Research Policy 41 (2012) 276–290

Contents lists available at SciVerse ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Governing knowledge in the scientific community: Exploring the role of retractions in biomedicine

Jeffrey L. Furman[a,b,*], Kyle Jensen[c,d,1], Fiona Murray[d,e,2]

[a] Boston University School of Management, Boston, MA 02215, United States
[b] National Bureau of Economic Research, Cambridge, MA 02138, United States
[c] Public Intellectual Property Resource for Agriculture, Davis, CA 95616, United States
[d] Harvard-MIT Health Sciences and Technology, Cambridge, MA 02139, United States
[e] MIT Sloan School of Management, Cambridge, MA 02142, United States

ARTICLE INFO

Article history:
Received 19 January 2010
Received in revised form 4 November 2011
Accepted 5 November 2011
Available online 4 January 2012

Keywords:
Retractions
Knowledge production
Scientific institutions
Science policy
False science

ABSTRACT

Although the validity of knowledge is critical to scientific progress, substantial concerns exist regarding the governance of knowledge production. While research errors are as relevant to the knowledge economy as defects are to the manufacturing economy, mechanisms to identify and signal "defective" or false knowledge are poorly understood. In this paper, we investigate one such institution – the system of scientific retractions. We analyze the universe of peer-reviewed scientific articles retracted from the biomedical literature between 1972–2006 and comparing with a matched control sample in order to identify the correlates, timing, and causal impact of scientific retractions. This effort provides insight into the workings of a distributed, peer-based system for the governance of validity in scientific knowledge. Our findings suggest that attention is a key predictor of retraction – retracted articles arise most frequently among highly-cited articles. The retraction system is expeditious in uncovering knowledge that is ever determined to be false (the mean time to retraction is less than two years) and democratic (retraction is not systematically affected by author prominence). Lastly, retraction causes an immediate, severe, and long-lived decline in future citations. Conditional on the obvious limitation that we cannot measure the absolute amount of false science in circulation, these results support the view that distributed governance systems can be designed to uncover false knowledge relatively swiftly and to mitigate the costs that false knowledge for future generations of producers.

1. Introduction

Knowledge accumulation, as Newton famously quipped, requires that new findings "stand on the shoulders of giants." The scientific community is occasionally, reminded that some shoulders serve as shaky foundations, when research once deemed accurate is uncovered as false (Lacetera and Zirulia, 2011; David and Pozzi, 2007). High profile instances of false science occur with frustrating regularity, involving researchers in the academic community and the private sector. In 2006, news that high-profile scientific papers describing the creation of human embryonic stem cells from cloning and somatic cell nuclear transfer were fraudulent roiled the academic community (Kennedy, 2006). In 2002, Bell Labs scientist Jan Hendrik Schön (rumored to have been a candidate for the Nobel Prize) was found to have falsified analysis in numerous publications, including in papers in top journals *Science* and *Nature*, over a four-year period (Reich, 2009). In 2009, Scott S. Reuben, a physician-scientist funded by Pfizer for research on analgesia, confessed to having fabricated data in more than 20 papers (Kowalczyk, 2009). In addition to shaking public confidence in research, false science can have broad-ranging social consequences, particularly when not swiftly identified. In 1998, a paper published in the British medical journal *Lancet* claimed to have identified a link between the MMR vaccine and autism. Not retracted until 2008, the paper contributed to dropping vaccination rates in the UK and continental Europe, caused several measles epidemics, and continues to inhibit vaccination efforts worldwide (Deer, 2004, 2009; Godlee, 2011). In addition to these prominent examples of fraud, the scientific community suffers dozens of cases of malfeasance (of various degrees) and errors, the extent and damage of which is difficult to measure.

In this paper, we examine the role of the system of retractions in the web of scientific institutions intended to identify and signal the

* Corresponding author. Tel.: +1 617 353 4656.
E-mail addresses: furman@bu.edu (J.L. Furman), kyle@pipra.org (K. Jensen), fmurray@mit.edu (F. Murray).
[1] Tel.: +1 530 754 2162.
[2] Tel.: +1 617 253 3681.

0048-7333/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.respol.2011.11.001

existence of false knowledge to the scientific community. Despite the scientific community's obvious interest in governing science and maintaining its veracity, and the growing interest of other distributed knowledge producing communities, the drivers and impact of retractions are not clear (Broad and Wade, 1983; Budd et al., 1998, 1999; Lacetera and Zirulia, 2011; David and Pozzi, 2007; Van Noorden, 2011). Indeed, while mechanisms to identify and signal "defective" or false knowledge are as or more important to the knowledge economy as systems to identify defects in the manufacturing economy, these practices remain poorly understood. Furthermore, we know relatively little about the effectiveness of institutions designed to ensure scientific accuracy (Lacetera and Zirulia, 2011; David and Pozzi, 2007). This stands in sharp contrast to the emerging analytical traditional examining other institutional arrangements that undergird scientific progress and allow for high levels of productivity and rapid knowledge accumulation (Dasgupta and David, 1994; Stephan, 1996; Mokyr, 2002). Our analysis is prompted more broadly by increasing concerns over governance – particularly of fraud or the falsification of information – among other communities engaging in distributed knowledge production (Hargadon and Bechky, 2006). For example, participants in Wikipedia have expressed rising concern spurred by the identification of the long-running fraud of one of its most frequent contributors (see Doran, 2007). Likewise, members of many open-source communities such as the Debian project use a set of "keyring" processes to overcome the threat of malicious contributors who introduce bugs or viruses into the code and utilize a complex series of governance mechanisms to ensure that formats are maintained (Ferraro and O'Mahony, forthcoming).

