Read it, understand it, believe it, use it:

Principles and proposals for a more credible research publication

Don Green, Macartan Humphreys, Jenny Smith

11 December 2013

EARLY DRAFT: COMMENTS WELCOME

Abstract

In recent years concerns have been raised that second-rate norms for analysis, reporting, and data access limit the gains that should follow from first-rate research strategies. At best, deficient norms slow the accumulation of knowledge; at worst, they result in a body of published work littered with results that are flawed, fragile, false, or in some cases, fraudulent. Scholars across disciplines have proposed a number of innovations for journal reform that seek to counter these problems. We review these and other ideas and offer a blueprint for a "best-practices" social science journal.

1 Introduction

Scholarship in the social sciences is undergoing a transformation that has been characterized as a "credibility revolution" (Angrist and Pischke 2010). Growing emphasis is now placed on the integrity of scientific claims, especially claims about cause and effect. Two manifestations of this fundamental change in scholarly outlook are the growing use of randomized experiments and the application of statistical methods that invoke minimal assumptions about the data generation process. By allocating treatments at random, researchers reduce their reliance on untestable assumptions about unobserved confounders that might systematically distort causal inferences. Moreover, the underlying simplicity of a well-conducted experiment allows researchers to present statistical results in an unusually transparent and intuitive manner. Many of these benefits can also be enjoyed by well-crafted observational research that pays close attention to selection processes and the nature of likely confounders.

The optimism stemming from more credible research design strategies is accompanied, however, by a growing recognition that these gains may not translate into commensurate gains in knowledge if standards of analysis, reporting, and access are deficient. This focus on credible analysis and reporting has become more urgent in recent years. Many findings that were previously regarded as authoritative are coming under fire on the grounds that they are false or fragile, not robust to seemingly innocuous modeling decisions or incapable of replication in other settings. Indeed, several scholars have recently expressed concern that existing findings cannot be reproduced even with the original data (Dafoe 2013; Ioannidis et al. 2009). In some cases lack of transparency has allowed elementary flaws to go unnoticed (for a recent example see the discussion in Herndon, Ash and Pollin (2013) of Reinhart and Rogoff (2010)). In others there are concerns that results are simply fraudulent (see Simonsohn 2012).

The credibility revolution has not only taken aim at the failings of specific pieces of research; it has led scholars to take a hard look at the institutions within which science is conducted and the perverse incentives that they can create for researchers. Because no institution is more prominent in the professional life of university professors than the peer-reviewed journal, many of the pathologies of scientific inquiry have been attributed to the way in which academic journals operate.

The problem of how publication incentives can produce false or fragile results is illustrated in Figure 1. Say that estimates get published only when they exceed some critical value; for example if they reach a certain magnitude or if they reach a certain threshold of significance. In that case the *expected* value of the published outcome—marked in the figure with a black dot—provides a biased estimate of the true value—marked in the figure with the white dot. Published results, on average, thus give a misleading impression of causal effects. Moreover, published results are vulnerable to being overturned through replication, since the expected outcome from a replication using an unbiased procedure will be the true value, not the expected published outcome.

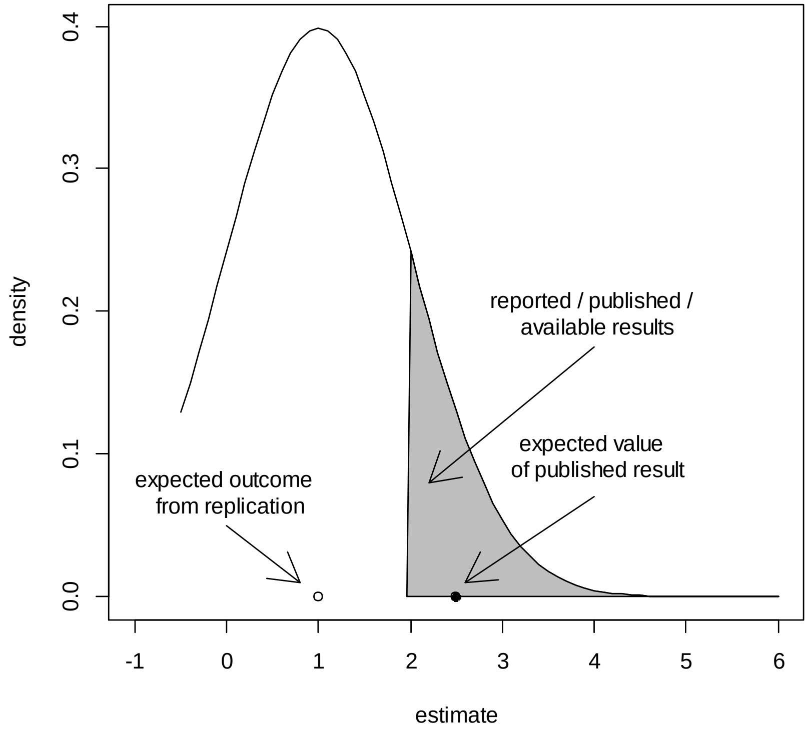

Figure 1: Illustration of the logic of reporting / publication bias. The histogram shows possible results from an analysis that uses some unbiased method to assess some quantity of interest. Say that estimates get published only when they exceed some critical value—such as all those in the shaded right tail of the histogram. In that case the expected value of the published outcome—marked with the black dot—provides a biased estimate of the true value—marked with the white dot. The expected outcome from a replication using an unbiased procedure will then in turn differ from the expected published result.

Such concerns cast doubt on the credibility of published research, especially in the social sciences. Although some of these concerns have been expressed for decades (see Sterling 1959), the extent of the biases has recently become more obvious. For a striking illustration see the analysis of publication bias in political science as documented by Gerber and Malhotra (2008), summarized in Figure 2.[1] Their evidence reveals a striking "surplus" in the number of coefficients that narrowly pass the 95% significance threshold in published articles. This pattern is what one would expect when journals favor manuscripts that report statistically significant results.

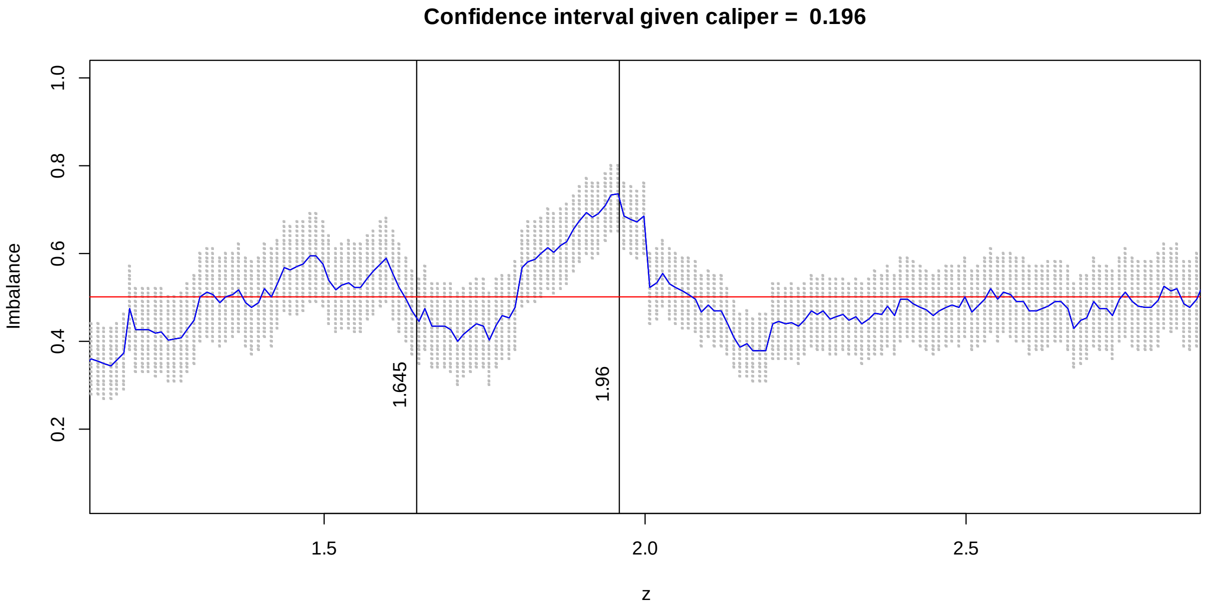

Figure 2: For each *z*-statistic in [1,3], the graph plots the probability that a statistic published in prominent political science journals is in the range [*z*,*z*+0.196] given that the statistic is in the range [*z*-0.196, *z*+0.196]. The graph peaks at 1.96, which is the value associated with significance at the 95% level; at its peak the probability is over 70%. Figure based on analysis and data from Gerber and Malhotra (2008).

It is only very recently that the research community has begun to seriously consider innovations for addressing the perverse incentives created by academic journals.

In this paper we identify a set of nine principles that help guide the choice of innovations to be adopted. These principles are intended to facilitate the central aims of research publication—consumers of research should be able to read it, understand it, believe it, and use it. In turn these principles guide the choice of innovations that should be prioritized to ensure the credibility and utility of social scientific research.

---

[1] Thanks to Grant Gordon and Al Fang for collaboration producing this graph.

2 Principles

Table 1 lists nine core principles for credible publication. The first two—that research be *accessible* and *uncensored*—relate to the availability of research findings and the role of analyst discretion. The third—*comprehensiveness*—is the requirement that published research report no less than the minimal amount of information needed to understand it. We classify three principles related to the believability of research: first that research be *free from error*, and second that it be *free from fraud*. Both of these are obvious principles, but they are ones that admit of relatively easy fixes that are not currently employed. The last of these is that research be *vetted*; again this is a standard principle of peer-reviewed research, but again innovations allow for fundamental improvements in the way that peer vetting is conducted. The last three principles relate to the utility of research. The first is that the research be *challengeable*, that is, that readers can stress test research findings; the second is that it be *reproducible*—that information provided to readers is sufficient to enable them to validate the claims independently; and the final principle is that the research be *productive* in the sense of contributing to the accumulation of knowledge. Again this is a widely shared goal of research but one for which there is scope for innovation.

Table 1: Goals, Principles and Innovations

Goal	Principle	Innovation
Read it	1 Uncensored	1. Registration of analysis plans
	1 Uncensored	2. Results blind review
	2 Accessible	3. Credibility before innovation
	2 Accessible	4. Open Access
	2 Accessible	5. Systematic archiving of data and materials
Understand it	3 Comprehensive	6. Content checklists
Believe it	4 Error free	7. Replication prior to publication
	5 Fraud free	8. Automated fraud checking
	6 Vetted	9. Enhanced powers for public peer review
	6 Vetted	10. Open evaluation prior to archiving
	6 Vetted	11. Labeling / Badges
Use it	7 Challengeable	12. Data availability
	8 Reproducible	13. Open materials
	9 Productive	14. Disciplinary posteriors

3 Innovations

We propose a strategy centered on fourteen recent innovations.

3.1 Get it

1 Registration of analysis plans

Researchers often base decisions about how to analyze their research results—or whether to report their analyses—on the patterns they observe in the data. While this sounds sensible in principle, in practice it opens the door to bias. Bias can enter at different stages: most obviously in terms of what results get published, and more subtly in terms of what results get reported in research write-ups. The scope for bias in at the reporting stage is surprisingly large. It is well appreciated that with 20 independent analyses on random data one should expect to report one significant result even though there is no true effect.[2] With latitude to adjust analyses by, for example, excluding outliers, selecting variables for covariate adjustment, or coding the outcome variable according to how modeling choices affect the statistical significance of the estimated treatment effect, the scope to generate "significant" findings increases enormously.

