Publication and other reporting biases in cognitive sciences: detection, prevalence and prevention

John P.A. Ioannidis[1], **Marcus R. Munafò**[2], **Paolo Fusar-Poli**[3], **Brian A. Nosek**[4], and **Sean P. David**[5]

[1]Departments of Medicine and Health Research and Policy, Stanford University School of Medicine, Department of Statistics, Stanford University School of Humanities and Sciences, and Meta-Research Innovation Center at Stanford (METRICS), Stanford, CA, USA

[2]MRC Integrative Epidemiology Unit, UK Centre for Tobacco and Alcohol Studies, and School of Experimental Psychology, University of Bristol, Bristol, UK

[3]Department of Psychosis Studies, Institute of Psychiatry, King's College London, UK

[4]Center for Open Science, and Department of Psychology, University of Virginia, VA, USA

[5]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

Abstract

Recent systematic reviews and empirical evaluations of the cognitive sciences literature suggest that publication and other reporting biases are prevalent across diverse domains of cognitive science. This review summarizes the various forms of publication and reporting biases and other questionable research practices, and overviews the available methods for probing into their existence. We discuss the available empirical evidence for the presence of such biases across the neuroimaging, animal, other pre-clinical, psychological, clinical trials, and genetics literature in the cognitive sciences. We also highlight emerging solutions (from study design to data analyses and reporting) to prevent bias and improve the fidelity in the field of cognitive science research.

Keywords

publication bias; reporting bias; cognitive sciences; neuroscience; bias

Introduction

The promise of science is that the generation and evaluation of evidence is done transparently to promote an efficient, self-correcting process. However, multiple biases may

Correspondence: John P.A. Ioannidis, MD, DSc, 1265 Welch Rd, Medical School Office Building Room X306, Stanford CA 94305, USA. jioannid@stanford.edu.

produce inefficiency in knowledge building. In this review, we discuss the importance of publication and other reporting biases, and we present some potential correctives that may reduce bias without disrupting innovation. Consideration of publication and other reporting biases is particularly timely for cognitive sciences. Cognitive sciences are undergoing a rapid expansion. As such, preventing or remedying these biases will have substantial impact on the efficient development of a credible corpus of published research.

Definitions of Biases and Relevance for Cognitive Sciences

The terms publication bias and selective reporting bias refers to the differential choice to publish studies or report particular results depending on the nature or directionality of findings [1]. There are several forms of such biases in the literature [2], including (1) *study publication bias*, where studies are less likely to be published when they reach non-statistically significant findings; (2) *selective outcome reporting bias*, where multiple outcomes are evaluated in a study and outcomes found to be non-significant are less likely to be published than statistically significant ones; and (3) *selective analysis reporting bias*, where certain data are analyzed using different analytical options and publication favors the more impressive, statistically significant variants of the results (Box 1).

Cognitive science employs a range of methods producing a multitude of measures and rich datasets. Given this quantity of data, the potential analyses available and pressures to publish, the temptation to use arbitrary (instead of planned) statistical analyses and to mine data are substantial and may lead to questionable research practices. Even extreme cases of fraud in the cognitive sciences, such as reporting of experiments that never took place, falsifying results and confabulating data have been recently reported [3–5]. Overt fraud is probably rare, but evidence of reporting biases in multiple domains of the cognitive sciences (discussed below) raise concern about the prevalence of questionable practices that occur via motivated reasoning without any intent to be fraudulent.

Tests for publication and other reporting biases

Tests for single studies, specific topics and wider disciplines

Explicit documentation of publication and other reporting biases requires availability of protocols, data, and results of primary analyses from conducted studies so that these can be compared against the published literature. However, these are not often available. A few empirical studies have retrieved study protocols from authors or trial data from FDA submissions [6–9]. These studies have shown that deviations in the analysis plan between protocols and published papers are common [6], and effect sizes of drug interventions are larger in the published literature compared with the corresponding data from the same trials submitted to FDA [7, 8]. It is challenging, but more common, to detect these biases using only the published literature. Several tests have been developed for this purpose; some are informal, others use statistical methods of variable rigor. These tests may be applied to single studies, multiple studies on specific questions, or wider disciplines.

Evaluating biases in each single study is attractive, but most difficult, since the data are usually limited, unless designs and analysis plans are registered *a priori* [9]. It is easier to

evaluate bias across multiple studies performed on the same question. When tests of bias are applied to a wider scientific corpus, it is difficult to pinpoint which single studies in this corpus of evidence are affected more by bias. The tractable goal is to gain insight into the average bias in the field.

Small-study effects

Tests of small-study effects have been popular since the mid-1990s. They evaluate studies included in the same meta-analysis and assess whether effect sizes are related to study size. When small studies have larger effects than large studies, this may reflect publication or selective reporting biases, but alternative explanations exist (as reviewed elsewhere [10]). Sensitivity and specificity of these tests in real life is unknown, but simulation studies have evaluated the performance in different settings. Published recommendations suggest cautious use of such tests [10]. In brief, visual evaluations of inverted funnel plots without statistical testing are precarious [11]; some test variants have better type I and type II error properties than others [12, 13]; and for most meta-analyses where there are a limited number of studies, the power of these tests is low [10].