To extend our understanding of governance of false science, we focus on two descriptive analyses and then proceed to a deeper causal question. In our analysis (and in contrast to prior literature) we employ a matched-sample control group, pairing retracted articles with objectively similar papers. In our first descriptive analysis, we investigate the covariates of retracted articles to identify whether particular features of articles or authors correlate with retraction. In our second analysis, we examine the time elapsed between publication and retraction in order to understand the amount of time ultimately retracted articles circulate before the errors are acknowledged. The results of our descriptive analysis provide new evidence that the significant predictors of retraction are indicators of prominence: (a) the corresponding author's affiliation with a Top 25 US-based university and (b) the number of citations an article receives in its first year after publication. Contrary to some evidence on corrupt behavior (Fisman and Miguel, 2007), our data does not suggest that specific countries or geographic regions are associated with retracted articles. Our analyses of time-to-retraction also provide new and robust econometric results: ultimately retracted articles tend to be rapidly acknowledged as false with a median time between publication and retraction of 19 months (suggesting that the Lancet article noted above is fortunately an outlier). However, none of the observable characteristics has a statistically significant impact on time-to-retraction.

In our third analysis, we ask a key causal question: What is the impact of retractions on research that builds on knowledge found to be false? Prior research has found that researchers continue to cite retracted articles suggesting that retractions are ineffective (Pfeifer and Snodgrass, 1990; Budd et al., 1998). In contrast to prior studies, we employ difference-in-differences estimation to quantify the impact of retraction comparing citation patterns for retracted articles to those of a matched control sample. Our results provide evidence that the system of journal retractions does provide important "news" to researchers – the impact of retraction is large and statistically significant. In our core specification, annual citations of an article drop by 65% following retraction, controlling for article age and calendar year. In the years *prior* to retraction, there is no such decline, implying that retractions are unanticipated by the scientific community. In extended analyses, we review the text of a set of post-retraction citations to understand why these articles persist in citing retracted research finding that many no longer rely on false knowledge. Overall, our results suggest that retractions effectively redirect research away from false paths and into alternative research paths not grounded in false science.

It is important to note that our analysis of retractions is necessarily limited as a study of the universe of false science, since we cannot ascertain the true underlying incidence of errors in published research.[3] This is an important limitation as it means that we cannot draw conclusions about the overall effectiveness of the system of retractions for preventing errors. However, our results enable us to clarify our understanding of a critical community-based system for signaling false science and avoiding false research paths. They also have important implications for those exploring the potential to expand the retraction system to other arenas: Notwithstanding the role of information networks, rapid online communication, blogs, etc., this centuries-old aspect of scientific governance remains salient and retractions serve an important "newsworthy" function.

Our paper proceeds as follows: In Section 2, we describe the institutional system of retractions and its role in the governance of scientific accuracy and outline our three research questions in more detail. In Section 3, we outline our empirical strategy and the data we used to analyze retractions. We then turn in Section 4 to our results and examine each of the questions in turn. We end with a discussion and suggestions for future research in Section 5.

2. False science and the institutional system of retractions

2.1. False science – definitions, extent and institutions

The fabrication of scientific data is widely agreed to constitute an outright form of scientific misconduct. Not surprisingly, detailed analyses of high profile fraud animate the history of science literature (e.g., Weiner, 1955; Pool, 1990; Mallove, 1991; Miller and Hersen, 1992; LaFollette, 1992; Kevles, 1998; Reich, 2009). False science can take additional forms, such as smaller-scale falsifications of results and data, plagiarism, unintended errors, and the inability to replicate research results for unknown reasons (Benos et al., 2005; Office of Science and Technology Policy, 2000; Resnick, 2003; Fanelli, 2009). False science and doubts regarding scientific ethics are not new phenomena. Newton was accused of introducing "fudge factors' into his analysis so as to make it more compelling (Westfall, 1973). Dalton, whose development of modern atomic theory was a critical building block in modern science, is thought to have misrepresented his experimental evidence in the early years of the seventeenth century (Nash, 1956; Broad and Wade, 1983). British mathematician Charles Babbage described the predilection of some of his colleagues for "cooking" the data as "an art of various forms, the object of which is to give to ordinary observations the appearance and character of those of the highest degree of accuracy" (Babbage, 1830). Observers have wrangled over whether Gregor Mendel falsified his data on pea breeding in the nineteenth century (Stern and Sherwood, 1966; Franklin et al., 2008). In our

[3] Other methods, including researcher surveys, provide additional information for forming estimates of the extent of fraud and error in the scientific community (Fanelli, 2009). Another caution comes from Lacetera and Zirulia (2011), whose model implies that detected frauds are likely to differ systematically from undetected frauds; specifically, their models predicts that frauds are more likely to be committed in incremental research, but are more likely to be detected in radical research. Some of our empirical evidence, namely that measures of article prominence are useful predictors of retraction, are consistent with these expectations.

analysis we will use an expansive conception of false science, in keeping with the broad definition encompassed by a *retraction*; the "removal" from the literature of a paper determined to be sufficiently fraudulent, falsified, mistaken or not reproducible that the authors or editors act to acknowledge its invalidity in the public record.

Evidence regarding the prevalence of false science is mixed and, of course, hard to gather. Reports from the US National Institutes of Health Office of Research Integrity suggest that fraud is perpetrated by between one in 100,000 scientists (Steneck, 2006) and one in 10,000 scientists (Marshall, 2000). Other estimates suggest that false science is more widespread: A meta-analysis of data on scientific misconduct reports "a pooled weighted average of 1.97% of scientists admitted to have fabricated, falsified or modified data or results at least once" (Fanelli, 2009, p. 1). Some surveys of scientists suggest even higher rates of misconduct and error. In a survey of accounting researchers, Bailey et al. (2001) found that high-output (>7 publications) researchers report 5.0% of their own articles are affected by some form of misconduct (including the massaging of data, ignoring certain results, concealing validity problems, claiming to have performed procedures that were not performed, or otherwise presenting false results). Interestingly, the same researchers estimate that 17.1% of others researchers' papers make false claims! In a similarly structured study, List et al. (2001) found that more than 4% of attendees surveyed at the 1998 American Economic Association reported having falsified research data. Although accounting and economics are fields in which the likelihood of replication is low (Hamermesh, 1997, 2007), these results support claims that fraud is prevalent despite norms and sanctions (Chubin, 1985; Glaeser, 2006).

