Hopewell S. Publishing protocols of systematic reviews. Comparing what was done to what was planned. *JAMA* 2002; **287**(21): 2831–2834.
64. Silagy CA, Middleton P and Hopewell S. Publishing protocols of systematic reviews: comparing what was done to what was planned. *JAMA* 2002; **287**(21): 2831–2834.
65. Patil P, Bachant Winner PO, Haibe Kains B, et al. Test set bias affects reproducibility of gene signatures. *Bioinformatics* 2015; **31**(14): 2318–2323. DOI: 10.1093/bioinformatics/btv157. Epub ahead of print 18 March 2015.
66. Quote Investigator. The stolen story and other newspaper stories by Jesse Lynch Williams in 1899. The adage was spoken by a fictional character named Billy Woods in a chapter called “The old reporter,” <https://quoteinvestigator.com/2013/11/22/dog-bites/> (1899, accessed 29 July 2018).

67. McVeigh ME and Mann SJ. The journal impact factor denominator: defining citable (counted) items. *JAMA* 2009; **302**(10): 1107–1109. DOI:10.1001/jama.2009.1301.
68. National Library of Medicine (NLM). Toxicology Data Network (TOXNET®), The Hazardous Substances Data Bank (HSDB), <http://wayback.archive-it.org/org350/20180312141530/https://www.nlm.nih.gov/pubs/factsheets/hsdbfs.html> (2016, accessed 8 January 2018).
69. U.S. Department of Health and Human Services (HHS). Agency for Toxic Substances and Disease Registry (ATSDR), <https://www.atsdr.cdc.gov/> (2017, accessed 8 January 2018).
70. Occupational Safety and Health Administration (OSHA). The OSHA Hazard Communication Standard (HCS), (Subpart Z, toxic and hazardous substances, 29 CFR 1910.1200), http://www.ilpi.com/msds/osha/1910_1200_APP_D.html (2012, accessed 29 July 2018).
71. Hansch C, McKarns SC, Smith CJ, et al. Comparative QSAR supports a free radical mechanism for toxic-substituted phenols. *Chem Biol Interact* 2000; **127**(1): 61–72.
72. Smith CJ and Hansch C. The relative toxicity of compounds in mainstream cigarette smoke condensate. *Food Chem Toxicol* 2000; **38**: 637–646.
73. Smith CJ, Payne V, Doolittle DJ, et al. Mutagenic activity of a series of synthetic and naturally occurring heterocyclic amines in Salmonella. *Mutat Res* 1992; **279**: 61–73.
74. Smith CJ, Hansch C and Morton MJ. QSAR treatment of multiple toxicities: the mutagenicity and cytotoxicity of quinolines. *Mutat Res* 1997; **379**: 167–175.
75. Smith CJ, Perfetti TA, Morton MJ, et al. The relative toxicity of substituted phenols reported in cigarette mainstream smoke. *Toxicol Sci* 2002; **69**: 265–278.
76. Smith CJ, Perfetti TA, Garg R, et al. IARC carcinogens reported in cigarette mainstream smoke and their calculated log P values. *Food Chem Toxicol* 2003; **41**(6): 807–817.
77. Smith CJ, Perfetti TA, Garg R, et al. Utility of the mouse dermal promotion assay in comparing the tumorigenic potential of cigarette mainstream smoke. *Food Chem Toxicol* 2006; **44**(10): 1699–1706.
78. Smith CJ, Perfetti TA, Garg R, et al. Percutaneous penetration enhancers in cigarette mainstream smoke. *Food Chem Toxicol* 2004; **42**: 9–15.
79. Maugh TH, II. (2011). *Corwin Hansch dies at 92; scientist whose advances led to new drugs and chemicals*. *Los Angeles Times*, 23 May, <http://www.latimes.com/local/obituaries/la-me-corwin-hansch-20110523-story.html> (accessed 6 April 2016).
80. Garg R and Smith CJ. Predicting the bioconcentration factor of highly hydrophobic organic chemicals. *Food Chem Toxicol* 2014; **69**: 252–259.
81. Garg R, Ko GM and Smith CJ. QSAR molecular parameters calculated for US EPA ToxCast phase 1 and 2 chemical compounds tested against embryonic zebrafish. *Toxicol Res Appl* 2017; **1**: 1–6. DOI: 10.1177/2397847317707371.
82. Smith CJ, Perfetti TA, Ko GM, et al. Ames mutagenicity, structural alerts of carcinogenicity, Hansch QSAR Parameters (ClogP, CMR, MgVol), tumor site concordance/multiplicity, and tumorigenicity rank in NTP 2-year rodent studies. *Toxicol Res Appl* 2018; **2**: 1–14. DOI: 10.1177/2397847318759327.
83. Button KS, Ioannidis JP, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013; **14**(5): 365–376.
84. Baker D, Lidster K, Sottomayor A, et al. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PLoS Biol* 2014; **12**(1): e1001756.