Registration of analysis plans seeks to address these two concerns. By providing a public record of planned research, registration can help interested consumers find results even if these do not enter the published record. In addition, registration helps to ensure the integrity of reported results. The idea here is also simple: researchers should describe, and make public, the analyses they plan to conduct before seeing the results of those analyses.

The case for registered research designs has long been recognized in the biomedical field, where many empirical studies take the form of randomized controlled trials of clinical interventions. Registration requirements for clinical trials have expanded rapidly in recent years, promoted by a variety of stakeholders. An early landmark was the 1997 FDA Modernization Act (FDAMA), which mandated registration of new-drug trials submitted for FDA review and led to the creation, in 2000, of the public online registry ClinicalTrials.gov. ClinicalTrials.gov allowed for the registration of any trial and enabled the disclosure of many study features not required by the FDAMA. In 2005, the International Committee of Medical Journal Editors (ICMJE) began to require timely, informative, and public

---

[2] To have a 95% chance of reporting at least one significant result you need latitude to implement about 58 tests (since $1 – 0.95^{58} \approx 0.95$).

registration of clinical trials "as a condition of consideration for publication" in its member journals; the ICMJE's registration requirements were broader than those of the FDAMA (De Angelis et al. 2004, 2005; Zarin et al. 2005).[3] Despite some investigator grousing about administrative burdens and forced disclosure of one's plans to research (and industry) competitors, the ICMJE's evaluation of its own initiative was positive (Laine et al. 2007). Since that time, the scope of registration in biomedical research has continued to grow: more trial registries have been established; the ICMJE policy has been expanded to cover more types of interventions and adopted by more journals; and formal checklists have been introduced to guide trial reporting and protocol design (Laine et al. 2007; Schulz et al. 2010; Chan et al. 2013). In 2007, a new FDA Act expanded statutory disclosure requirements to include a limited presentation of results (Wood 2009). By September 2010, ClinicalTrials.gov contained 79,413 trial registrations and 2,178 results records, with approximately 330 new registrations and 30 new results submissions arriving each week (Zarin et al. 2011).

Although formal registration of biomedical studies is now very common (Huser and Cimino 2013), formally "complete" registrations are not necessarily informative or useful. Studies of ClinicalTrials.gov and other trial registries find that many investigators provide only sketchy descriptions of outcome measures and interventions, that key methodological details may be omitted (protocol for assignment-to-treatment, anticipated controls for the results analysis, etc.), and that researchers often fail to register trials in a timely fashion and to update those registrations with trial results (Reveiz et al. 2010; Zarin et al. 2011; Gill 2012; Pino et al. 2012). Registration, even with mandated results reporting, has not eliminated the problem of "invisible and abandoned trials" (Doshi et al. 2013). Many registries, finally, provide only partially-adequate guidance even to conscientious investigators, although the major international registries perform well (Moja et al. 2009).

If the universal registration of informative protocols remains a work in progress in biomedical research, what is the outlook for registration in the social sciences? Calls for a social-science analogue to medical investigators' trial registries and published analysis plans have recently been heard from development economists (Rasmussen et al. 2011; Casey et al. 2012), psychologists (Wagenmakers et al. 2012), and

---

[3] Illustrating the iterative fashion in which medical registration requirements have expanded, the ICMJE's requirements were supported (initially exclusively) by ClinicalTrials.gov but broader than the FDAMA mandate. The editors adopted a 20-item "minimal data set" recently developed by the World Health Organization for its International Clinical Trial Registry Platform (ICTRP) (De Angelis et al. 2005; Sim et al. 2006). A second notable international initiative in clinical-trial registration is the work of the Ottawa Group (Krleža-Jerić et al. 2005; Krleža-Jerić and Lemmens 2009).

political scientists (Gerber and Malhotra 2008; Humphreys et al. 2013; Monogan 2013). Humphreys et al. (2013) illustrate the use of a full “mock report” as a pre-specified framework for evaluating a development intervention implemented via RCT. A more limited—and less labor-intensive—form of researcher pre-commitment is an archived “preanalysis plan” or “statistical analysis plan.” Pre-specification of a planned statistical analysis has been identified as a best practice in registration of randomized controlled trials (McPhail et al. 2006; Finfer and Bellomo 2009) and has also been advocated for observational research (Thomas and Peterson 2012; Onukwugha 2013); for sample implementations in the social sciences, see Casey et al. (2012) and Monogan (2013). Whether the study in question is experimental or observational, a formal analysis plan should be placed in the public domain before data analysis begins and, if possible, before the realization of outcomes.[4]

Proposals to institute registries have raised various concerns about the practical implications of registration requirements in social science journals. Perhaps the most important of these is that registration will place a brake on creativity by preventing scholars from developing hypotheses in tandem with data analysis. Our view is that exploration and transparency are not incompatible and that options for researchers who do not have well-formed hypotheses include (a) “null” registrations that signal that there are no specific analysis plans and (b) registered exploratory plans, in which researchers register the *strategies* they intend to use for exploration. Exploratory research plays an important scientific role and registration requirements should be structured so as not to prevent it; the point of preregistration, rather, is to clarify for the reader which research findings emerged from pre-specified analyses and which emerged as the analyst sifted through the results.

2 Results-blind review

Reviewers and publishers determine what to publish as a function of patterns of findings. This selectivity can produce biases that are as severe as biases arising from tendentious forms of data analysis. A response is to use results-blind peer review—a procedure in which reviewers determine

---

[4] This raises the question of whether registration of analysis plans is appropriate to observational research. This issue, which will be of central importance in many social sciences, has thus far been debated most vigorously by epidemiologists. A workshop report in favor of registering observational epidemiological studies (ECETOC 2009) was received favorably by the *Lancet* (2010) and the *British Medical Journal* (Loder et al. 2010; see also Williams et al. 2010; Swaen et al. 2011), but negatively by the editors of *Epidemiology* (2010) and by a number of scholars in the field (including De Jonge et al. 2011; Pearce 2011; Savitz 2011; Lash and Vandenbroucke 2012).

whether to publish research based on the design of the research, not the results. This innovation allows (though it does not require) the review process to occur *prior to research implementation*, a stage at which reviewers can focus directly on the importance of the research question and the adequacy of the proposed research design.

Proposals for results-blind review—either by reviewing research protocols at the design stage, as we suggest, or by withholding results when articles are submitted—have been around for some time (Walster and Cleary 1970; Newcombe 1987; Lawlor 2007; Smulders 2013). Those who have called for results-blind review share the goal of minimizing the biases in published research findings that can arise from reviewers' or editors' preference for splashy or statistically significant results. Notably, none argue for manuscript acceptance or rejection on the basis of research design narrowly conceived: while an appropriate design is a must, reviewers are also urged to consider the anticipated value of the proposed research.

At least one results-blind review model exists in practice: the "registered report" recently introduced by the neuroscience journal *Cortex* (Chambers 2013). Manuscripts for registered reports are invited for review while the associated research is in the design stage. Positive peer review of the proposed hypotheses, procedures, and analysis plan leads to "in principle acceptance"; manuscripts clearing this hurdle are slated to be published regardless of results, "pending quality checks" for which authors will have submitted their raw data and lab logs. While the *Cortex* initiative is unusual, other journals also offer some advantage in the manuscript pipeline to studies they have endorsed in protocol.[5]

There are a number of potential objections to results-blind review. Some are practical. If research has already been conducted, for example, it is likely that at least some reviewers will either know (say from a conference presentation) or be able to access research findings that have been excluded from a manuscript sent for review (Smulders 2013). Another objection is more substantive. For many research questions, the scientific value of a particular result is conditional on what that result is (Lawlor 2007). Figure 2 illustrates the general principle that how much is learned from a study depends on the result of the study. Each panel in the figure considers a situation in which there is a given prior belief over some

---

[5] *The Lancet*, for example, offers to review randomized-trial protocols and to "publish on our website summaries of those protocols that survive review." With respect to the final papers, the journal commits to send a manuscript for peer review if it originally published the protocol, though not necessarily to its publication (http://www.thelancet.com/lancet-information-for-authors/protocol-review).

quantity. This prior belief is itself based on a set of past observations ("Past *n*"). If the belief has been formed on the basis of a large number of past observations then there may be great confidence in the belief, and new knowledge will have relatively weak effects on beliefs; if priors are based on limited past observations (past *n* is low), then there is scope for larger shifts in belief. How large a shift there is in each situation, however, depends on both the findings of new research and the confidence researchers have in those findings.

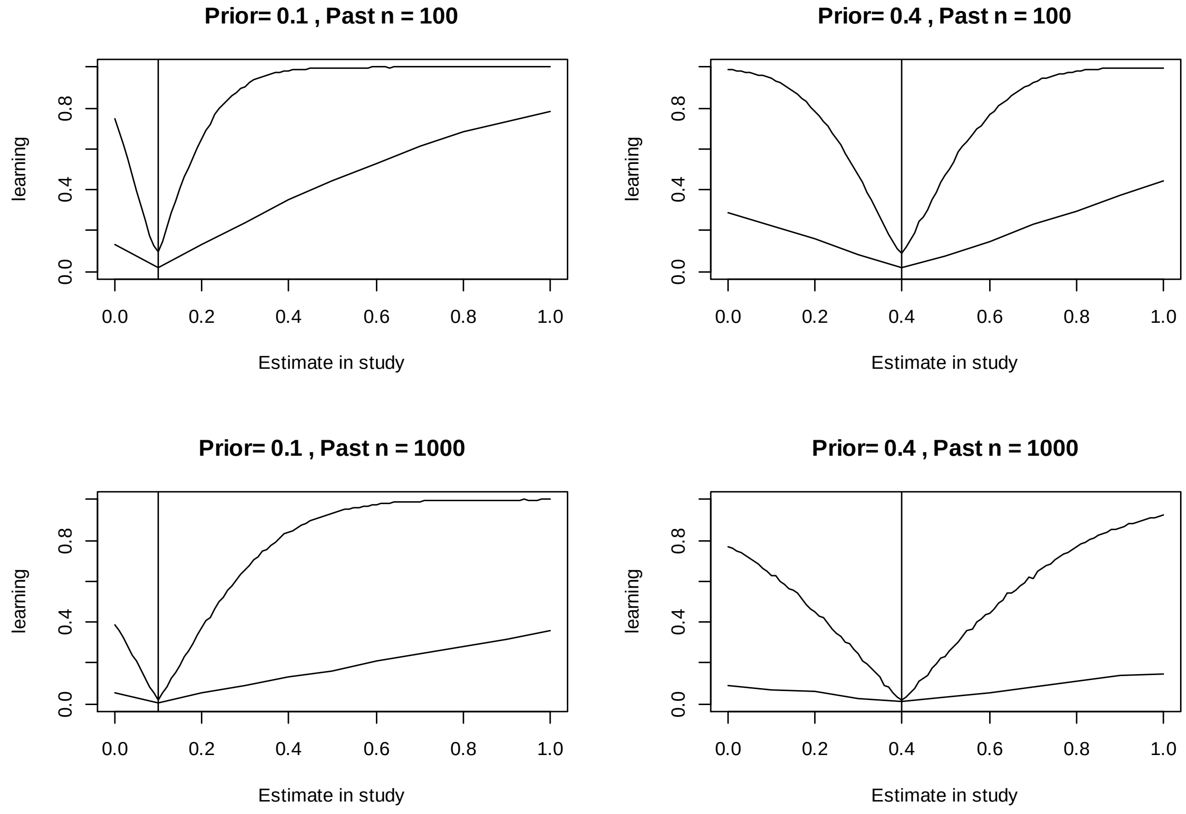

Figure 3. Learning, defined as the KS statistic of difference between prior and posterior distributions as a function of prior beliefs, study power, and study results. Priors are assumed to be given by a Beta distribution with mean of .1 (left column) or .4 (right column) and based on 100 (top row) or 1000 (bottom row) prior observations. Each graph shows the learning for given estimates (horizontal axis) from a study with 100 observations (top curve) and 10 observations (bottom curve). Learning is greater when results are distant from priors and when past priors are more diffuse.