The publication of false science, whether due to mendacity or error, can cause the dramatic loss of reputation for the individual scientist associated with the falsification (see Merton, 1957, 1973). Far beyond this individual loss, false science has an impact that ripples throughout the research community: Other scientists may develop faulty projects based on spurious findings, thus losing years of effort trying to develop products from false findings. In the process they can squander public and private funds: a recent report on the dramatic fraud perpetuated by Schön in super-conducting plastics described how "scientists in a dozen top physics laboratories wasted years of effort and millions of dollars trying to replicate and explain the discoveries Schön" (Cookson, 2009; reviewing Reich, 2009). In medical science, when physicians follow the false advice of members of the medical community, patients' wellbeing and mortality are severely impacted. In addition, fraud and even honest errors in publications as well as other knowledge sources can shake public confidence in the scientific system (Broad and Wade, 1983). With its wide reaching consequences for the scientific community, the public, and funders, a well-functioning institutional complex for the governance of scientific knowledge is of central importance in any system of scientific knowledge production.

An effective system for knowledge governance requires both formal and informal institutions to limit the production of false science, to police knowledge that is produced, and, finally, to make scientists aware of false science in order to limit its detrimental impact. A variety of such institutions exists – most relying on the scientific community (journal referees and editors, researchers' host institutions, scientific communities, and funders). At the start of scientific knowledge production, incentives limit scientific fraud through the threat of loss of reputation that follows revealed transgressions (Merton, 1957). However, as Lacetera and Zirulia (2011) have recently modeled, by rewarding novel contributions to knowledge with only a limited probability of discovering misconduct, the reputation-based incentive system is surprisingly vulnerable to false science. Once produced, the peer review system limits falsification (Zuckerman and Merton, 1971) and potentially serves to limit errors in distributed knowledge production such as software. This is a system almost as old as the practice of academic publishing, although until the middle of the eighteenth century it mainly enable scientific societies to meet the censorship requirements of the state over printing presses (Biagoli, 2000). Reviewers, however, confront considerable challenges in trying to ensure the validity and accuracy of the papers (Martin et al., 2007) and many have argued for the reporting of primary data (Couzin, 2006, p. 1853) and the broader use of repositories for data and materials (Furman and Stern, 2011).

Beyond peer review, several formal institutional mechanisms exist to find and deal with false science. At one extreme, egregious forms of scientific misconduct can (rarely) face scrutiny from criminal courts in the United States. In June 2006, researcher Eric Poehlman, who had published widely on menopause, aging, and obesity, pled guilty to falsifying information on federal research grants (Couzin and Unger, 2006; Sox and Rennie, 2006) and was sentenced to a jail term of 366 days.[4] Several cases of fraud including that of Darsee at Harvard (Culliton, 1983) inspired a series of Congressional hearings on misconduct in the 1980s, including investigation of work co-authored by Nobel Prize Winner David Baltimore (LaFollette, 1992). In the aftermath of these inquiries, the Department of Health and Human Services consolidated the NIH's Office of Scientific Integrity and the Assistant Secretary for Health's Office of Scientific Integrity Review into the Office of Research Integrity (David and Pozzi, 2007).

2.2. Institutional system of retractions

Among the most enduring institution for governing false science is the use of retractions to alert the scientific community to false research. As Garfield elaborates, "the timely notification of errors and retractions is essential for the efficiency of the scientific enterprise" (1989, p. 333). Introduced in the sixteenth century with the rise of the system of journal publications (Biagoli, 2000), retractions enable authors and representatives of journals or universities to publicize false claims previously made in their journal thus "removing" them from the scientific literature. Less controversial than full-blown investigations, retractions provide a clear signal of false science and go some way towards creating a norm of regular corrections in scientific journals (for mistakes and difficulties in replication) (Garfield, 1988). Nonetheless, the procedures and standards associated with article retraction are idiosyncratic. Atlas (2004) reports being able to obtain information on retraction policies for fewer than 20% of the 122 biomedical journals he studied. Indeed, 76 of the journals acknowledged that they had no formal policy. In general, however, "full retractions" invalidate the entire content of an article, while "partial retractions" acknowledge sections of a paper or sets of analyses as inaccurate. Whereas "errata," "corrections," or "comments," identify isolated inaccuracies in a paper, retractions are reserved for circumstances in which significant portions of an article are incorrect or cannot be substantiated.

Before proceeding to review the literature on the determinants of retraction, it is important to note that article retraction remains rare. By the mid-1980s, less than 1% of articles indexed in MEDLINE in 1986 had been retracted (Kotzin and Schuyler, 1989). Redman et al. (2008) report than retractions constituted less than 0.01% of all records in PubMed between 1995 and 2004, although they report that figure has risen over time. Both of these facts are consistent

[4] In a bizarre case of research misconduct, graduate student Scott Doree received a 10-month jail sentence for faking the theft of his own research materials in an attempt to cover up the fact that he had been fabricating data for approximately four years (Potter, 2003).

with the percentage of articles retracted from journals that appear in our sample. Of course, one interesting, open question in this area regards the total number of articles that should have been retracted, which would reflect the true number of instances of false science articles (Cokol et al., 2007); this, however, has not been and cannot easily be examined (Lacetera and Zirulia, 2011).

3. Research questions and empirical approach

3.1. Research questions

With the goal of bringing broader insights into the institutions of science and other forms of distributed knowledge production, we ask three questions about the system of retractions: (1) Are specific factors systematically associated with the likelihood of article retraction? (2) Which factors, if any, affect the lag between publication and retraction? (3) What is the impact of retraction on follow-on research in the published literature? Despite the importance of these questions, research has not investigated these issues using modern econometric techniques and a control sample methodology.