Selection models

Selection model approaches evaluate whether the pattern of results that have been accumulated from a number of studies suggests an underlying filtering process, such as the non-publication of a given percentage of results that had not reached formal statistical significance [14–16]. These methods have been less widely applied than small-study effect tests, even though they may be more promising. One application of a selection model approach examined an entire discipline (Alzheimer's disease genetics) and found that selection forces may be different for first/discovery results versus subsequent replication/refutation results or late replication efforts [17]. Another method that probes for data "fiddling" has been proposed [18] to identify selective reporting of extremely promising p-values in a body of published results.

Excess significance

Excess significance testing evaluates whether the number of statistically significant results in a corpus of studies it too high, under some plausible assumptions about the magnitude of the true effect size [19]. The Ioannidis test [19] can be applied to meta-analyses of multiple studies and also to larger fields and disciplines where many meta-analyses are compiled. The number of expected studies with nominally statistically significant results is estimated by summing the calculated power of all the considered studies. Simulations suggest that the most appropriate assumption is the effect size of the largest study in each meta-analysis [20].

Francis has applied a variant of the excess significance test on multiple occasions on discrete psychological science studies, where multiple experiments have statistically significant results [21]. The probability is estimated that all experiments in the study would have had statistically significant results. The appropriateness of applying excess significance testing in single studies has been questioned. Obviously this application is more charged, because specific studies are pinpointed as being subject to bias.

Other field-wide assessments

In some fields such as neuroimaging, where identification of foci rather than effect sizes is the goal, one approach evaluates whether the number of claimed discovered foci is related to the sample size of the performed studies [22]. In the absence of bias, one would expect larger studies to have larger power and thus detect more loci. Lack of such a relationship, or even worse, an inverse relationship with fewer foci discovered with larger studies, offers indirect evidence of bias.

Methods for correction of bias

Both small-study effects approaches and selection models allow extensions to correct for the potential presence of bias. For example, one can estimate the extrapolated effect size for a study with theoretically infinite sample size that is immune from small-study bias [23]. For selection models, one may impute missing non-significant studies in meta-analyses to estimate corrected effect sizes [16]. Another popular approach is the trim-and-fill method, and variants thereof, for imputing missing studies [24]. Other methods try to explicitly model the impact of multiple biases on the data [25, 26]. While these methods are a focus of active research and some are even frequently used, their performance is largely unknown. They may generate a false sense of security that we can remedy distorted/selectively reported data. More solid approaches for correction of bias would require access to raw data, protocols, analyses codes, and registries of unpublished information, but this is usually not available.

Empirical Evidence for Publication and Other Reporting Biases in Cognitive Sciences

Neuroimaging

With over three-thousand publications annually in the past decade, including several meta-analyses [27] (Figure 1), the identification of potential publication or reporting biases is of crucial interest for the future of neuroimaging. The first study to evaluate evidence for an excess of statistically significant results in neuroimaging focused on brain volumetric studies based on Region of Interests (ROIs) analyses of psychiatric conditions [28]. The study analyzed 41 meta-analyses (461 datasets) and concluded that there were too many studies with statistically significant results, probably due to selective outcome or analysis reporting [28]. Another evaluation assessed studies employing voxel-based-morphometry (VBM) methods, which are less subject to selective reporting of selected ROIs [29]. Still, in a dataset of 47 whole-brain meta-analyses including 324 individual VBM studies, the number of foci reported in small VBM studies and even in meta-analyses with few studies was often inflated, consistent with reporting biases [30]. Interestingly, this and another investigation suggested that studies with fewer coauthors tend to report larger brain abnormalities [31]. Similar biases were also detected in the use of functional magnetic resonance imaging (fMRI) in an evaluation of 94 whole brain fMRI meta-analyses with 1788 unique datasets of psychiatric or neurological conditions and tasks [32]. Reporting biases seemed to affect small fMRI studies, which may be analyzed and reported in ways that may generate a larger number of foci (many foci especially in smaller studies may represent false-positives) [32].

Animal studies and other pre-clinical studies

The poor reproducibility of preclinical studies, including animal studies, has recently been highlighted by a number of failures to replicate key findings [33–35]. Irreproducible results indicate that various biases may be operating in this literature.

In some fields of *in vitro* and cell-based preclinical research direct replication attempts are uncommon and thus bias is difficult to assess empirically. Conversely, there are often multiple animal model studies performed on similar questions. Empirical evaluations have shown that small studies consistently give more favorable results than larger studies [36] and study quality is inversely related to effect size [37–40]. A large-scale meta-epidemiological evaluation [38, 41] used data from the Collaborative Approach to Meta-Analysis and Review of Animal Data in Experimental Studies consortium (CAMARADES; http://www.camarades.info/). It assessed 4,445 datasets synthesized in 160 meta-analyses of neurological disorders, and found clear evidence of an excess of significance. This observation suggests strong biases, with selective analysis and outcome reporting biases being plausible explanations.

Psychological science

Over fifty years ago, Sterling [42] offered evidence that almost all published psychological studies reported positive results. This trend has been persistent and perhaps even increasing over time [43], and particularly in the United States [44]. Evidence suggests perceived bias against negative results in peer-review [45], and a desire for aesthetically pleasing positive effects [46] leads authors to be less likely to even attempt to publish negative studies [45]. Instead, many psychological scientists report taking advantage of selective reporting and flexibility in analysis to make their research results more publishable [47, 48]. Such practices may contribute to unusually elevated likelihoods of positive results that are just below the nominal .05 threshold for significance (p-hacking) [49, 50].