For example, a study that finds that an object travels slower than the speed of light would add little to existing knowledge; if instead the study concluded that the object moved faster than the speed of light, the results would add considerably to knowledge. Results-blind reviewing treats the two studies equally. To address this problem, journals might adopt one of two solutions. They could accept or reject conditional on the *expected* gains to knowledge, which can be assessed *ex ante*, given priors on the issue in question. Such a strategy is results-blind, but the determination to publish is not a functiononly of the quality of the research design but also of the likelihood that it will shift beliefs. Alternatively, journals could develop a policy in which article placement or format is conditional on results: articles could be published in long form or short form, online or offline, or in different series; perhaps in a top-tier or second-tier outlet.

Some exploratory results will prove to be both interesting and true. For this reason—and to encourage researchers to report honestly how their results came about—even journals that prioritize results-blind review might set aside space for what Greenland (2007) terms "hypothesis-screening" analyses.[6] Reporting on such unanticipated findings would not be precluded by the failure of results-blind peer review. The findings from these analyses would of course require confirmation in future research, for which designs could be subjected to results-blind review.

3 Credibility before innovation

Key to overcoming biases is the need to reassess the weights placed on credibility and novelty when determining what research is published. Currently, top-tier journals seek to publish results that are both high on innovation and high on quality of implementation. The second tier publishes results that are weaker on one or both of these dimensions but typically favor novelty. Countering bias requires, however, a reversal of these priorities with the publication of credible research findings even if the contribution to knowledge is modest.

In the context of online publishing, the space constraint that formerly excluded second-tier manuscripts is less pressing; the operative constraints are the resources needed to maintain quality control over an expanding set of articles. Journals in some disciplines have created "proceedings" or "letters" publications (e.g., *Statistics and Probability Letters*) to facilitate brief, fast-track presentation of second-tier studies that might otherwise be consigned to the file drawer.

---

[6] As opposed to the more usual "hypothesis-generating." If a researcher thinks to test a hypothesis, it has already been "generated."

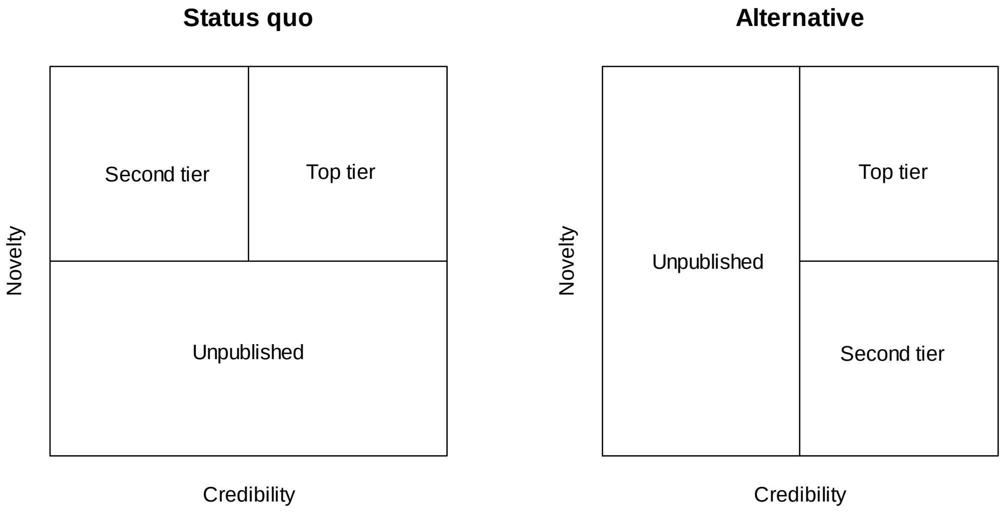

Figure 4: Top-tier journals seek to publish on the basis of credibility and novelty; second-tier journals face a tradeoff. We argue journals should compromise on the novelty margin but not the credibility margin.

4 Open Access

Results should not just be formally published; they should be accessible to the public. Open-access publishing—understood as free availability of published articles through the internet, with no rights, other than the right to citation, reserved—has grown exponentially in recent years, with one recent study estimating that 43% of peer-reviewed journal articles published from 2008-2011, and 36% of articles in the social sciences, were available through open access (Archambault et al. 2013: 14). Open access ensures free worldwide distribution of scholarly research, facilitates dialogue among researchers, and is a prerequisite for the kinds of crowdsourced vetting strategies described below. Recent critiques arguing that the rapid expansion of author-funded publishing may lead to declining standards (e.g., Jeon and Rochet 2010; Bohannon 2013; Butler 2013; Haug 2013) highlight the need for the open-source format to be coupled with other innovations.

Open-access scientific publishing is divided into two broad categories, “green” and “gold” (Albert 2006; Björk and Solomon 2012; Nosek and Bar-Anan 2012). “Green OA” refers to articles published in subscription journals that are also freely available online. Many green OA articles can be found in online repositories—including non-university sites like arXiv, PubMed Central, and SSRN (the Social Science

Research Network)—but the largest share continues to be "self-archived" on the websites of individual scientists. "Gold OA" refers to articles published in open-access journals.[7] The earliest OA journals were launched by individual scientists in the late 1990s; soon, some established journals began to convert their online editions to OA, and the two major OA-native publishers were founded in 2000. These publishers, the Public Library of Science (PLoS) and BioMed Central, introduced a funding model based on article-publishing fees, paid by authors or funders, in place of subscription revenue.[8]

Both OA "paths" have expanded rapidly over time, in absolute size and also as a share of scholarly publishing—though green has remained consistently larger than gold. Archambault et al. (2013) estimate gold OA at 10% of articles across all disciplines (9% in the social sciences) for 2008-2011; green at 33%. Laakso and Björk (2012) estimate that 12% of new peer-reviewed articles were published to immediate OA in 2011, with an additional 5% available within twelve months. Fully 49% of the immediate-OA articles appeared in outlets supported by article-processing charges.

How influential is open-access publishing? Studies that combine all forms of open access generally find that OA articles are *more*, not less, likely to be cited than their subscription-access counterparts (Lawrence 2001; Antelman 2004; Eysenbach 2006; Norris et al. 2008; see also Gleditsch et al. 2003 and Piwowar et al. 2007 on the citation effects of sharing data; Vandewalle 2012 on sharing code). However, the interpretation of this correlation has sparked debate. Studies investigating whether the OA citation edge is an artifact of researchers' choosing open access for their higher-quality work have produced conflicting results. Gargouri et al. (2010) compare "self-selective" with "mandatory" self-archiving and find an OA advantage in both categories; by contrast, a randomized controlled trial organized by Davis et al. (2008) found no difference in citation counts for articles assigned to OA and subscription conditions, although the OA articles did attract significantly more readers and downloads. Whatever its source, the OA citation advantage seems to disappear for gold OA alone. Archambault et al. (2013) find that the net OA citation increase is composed of a significant citation boost for green OA combined with a drop in citations for gold OA. At the same time, studies of journal-level citation counts (impact factors) indicate that some parts of the OA publishing universe are now highly competitive with

---

[7] A third type, "hybrid OA," describes journals that are primarily subscription-access but permit authors to choose immediate open access for their articles against payment of a fee. Among Elsevier journals, for example, almost two-thirds (~1500 of ~2400) are now hybrids (Liesegang 2013).

[8] BioMed Central was purchased by Springer in 2008, and the commercial share of open-access publishing has grown steadily over time (Laakso and Björk 2012).

traditional publishing. New OA journals (founded 2002-2011) and OA journals funded by publication fees are close to converging on their subscription equivalents in average citation rates (Björk and Solomon 2012; Solomon et al. 2013).[9]

The recent "sting" operation conducted by Bohannon (2013) suggests that extremely low standards of review are present in very many open-access journals. Although the study did not compare the quality of review in OA publications to traditional publications, it is possible that the business model employed in many OA journals—in which authors pay journals to publish their work—gives rise to weak incentives for journals to filter. The study also highlighted however the great heterogeneity of the group, with many journals filtering research efficiently while others were slow even to accept a retraction of a bogus article.

Findings on article citation and journal impact suggest that research publications are not doomed to obscurity or audience skepticism simply because they appeared in "gold open access" rather than in a traditional, subscription journal. Unfortunately, open-access journals have not necessarily been leaders in implementing other best practices in research publishing—as Bohannon's results amply demonstrate.[10] The flexibility of the online platform, however, means that open-access journals are especially well-positioned to adopt many of the innovations discussed here. They would increase their credibility and appeal by doing so.

5 Systematic Archiving of Data and Materials

Plans, materials, and data need to be archived in a way that makes them publicly accessible. Moreover, archiving must ensure that this information remains intelligible over the long-term without further support from the original investigators. Standards set by the Open Archival Information System (OAIS) provide a useful guide to best practices (OAIS 2012). Apart from the Inter-University Consortium for Political and Social Research (ICPSR), few archives in the social sciences currently meet these standards,

---

[9] The greatest impact-factor advantage was enjoyed by subscription journals established before 1996; no-fee OA journals lagged farthest behind. The latter, however, are disproportionately likely to be published in languages other than English, outside the major publishing countries (Solomon et al. 2013; Björk and Solomon 2012).

[10] McCullough (2009) asks why OA journals in economics have not followed (or led) the top journals in instituting mandatory data/code archives. Meerpohl et al. (2011) find that OA journals in pediatrics are somewhat but not much more likely than subscription journals to recommend practices like trial registration (23% of 69 OA journals) and adherence to the CONSORT guidelines (20%) in their author instructions.

and there is a clear need to professionalize the way in which datasets aredocumented and preserved (Albright and Lyle 2010; Gutmann et al. 2009). For more on data archiving and data access, see item 12 below.

3.2 Understand it

6 Content checklists for minimal reporting requirements

It should be possible not just to read results but also to understand them. In practice this requires that sufficient information be made available in published research to allow readers to assess the meaning and credibility of the statistical results. For experimental work, key information that should be provided systematically, but is often lacking, includes information on how random allocation sequences were generated, who implemented them, and whether administrators and subjects were blinded. Also crucial is information about whether subjects complied with the assigned treatment, whether they dropped out of the study, and, if so, how many attrited in each experimental condition. Published research that fails to conform to basic reporting standards is often unsuitable for meta-analysis.

Core requirements such as these are increasingly codified in formal checklists designed to guide trial reporting, meta-analysis, protocol design, and other research tasks. The oldest of these checklists is known as the Consolidated Standards of Reporting Trials, or CONSORT, revised most recently in 2007. CONSORT consists of a 25-item checklist for reporting the results of parallel trials, together with a flowchart for tracking the progress of individual participants through a study (Schulz et al. 2010; for more detail, see Moher et al. 2010 and consort-statement.org). Revisions to CONSORT have extended its scope several times, notably to nonpharmacologic interventions, and the CONSORT standards are currently in use at some 300 medical journals (Boutron et al. 2008, 2010).

The social sciences have been slow to adopt CONSORT and its descendants, which include PRISMA for meta-analysis (Moher et al. 2009), STROBE for the reporting of observational epidemiology studies (Von Elm et al. 2007), and SPIRIT for clinical trial protocols (Chan et al. 2013). Nevertheless, CONSORT does provide the foundation for the American Psychological Association’s Journal Article Reporting Standards (Cooper et al. 2008), and an initiative is currently underway to develop a CONSORT extension for social and psychological interventions (CONSORT-SPI; see Montgomery et al. 2013). Boutron et al. (2010) illustrate the application of CONSORT to a get-out-the-vote field experiment in political science.