Although an understanding of the factors predicting retraction would seem to be of potential value for public policy and prevention, we are not aware of prior research on the statistical correlates of retracted articles. Regarding question two, the factors that affect the lag between publication and the timing of retraction are critical if they alert scientists to the slow pace of finding false science in particular fields, researchers or universities. At least among scientific articles that are ultimately retracted time-to-retraction is a guide to the efficiency of this governance mechanism.

Our third and final question seeks a deeper understanding of the system of retractions by examining the impact of retraction on follow-on (published) research. This is a concern because the integrity and productivity of science depends on "standing on the shoulders" of insightful prior works. Pfeifer and Snodgrass (1990) conclude that, "[m]ethods currently in place to remove invalid literature from use appear to be grossly inadequate" (1990, p. 1423). Indeed, prior research on this topic raises serious concerns about the effectiveness of retractions in steering subsequent research projects away from the shaky foundations of false science (Budd et al., 1998; Redman et al., 2008). However, none of the prior work uses effective control groups or econometric methods.

Our approach allows us to address these three questions by conducting analyses that are more extensive than those in previous studies. We use recent empirical advances to address this question. Our aim is to provide a thorough, systematic and causal analysis of the most widely used governance system for false knowledge. First, we define and incorporate a control group of articles into our analysis whose features help identify the characteristics of articles that are ultimately retracted from the scientific literature. Second, by making use of several different forms of multivariate analyses, we supplement descriptive statistics with precise estimates and more clear causal identification. Finally, we complement our econometric approach with qualitative analyses of citations to a subset of the retracted articles. This allows us to get more insight into the reasons that researchers cite not-yet-retracted papers and already-retracted papers.

3.2. Retractions dataset

We derive our data for this study principally from two sources of bibliometric materials: the National Center for Biotechnology Information's (NCBI) PubMed database and the Institute for Scientific Information's Web of Science. PubMed (accessed via www.pubmed.gov) provides access to MEDLINE, a database

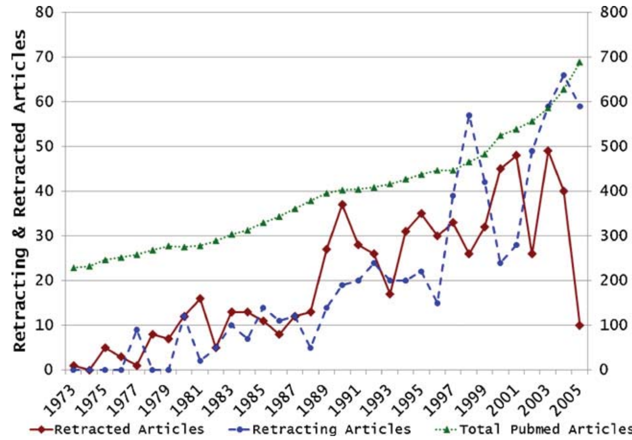

Fig. 1. Retracted and retracting articles, by year.

cataloging references to biomedical and life sciences journals and which is the core database at the NIH's National Library of Medicine (NLM). We used PubMed in order to identify retractions from the biomedical literature, which we refer to as "retracting articles", and to link them to the articles they retract, "retracted" or "root" articles. The National Library of Medicine classifies an article as a *retraction* if three conditions are satisfied: (1) "only if it clearly states that the article in question is being retracted or withdrawn in whole or in part," (2) if the article is signed by an authorized representative, e.g., one or more of the authors, one of the authors' legal representatives, or the journal editor, and (3) the retraction is appropriately labeled and is available in citable form (National Library of Medicine, 2009). Published statements that possess these criteria are labeled by the NLM as retractions and are then cross-linked with the original publications that they retract. Sometimes, a single retraction may retract multiple articles. Our dataset is based on 677 *retracted articles* that could be linked to Web of Science publications and the *retracting articles* that identify them as false. The earliest retracting article appeared in April 1977, while the most recent was published in January 2006. The earliest retracted article in the sample was originally published in 1972, while the most recent was published in October 2005. Fig. 1 shows the annual numbers of retracted and retracting articles and the total count of publications indexed in Pubmed. The rate of retraction increases over time, from 3.6/year in the 1970s to 36/year in this 2000s. This corresponds to an increase from 2 retractions per 100,000 Pubmed publications in the 1970s to approximately 8 per 100,000 in the early 2000s.

We rely on the ISI Web of Science's *Science Citation Index – Expanded* (SCI) database in order to obtain more detailed bibliometric information about each retracting and retracted article and for all of the articles citing the retracted article. The SCI catalogs more than 6600 science and technology journals, identifying details about most journal articles, including author(s), journal name, paper title, authors, addresses and affiliations, abstracts, and keywords. In addition, the SCI records information about each of the articles cited by a particular paper and each article that cites that paper. Some variables, including the year of publication and number of authors and addresses, are directly observable in the SCI data, while others, including the count of citations received, can be computed from it. While complete bibliometric information is available for most articles indexed in the SCI in recent years, some observations lack information for certain fields, including addresses. Observations with missing fields are concentrated in the 1970s and 1980s.

One important feature to note is how we link articles to institutional affiliations and geographic regions. Although the Web of Science identifies the names of all of the authors listed on a paper and all of the addresses and institutional affiliations of these authors, it does not match affiliations and addresses to authors during our sample period. Thus, the database does not support the ability to consider the institutional affiliation of a paper's first or last author. However, the Web of Science does identify each paper's reprint author. For the purposes of our analysis, we consider the reprint author to be the representative author on the paper and assume that examining her institutional affiliation and geographic location will provide insight into the role that these factors play in retraction.[5] We derive the variables associated with articles' Reprint Authors (RP) from the text in the Reprint Author column of the SCI data.