85. Lazic SE. The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? *BMC Neurosci* 2010; **11**: 5.
86. Strasak AM, Zaman Q, Marinell G, et al. The use of statistics in medical research: a comparison of the New England journal of medicine and nature medicine. *Am Stat* 2007; **61**(1): 47–55.
87. Weissgerber T, Milic N, Winham S, et al. Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol* 2015; **13**(4): e1002128.
88. Weissgerber TL, Garovic VD, Winham SJ, et al. Editorial: transparent reporting for reproducible science. *J Neurosci Res* 2016; **94**: 859–864.
89. Fosang AJ and Colbran RJ. Transparency is the key to quality. *J Biol Chem* 2015; **290**(50): 29692–29694. DOI: 10.1074/jbc.E115.000002.
90. Feinstein AR and Esdaile JM. Incidence, prevalence, and evidence: scientific problems in epidemiologic statistics for the occurrence of cancer. *Am J Med* 1987; **82**(1):113–123.
91. Grubbs FE. Sample criteria for testing outlying observations. *Ann Math Stat* 1950; **21**: 27–58.
92. DeRoo D. *GLPs and GMPs: when are they necessary? A white paper*. North American Science Associates, Inc. (NAMSA) Whitepaper **13** 12/2014, https://www.namsa.com/wp-content/uploads/2015/10/WP_Requirements-for-GLP-and-GMP-Testing.pdf (2014, accessed 30 July 2018).
93. Center for Drug Evaluation and Research (CDER). Guidance for industry: investigating out-of-specification (OOS) test results for pharmaceutical production. Rockville, MD: FDA (October 2006), www.fda.gov/downloads/Drugs/Guidances/ucm070287.pdf (2006, accessed 30 July 2018).
94. Center for Devices and Radiological Health (CDRH), Center for Biologics Evaluation and Research. The applicability of good laboratory practice in premarket device submissions: questions and answers - draft guidance for industry and food and drug administration staff. Rockville, MD: FDA, <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm366338.Htm> (2013, accessed 29 July 2018).
95. Code of Federal Regulations (CFR). Title 21: food and drugs: subchapter A—general, “protection of human subjects,” part 50 (US Government Printing Office, Electronic Code of Federal Regulations, current as of 4 December 2014a), <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=50> (accessed 29 July 2018).

96. Code of Federal Regulations (CFR). Title 21: food and drugs: subchapter A—general, institutional review boards, part 56 (US Government Printing Office, Electronic Code of Federal Regulations, current as of 4 December 2014b), http://www.ecfr.gov/cgi-bin/text-idx?SID=60af1636c5e74201b3aa09cdb1677a1e&tpl=/ecfrbrowse/Title21/21cfr56_main_02.tpl (accessed 29 July 2018).
97. Code of Federal Regulations (CFR). Title 21: food and drugs: subchapter C—drugs: general, “current good manufacturing practice for finished pharmaceuticals,” parts 210–211 (US Government Printing Office, Electronic Code of Federal Regulations, current as of 4 December 2014c), <http://www.ecfr.gov/cgi-bin/text-idx?SID=ed60e423470b57343b8fba479006f514&node=pt21.4.210&rgn=div5> (accessed 29 July 2018).
98. Code of Federal Regulations (CFR). Title 21: food and drugs: subchapter F—biologics parts 600–680 (US Government Printing Office, Electronic Code of Federal Regulations, current as of 4 December 2014d), <http://www.ecfr.gov/cgi-bin/text-idx?SID=ed60e423470b57343b8fba479006f514&tpl=/ecfrbrowse/Title21/21CISubchapF.tpl> (accessed 29 July 2018).
99. Code of Federal Regulations (CFR). Title 21: food and drugs: subchapter H—medical devices, “quality system regulation,” part 820 (US Government Printing Office, Electronic Code of Federal Regulations, current as of 4 December 2014e), <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=820&showFR=1&subpartNode=21:8.0.1.1.12.8> (accessed 29 July 2018).
100. Code of Federal Regulations (CFR). Title 21: food and drugs, subchapter L—regulations under certain other acts administered by the food and drug administration, “human cells, tissues and cellular and tissue-based products,” part 1271 (US Government Printing Office, Electronic Code of Federal Regulations, current as of 4 December 2014f), <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=1271> (accessed 29 July 2018).
101. Code of Federal Regulations (CFR). Title 21: food and drugs: subchapter A—general, “good laboratory practice for nonclinical laboratory studies,” part 58 (US Government Printing Office, Electronic Code of Federal Regulations, current as of 4 December 2014g), <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?cfrpart=58> (accessed 29 July 2018).
102. Ralston A and Reilly ED (ed). *Encyclopedia of computer science*, 3rd ed. New York: Van Nostrand Reinhold, 1993. ISBN 0-442-27679-6, article digital computers history.