3.3 Believe it

7 Replication prior to publication

There is growing interest in analytic replication of published results to determine whether reported results can be independently generated by other researchers *using the same data*. Calls for analytic replicability—known in many fields as the move to "reproducible research"—have recently been heard in a variety of disciplines.[11] While analytic replication by individual researchers serves a useful checking function—and often a pedagogical function as well—there is no reason this kind of replication should be conducted only in a decentralized and *ad hoc* way. Analytic replication ought to be publicly accessible and should be undertaken routinely by journals or outside certification groups prior to publication. Researchers' aspirations to expand knowledge via self-correction (Ioannidis 2012) and to influence a policy world increasingly committed to evidence-based decision-making (Cooper et al. 2008) require public, credible commitments to research accuracy and transparency.

Investigations in a variety of disciplines show that scholars are correct to be concerned about the analytic replicability of published research. In a notable, early study at the *Journal of Money, Credit and Banking*, Dewald et al. (1986) were able to reproduce the results of just two of 54 *JMCB* articles on the strength of data sets provided by authors.[12] Data availability is a major hurdle for many would-be replicators. When McCullough et al. (2006) revisited the *JMCB*'s data/code archive for 1996-2003, they found that only 69 (36%) of 193 empirical articles that should have had archive entries in fact did so. After struggles with missing or unlabeled data and idiosyncratic code, McCullough et al. were able to attempt replication of 62 studies; 14 (23%) replicated completely. Ioannidis et al. (2009) attempted to

---

[11] These include computing and computational science (Fomel and Claerbout 2009; LeVeque et al. 2012; YLS Round Table on Data and Code Sharing 2010), biomedical sciences (Delescluse et al. 2012; Huang and Gottardo 2013), epidemiology (Peng et al. 2006), environmental science (Boose et al. 2007), economics (Hamermesh 2007; Anderson et al. 2008), sociology (Freese 2007), and political science (King 1995; *International Studies Perspectives* 2003; Dafoe 2013). In image-intensive biosciences, image manipulation has been a major concern (Rossner and Yamada 2004; Rossner 2006).

[12] In 1982, the *JMCB* launched a pioneering program to request and store authors' original code and data for the use of interested readers. For its replicability study, this *JMCB* Project requested code and data from 92 authors covered by the new policy and a comparison group of 62 who had recently published in the journal (Dewald et al. 1986). 26% of the first group and 65% of the second either did not reply or could not provide data. Of the first 54 data sets received, Dewald et al. judged that eight (15%) were "complete enough to allow an attempt at replication"; two of the eight replicated successfully, and two more with minor inconsistencies. Perhaps ironically, this early *JMCB* archive was discontinued and destroyed following a later change of editors.

reproduce a key table or figure in each of 18 articles in *Nature Genetics*. Two could be reproduced "in principle," six "partially or with some discrepancies," and ten not at all—mostly due to data problems, although 16 of the articles linked to data in a public repository. A survey of 123 political scientists who had attempted replication of published results found that 65% reported important elements of the replication files missing (Dafoe 2013). Although 48% of respondents had nonetheless been "able to precisely reproduce the main results," only 25% agreed that "most or all of the key results were robust."

How can this dismal record of non-replicable research be improved? The reporting checklists described in item 6, such as CONSORT, are an important contribution to research transparency, as are disciplinary efforts like the American Psychological Association's Journal Article Reporting Standards (Cooper et al. 2008). Increasing numbers of journals now request or require replication data from authors, and some high-profile journals have instituted pro-active policies to monitor compliance. The *Nature* journal family instituted "spot-checks" in 2007; the *Journal of Cell Biology* screens all accepted manuscripts in this way (*Nature Immunology* 2007; Mellman 2007). Still, replication policies remain far from universal—Gherghina and Katsanidou (2013) found that only 18 of 120 political science journals had a replication policy posted online—and journals themselves rarely do more than retain the data forwarded by researchers. Dissatisfaction with the status quo has prompted transparency initiatives from individual scholars, including a pilot "Replication Audit" in political science (Dafoe 2013) and the creation of the website PsychDisclosure.org, a platform for authors to publicly disclose methodological details not required by current reporting standards (LeBel et al. 2013).

On the more optimistic side of the ledger, the rapid development of computing power and networking capability has made possible a wide variety of new tools for reproducible research. These tools close the credibility gap in ways that go well beyond preventing misrepresentation or outright fraud. They include online platforms for data/code storage and reuse, such as RunMyCode.org (Stodden et al. 2012); online research environments, like the open-access biomedical research platform Galaxy (Goecks et al. 2010; galaxyproject.org) and the bioinformatics workflow repository myExperiment (Goble et al. 2010; myexperiment.org); cloud-supported "snapshot" reproduction and storage of research computing environments (Dudley and Butte 2010); and "literate" data-analysis tools, such as Sweave, that enable the integration of statistical analysis and report-writing (Leisch 2002). A particularly ambitious goal is the "executable paper," which is designed to replace static research articles with digital objects fully capable of reproduction and validation by other scholars (Brammer et al. 2011; Li-Thiao-Té 2012; Ciepiela et al. 2013). Yet another group of scholars is seeking ways to increase the transparency and replicability of qualitative and mixed-methods research (Lieberman 2010; Moravcsik 2010). Our proposal for independent or journal-based replication prior to publication complements these other tools for producing and validating reproducible—analytically replicable—research.

What are the prospects for institutionalizing analytic replicability in the social sciences? One important concern—and one that has especially interested economists and political scientists (Hamermesh 2007; Anderson et al. 2008; Dafoe 2013)—has to do with scholars' incentives to provide the data and documentation required to make their own research reproducible.[13] The shift to an analytic-replicability regime imposes costs—in time and trouble, not to mention the potential risk of exposing one's mistakes—on individual researchers in exchange for benefits to the scientific community at large. For this reason, reform seems more likely to succeed if initiated at the institutional rather than the individual level. Over time, as analytic replication becomes a routine early-stage event in the life-cycle of a scholarly publication rather than an infrequent late-stage one (Nosek and Bar-Anan 2012), new norms and expectations surrounding the practice of replication may become self-reinforcing (Dafoe 2013).

8 Automated fraud checking

Just as journals can routinize analytic replication, so too can they routinize fraud checking. Academic fraud is widely believed to be rare, but recent years have witnessed several high-profile cases that caused widespread concern about the extent and severity of the problem (Callaway 2011; Rossner 2007; Titus et al. 2008). Although the transparency-promoting tools and standards described above do not have the primary purpose of preventing fraud, that is certainly one of their effects. For example, a would-be fraudster confronted with a pre-publication replication requirement of the type just proposed would need to create a plausible fabrication of raw data—a task requiring considerable knowledge and skill (Simonsohn 2012). Researchers have also developed a variety of statistical techniques that indicate the likely presence of fraud in published findings (see, for example, Simonsohn et al. 2012 on left-skewed *p*-value distributions as evidence of "*p*-hacking"). Clever fraudsters may devise schemes to fool such tests, of course, and the tests in turn may need to become ever more sophisticated in order to

---

[13] A second family of concerns is legal. For a discussion of these issues, see for example Stodden's (2009) proposal for a copyright-displacing Reproducible Research Standard.

raise the cost of committing fraud. Since fraud-checking algorithms are inexpensive for journals and individual investigators to implement, their use represents a natural step to provide reassurances to readers on the credibility of results.

9 Intensive and transparent peer review

The current system of peer review invites 2-5 reviewers to evaluate a manuscript without access to its data. This content-focused screening method has advantages inasmuch as it focuses attention on the connections the author draws between the new research and existing literature and theory. We propose supplementing this system with at least one additional reviewer who is instructed to devote primary attention to the data analysis with the original data in hand. This kind of reviewing model is not, to our knowledge, used in the social sciences,[14] perhaps because of the demands it places on reviewers and because of authors' concerns that their data and ideaswill be stolen by unscrupulous reviewers. A remedy to both problems is to assign this form of intensive review to an associate editor, who has a recognized obligation to maintain an arms-length relationship to the data while the manuscript remains unpublished.[15]

10 Extensive peer review: Crowdsourced evaluation prior to archiving

The number of minds tasked with reviewing a manuscript prior to publication is generally small. But for prominent research, the number of knowledgeable readers may be very large. For some kinds of problems, broad knowledge can complement focused scrutiny. For example, a given reader may recognize immediately that a seemingly novel claim is a special case of a more general result; a scrupulous intensive reviewer might faithfully confirm the special case but be unaware of the general

---

[14] Statistical review is in use in other disciplines, however; *The Lancet* was submitting provisionally-accepted papers to external statistical review as early as 1990. Its outside statisticians' recommended "reject" rate was twice that produced by the statistical checklists then in use at the *British Medical Journal* (Gore and Jones 1992).

[15] Alternatively, if this task were entrusted to reviewers, one method of reducing reviewer workload (and perhaps of increasing author security) would be to develop a web-based platform on which reviewers could inspect and re-execute a manuscript's statistical results without needing to install new software, download data and code, and figure out how to run the author's analysis on their own. Leisch et al.'s (2011) contribution to the "executable papers" challenge is a proposal along these lines, for R and Sweave.

result. Crowdsourced reviewing may provide a way to exploit decentralized knowledge of this form. We propose a mechanism to integrate moderated crowdsourcing into the review process. The core idea is to introduce a probation period, for example of six months, between conditional acceptance and formal publication. During this period papers and data would be published online, accompanied by a discussion board open for comment. Comments would be moderated by the associate editor responsible for the piece, who would determine which submitted comments would be opened as "cases" to which the author has the opportunity to respond. Registered users could offer signed comments on issues they believe require resolution prior to formal publication, and associate editors would make a ruling on these prior to publication and archiving.

A review system along these lines is currently in use at the open-access journal *Atmospheric Chemistry and Physics* (Pöschl 2004; atmospheric-chemistry-and-physics.net). Manuscripts submitted to ACP first undergo "rapid access peer-review" by a journal co-editor; technical corrections may be made at this stage. Manuscripts that survive this review are immediately published on Atmospheric Chemistry and Physics Discussions (ACPD), an online platform for what the journal calls "interactive public discussion." Reviewer comments (anonymous or attributed, at reviewers' option), author responses, and short attributed comments from other scholars are all published on ACPD (which, like ACP itself, is permanently archived). Once authors have revised their manuscripts in response to this discussion, the journal editors decide on final publication, and accepted papers appear in ACP. An ACP editor (Pöschl 2004) sees advantages to this process on all sides. Readers gain from early access to new research, from the opportunity for discussion, and from high-quality papers at the final stage. Authors gain from rapid and widely-sourced feedback and from the opportunity to reply to criticism directly as well as via manuscript revisions. Editors and reviewers, finally, benefit from added administrative efficiency due to the reduced risk that manuscripts will be assigned to inappropriate or uninterested reviewers. The entire structure creates disincentives for the submission of sloppy or fraudulent work.

Additional suggestions and models for post-publication peer review have recently been outlined by a variety of scholars (Le Borgne and Campo 2011; Kriegeskorte 2012; Nosek and Bar-Anan 2012; Wicherts et al. 2012; Witten and Tibshirani 2012).[16] These authors favor open, online post-publication discussion

---

[16] Lawlor's (2007) proposal for "prior to results submission" also has a post-publication dimension. She suggests that papers accepted prior-to-results might be published without a discussion section, and that the authors'

of published research, and they make a variety of suggestions for moderating discussion, reviewing the reviewers, disclosing the identities of reviewers, and other issues. Some take the ACP model a step further, advocating a publishing model with very low barriers to "publication" followed by intensive, open-ended "evaluation" (Nosek and Bar-Anan 2012; see also some of the contributors to Kriegeskorte 2012). Creating a platform for crowdsourced, post-publication peer review overcomes one important limitation of the current journal system, which imposes a high barrier to publishing any type of reply or comment in the place where the original article appeared (Witten and Tibshirani 2012). The prospect of encountering ongoing evaluation may encourage authors to be more careful in the pre-publication phase of manuscript preparation as well.