Clinical trials

Publication and other reporting biases in the clinical trials literature have often been ascribed to financial vested interests from manufacturer sponsors [51]. For example, in a review [52] of all randomized controlled trials of nicotine replacement therapy (NRT) for smoking cessation, more industry-supported trials (51%) reported statistically significant results than non-industry trials (22%); this difference was unexplained by trial characteristics. Moreover, industrysupported trials indicated a larger effect of NRT (summary odds ratio 1.90, 95% CI 1.67 to 2.16) than non-industry trials (summary odds ratio 1.61, 95% CI 1.43 to 1.80). Evidence of excess significance has also been documented in trials of neuroleptics [19]. Comparisons of published results against FDA records shows that, while almost half of the trials on antidepressants for depression have negative results in the FDA records, these negative results either remain unpublished or are published with distorted reporting that shows them as positive [8]; thus, the published literature shows larger estimates of treatment effects for antidepressants than the FDA data. A similar pattern has been recorded also for trials on antipsychotics [7].

Financial conflicts are not the only explanation for such biases. For example, similar problems to antidepressants seem to exist also in the literature of randomized trials on psychotherapies [53–55]. Allegiance bias towards favoring specific cognitive, behavioral, or social theories that underlie these interventions may be responsible.

Genetics

Over the past couple of decades candidate gene studies have exemplified many of the problems of poor reproducibility; few of the early studies in this literature could be replicated reliably [56, 57]. This problem has been even worse in the literature of genetics of cognitive traits and personality, where phenotype definitions are flexible [58].

The development of genome-wide association methods necessitated greater statistical stringency (given the number of statistical tests conducted) and much larger sample sizes. This has transformed the reliability of genetic research, with fewer but more robust genetic signals now being identified [59]. The successful transformation has required international collaboration, the combination of data in large consortia, and the adoption of rigorous replication practices in meta-analysis within consortia [60, 61].

Approaches to Prevent Bias

Most scientists embrace fundamental scientific values such as disinterestedness and transparency even while believing that other scientists do not [62]. However, noble intentions may not be sufficient to decrease biases when people are not able to recognize or control their own biases [63, 64], rationalize biases through motivated reasoning [65], and are embedded in a culture that implicitly rewards the expression of biases for personal career advancement – publication over accuracy. Actual solutions should recognize the limits of human capacity to act on intentions, and nudge the incentives that motivate and constrain behavior. Figure 2 illustrates common practices and possible solutions across the workflow for addressing multiple biases, which are discussed below.

Study registration

Study registration has been used primarily for clinical trials. Most leading journals now require protocol registration for clinical trials at inception as a condition of publication. Using public registries such as http://clinicaltrials.gov/, one can discover studies that were conducted and never reported. However, still only the minority of clinical trials is registered and registration is very uncommon for other types of designs for which registration may also be useful [66–68].

Pre-specification

Pre-specification of the project design and analysis plan can allow separating pre-specified from post hoc exploratory analyses. Currently, registered studies do not necessarily include such detailed pre-specification. Exploratory analysis can be very useful, but they are tentative. Pre-specified analyses can be interpreted with greater confidence [69]. Researchers can register and pre-specify analyses through services like the Open Science Framework (OSF; http://osf.io/).

Data, protocol, code and analyses availability

For materials that can be shared digitally, transparency of the materials, data, and research process is now straightforward. There are hundreds of data repositories available for data storage – some generic (e.g., http://thedata.org/), others specific for particular types of data (e.g., http://openfmri.org/), and others offering additional services for collaboration and project management (e.g., OSF and http://figshare.com/). There are also an increasing number of options for registering protocols (e.g., PROSPERO for systematic reviews, Trials, F1000 journals).

Replication

Another core value of science is reproducibility [70]. Present incentives for innovation vastly outweigh those for replication, making replications rare in many fields [71, 72]. Journals are altering those incentives with a "registered reports" publishing option in order to avoid bias on the part of authors or reviewers who might otherwise be swayed by statistical significance or the "information gain" [73] arising from a study's results [74]. Authors propose replicating an existing result and provide justification for importance and description of the protocol and analysis plan. Peer review and conditional acceptance prior to data collection lowers the barrier for authors to pursue replication, and increases confidence in the results of the replication via peer review and pre-specification. Multiple journals in psychology and neuroscience have adopted this format [75–77]; http://www.psychologicalscience.org/index.php/replication), and funders like NIH are considering support for replication research [78].

Collaborative consortia

Incentives for replication research are enhanced further through crowdsourcing and collaborative consortia to increase the feasibility and impact of replication. The Reproducibility Project is an open collaboration of more than 180 researchers aiming to estimate the reproducibility of psychological science. Replication teams each select a study from a sample of articles published in three psychology journals in 2008. All teams follow a protocol that provides quality assurance and adherence to standards. The effort distributes workload among many contributors and aggregates the replications for maximal value [79, 80]. The Many Labs Project implemented a single study session including thirteen experimental effects in psychology [81]. The same protocol (with translation) was administered in 36 laboratory and web samples around the world (total N = 6344).

Commentators also note untapped opportunities for replication research in the context of advanced undergraduate and graduate pedagogy [82, 83]. The Collaborative Replications and Education Project (CREP; https://osf.io/fLAuE/) identified 10 published studies from 2010 of those that were amenable to integration and execution as a course-based research project. Project coordinators provided instructional aids and defined standards for documenting and aggregating the results of replications across institutions.

Some fields such as genetics have been transformed by the widespread adoption of replication practices as a sine qua non in the setting of large-scale collaborative consortia

where tens and hundreds of teams participate. Such consortia allow ensuring that maximal standardization or harmonization exists in the analysis plan across the participating teams.