3.3. Control sample

To aid in our analysis of the determinants and impact of retraction on follow-on citations, we match our sample of retracted scientific articles with a sample of articles that serve as key controls in two of our three analyses (the determinants of retraction and the impact of retraction on follow-on scientific citations). This allows us to deal with some of the key difficulties in the interpretation of the existing literature in this field. More specifically, by employing a matched control sample, we are able to infer the degree to which particular observable article characteristics are associated with retractions relative to their overall prevalence in the scientific literature. Similarly, in our examination of the impact of retraction on follow-on research, we are able to compare the citation patterns of ultimately retracted articles with those of similar but never-retracted control articles, thus identifying the impact of retraction on citations relative to a counterfactual estimate. This allows us to identify the causal influence of retractions on the rate of citation – a characteristic that would otherwise be more difficult to infer precisely from the overall citation trend of retracted articles alone.

There are a number of paths that could be pursued in order to develop a useful control group as a baseline from which to investigate the drivers of retraction. One possibility would be to consider the universe of all published articles to be "at risk" of retraction and to conduct an analysis of the covariates associated with year-to-year probabilities of retraction. This option is foreclosed by restrictions on the use of the Science Citation Index and by impracticality of the scope of analysis. A more computationally tractable approach involves comparing retracted articles to a random but representative set of not-yet-retracted articles . Selecting a sample of control articles at random may lead to bias, however, as the propensity to retract articles may differ across journals for reasons unrelated to the probability of those journals publishing false science.

We therefore develop a control sample based on *nearest neighbors* – a set of articles published immediately before and after the retracted articles in the journal in which they appear –gathering bibliometric features of those articles, including the number of annual citations, the institutional affiliation of their key authors, and their geographic origins. We believe that this minimizes observable variation between the retracted and control articles to the greatest possible degree and controls implicitly for factors associated with the behavior of particular journals. In extended analyses, we refine the match between the treated (retracted) articles and the controls, analyzing only on those pairs of retracted and control articles that match on citation counts as well as publication date and either journal or subject field.

4. Empirical analysis

4.1. Descriptive statistics

Descriptive statistics for the key variables in the analysis appear in Table 1. On most observable dimensions, the key variables are similar for control and retracted articles. For example, the means and standard deviations of each sample's number of authors, number of addresses (institutions), Reprint Author's institution type (university, hospital, firm), and even US vs. Non-US origin are statistically indistinguishable. The only statistically significant difference between the two root articles samples is the substantially greater fraction of retracted articles associated with a Top 25 US-university Reprint Author (12% vs. 5%). Consistent with the prospect of citations being curtailed following retraction, the annual average number of citations received by the mean control sample article (3.62 citations) exceeds that of the retracted sample (2.47) and the associated means and standard deviations of the other citations received variables reflect this as well.

Because our research design compares retracted articles with control articles matched based on journal of publication and time of publication, analyses that compare the control articles with the retracted articles cannot tell us about the factors that affect the likelihood of retraction *across* journals. By incorporating data on the total number of published items for each journal that appear in the PubMed database, we can compute and compare baseline retraction rates across journals. Table 2 reports the number of retractions and rate of retractions for the 20 journals in our sample with the largest number of retracted articles. It should be noted that prestige is also critical when looking at retractions by journal. We also report ISI Journal Impact Factor scores. Within this set, the number of retractions is higher among those journals that have high Journal Impact Factors, such as *Science*, *Nature*, and *PNAS*. The *rate* of retraction, however, is higher among journals with relatively lower Journal Impact Factors.

Of the 677 retracted articles, 112 (16.8%) appeared in three high-prestige journals – *Science, Nature* and the *Proceedings of the National Academy of Sciences (PNAS)*. Despite the high relative count of retractions in these journals, their *incidence* of retraction is actually relatively low – retracted articles constitute approximately 0.07% of the total number of PubMed-indexed publications in these journals, substantially less than the median rate of 0.12% among journals with at least one retracted article. The rate of retraction is relatively low among high impact journals such as *Science*, *Nature*, and *PNAS* (which have 2007 Journal Impact Factors of 26.4, 28.8, and 9.6, respectively and retraction rates below 0.10%). In contrast, the journals with the five highest rates of retraction have JIFs below 4.0 (*Mol Cancer*, 3.61% of PubMed-indexed publications retracted; *Indoor Air,* 1.46%; *Medicina*, 0.68%; *Surg Laparosc Endosc*, 0.24%; and *Jpn J Med Sci Biol*, 0.23%). While descriptive statistics provide useful insights into the overall functioning of the retraction system, a more thorough examination of this critical institution requires more advanced econometrics. In the following sections we present the results of these analysis organized to address the three key question posed by our study.

4.2. Correlates of retraction

Our first research question asks, "What are the drivers of false science?" This is particularly difficult to resolve, as it ideally requires both the universe of false science and a vector of exogenous factors that can help explain the circumstances under which false science arises. Instead, we investigate the correlates of retracted articles. To do this, we conduct probit regressions, examining whether specific article characteristics predispose articles to

[5] Brief interviews with a sample of biomedical researchers at our home institutions and our review of the data suggest reprint authors are usually associated with the lab of the paper's principal investigator (PI) or Co-PI.

Table 1
Descriptive statistics, for retracted and control articles.