There are many obstacles to high-quality pre-publication manuscript reviews. Reviewers, especially those who write expeditiously and insightfully, are overworked—not for nothing do Hochberg et al. (2009) refer to a "tragedy of the reviewer commons." Moreover, appropriate reviewers may be difficult to identify; only a small number of reviews can reasonably be solicited for any given manuscript; and the promise of anonymity may create opportunities for bad behavior by individual reviewers and deference to that behavior by editors. To create incentives for thoughtful and judicious evaluation of manuscripts may require some form of what Wicherts and coauthors (2012) dub "peer-reviewed peer review."

11 Labelling

A last component of vetting is the *communication* that vetting has taken place. There can be considerable confusion over whether peer review has been undertaken. The Reinhart and Rogoff (2010) study called into question by Herndon, Ash and Pollin (2013), for example, was published in the *American Economic Review*'s annual *Papers and Proceedings* issue, which unlike other issues of the *Review* is not peer reviewed.

A set of recognized labels could provide rapid information as to whether a given piece of research has been vetted in various ways. The Center for Open Science has proposed a set of "badges" to communicate quickly whether a piece has been preregistered and whether relevant data and materials are made available to the public (see Figure 5). Labelling has the additional advantage of publicly

---

original manuscript discussion of how the various possible results could be interpreted would serve as the first entry of an online discussion board on which all readers could discuss the eventual findings.

identifying the purpose of research: for example, whether a given article is providing primarily exploratory or primarily confirmatory evidence.

Figure 5: COS Badges

Note: Badges generated by the Center for Open Science (openscienceframework.org). The project's "aim is to specify a standard by which we can say that a scientific study has been conducted in accordance with open-science principles and provide visual icons to allow advertising of such good behaviors" (Peirce et al. 2013).

Such labelling could even be incorporated in the citation process, providing readers with immediate information to help assess competing claims. For example, by citing "Murphy (2010 ☑🔓)" or "Kelly (2009 ☒🔒)" the author indicates whether the cited articles are peer reviewed and have open data.

3.4 Use it

12 Access to data. Readers can best assess the quality of research when they can analyze the data themselves. Whereas journals and reviewers need access to data in order to ensure error-freeness, readers need access in order to probe arguments: data availability makes published findings challengeable. Given reasonable disagreements over what types of analysis are appropriate for different types of problems, the ability to interact with published analyses gives readers direct access to the knowledge generated by researchers and reduces the extent to which that knowledge is constrained by researchers' analysis and interpretation choices. The value of first-hand inspection of results provides a rationale for complete data access independent of the need to verify that the data generate the results claimed by authors. This rationale is already recognized by researchers: In a survey conducted for the NSF-funded DataONE project, 67.2% of responding scientists agreed that "lack of access to data generated by other researchers or institutions is a major impediment to progress in science"; 83.5% (including 80% of social scientists) agreed that "I would use other researchers' datasets if their datasets were easily accessible" (Tenopir et al. 2011; N=1329, including 204 social scientists).

The principle of access to data extends also to qualitative research, with some scholars calling for "active citation" standards that are more "precise, annotated, and primary," a goal more easily achieved with internet based publishing (Moravcsik 2010).

Pleas for increased data access can be heard from scholars in a wide variety of disciplines, especially in the life sciences.[17] Journal policies on data availability vary widely, as does authors' adherence to those policies. In a study of 500 articles published in the highest-impact scientific journals in 2009, Alsheikh-Ali et al. (2011) found that 70% were subject to a data-availability policy, with which 59% of that subset did not fully comply. Article-integrated links to supplementary data resources may not work, even just a few years after publication (Evangelou et al. 2005; Anderson et al. 2006). Replication data underlying published studies may be impossible to obtain even if authors have pledged to adhere to a journal's data-sharing policy. Wicherts et al. (2006) made energetic attempts to obtain data from the authors of recent articles in four American Psychological Association journals; after repeated contacts and much forwarding of requested information (analysis plans, signed confidentiality pledges, ethics-board approvals), they were able to obtain 64 (25.7%) of the 249 data sets they had sought. Savage and Vickers (2009), whose pursuit was less dogged but who did prompt authors with a reminder of journal data-sharing guidelines, received only one of ten raw data sets they requested from authors of articles in *PLOS Medicine* and *PLOS Clinical Trials*.

More optimistically, these studies suggest that journal and institutional policies on data access can make a significant difference in the availability of data for reanalysis by other scholars. While only 47 (9%) of the articles examined by Alsheikh-Ali et al. (2011) made primary raw data fully available online, all 47 had appeared in journals with formal data-availability policies. Working links to online supplementary materials were much more likely to have been maintained when journals rather than authors were responsible for storage (Anderson et al. 2006). In a study of published research in population genetics, Vines et al. (2013) found that online data availability was only marginally higher for journals that

---

[17] See for example Hanson et al. (2011); Wolkovich et al. (2012) and Reichman et al. (2011) on ecology; Koslow (2000) and Pitt and Tang (2013) on cognitive science; Walport and Brest (2011) on public health; Ross et al. (2012) on cardiovascular medicine; and Beck and Neylon (2012) on archaeology.

recommended data archiving as against those with no policy. Mandatory archiving, however, was associated with greater data accessibility, especially when journals also required authors to include a data-availability statement within accepted manuscripts.

In addition to the stimulus provided by journals' data-access mandates, observers in this area point to researcher incentives—more precisely, researcher disincentives—as a primary influence on scientists' data-sharing choices (Wicherts et al. 2006; Nelson 2009). In a recent survey of biodiversity researchers, only 37% expressed willingness to release data upon article publication, with an additional 30% prepared to share after an embargo period of 1-3 years (Huang et al. 2012; N=372). In the survey for DataONE conducted by Tenopir et al. (2011), 78% of respondents (and 70% of social scientists) expressed willingness to "place at least some of my data into a central data repository with no restrictions"; 41% (32% of social scientists) were willing to do this with all data. (With respect to current practice, 75% agreed that "I share my data with others," although only 36% also agreed that "others can access my data easily.") What motivates the decision not to share? The top reasons provided in the DataONE survey were lack of time (54%), lack of funds (40%), "do not have rights to make data public" (24%), and "no place to put data" (24%). Among the biodiversity researchers, the most commonly cited concern was the "need to do more analyses of my published dataset" (51%), although other expressed concerns would admit of institutional solutions (29% agreed with the statement "I do not know any properly public database to archive my data").

Where should responsibility for data access rest? One approach is to formulate standards directed primarily to authors. For example, the American Political Science Association's draft "Guidelines for Data Access and Research Transparency" (APSA 2013) lay out the professional responsibilities of individual researchers who publish in the Association's journals. What the research just described suggests, however, is that both short-term compliance and long-term access will be better guaranteed by institutional commitments than by exhortations to individual scholars. Requirements established by journals, especially high-profile journals (Dafoe 2013), have a profound effect on whether researchers in fact share their data.

At the same time, new collaborations and information systems may lower the cost of data archiving, documentation, and long-term preservation. One model for consideration is the life-sciences repository Dryad (Vision 2010; datadryad.org). Dryad serves a consortium of journals as an open-access data repository, with the option of making data privately available to reviewers while a manuscript is under consideration; at the same time, it also serves individual researchers as a platform on which to deposit and publicize “orphan” data. By providing a unified infrastructure for data storage and access—automated solicitation of data files from authors, integration in the review process, version control, permanent archiving after publication—Dryad reduces the burden of maintaining data availability for both authors and journals. By ensuring open access and attempting to maximize discoverability for the materials in its repository, Dryad provides a service to the scientific community as well.

There are, however, some important limitations to the Dryad model and data repositories like it. One is that the documentation of the archived datasets is only as good as the meta-data provided by the researchers. The quality of meta-data for self-archived datasets in the social sciences is often poor and may not support independent reuse of data or replication of published studies. One of the reasons offered by psychologists for their inability to make data available was the necessity of writing a codebook (Wicherts et al. 2006, 2011). If codebooks and other meta-data are produced more easily and completely during the research process and then provided alongside primary data as a matter of course, researchers will find that other scholars’ interest in their data imposes little additional cost to themselves. Another limitation of most data repositories is that, unlike data archives, which conduct a full review of incoming data and documentation, repositories do not check for sensitive or missing information, nor do they verify that replication code runs properly with the accompanying data. Efforts are under way to address these shortcomings by establishing dedicated replication archives committed to data review (Peer 2013; Peer and Green 2012), by building partnerships among ICPSR and other data repositories to create services and tools for data curation and review, and by articulating best practices for the archiving and preservation of social-science datasets (UKDA 2013; ICPSR 2013). Until this infrastructure is in place, however, it will be up to journals to maintain quality control over replication data, computer code and meta-data deposited by authors.

13 Open materials. No less important than access to data is access to the materials used to generate that data in the first place, including protocols, instruments, and field notes that describe implementation or shed light on the context within which the study occurred. A repository for data in the social sciences must have sufficient flexibility to accommodate the many kinds of meta-data that accompany social-scientific projects.[18] Although preparing these materials for general consumption would undoubtedly impose a burden on researchers in terms of time and trouble—reprising one of the researcher disincentives to data-sharing, above—it would reduce barriers to sharing by another route. The Open Science Framework facilitates this kind of documentation by maintaining complete sets of research materials and real-time logs of research activities (opensciencframework.org; see also Nosek and Bar-Anan 2012; Nosek et al. 2012). This kind of pay-as-you-go approach lowers the costs that would ordinarily be borne by the researcher at the very end of the research process.

14 Updating disciplinary posteriors. The final innovation needed is a method for determining the gains to knowledge from any piece of research. Research findings should not simply stand on their own; they should fit into something and add to something. We propose a mechanism whereby journals systematically gather data on "disciplinary" priors—the priors of readers and reviewers for a given claim. Elicitation of priors is most meaningful and easily done in cases in which journals review prior to the generation of results. With priors in hand, journals can post alongside published articles the shift in posterior that is implied by the research, ultimately allowing readers to assess what has been learned. A nice feature of prior elicitation for not-yet-generated results is that it is possible to assess the predictive skills of different assessors and over time accord greater weight to the beliefs of well-informed assessors.

4 Conclusion

We have reviewed a broad array of innovations that could be employed by journals seeking to improve standards of research publication in social science.

In Table 2 we propose specific recommendations that could be employed by a social-science journal. None of these recommendations appear to be in conflict with each other (except insofar as theymake competing demands on resources)—in fact, several recommendations are clearly complementary. For example, results-blind peer review, and specifically peer review prior to research implementation (Recommendation #2), provides a forum for piloting a mechanism for the elicitation of priors

---

[18] Much of this meta-data would presumably be qualitative; on archiving qualitative data, see Elman et al. (2010).

(Recommendation #14). Open access (#4) facilitates open evaluation (#10), and pre-publication replication (#7) can facilitate open data and archiving (#5, #12).