Pre-publication review practices

Authors follow standards and checklists for formatting manuscripts for publication, but there are few standards for the content of what is reported in manuscripts in basic science. In January 2014, *Psychological Science* introduced a four-item checklist of disclosure standards for submission: rationale for sample size, outcome variables, experimental conditions, and use of covariates in analysis. Enriched checklists and reporting standards such as CONSORT and ARRIVE are in use in clinical trials and animal research respectively (http://www.consortstatement. org/; http://www.nc3rs.org.uk/page.asp? id=1357). And, specialized reporting standards have been introduced for some methods, such as fMRI [84]. Such practices may improve the consistency and clarity of reporting.

Publication practices

Beyond disclosure requirements and replication reports, journals can encourage or require practices that shift incentives and decrease bias. For example, in 2014 eight charter journals in psychology and neuroscience adopted badges to acknowledge articles that met criteria for Open Data, Open Materials, and Preregistration (http://osf.io/TVyXZ). Further, journals can make explicit that replications are welcome, not discouraged [85], particularly of research that had been published previously in that journal. Finally, journals like *Psychological Science* are changing to allow fulsome description of methods [86]. Full description may facilitate the adoption of standards in contexts that allow substantial flexibility. For example, a study of 241 studies using fMRI found 223 unique analysis pipelines with little justification for this variation [87]. Further, activation patterns were strongly dependent on the analysis pipeline applied.

Post-publication review practices

Retractions are an extremely rare event [88] as are full re-analyses of published results. Increasing opportunities and exposure to post-publication review commentary and re-analysis may be useful. Post-publication comments can now be made in PubMed Commons and have been possible for a long time in the *BMJ*, *Frontiers*, *PLoS* and other journal groups.

Incentive structures

The primary currency of academic science is publication. Incentives need to be re-aligned so that emphasis is given not to quantity of publications, but quality and impact of the work, reproducibility track-record and adherence to sharing culture (for data, protocols, analyses, and materials) [63, 89].

Concluding Remarks

Overall, publication and other selective reporting biases are probably prevalent and influential in diverse cognitive sciences. Different approaches have been proposed to remedy these biases. Box 2 summarizes some outstanding questions in this process. As

shown in Figure 2, there are multiple entry points for each of diverse 'problems' from designing to conducting studies, analyzing data and reporting results. 'Solutions' are existing, but uncommon, practices that should improve the fidelity of scientific inference in the cognitive sciences. We have limited evidence on effectiveness of these approaches. Moreover, improvements may require interventions at multiple levels. Such interventions should aim at improving the efficiency and reproducibility of science without stifling creativity and the drive for discovery.

Acknowledgments

References

1. Green S, et al. Cochrane Handbook for Systematic Reviews of Interventions, version 5.1.0. 2011

2. Dwan K, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One. 2008; 3:e3081. [PubMed: 18769481]

3. Vogel G. Scientific misconduct. Psychologist accused of fraud on 'astonishing scale'. Science. 2011; 334:579. [PubMed: 22053018]

4. Vogel G. Scientific misconduct. Fraud charges cast doubt on claims of DNA damage from cell phone fields. Science. 2008; 321:1144–1145. [PubMed: 18755946]

5. Vogel G. Developmental biology. Fraud investigation clouds paper on early cell fate. Science. 2006; 314:1367–1369. [PubMed: 17138872]

6. Saquib N, et al. Practices and impact of primary outcome adjustment in randomized controlled trials: meta-epidemiologic study. BMJ. 2013; 347:f4313. [PubMed: 23851720]

7. Turner EH, et al. Publication bias in antipsychotic trials: an analysis of efficacy comparing the published literature to the US Food and Drug Administration database. PLoS Med. 2012; 9:e1001189. [PubMed: 22448149]

8. Turner EH, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med. 2008; 358:252–260. [PubMed: 18199864]

9. Dwan K, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. PLoS One. 2013; 8:e66844. [PubMed: 23861749]

10. Sterne JA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011; 343:d4002. [PubMed: 21784880]

11. Lau J, et al. The case of the misleading funnel plot. BMJ. 2006; 333:597–600. [PubMed: 16974018]

12. Peters JL, et al. Comparison of two methods to detect publication bias in metaanalysis. JAMA. 2006; 295:676–680. [PubMed: 16467236]

13. Harbord RM, et al. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Stat Med. 2006; 25:3443–3457. [PubMed: 16345038]

14. Copas JB, Malley PF. A robust P-value for treatment effect in meta-analysis with publication bias. Stat Med. 2008; 27:4267–4278. [PubMed: 18384185]

15. Copas J, Jackson D. A bound for publication bias based on the fraction of unpublished studies. Biometrics. 2004; 60:146–153. [PubMed: 15032784]

16. Hedges LV, Vevea JL. Estimating effect size under publication bias: small sample properties and robustness of a random effects selection model. J Educ Behavioral Stat. 1996; 21:299–332.

17. Pfeiffer T, et al. Quantifying selective reporting and the Proteus phenomenon for multiple datasets with similar bias. PLoS One. 2011; 6:e18362. [PubMed: 21479240]

18. Gadbury GL, Allison DB. Inappropriate fiddling with statistical analyses to obtain a desirable p-value: tests to detect its presence in published literature. PLoS One. 2012; 7:e46363. [PubMed: 23056287]

19. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. Clin Trials. 2007; 4:245–253. [PubMed: 17715249]

20. Ioannidis JPA. Clarifications on the application and interpretation of the test for excess significance and its extensions. J Math Psychol. 2013 *(in press)*.