103. Einstein ME and Franklin JC. Computer manufacturing enters new era of growth. *Mon Labor Rev* 1986; September: 9–16, <https://www.bls.gov/opub/mlr/1986/09/art2full.pdf> (accessed 20 July 2018).
104. Diggle PJ. Statistics: a data science for the 21st century. *J R Stat Soc Ser A (Stat Soc)* 2015; **178**(4): 793–813.
105. Rutgers University. <https://www.payscale.com/research/US/Job=Statistician/Salary>; [http://statistics.rutgers.edu/consulting-fees\(2018\)](http://statistics.rutgers.edu/consulting-fees(2018)).
106. Nayak BK and Hazra A. How to choose the right statistical test? *Indian J Ophthalmol* 2011; **59**(2): 85–86. DOI:10.4103/0301-4738.77005.
107. Thiese MS, Arnold ZC and Walker SD. The misuse and abuse of statistics in biomedical research. *Biochemia Medica* 2015; **25**(1): 5–11. DOI:10.11613/BM.2015.001.
108. Feng C, Wang H, Lu N, et al. Log-transformation and its implications for data analysis. *Shanghai Arch Psychiatry* 2014; **26**(2): 105–109. DOI: 10.3969/j.issn.1002-0829.2014.02.009.
109. Sedgwick P. Multiple hypothesis testing and Bonferroni’s correction. *BMJ* 2014; **349**. <https://www.bmj.com/content/349/bmj.g6284/article-info> (accessed 6 July 2016). DOI: 10.1136/bmj.g6284.
110. Kwak SK and Kim JH. Statistical data preparation: management of missing values and outliers. *Korean J Anesthesiol* 2017; **70**(4): 407–411. DOI:10.4097/kjae.2017.70.4.407.
111. Occupational Safety and Health Act. 29 U.S.C. *United States Code, 2016 edition, title 29 – labor chapter 15 – occupational safety and health sec. 651 – congressional statement of findings and declaration of purpose and policy*. U.S. Government Publishing Office, www.gpo.gov (2016, accessed 30 July 2018).
112. Bobst S (ed). Chapter 4 – The history of risk assessment within OSHA and ACGIH: asbestos case study. In: Bobst S (ed) *History of risk assessment in toxicology, history of toxicology and environmental health*. Cambridge: Elsevier, 2017, pp. 27–30. ISBN 978-0-12-809532-4.
113. Corn M. Role of the American Conference of Governmental Industrial Hygienists and the American Industrial Hygiene Association in OSHA affairs. *Am Ind Hyg Assoc J* 1976; **37**(7): 391–394. DOI: 10.1080/0002889768507480.
114. American Conference of Governmental Industrial Hygienists (ACGIH). [https://www.acgih.org/about-us/history\(2018\)](https://www.acgih.org/about-us/history(2018)).
115. American Conference of Governmental Industrial Hygienists (ACGIH). TLVs and BEIs, [https://www.acgih.org/\(2018\)](https://www.acgih.org/(2018)).
116. Allen DT and Shonnard DR. *Green engineering: environmentally conscious design of chemical processes*. Upper Saddle River: Prentice-Hall PTR, 2001. ISBN 0132441853, 9780132441858, 576.
117. Occupational and Safety Administration (OSHA). [https://www.osha.gov/\(2018\)](https://www.osha.gov/(2018)).
118. Johnson D. OSHA’s exposure limits are dangerously out of date, industrial safety and hygiene news, <https://www.ishn.com/articles/103083-oshas-exposure-limits-are-dangerously-out-of-date> (accessed 14 July 2016).
119. United States Department of Labor (US DOL). Occupational safety and health administration US DOL, OSHA permissible exposure limits – annotated tables, <https://www.osha.gov/dsg/annotated-pels/index.html> (2018).

120. Paustenbach DJ, Cowan DM and Sahmel J. Chapter 20: The history and biological basis of occupational exposure limits for chemical agents. In: Rose VE and Cohns B (eds) *Patty's Industrial Hygiene*. 6th ed. New York: Wiley & Sons, 2011, pp. 865–954. DOI: 10.1002/0471435139.hyg041.
121. Popenorf W. *Industrial hygiene control of airborne chemical hazards*. Boca Raton: CRC Press, 2006, pp. 1–728. ISBN 1420009400, 9781420009408.
122. Smith CJ, Johnson GT, Harbison RD, et al. Dose-dependent neurologic abnormalities in workers exposed to 1-bromopropane. *Int J Occup Environ Med* 2011; **53**: 707–708.
123. The Freedom of Information Act (FOIA). 5 U.S.C. § 552 and FOIA improvement act of 2016 (Public Law No. 114-185) § 552, <https://www.justice.gov/oip/freedom-information-act-5-usc-552> (2016, accessed 30 July 2018).