Variable	Retracted articles ($n = 677$)		Controls ($n = 1340$)	
	Mean	Std. dev.	Mean	Std. dev.
Root article characteristics				
Publication year	1994.58	7.30	1994.46	7.37
# authors	4.26	2.80	4.20	2.90
# addresses	2.20	1.60	2.30	1.68
Root article – reprint (RP) author characteristics				
% US reprint author	0.34	0.47	0.30	0.46
% not US reprint author	0.29	0.45	0.31	0.46
% Top 25 US-university RP author	0.12	0.33	0.05	0.22
% university reprint author	0.13	0.34	0.08	0.27
% hospital reprint author	0.46	0.50	0.45	0.50
% firm reprint author	0.14	0.35	0.13	0.33
Root article – scientific field characteristics				
Clinical medicine	0.41	0.49	0.42	0.49
Multidisciplinary	0.17	0.37	0.16	0.37
Biology and biochemistry	0.11	0.32	0.11	0.31
Molecular biology and genetics	0.09	0.29	0.08	0.27
Neuroscience and behavior	0.06	0.23	0.05	0.21
Unclassified	0.05	0.22	0.06	0.24
Immunology	0.03	0.17	0.03	0.17
Microbiology	0.02	0.14	0.02	0.14
Annual citation characteristics (9079 and 18,294 paper-year observations, respectively)				
Cites received	2.52	7.06	3.68	10.82
Cites received in 1st year	1.35	3.33	0.94	3.45
Cites from US RP authors	0.87	2.80	1.34	4.44
Cites from non-US RP authors	1.13	3.67	1.72	5.79
Cites from Top 25 US-Uni RP author	0.16	0.65	0.24	1.05
Cites from university RP author	1.29	3.80	2.07	6.70
Cites from hospital reprint author	0.33	1.24	0.48	1.79
Cites from firm reprint author	0.04	0.28	0.06	0.38

be in our retracted sample rather than in the control sample of not-retracted articles. Specifically, we estimate Eq. (1):

$$P_{ijt} = f(\alpha_j, \beta_t, \delta_{t\text{-}pubyear}, X_i, \varepsilon_{i,j,t}) \quad (1)$$

where P_{ijt} reflects the probability of retraction of article i, in journal j, in year t, α_j represents a fixed effect for each journal in which a retracted article appears, β_t is a calendar year fixed effect, $\delta_{t\text{-}pubyear}$ captures the age of the article, and X_i represents a vector of article characteristics that may be associated with retraction, including time-invariant article characteristics reflecting the number of authors and institutions associated with the paper, the type of institution with which the paper is associated (university, firm, hospital, etc.), the institution's status (e.g., top 25 universities), and geographic location (e.g., country and region). We employ robust standard errors in our analysis and cluster these at the level of each journal. In robustness checks, we also take advantage of the matched nearest neighbor articles, including fixed effects for each "family" of retracted articles and nearest neighbor articles and clustering the standard errors by article family, rather than including journal fixed effects and clustering by journal. In other robustness

Table 2
Fraction of publications retracted, by journal, 1972–2006.[b]

Rank	Journal	Field	Publications retracted, 1972–2006		JIF'07[a]
			%	#	
1	Mol Cancer	Molecular Biology and Genetics	3.61	3	3.69
2	Indoor Air	Engineering	1.46	4	2.89
3	Medicina (B Aires)	Clinical Medicine	0.68	2	0.19
4	Surg Laparosc Endosc	Clinical Medicine	0.24	2	0.58
5	Jpn J Med Sci Biol	Biology and Biochemistry	0.23	2	1.07
6	Contraception	Clinical Medicine	0.19	7	2.26
7	RNA	Biology and Biochemistry	0.14	2	5.84
8	Nippon Seikeigeka Gakkai Zasshi	Unclassified	0.13	2	0.16
9	Ginekol Pol	Clinical Medicine	0.12	6	n/a
10	Nat Med	Clinical Medicine	0.12	5	26.38
11	Glia	Neuroscience and Behavior	0.11	2	5.38
12	Cell	Biology and Biochemistry	0.11	13	29.89
13	J Clin Invest	Clinical Medicine	0.11	17	16.92
14	Wiad Lek	Unclassified	0.09	8	n/a
15	Resuscitation	Clinical Medicine	0.09	2	2.55
16	Virus Res	Microbiology	0.09	2	2.81
17	EMBO J	Molecular Biology and Genetics	0.09	12	8.66
18	Science	Multidisciplinary	0.07	48	26.37
19	Nature	Multidisciplinary	0.07	30	28.75
20	Neuron	Neuroscience and Behavior	0.07	3	13.41

[a] JIF'07 represents the ISI Web of Science Journal Impact Factor in 2007.
[b] Reports (a) the fraction of publications retracted, using the total number of Pubmed-indexed items as the denominator; (b) the total number of retracted articles, and (c) the 2007 ISI Journal Impact Factor (JIF) for the twenty journals with the highest fraction of retracted articles.

Table 3

Probit models of characteristics associated with Retracted Articles. Table reports marginal effects with all covariates evaluated at the sample mean; all models include Journal and year FEs. Robust standard errors, clustered by journal, are reported in parentheses.

	(3-1) Base model: cites in publication year	(3-2) Cites in publication year and top	(3-3) With institutional affiliation	(3-4) With region FEs	(3-5) With selected country FEs	(3-6) Specialist vs. generalist journals	(3-7) With subject field FEs
Characteristics of retracted article							
Citations received in 1st year after publication	0.006* (0.004)	0.006* (0.003)	0.006** (0.003)	0.005 (0.003)	0.004 (0.003)	0.005* (0.003)	0.005* (0.003)
# authors			−0.002 (0.005)				
Reprint author institutional affiliation							
US-Top 25 universities		0.301*** (0.058)	0.296*** (0.059)	0.326*** (0.063)	0.324*** (0.063)	0.237*** (0.047)	0.237*** (0.048)
Firm affiliation			−0.146 (0.111)				
Hospital affiliation			0.012 (0.046)				
Other reprint author characteristics							
RP region (other regions omitted)[a]							
USA				−0.010 (0.141)	−0.062 (0.039)		
Europe				−0.038 (0.145)			
Asia-Pacific				0.144 (0.162)			
Middle East				−0.284 (0.121)			
Country (non-reported countries listed below)[a]							
Israel					−0.290 (0.094)		
Germany					−0.024 (0.076)		
England					−0.087 (0.070)		
Japan					0.035 (0.083)		
China					0.288 (0.229)		
South Korea					0.347 (0.254)		
Additional controls							
Journal FEs	Yes	Yes	Yes	Yes	Yes		
Subject FEs[b]						General vs. specific *Not sig.*	Narrow subjects *Not sig.*
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1849	1617	1614	1617	1617	1731	1731
Log Likelihood	−1179.85	−1019.91	−1016.77	−1010.44	−1008.08	−1096.59	−1096.53

Constant and dummy variable for missing RP address not reported.