21. Francis G. The psychology of replication and replication in psychology. Perspectives in Psychological Science. 2012; 7:585–594.

22. David SP, et al. Potential reporting bias in FMRI studies of the brain. PLoS One. 2013; 8:e70104. [PubMed: 23936149]

23. Rücker G, et al. Detecting and adjusting for small-study effects in meta-analysis. Biom J. 2011; 53:351–368. [PubMed: 21374698]

24. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000; 56:455–463. [PubMed: 10877304]

25. Ioannidis JP. Commentary: Adjusting for bias: a user's guide to performing plastic surgery on meta-analyses of observational studies. Int J Epidemiol. 2011; 40:777–779. [PubMed: 21233141]

26. Thompson S, et al. A proposed method of bias adjustment for meta-analyses of published observational studies. Int J Epidemiol. 2011; 40:765–777. [PubMed: 21186183]

27. Jennings RG, Van Horn JD. Publication bias in neuroimaging research: implications for meta-analyses. Neuroinformatics. 2012; 10:67–80. [PubMed: 21643733]

28. Ioannidis JP. Excess significance bias in the literature on brain volume abnormalities. Arch Gen Psychiatry. 2011; 68:773–780. [PubMed: 21464342]

29. Borgwardt S, et al. Why are psychiatric imaging methods clinically unreliable? Conclusions and practical guidelines for authors, editors and reviewers. Behav Brain Funct. 2012; 8:46. [PubMed: 22938624]

30. Fusar-Poli P, et al. Evidence of Reporting Biases in Voxel-Based Morphometry (VBM) Studies of Psychiatric and Neurological Disorders. Human Brain Mapping. 2013 Epub.

31. Sayo A, et al. Study factors influencing ventricular enlargement in schizophrenia: a 20 year follow-up meta-analysis. Neuroimage. 2012; 59:154–167. [PubMed: 21787868]

32. David SP, et al. Potential reporting bias in FMRI studies of the brain. PLoS One. 2013; 8:e70104. [PubMed: 23936149]

33. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012; 483:531–533. [PubMed: 22460880]

34. Prinz F, et al. Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov. 2011; 10:712. [PubMed: 21892149]

35. ter Riet G, et al. Publication bias in laboratory animal research: a survey on magnitude, drivers, consequences and potential solutions. PLoS One. 2012; 7:e43404. [PubMed: 22957028]

36. Sena ES, et al. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. PLoS Biol. 2010; 8:e1000344. [PubMed: 20361022]

37. Ioannidis JP. Extrapolating from animals to humans. Sci Transl Med. 2012; 4:151ps115.

38. Jonasson Z. Meta-analysis of sex differences in rodent models of learning and memory: a review of behavioral and biological data. Neurosci Biobehav Rev. 2005; 28:811–825. [PubMed: 15642623]

39. Macleod MR, et al. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke. 2008; 39:2824–2829. [PubMed: 18635842]

40. Sena E, et al. How can we improve the pre-clinical development of drugs for stroke? Trends Neurosci. 2007; 30:433–439. [PubMed: 17765332]

41. Button KS, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013; 14:365–376. [PubMed: 23571845]

42. Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. Journal of the American Statistical Association. 1959; 54:30–34.

43. Fanelli D. Negative results are disappearing from most disciplines and countries. Scientometrics. 2012; 90:891–904.

44. Fanelli D, Ioannidis JP. US studies may overestimate effect sizes in softer research. Proc Natl Acad Sci U S A. 2013; 110:15031–15036. [PubMed: 23980165]

45. Greenwald AG. Consequences of prejudice against the null hypothesis. Psychological Bulletin. 1975; 82:1–20.

46. Giner-Sorolla. Science or art? How esthetic standards grease the way through the publication bottleneck but undermine science. Psychological Science. 2012; 7:562–571.

47. John LK, et al. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol Sci. 2012; 23:524–532. [PubMed: 22508865]

48. Simmons JP, et al. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci. 2011; 22:1359–1366. [PubMed: 22006061]

49. Masicampo EJ, Lalande DR. A peculiar prevalence of p values just below .05. Q J Exp Psychol (Hove). 2012; 65:2271–2279. [PubMed: 22853650]

50. Simonsohn U, et al. P-Curve: A Key to the File-Drawer. J Exp Psychol Gen. 2013 *(in press)*.

51. Stamatakis E, et al. Undue industry influences that distort healthcare research, strategy, expenditure and practice: a review. Eur J Clin Invest. 2013; 43:469–475. [PubMed: 23521369]

52. Etter JF, et al. The impact of pharmaceutical company funding on results of randomized trials of nicotine replacement therapy for smoking cessation: a metaanalysis. Addiction. 2007; 102:815–822. [PubMed: 17493109]

53. Cuijpers P, et al. Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis. Psychol Med. 2013:1–11.

54. Cuijpers P, et al. Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis. Psychol Med. 2013; 44:685–695. [PubMed: 23552610]

55. Cuijpers P, et al. Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. Br J Psychiatry. 2010; 196:173–178. [PubMed: 20194536]

56. Munafò MR. Reliability and replicability of genetic association studies. Addiction. 2009; 104:1439–1440. [PubMed: 19686515]

57. Ioannidis JP, et al. Replication validity of genetic association studies. Nat Genet. 2001; 29:306–309. [PubMed: 11600885]