Table 2: Summary of Innovations

Innovation	Costs and risk	Requires inter-journal coordination?
1. Registration of analysis plans should be formally recognized by journal	Requires social-science registry open to experimental and observational research	Most effective with common registry
2. Results-blind review should be available to authors, though this should not preclude ex post review	Minimal	No
3. Credibility before innovation: online space should be provided for credible results even if these exhibit little innovation	Minimal	No
4. Articles should be available via open access	Requires financing structure	No
5. Ssystematic archiving of article data and materials should occur prior to publication	Requires financing structure	Most effective with common archiving standards
6. CONSORT-style checklists should be required of all publications	Additional burden on researchers	No
7. Replication prior to publication should be implemented by in-house statistician or associate editor	Minimal	No
8. Automated fraud checking should be implemented by in-house statistician or associate editor	Minimal	No
9. Peer reviewers should have access to data	Minimal	No
10. An open-evaluation platform should be created for the identification of errors or redundancy, and for the discussion of research findings	Possible risk of non-participation by readers	No
11. Badges should be employed to signal peer review, open data and materials, and registration (when applicable)	Minimal	Most effective with common standards
12. Open access to data and accompanying documentation should be a precondition of publication	Minimal	No
13. Open materials should be a precondition of publication	Minimal	No
14. A pilot mechanism should be developed for the collation of disciplinary priors for all studies that are conditionally pre-accepted before implementation	Requires support for a central exchange Possible risk of non-participation by readers	Most effective with centralized exchange

We believe that most of these innovations carry little or no risk. Some create burdens on researchers and journal staff, and some require effective institutions that would have to be supported and staffed—but it should not be surprising that refurbishing the journal system will require investment in basic infrastructure and personnel. This financial hurdle raises the question of how to fund the transition to a new journal system and make it sustainable over time. Although this question goes beyond the scope of this document, we suspect that the answer lies in some combination of support from public and private foundations, subsidies from professional associations, and fees charged on a sliding-scale basis to publishing authors. These practical questions—as well as the broader question of how to implement and sustain a best-practices model—are perhaps best addressed through the creation of a new flagship journal in the social sciences. Defining the scope and format of this journal is currently the subject of discussion among potential editors, funders, and publishers.

We close with a word on the cultural implications of the innovations we advocate. The core aim of the innovations described here is to facilitate the process by which scientifically sound knowledge is created and disseminated. Separating good science from bad requires close attention to detail, but the current journal system keeps reviewers and readers at an arm's length from these details. It is often difficult to figure out precisely how a study was conducted or how the data were analyzed. One can well understand why this equilibrium has come about. For those doing shoddy work, the current system prevents detection. But even those doing first-rate work might be reluctant to embrace a system of transparency that makes it easier to discover and publicize errors, especially in a culture that sensationalizes such discoveries. The innovations described above help propagate procedures and incentives that prevent errors, normalize the process by which errors are detected and corrected, and greatly reduce the sensationalism that surrounds the corrective process on which science depends.

5 References

Albert, Karen M. 2006. “Open Access: Implications for Scholarly Publishing and Medical Libraries.” *Journal of the Medical Library Association* 94(3): 253-262.

Albright, Jeremy J., and Jared A. Lyle. 2010. “Data Preservation Through Data Archives.” *PS: Political Science and Politics* 43(1): 17-21.

Alsheikh-Ali, Alawi A., Waqas Qureshi, Mouaz H. Al-Mallah and John P.A. Ioannidis. 2011. “Public Availability of Published Research Data in High-Impact Journals.” *PLOS One* 6(9): e24357.

American Political Science Association. 2013. Draft Guidelines for Data Access and Research Transparency for Quantitative Research in Political Science. Manuscript (July 28).

Anderson, Nicholas R., Peter Tarczy-Hornoch and Roger E. Bumgarner. 2006. “On the Persistence of Supplementary Resources in Biomedical Publications.” *BMC Bioinformatics* 7: 260.

Anderson, Richard G., William H. Greene, B.D. McCullough and H.D. Vinod. 2008. “The Role of Data/Code Archives in the Future of Economic Research.” *Journal of Economic Methodology* 15(1): 99-119.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. “The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics.” *Journal of Economic Perspectives* 24(2): 3-30.

Antelman, Kristin. 2004. “Do Open-Access Articles Have a Greater Research Impact?” *College & Research Libraries* 65(5): 372-382.

Archambault, Eric, Didier Amyot, Philippe Deschamps, Aurore Nicol, Lise Rebout and Guillaume Roberge. 2013. *Proportion of Open Access Peer-Reviewed Papers at the European and World Levels – 2004-2011*. Science-Metrix report for the European Commission DG Research & Innovation.

Beck, Anthony, and Cameron Neylon. 2012. “A Vision for Open Archaeology.” *World Archaeology* 44(4): 479-497.

Björk, Bo-Christer, and David Solomon. 2012. “Open Access Versus Subscription Journals: A Comparison of Scientific Impact.” *BMC Medicine* 10(73).

Bohannon, John. 2013. “"Who's Afraid of Peer Review?” *Science* 342 (6154): 60-65.

Boose, Emery R., Aaron M. Ellison, Leon J. Osterweil, Lori A. Clarke, Rodion Podorozhny, Julian L. Hadley, Alexander Wise and David R. Foster. 2007. “Ensuring Reliable Datasets for Environmental Models and Forecasts.” *Ecological Informatics* 2: 237-247.

Boutron, Isabelle, David Moher, Douglas G. Altman, Kenneth F. Schulz and Philippe Ravaud for the CONSORT Group. 2008. “Extending the CONSORT Statement to Randomized Trials of Nonpharmacologic Treatment: Explanation and Elaboration.” *Annals of Internal Medicine* 148: 295-309.

Boutron, Isabelle, Peter John and David J. Torgerson. 2010. “Reporting Methodological Items in Randomized Experiments in Political Science.” *Annals of the American Academy of Political and Social Science* 628: 112-131.

Brammer, Grant R., Ralph W. Crosby, Suzanne J. Matthews and Tiffani L. Williams. 2011. “Paper Mâché: Creating Dynamic Reproducible Science.” *Procedia Computer Science* 4: 658-667.

Butler, Declan. 2013. “The Dark Side of Publishing.” *Nature* 495 (March 28): 433-435.

Callaway, Ewen. 2011. “Report Finds Massive Fraud at Dutch Universities.” *Nature* 479 (November 3): 15.

Casey, Katherine, Rachel Glennerster and Edward Miguel. 2012. “Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan.” *Quarterly Journal of Economics* 127(4): 1755-1812.

Chambers, Christopher D. 2013. “Registered Reports: A New Publishing Initiative at *Cortex*.” *Cortex* 49: 609-610.

Chan, An-Wen, Jennifer M. Tetzlaff, Douglas G. Altman, Andreas Laupacis, Peter C. Gøtzsche, Karmela Krleža-Jerić, Asbjørn Hróbjartsson, Howard Mann, Kay Dickersin, Jesse A. Berlin, Caroline J. Doré, Wendy R. Parulekar, William S.M. Summerskill, Trish Groves, Kenneth F. Schulz, Harold C. Sox, Frank W. Rockhold, Drummond Rennie and David Moher. 2013. “SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials.” *Annals of Internal Medicine* 158(3): 200-207.

Ciepiela, Eryk, Daniel Harężlak, Marek Kasztelnik, Jan Meizner, Grzegorz Dyk, Piotr Nowakowski and Marian Bubak. 2013. “The Collage Authoring Environment: From Proof-of-Concept Prototype to Pilot Service.” *Procedia Computer Science* 18: 769-778.

Cooper, Harris, and the APA Publications and Communications Board Working Group on Journal Article Reporting Standards. 2008. “Reporting Standards for Research in Psychology: Why Do We Need Them? What Might They Be?” *American Psychologist* 63(9): 839-851.

Dafoe, Allan. 2013. “Science Deserves Better: The Imperative to Share Complete Replication Files.” Working paper, Yale University.

Davis, Philip M., Bruce V. Lewenstein, Daniel H. Simon, James G. Booth and Matthew J.L. Connolly. 2008. “Open Access Publishing, Article Downloads, and Citations: Randomized Controlled Trial.” *British Medical Journal* 337(7665): 343-345.

De Angelis, Catherine, Jeffrey M. Drazen, Frank A. Frizelle, Charlotte Haug, John Hoey, Richard Horton, Sheldon Kotzin, Christine Laine, Ana Marusic, A. John P.M. Overbeke, Torben V. Schroeder, Hal C. Sox and Martin B. Van Der Weyden. 2004. “Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors.” *New England Journal of Medicine* 351(12): 1250-1251.

De Angelis, Catherine, Jeffrey M. Drazen, Frank A. Frizelle, Charlotte Haug, John Hoey, Richard Horton, Sheldon Kotzin, Christine Laine, Ana Marusic, A. John P.M. Overbeke, Torben V. Schroeder, Hal C. Sox and Martin B. Van Der Weyden. 2005. “Is This Clinical Trial Fully Registered? A Statement from the International Committee of Medical Journal Editors.” *Annals of Internal Medicine* 143(2): 146-148.

de Jonge, P., H.J. Conradi, B.D. Thombs, J.G.M. Rosmalen, H. Burger and J. Ormel. 2011. “Prevention of False Positive Findings in Observational Studies: Registration Will Not Work But Replication Might.” *Journal of Epidemiology and Community Health* 65(2): 95-96.

Delescluse, Matthieu, Romain Franconville, Sébastien Joucla, Tiffany Lieury and Christophe Pouzat. 2012. “Making Neurophysiological Data Analysis Reproducible: Why and How?” *Journal of Physiology – Paris* 106: 159-170.

Dewald, William G., Jerry G. Thursby and Richard G. Anderson. 1986. “Replication in Empirical Economics: The *Journal of Money, Credit and Banking Project*.” *American Economic Review* 76(4): 587-603.

Doshi, Peter, Kay Dickersin, David Healy, S. Swaroop Vedula and Tom Jefferson. 2013. “Restoring Invisible and Abandoned Trials: A Call for People to Publish the Findings.” *British Medical Journal* 346: f2865.

Dudley, Joel T., and Atul J. Butte. 2010. “*In Silico* Research in the Era of Cloud Computing.” *Nature Biotechnology* 28(11): 1181-1185.

ECETOC [European Centre for Ecotoxicology and Toxicology of Chemicals]. 2009. *Enhancement of the Scientific Process and Transparency of Observational Epidemiology Studies*. Brussels; Workshop Report No. 18.

Elman, Colin, Diana Kapiszewski and Lorena Vinuela. 2010. “Qualitative Data Archiving: Rewards and Challenges.” *PS: Political Science and Politics* 43(1): 23-27.

Epidemiology editors. 2010. “The Registration of Observational Studies – When Metaphors Go Bad.” *Epidemiology* 21(5): 607-609.

Evangelou, Evangelos, Thomas A. Trikalinos and John P.A. Ioannidis. 2005. “Unavailability of Online Supplementary Scientific Information from Articles Published in Major Journals.” *FASEB Journal* 19 (December): 1943-1944.

Eysenbach, Gunter. 2006. “Citation Advantage of Open Access Articles.” *PLOS Biology* 4(5): 692-698.

Finfer, Simon, and Rinaldo Bellomo. 2009. “Why Publish Statistical Analysis Plans?” *Critical Care and Resuscitation* 11(1): 5-6.

Fomel, Sergey, and Jon F. Claerbout. 2009. “Reproducible Research.” *Computing in Science & Engineering* 11(1): 5-7.

Freese, Jeremy. 2007. “Replication Standards for Quantitative Social Science: Why Not Sociology?” *Sociological Methods & Research* 36(2): 153-172.

Gargouri, Yassine, Chawki Hajjem, Vincent Larivière, Yves Gingras, Les Carr, Tim Brody and Stevan Harnad. 2010. “Self Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research.” *PLOS One* 5(10): e13636.

Gerber, Alan, and Neil Malhotra. "Do statistical reporting standards affect what is published? Publication bias in two leading political science journals." *Quarterly Journal of Political Science* 3.3 (2008): 313-326.