[a] In (3-5), the following country FEs were included but not reported: France, Canada, Australia, Italy, Switzerland, Netherlands, China, South Korea, Spain, Taiwan; other countries constitute the omitted category.

[b] In (3-7), the following subject FEs are included but not reported separately: clinical medicine, general/multidisciplinary, biology and biochemistry, molecular biology and genetics, and neuroscience and behavior.

* Significance at 10% level.

** Significance at 5% level.

*** Significance at the 1% level.

checks, we re-estimate our core model, restricting the sample to just those pairs of control and retracted article that match well on pre-retraction citation counts as well as publication date and journal/subject field. The nature of our control approach means that our analysis assesses how differences among articles within a journal affect the likelihood that an article in our sample will appear in either the treated (retracted) sample or the control (non-retracted sample). Since we directly control for journal-specific characteristics, we cannot draw any conclusions about which factors affect the differential likelihood of retraction across journals.

Table 3 presents the results of our probit estimations, reporting marginal effects evaluated at the mean of the independent variables, which enables a straightforward interpretation of the magnitude of the effects. Each model includes robust standard errors, which are clustered by article family (i.e., triad of retraction article and two nearest neighbor articles).

Our findings contrast with some existing literature on other types of negative outcomes outputs (Huckman and Pisano, 2006) and fraud (e.g., Fisman and Miguel, 2007). Our main finding is that the most robust and statistically significant predictors of retraction are (a) the association of the corresponding author with a US-Top 25 (or US-Top 50) research university and (b) the number of citations that an article receives in its first year after publication. Within our sample of retracted and nearest neighbor articles, articles associated with US-Top 25 universities are between 24% and 33% more likely to be in the retracted group, depending on the other covariates.[6] Controlling for all other factors, the number of citations that an article receives is also positively associated with the probability of article retraction, although the elasticity is substantially lower evaluating at the mean of that variable and there are some specifications in which that result is statistically insignificant.

[6] The result that papers with reprint authors at Top 25 U.S. universities face a higher probability of falling into the retracted sample than the control sample persists in two different types of robustness checks. In the first check, we include article family fixed effects rather than journal fixed effects and cluster standard errors at the level of the article family, where each article family represents a pairing between one retracted article and the nearest neighbor control articles before and after that article, if both can be found. In the second type of robustness check, we re-estimate the core models using a sample selected via the methods of coarsened exact matching that we described in Section 4.4 (*Drivers of Retraction*). These methods involve selecting a sub-sample in which we match treated (retracted) and control articles in the same journal (or field), age, and strata of citations received in the first year. Because such models match on Citations in First Year after Publication, early citations are, by design, not statistically significant predictors of retraction in these regressions. In such models, which we do not report in the paper, articles associated with US-Top 25 universities are between 33% and 41% more likely to be in the retracted group.

There are a number of ways to interpret this finding. One possibility is that papers by authors at top universities and papers that receive a high number of citations soon after publication are subject to greater scrutiny (and, possibly, a higher probability of replication) than less prominent articles. This interpretation is consistent with Lacetera and Zirulia's (2011) prediction that misconduct is more likely to be detected in radical research rather than incremental research, although it is worth noting that their model expects fraud to be more likely in incremental rather than radical research. Another possible interpretation of this result is that the pressure to retract false science featuring high-profile individuals or papers is greater, so that the "bar" for retraction is lower for such papers. A third possibility is that ultimately retracted articles attract early debate about their veracity and that such debates drive this citation result.[7] We read the first year citations to ultimately retracted articles to explore this possibility, but found relatively few examples in which first year citations specifically questioned or disavowed the findings of ultimately retracted articles and, in these cases, the number of such articles was low relative to the total number of early citations.[8] Instead, it appeared as if first year citations acknowledge the importance of the research area and the general uncertainty regarding the subject of the ultimately retracted article.

As the descriptive statistics suggest, none of the other observable correlates are associated with a statistically higher likelihood of article retraction, including co-author team size, such as the number of authors or number of institutions (the second of which is not reported in the table but which was included in robustness checks) and measures reflecting institutional affiliation or geographic location of reprint author. The absence of regional or national effects is interesting in light of other findings on national origin and corrupt behavior (e.g., Fisman and Miguel, 2007); however, this non-finding may be the result of the relatively low number of retractions per country outside the United States (e.g., Japan, England, Germany, and Canada are associated with 31, 22, 18, and 15 Reprint Authors, respectively).

4.3. Time-to-retraction

In our second question, we examine the delay between publication and retraction and investigate factors predicting this delay. Prior research offers varying views on the time elapsed between publication and retraction. Redman et al. (2008) report mean time to retraction in their 1995–2004 PubMed sample is 20.75 months. By contrast, David and Pozzi (2007) find a lag of three years for articles associated with cases prosecuted by the Office of Research Integrity from 1994 to 2006. In Fig. 2, we present a histogram of the time-to-retraction for the articles retracted during our sample period. While some articles remained unacknowledged for more than ten years before being retracted, nearly 15% of the articles retracted were retracted during their year of publication, and more than 50% were retracted within 24 months.[9] This suggests that most research that is ever acknowledged publicly as false is identified relatively soon after initial publication. Of course, un-retracted articles remain indefinitely "at risk" for retraction. As a consequence, it is worth considering the time to retraction as a lower bound estimate of the time to official public acknowledgement of false science.