58. Munafò MR, Flint J. Dissecting the genetic architecture of human personality. Trends Cogn Sci. 2011; 15:395–400. [PubMed: 21831694]

59. Hirschhorn JN. Genomewide association studies--illuminating biologic pathways. N Engl J Med. 2009; 360:1699–1701. [PubMed: 19369661]

60. Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet. 2013; 14:379–389. [PubMed: 23657481]

61. Panagiotou OA, et al. The power of meta-analysis in genome-wide association studies. Annu Rev Genomics Hum Genet. 2013; 14:441–465. [PubMed: 23724904]

62. Anderson MS, et al. Normative dissonance in science: results from a national survey of u.s. Scientists. J Empir Res Hum Res Ethics. 2007; 2:3–14. [PubMed: 19385804]

63. Nosek BA, et al. Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. Perspectives on Psychological Science. 2012; 7:615–631.

64. Nosek, BA., et al. Implicit social cognition. In: Fiske, S.; Macrae, CN., editors. Handbook of Social Cognition. Sage; 2012.

65. Kunda Z. The case for motivated reasoning. Psychol Bull. 1990; 108:480–498. [PubMed: 2270237]

66. Ioannidis JP. The importance of potential studies that have not existed and registration of observational data sets. JAMA. 2012; 308:575–576. [PubMed: 22871867]

67. World-Medical-Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013; 310:2191–2194. [PubMed: 24141714]

68. Dal-Ré R, et al. Should prospective registration of observational research be only a matter of willingness? Sci Transl Med. 2013 (in press).

69. Wagenmakers E-J, et al. An agenda for purely confirmatory research. Perspectives on Psychological Science. 2012; 7:632–638.

70. Schmidt S. Shall we really do it again? The powerful concept of replication is neglected in the social sciences. Review of General Psychology. 2009; 13:90–100.

71. Collins, HM. Changing order. Sage; 1985.

72. Makel MC, et al. Replications in psychology research: How often do they really occur? Perspectives on Psychological Science. 2012; 7:537–542.

73. Evangelou E, et al. Perceived information gain from randomized trials correlates with publication in high-impact factor journals. J Clin Epidemiol. 2012; 65:1274–1281. [PubMed: 22959593]

74. Emerson GB, et al. Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. Arch Intern Med. 2010; 170:1934–1939. [PubMed: 21098355]

75. Chambers CD. Registered reports: a new publishing initiative at Cortex. Cortex. 2013; 49:609–610. [PubMed: 23347556]

76. Nosek BA, Lakens DE. Call for Proposals: Special issue of Social Psychology on"Replications of important results in social psychology". Soc Psychol. 2013; 44:59–60.

77. Wolfe J. Registered Reports and Replications in Attention, Perception, and Psychophysics. Atten Percept Psychophys. 2013; 75:781–783.

78. Wadman M. NIH mulls rules for validating key results. Nature. 2013; 500:14–16. [PubMed: 23903729]

79. Open-Science Collaboration. An open, large-scale collaborative effort to estimate the reproducibility of psychological science. Perspectives on Psychological Science. 2012; 7:657–660.

80. Open-Science-Collaboration. The Reproducibility Project: A model of large-scale collaboration for empirical research on reproducibility. In: Stodden, V.; Leisch, F.; Peng, R., editors. Implementing Reproducible Computational Research (A Volume in the R Series). Taylor & Francis; 2013.

81. Klein RA, Ratliff KA. Investigating variation in replicability: A"Many Labs" replication project. Registered design. Social Psychology. 2014 (in press).

82. Frank MC, Saxe R. Teaching replication. Perspectives on Psychological Science. 2012; 7:600–604.

83. Grahe JE, et al. Harnessing the undiscovered resource of student research projects. Perspectives on Pyschological Science. 2012; 7

84. Poldrack RA, et al. Guidelines for reporting an fMRI study. Neuroimage. 2008; 40:409–414. [PubMed: 18191585]

85. Neuliep JW, Crandell R. Editorial bias against replication research. Journal of Social Behavior and Personality. 1990; 5:85–90.

86. Eich E. Business not as usual. Psychological Science. 2014 (in press).

87. Carp J. The secret lives of experiments: methods reporting in the fMRI literature. Neuroimage. 2012; 63:289–300. [PubMed: 22796459]

88. Steen RG. Retractions in the scientific literature: do authors deliberately commit research fraud? J Med Ethics. 2010; 37:113–117. [PubMed: 21081306]

89. Koole SL, Lakens D. Rewarding replications: A sure and simple way to improve psychological science. Perspectives on Psychological Science. 2012; 7:608–614.

90. Rosenthal R. The file drawer problem and tolerance for null results. Psychol Bull. 1979; 86:638–641.

91. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA. 1998; 279:281–286. [PubMed: 9450711]

92. Ioannidis JP, 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; 58:543–549. [PubMed: 15878467]

93. Kirkham JJ, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ. 2010; 340:c365. [PubMed: 20156912]

94. Hahn S, et al. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. Stat Med. 2000; 19:3325–3336. [PubMed: 11122498]

95. Moreno SG, et al. Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. BMJ. 2009; 339:b2981. [PubMed: 19666685]

96. Vedula SS, et al. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. N Engl J Med. 2009; 361:1963–1971. [PubMed: 19907043]

97. Ioannidis JP. Why most discovered true associations are inflated. Epidemiology. 2008; 19:640–648. [PubMed: 18633328]

98. Young NS, et al. Why current publication practices may distort science. PLoS Med. 2008; 5:e201. [PubMed: 18844432]

Box 1: Various publication and reporting biases

- **Study publication bias** can distort meta-analyses and systematic reviews of a large number of published studies when authors are more likely to submit and/or editors more likely to publish studies with "positive" (i.e., statistically-significant) results or suppress "negative" (i.e., non-significant) results.