Gherghina, Sergiu, and Alexia Katsanidou. 2013. “Data Availability in Political Science Journals.” *European Political Science* 12: 333-349.

Gill, Christopher J. 2012. “How Often Do US-Based Human Subjects Research Studies Register on Time, and How Often Do They Post their Results? A Statistical Analysis of the ClinicalTrials.gov Database.” *BMJ Open* 2: e001186.

Gleditsch, Nils Petter, Claire Metelits and Håvard Strand. 2003. “Posting Your Data: Will You Be Scooped or Will You Be Famous?” *International Studies Perspectives* 4(1): 89-97.

Goble, Carole A., Jiten Bhagat, Sergejs Aleksejevs, Don Cruickshank, Danius Michaelides, David Newman, Mark Borkum, Sean Bechhofer, Marco Roos, Peter Li and David De Roure. 2010. “myExperiment: A Repository and Social Network for the Sharing of Bioinformatics Workflows.” *Nucleic Acids Research* 38: W677-W682.

Goecks, Jeremy, Anton Nekrutenko, James Taylor and The Galaxy Team. 2010. “Galaxy: A Comprehensive Approach for Supporting Accessible, Reproducible, and Transparent Computational Research in the Life Sciences.” *Genome Biology* 11: R86.

Gore, Sheila M., and Gerald Jones. 1992. “*The Lancet*’s Statistical Review Process: Areas for Improvement by Authors.” *Lancet* 340(8811): 100-102.

Greenland, Sander. 2007. “Commentary on ‘Quality in Epidemiological Research…’” *International Journal of Epidemiology* 36: 944-945.

Gutmann, Myron P., Mark Abrahamson, Margaret O. Adams, Micah Altman, Caroline Arms, Kenneth Bollen, Michael Carlson, Jonathan Crabtree, Darrell Donakowski, Gary King, Jared Lyle, Marc Maynard, Amy Pienta, Richard Rockwell, Lois Timms-Ferrara and Copeland H. Young. 2009. “From Preserving the Past to Preserving the Future: The Data-PASS Project and the Challenges of Preserving Digital Social Science Data.” *Library Trends* 57(3): 315-337.

Hamermesh, Daniel S. 2007. “Viewpoint: Replication in Economics.” *Canadian Journal of Economics* 40(3): 715-733.

Hanson, Brooks, Andrew Sugden and Bruce Alberts. 2011. “Making Data Maximally Available.” *Science* 331 (February 11): 649.

Haug, Charlotte. 2013. “The Downside of Open-Access Publishing.” *New England Journal of Medicine* 368(9): 791-793.

Herndon, Thomas, Michael Ash and Robert Pollin. 2013. “Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff.” *Political Economy Research Institute Working Paper Series* (322).

Hochberg, Michael E., Jonathan M. Chase, Nicholas J. Gotelli, Alan Hastings and Shahid Naeem. 2009. “The Tragedy of the Reviewer Commons.” *Ecology Letters* 12: 2-4.

Huang, Xiaolei, Bradford A. Hawkins, Fumin Lei, Gary L. Miller, Colin Favret, Ruiling Zhang and Gexia Qiao. 2012. “Willing or Unwilling to Share Primary Biodiversity Data: Results and Implications of an International Survey.” *Conservation Letters* 5: 399-406.

Huang, Yunda, and Raphael Gottardo. 2013. "Comparability and Reproducibility of Biomedical Data." *Briefings in Bioinformatics* 14(4): 391-401.

Humphreys, Macartan, Raul Sanchez de la Sierra and Peter van der Windt. 2013. "Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration." *Political Analysis* 21: 1-20.

Huser, Vojtech, and James J. Cimino. 2013. "Evaluating Adherence to the International Committee of Medical Journal Editors' Policy of Mandatory, Timely Clinical Trial Registration." *Journal of the American Medical Informatics Association* 20: e169-e174.

ICPSR. 2013. Guide to Social Science Data Preparation and Archiving. Phase 5: Preparing Data for Sharing. http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter5.html Retrieved 12/11/13.

[Various authors.] 2003. Symposium on Replication in International Studies Research. *International Studies Perspectives* 4: 72-107.

Ioannidis, John P.A. 2012. "Why Science Is Not Necessarily Self-Correcting." *Perspectives on Psychological Science* 7(6): 645-654.

Ioannidis, John P.A., David B. Allison, Catherine A. Ball, Issa Coulibaly, Xiangqin Cui, Aedín C. Culhane, Mario Falchi, Cesare Furlanello, Laurence Game, Giuseppe Jurman, Jon Mangion, Tapan Mehta, Michael Nitzberg, Grier P. Page, Enrico Petretto and Vera van Noort. 2009. "Repeatability of Published Microarray Gene Expression Analyses." *Nature Genetics* 41(2): 149-155.

Jeon, Doh-Shin, and Jean-Charles Rochet. 2010. "The Pricing of Academic Journals: A Two-Sided Market Perspective." *American Economic Journal: Microeconomics* 2(2): 222-255.

King, Gary. 1995. "Replication, Replication." *PS: Political Science and Politics* 28(3): 444-452.

Koslow, Stephen H. 2000. "Should the Neuroscience Community Make a Paradigm Shift to Sharing Primary Data?" *Nature Neuroscience* 3(9): 863-865.

Kriegeskorte, Nikolaus, Alexander Walther and Diana Deca. 2012. "An Emerging Consensus for Open Evaluation: 18 Visions for the Future of Scientific Publishing." *Frontiers in Computational Neuroscience* 6: 94.

Krleža-Jerić, Karmela, An-Wen Chan, Kay Dickersin, Ida Sim, Jeremy Grimshaw and Christian Gluud for the Ottawa Group. 2005. "Principles for International Registration of Protocol Information and Results from Human Trials of Health Related Interventions: Ottawa Statement (Part 1)." *British Medical Journal* 330(7497): 956-958.

Krleža-Jerić, Karmela, and Trudo Lemmens for the Ottawa Group. 2009. "7th Revision of the Declaration of Helsinki: Good News for the Transparency of Clinical Trials." *Croatian Medical Journal* 50: 105-110.

Laakso, Mikael, and Bo-Christer Björk. 2012. "Anatomy of Open Access Publishing: A Study of Longitudinal Development and Internal Structure." *BMC Medicine* 10(124).

Laine, Christine, Richard Horton, Catherine D. De Angelis, Jeffrey M. Drazen, Frank A. Frizelle, Fiona Godlee, Charlotte Haug, Paul C. Hébert, Sheldon Kotzin, Ana Marusic, Peush Sahni, Torben V. Schroeder, Harold C. Sox, Martin B. Van Der Weyden and Freek W.A. Verheugt. 2007. "Clinical Trial Registration – Looking Back and Moving Ahead." *New England Journal of Medicine* 356(26): 2734-2736.

Lancet editors. 2010. "Should Protocols for Observational Research Be Registered?" *Lancet* 375(9712): 348.

Lash, Timothy L., and Jan P. Vandenbroucke. 2012. "Should Preregistration of Epidemiologic Study Protocols Become Compulsory? Reflections and a Counterproposal." *Epidemiology* 23(2): 184-188.

Lawlor, Debbie A. 2007. "Quality in Epidemiological Research: Should We Be Submitting Papers Before We Have the Results and Submitting More Hypothesis-Generating Research?" *International Journal of Epidemiology* 36: 940-943.

Lawrence, Steve. 2001. "Free Online Availability Substantially Increases a Paper's Impact." *Nature* 411 (May 31): 521.

LeBel, Etienne P., Denny Borsboom, Roger Giner-Sorolla, Fred Hasselman, Kurt R. Peters, Kate A. Ratliff and Colin Tucker Smith. 2013. "PsychDisclosure.org: Grassroots Support for Reforming Reporting Standards in Psychology." *Perspectives on Psychological Science* 8(4): 424-432.

Le Borgne, Y.-A., and A. Campo. 2011. "Open Review in Computer Science." *Procedia Computer Science* 4: 778-780.

Leisch, Friedrich. 2002. "Sweave: Dynamic Generation of Statistical Reports Using Literate Data Analysis." In W. Härdle and B. Rönz, eds., *Compstat 2002: Proceedings in Computational Statistics* (Heidelberg: Physika Verlag), pp 575-580.

Leisch, Friedrich, Manuel Eugster and Torsten Hothorn. 2011. "Executable Papers for the R Community: The R^2 Platform for Reproducible Research." *Procedia Computer Science* 4: 618-626.

LeVeque, Randall J., Ian M. Mitchell and Victoria Stodden. 2012. "Reproducible Research for Scientific Computing: Tools and Strategies for Changing the Culture." *Computing in Science & Engineering* 14(4): 13-17.

Lieberman, Evan S. 2010. "Bridging the Qualitative-Quantitative Divide: Best Practices in the Development of Historically Oriented Replication Databases." *Annual Review of Political Science* 13: 37-59.

Liesegang, Thomas J. 2013. "Authors May Select Immediate Open Access to Their Articles." *American Journal of Opthalmology* 156(3): 417-419.

Li-Thiao-Té, Sébastien. 2012. "Literate Program Execution for Reproducible Research and Executable Papers." *Procedia Computer Science* 9: 439-448.

Loder, Elizabeth, Trish Groves and Domhnall MacAuley. 2010. "Registration of Observational Studies: The Next Step Towards Research Transparency." *British Medical Journal* 340 (February 20): 375-376.

McCullough, B.D. 2009. "Open Access Economics Journals and the Market for Reproducible Economic Research." *Economic Analysis & Policy* 39(1): 117-126.

McCullough, B.D., Kerry Anne McGeary and Teresa D. Harrison. 2006. "Lessons from the *JMCB* Archive." *Journal of Money, Credit and Banking* 38(4): 1093-1107.

McPhail, Deborah, Isy Goodwin and Kerry Gordon. 2006. "Reviewing Statistical Analysis Plans: A Guide for Medical Writers." *Drug Information Journal* 40: 197-202.

Meerpohl, Joerg J., Robert F. Wolff, Gerd Antes and Erik von Elm. 2011. "Are Pediatric Open Access Journals Promoting Good Publication Practice? An Analysis of Author Instructions." *BMC Pediatrics* 11:27.

Mellman, Ira. 2007. "Incomplete Screening?" *Nature Immunology* 8(5): 473.

Moher, David, Sally Hopewell, Kenneth F. Schulz, Victor Montori, Peter C. Gøtzsche, P.J. Devereaux, Diana Elbourne, Matthias Egger and Douglas G. Altman. 2010. "CONSORT 2010 Explanation and Elaboration: Updated Guidelines for Reporting Parallel Group Randomized Trials." *British Medical Journal* 340: c869.

Moher, David, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman and the PRISMA Group. 2009. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *Journal of Clinical Epidemiology* 62: 1006-1012.

Monogan, James E. III. 2013. "A Case for Registering Studies of Political Outcomes: An Application in the 2010 House Elections." *Political Analysis* 21: 21-37.

Montgomery, Paul, Sean Grant, Sally Hopewell, Geraldine Macdonald, David Moher, Susan Michie and Evan Mayo-Wilson. 2013. "Protocol for CONSORT-SPI: An Extension for Social and Psychological Interventions." *Implementation Science* 8: 99.

Moravcsik, Andrew. 2010. "Active Citation: A Precondition for Replicable Qualitative Research." *PS: Political Science and Politics* 43(1): 29-35.

Moya, Lorenzo P., Ivan Moschetti, Munira Nurbhai, Anna Compagnoni, Alessandro Liberati, Jeremy M. Grimshaw, An-Wen Chan, Kay Dickersin, Karmela Krleža-Jerić, David Moher, Ida Sim and Jimmy Volmink. 2009. "Compliance of Clinical Trial Registries with the World Health Organization Minimum Data Set: A Survey." *Trials* 10: 56.