- Study publication bias is also called the **"file drawer problem"** [90], whereby most studies, with null effects, are never submitted and thereby left in the file drawer. However, study publication bias is probably also accentuated by peer-reviewers and editors, even for "negative" studies that are submitted by their authors for publication.

- Sometimes studies with "negative" results may be published, but with greater delay, than studies with "positive" results, causing the so-called **"time-lag bias"** [91].

- In the **"Proteus phenomenon"**, "positive" results are published very fast and then there is a window of opportunity for the publication also of "negative" results contradicting the original observation that may be tantalizing and interesting to editors and investigators. The net effect is rapidly alternating extreme results [17, 92].

- **Selective outcome or analyses reporting biases** can occur in many forms including selective reporting of some of the analyzed outcomes in a study when others exist; selective reporting of a specific outcome; and incomplete reporting of an outcome [93].

- Examples of selective analyses reporting include selective reporting of subgroup analyses [94], reporting of per protocol instead of intention to treat analyses [95] and relegation of primary outcomes to secondary outcomes [96] so that non-significant results for the pre-specified primary outcomes are subordinated to non-primary positive results [8].

- As a result of all these selective reporting biases, the published literature is likely to have an excess of statistically significant results. Even when associations and effects are genuinely non-null, the published results are likely to over-estimate their magnitude [97, 98].

Box 2: Outstanding questions

There are several outstanding questions about how to minimize the impact of biases in the literature. These include who should take the lead in changing various practices (journals, scientists, editors, funding agencies, regulators, patients, the public, or combinations of the above); how to ensure that more stringent requirements to protect from bias do not stifle creativity by increasing the burden of documentation; the extent to which new measures to curtail bias may force investigators, authors and journals to adapt their practices in a normative fashion, satisfying new checklists of requirements, but without really changing the essence of the research practices; how to avoid excessive, unnecessary replication happening at the expense of not investing in new research directions; how to optimize the formation and efficiency of collaborative consortia of scientists; and how to tackle potential caveats or collateral problems that may ensue while trying to maximize public access to raw data (e.g. in theory analyses may be performed by ambitious re-analysts with the explicit purpose of showing that they get a different [and thus more easily publishable] result).

Highlights

- Publication and other reporting biases can have major detrimental effects on the credibility and value of research evidence.

- There is substantial empirical evidence of publication and reporting biases in cognitive science disciplines

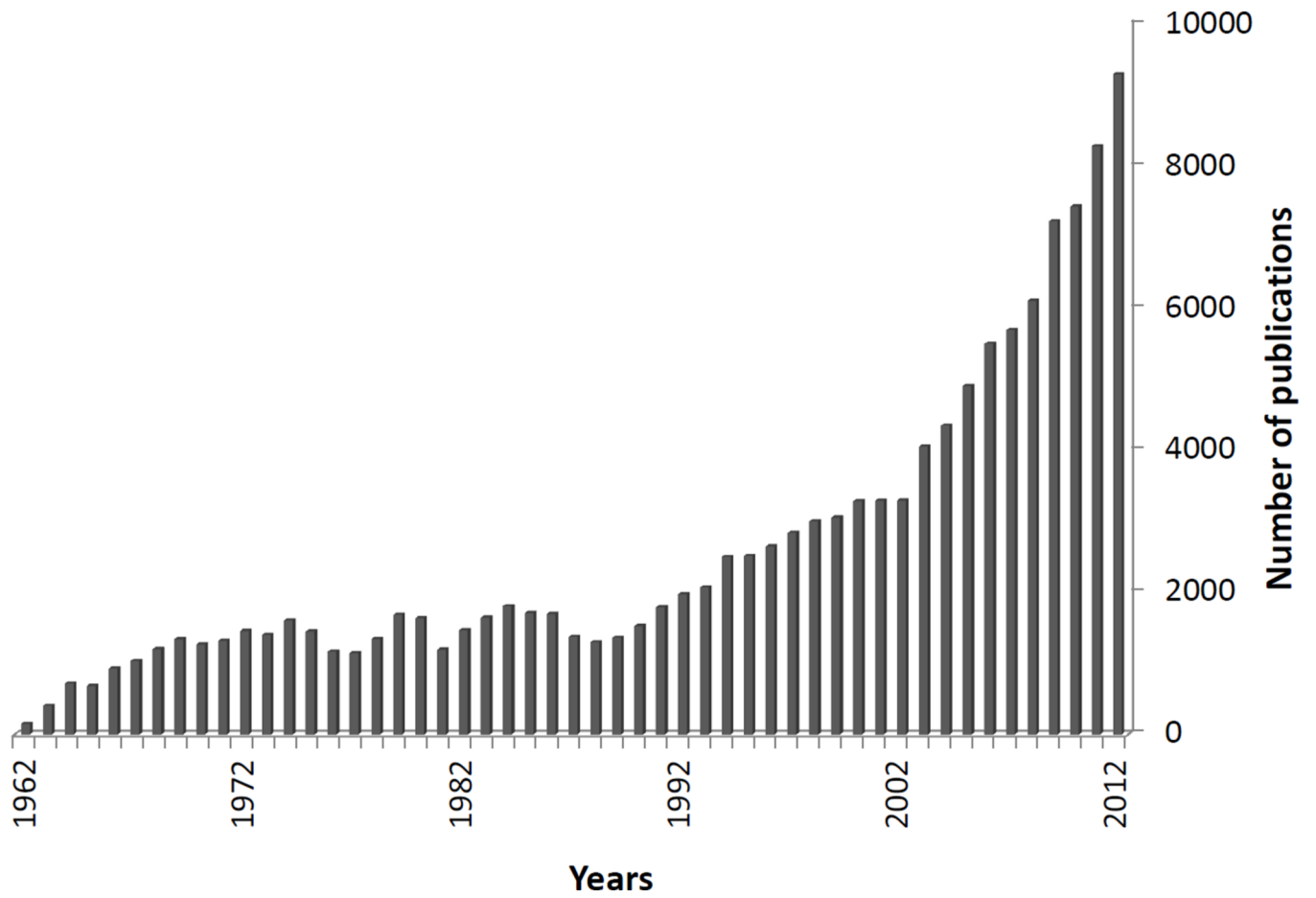- Common types of bias are discussed and potential solutions are proposed

Figure 1. Brain Imaging Studies over the Last 50 Years
Number of brain imaging studies published in PubMed (search term "neuroimaging",
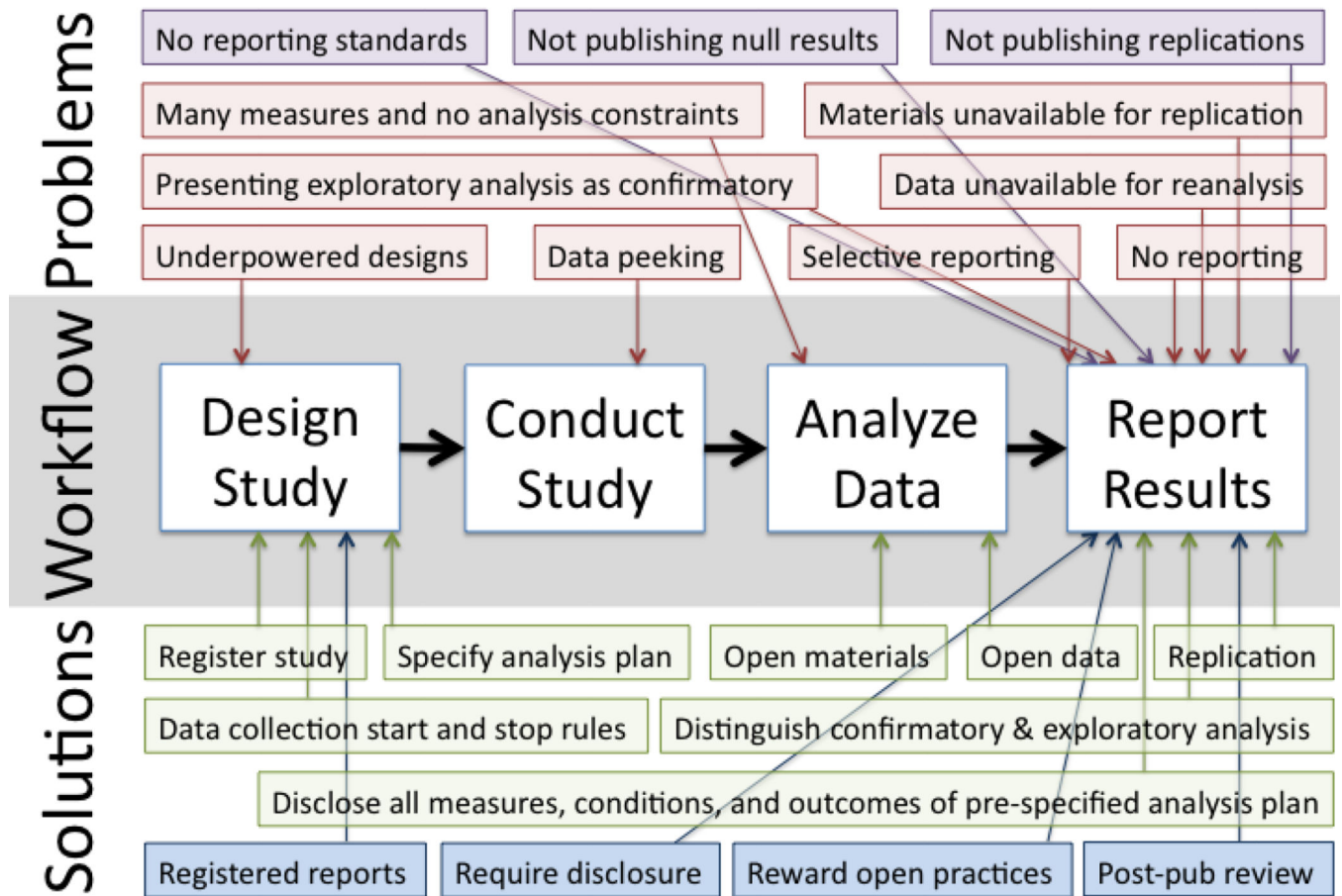updated up to 2012) over the past 50 years.

Figure 2. Common Practices and Possible Solutions
Common practices and possible solutions across the workflow for addressing publication biases. Red problems and green solutions are mostly controllable by researchers; Purple problems and blue solutions are mostly controllable by journal editors. Funding agencies may also be major players in shaping incentives and reward systems.