Nature Immunology editors. 2007. "Spot Checks." *Nature Immunology* 8(3): 215.

Nelson, Bryn. 2009. "Empty Archives." *Nature* 461 (September 10): 160-163.

Newcombe, Robert G. 1987. "Towards a Reduction in Publication Bias." *British Medical Journal (Clinical Research Edition)* 295(6599): 656-689.

Norris, Michael, Charles Oppenheim and Fytton Rowland. 2008. "The Citation Advantage of Open-Access Articles." *Journal of the American Society for Information Science and Technology* 59(12): 1963-1972.

Nosek, Brian A., and Yoav Bar-Anan. 2012. "Scientific Utopia: I. Opening Scientific Communication." *Psychological Inquiry* 23(3): 217-243.

Nosek, Brian A., Jeffrey R. Spies and Matt Motyl. 2012. “Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability.” *Perspectives on Psychological Science* 7(6): 615-631.

OAIS. 2012. *Reference Model for an Open Archival Information System (OAIS), Recommended Practice*. CCSDS 650.0-M-2 (Magenta Book) Issue 2, June 2012.

Onukwugha, Eberechukwu. 2013. “Improving Confidence in Observational Studies: Should Statistical Analysis Plans Be Made Publicly Available?” *PharmacoEconomics* 31: 177-179.

Pearce, Neil. 2011. “Registration of Protocols for Observational Research is Unnecessary and Would Do More Harm than Good.” *Occupational and Environmental Medicine* 68(2): 86-88.

Peer, Limor. 2013. “The Role of Data Repositories in Reproducible Research. Yale Institution for Social and Policy Studies.” Blog Post. July 24, 2013. http://isps.yale.edu/news/blog/2013/07/the-role-of-data-repositories-in-reproducible-research#.UqhkdKWdkds Retrieved 12/11/13.

Peer, Limor and Ann Green. 2012 “Building an Open Data Repository for a Specialized Research Community: Process, Challenges and Lessons.” *International Journal of Digital Curation*. 7(1): 151-162.

Peirce, Jonathan, Brian A. Nosek, Alex O. Holcombe, Jeffrey R. Spies, Lee de-Wit, Melissa Lewis, Roger Giner-Sorolla, Johanna Cohoon, Macartan Humphreys, Ben B. Blohowiak, Eric Eich, Frank Farach and Fred Hasselman. 2013. “Badges to Acknowledge Open Practices.” At https://openscienceframework.org/project/TVyXZ/wiki/faq/ (updated December 2, 2013).

Peng, Roger D., Francesca Dominici and Scott L. Zeger. 2006. “Reproducible Epidemiologic Research.” *American Journal of Epidemiology* 163(9): 783-789.

Pino, Cécile, Isabelle Boutron and Philippe Ravaud. 2012. “Inadequate Description of Educational Interventions in Ongoing Randomized Controlled Trials.” *Trials* 13: 63.

Pitt, Mark A., and Yun Tang. 2013. “What Should Be the Data Sharing Policy of Cognitive Science?” *Topics in Cognitive Science* 5: 214-221.

Piwowar, Heather A., Roger S. Day and Douglas B. Fridsma. 2007. “Sharing Detailed Research Data Is Associated with Increased Citation Rate.” *PLOS One* 3 (March): e308.

Pöschl, Ulrich. 2004. “Interactive Journal Concept for Improved Scientific Publishing and Quality Assurance.” *Learned Publishing* 17(2): 105-113.

Rasmussen, Ole Dahl, Nikolaj Malchow-Møller and Thomas Barnebeck Andersen. 2011. “Walking the Talk: The Need for a Trial Registry for Development Interventions.” *Journal of Development Effectiveness* 3(4): 502-519.

Reichman, O.J., Matthew B. Jones and Mark P. Schildhauer. 2011. “Challenges and Opportunities of Open Data in Ecology.” *Science* 331 (February 11): 703-705.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2010. “Growth in a Time of Debt.” *American Economic Review: Papers & Proceedings* 100(2): 573-578.

Reveiz, Ludovic, An-Wen Chan, Karmela Krleža-Jerić, Carlos Eduardo Granados, Mariona Pinart, Itziar Etxeandia, Diego Rada, Monserrat Martinez, Xavier Bonfill and Andrés Felipe Cardona. 2010. "Reporting of Methodologic Information on Trial Registries for Quality Assessment: A Study of Trial Records Retrieved from the WHO Search Portal." *PLOS One* 5(8): e12484.

Ross, Joseph S., Richard Lehman and Cary P. Gross. 2012. "The Importance of Clinical Trial Data Sharing: Toward More Open Science." *Circulation Cardiovascular Quality and Outcomes* 5: 238-240.

Rossner, Mike. 2006. "How to Guard Against Image Fraud." *The Scientist* 20(3): 24-25.

Rossner, Mike. 2007. "Hwang Case Review Committee Misses the Mark." *Journal of Cell Biology* 176(2): 131-132.

Rossner, Mike, and Kenneth M. Yamada. 2004. "What's in a Picture? The Temptation of Image Manipulation." *Journal of Cell Biology* 166(1): 11-15.

Savage, Caroline J., and Andrew J. Vickers. 2009. "Empirical Study of Data Sharing by Authors Publishing in PLoS Journals." *PLOS One* 4(9): e7078.

Savitz, D.A. 2011. "Registration of Observational Studies Does Not Enhance Validity." *Nature Clinical Pharmacology & Therapeutics* 90(5): 646-648.

Schulz, Kenneth F., Douglas G. Altman and David Moher, for the CONSORT Group. 2010. "CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials." *British Medical Journal* 340 (March 27): 698-702.

Sim, Ida, An-Wen Chan, A. Metin Gülmezoglu, Tim Evans and Tikki Pang. 2006. "Clinical Trial Registration: Transparency is the Watchword." *Lancet* 367 (May 20): 1631-1633.

Simonsohn, Uri. 2012. "Just Post It: The Lesson from Two Cases of Fabricated Data Detected by Statistics Alone." Working paper, University of Pennsylvania.

Simonsohn, Uri, Leif D. Nelson and Joseph P. Simmons. 2012. "*P*-curve: A Key to the File Drawer." Working paper, University of Pennsylvania.

Smulders, Yvo M. 2013. "A Two-Step Manuscript Submission Process Can Reduce Publication Bias." *Journal of Clinical Epidemiology* 66: 946-947.

Solomon, David J., Mikael Laakso and Bo-Christer Björk. 2013. "A Longitudinal Comparison of Citation Rates and Growth Among Open Access Journals." *Journal of Informetrics* 7: 642-650.

Sterling, Theodore D. 1959. "Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance – Or Vice Versa." *Journal of the American Statistical Association* 54(285): 30-34.

Stodden, Victoria. 2009. "The Legal Framework for Reproducible Scientific Research: Licensing and Copyright." *Computing in Science & Engineering* 11(1): 35-40.

Stodden, Victoria, Christophe Hurlin and Christophe Pérignon. 2012. "RunMyCode.org: A Novel Dissemination and Collaboration Platform for Executing Published Computational Results." SSRN working paper 2147710.

Swaen, Gerard M.H., Neil Carmichael and John Doe. 2011. “Strengthening the Reliability and Credibility of Observational Epidemiology Studies by Creating an Observational Studies Register.” *Journal of Clinical Epidemiology* 64: 481-486.

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff and Mike Frame. 2011. “Data Sharing by Scientists: Practices and Perceptions.” *PLOS One* 6(6): e21101.

Thomas, Laine, and Eric D. Peterson. 2012. “The Value of Statistical Analysis Plans in Observational Research: Defining High-Quality Research From the Start.” *Journal of the American Medical Association* 308(8): 773-774.

Titus, Sandra L., James A. Wells and Lawrence J. Rhoades. 2008. “Repairing Research Integrity.” *Nature* 453 (June 19): 980-982.

UK Data Archive. Data Management Checklist. http://www.data-archive.ac.uk/create-manage/planning-for-sharing/data-management-checklist. Retreived 12/11/13

Vandewalle, Patrick. 2012. “Code Sharing is Associated with Research Impact in Image Processing.” *Computing in Science & Engineering* 14(4): 42-47.

Vines, Timothy H., Rose L. Andrew, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Nolan C. Kane, Jean-Sébastien Moore, Brook T. Moyers, Sébastien Renaut, Diana J. Rennison, Thor Veen and Sam Yeaman. 2013. “Mandated Data Archiving Greatly Improves Access to Research Data.” *FASEB Journal* 27 (April): 1304-1308.

Vision, Todd J. 2010. “Open Data and the Social Contract of Scientific Publishing.” *BioScience* 60(5): 330-331.

von Elm, Erik, Douglas G. Altman, Matthias Egger, Stuart J. Pocock, Peter C. Gøtzsche and Jan P. Vandenbroucke for the STROBE Initiative. 2007. “The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies.” *PLOS Medicine* 4(10): e296.

Wagenmakers, Eric-Jan, Ruud Wetzels, Denny Borsboom, Han L.J. van der Maas and Rogier A. Kievit. 2012. “An Agenda for Purely Confirmatory Research.” *Perspectives on Psychological Science* 7(6): 632-638.

Walport, Mark, and Paul Brest. 2011. “Sharing Research Data to Improve Public Health.” *Lancet* 377 (February 12): 537-539.

Walster, G. William, and T. Anne Cleary. 1970. “A Proposal for a New Editorial Policy in the Social Sciences.” *The American Statistician* 24(2): 16-19.

Wicherts, Jelte M., Marjan Bakker and Dylan Molenaar. 2011. “Willingness to Share Research Data is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results.” *PLOS One* 6(11): e26828.

Wicherts, Jelte M., Denny Borsboom, Judith Kats and Dylan Molenaar. 2006. “The Poor Availability of Psychological Research Data for Reanalysis.” *American Psychologist* 61(7): 726-728.

Wicherts, Jelte M., Rogier A. Kievit, Marjan Bakker and Denny Borsboom. 2012. “Letting the Daylight In: Reviewing the Reviewers and Other Ways to Maximize Transparency in Science.” *Frontiers in Computational Neuroscience* 6: 20.

Williams, Rebecca J., Tony Tse, William R. Harlan and Deborah A. Zarin. 2010. “Registration of Observational Studies: Is It Time?” *Canadian Medical Association Journal* 182(15): 1638-1642.

Witten, Daniela M., and Robert Tibshirani. 2012. “Scientific Research in the Age of Omics: The Good, the Bad, and the Sloppy.” *Journal of the American Medical Informatics Association* 20: 125-127.

Wolkovich, Elizabeth M., James Regetz and Mary I. O’Connor. 2012. “Advances in Global Change Research Require Open Science by Individual Researchers.” *Global Change Biology* 18: 2102-2110.

Wood, Alastair J.J. 2009. “Progress and Deficiencies in the Registration of Clinical Trials.” *New England Journal of Medicine* 360(8): 824-830.

Yale Law School Roundtable on Data and Code Sharing. 2010. “Reproducible Research: Addressing the Need for Data and Code Sharing in Computational Science.” *Computing in Science & Engineering* 12(5): 8-13.

Zarin, Deborah A., Nicholas C. Ide, Tony Tse, William R. Harlan, Joyce C. West and Donald A.B. Lindberg. 2007. “Issues in the Registration of Clinical Trials.” *Journal of the American Medical Association* 297(19): 2112-2120.

Zarin, Deborah A., Tony Tse, Rebecca J. Williams, Robert M. Califf and Nicholas C. Ide. 2011. “The ClinicalTrials.gov Results Database – Update and Key Issues.” *New England Journal of Medicine* 364(9): 852-860.