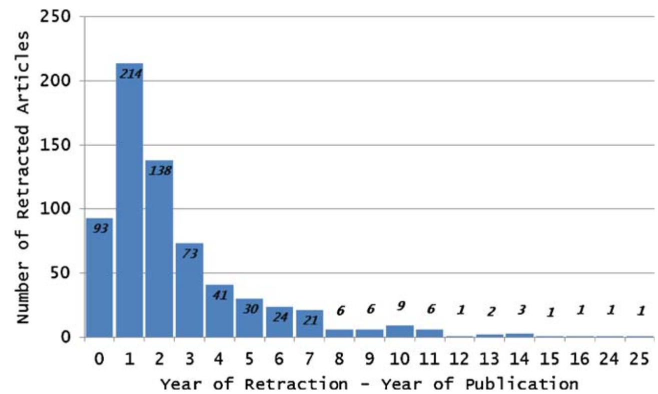

Fig. 2. Histogram, years to retraction.

To provide deeper insights into the drivers of the elapsed time between an article's publication and its ultimate retraction, we conduct OLS regressions using the natural logarithm of the months-to-retraction as the dependent variable. This approach is equivalent to a hazard model approach, but enables a more straightforward interpretation of the coefficients. Specifically, we estimate the equation:

$$\ln(\text{months-to-retraction})_{it} = f(\beta_t, \delta_{t\text{-}pubyear}, X_i; \varepsilon_{i,j,t}) \quad (2)$$

The variables in this model are the same as in (1).

Table 4 reports the results of OLS regressions that model the logarithm of the number of months of time-to-retraction for each of the 677 retracted papers in our sample. In this framework, the regression coefficients can be interpreted as elasticities. Columns (4-1)–(4-3) report the results of estimates using the entire sample of retracted articles. These results suggest that, although time-to-retraction declines as a function of calendar year, few other observable characteristics have a systematic and statistically significant impact on time-to-retraction: Reprint author institution type, geographic region-of-origin and US-region-of-origin are not systematic predictors of retraction. While paper and author prominence are correlated with a greater likelihood of retraction, they are not associated with more rapid retraction.

Hospital affiliation is significant in (4-2), implying that papers with hospital-based reprint authors are retracted approximately 30% more quickly than other articles. This result, however, is quite sensitive to the inclusion of other regressors. Similarly, some regressions including region- or country-specific dummy variables do yield significant results – e.g., retractions from Canada and Mexico occur 60% more quickly than other retractions in (4-3). In most cases, however, these results are due to unusual circumstances. For example, the results suggest that papers with Australian reprint authors experience a longer time-to-retraction. We do not interpret this statistical result as particularly insightful, however, as it is based on the fact that the two papers in the sample with the longest time to retraction are associated with one specific Australian reprint author (Swift), whose publications were not retracted for 24 and 25 years after initial publication. This yields a statistical result, but does not appear to be representative. Each of the 11 other Australian retractions are acknowledged in four years or less.

Our finding that the time-to-retraction declines as a function of publication year suggests that the retractions system is becoming

[7] We are grateful to a referee for suggesting this explanation, the examination of which provided substantial insight into the nature of debate about retracted articles.

[8] One illustrative example, however, of debate about a paper driving early citations is the case of Arnold et al.'s June 1996 Science paper, which was questioned in 1996, prompted a failed attempt at replication in early 1997, and led to retraction in July 1997.

[9] The Redman et al. (2008) sample of articles retracted from 1994–2004 includes 328 articles (of which they analyze the 315 articles published in English); our sample includes 367 articles. The difference likely lies in the fact that our data was downloaded more recently and includes additional articles, which PubMed subsequently added into its database.

Table 4
Drivers of time-to-retraction (OLS regressions).

	(4-1)	(4-2)	(4-3)	(4-4)	(4-5)	(4-6)	(4-7)	(4-8)
	Sample = all retracted articles (1972–2006)						Sample = articles published 1972–2000	
Characteristics of retracted article								
Publication year	−0.047*** (0.005)	−0.048*** (0.005)	−0.040*** (0.006)	−0.048*** (0.005)	−0.048*** (0.005)	−0.040*** (0.006)	−0.018** (0.007)	−0.015* (0.008)
Cites in 1st year	0.005 (0.010)	0.004 (0.010)	0.006 (0.010)	0.005 (0.010)	0.006 (0.010)	0.007 (0.010)	0.006 (0.016)	0.005 (0.016)
Journal impact factor	−0.002 (0.004)	−0.003 (0.004)	−0.003 (0.004)	−0.001 (0.004)	−0.000 (0.005)	−0.002 (0.005)	−0.003 (0.005)	−0.001 (0.005)
# authors	0.014 (0.014)	0.014 (0.014)	0.014 (0.015)	0.017 (0.014)	0.013 (0.014)	0.012 (0.015)	0.013 (0.018)	0.022 (0.020)
Reprint Author affiliation								
University affiliation		0.065 (0.098)				0.064 (0.098)		0.024 (0.127)
Hospital affiliation		−0.327** (0.118)				−0.324 ** (0.121)		−0.365 ** (0.145)
Reprint Author geographic location (other regions omitted)								
USA			0.065 (0.160)			0.089 (0.162)		−0.190 (0.198)
Europe			−0.031 (0.171)			−0.026 (0.174)		−0.120 (0.221)
Asia-Pacific			−0.229 (0.202)			−0.274 (0.202)		−0.580** (0.274)
Subject field (other fields omitted)								
General interest				−0.036 (0.122)	−0.061 (0.168)	−0.062 (0.172)		−0.158 (0.199)
Clinical medicine					−0.060 (0.128)	0.015 (0.135)		0.040 (0.165)
Biology and biochemistry					0.130 (0.159)	0.185 (0.162)		0.163 (0.194)
Molecular biology and Genetics					0.041 (0.171)	0.065 (0.175)		0.009 (0.205)
Neuroscience and behavior					−0.152 (0.194)	−0.102 (0.203)		−0.255 (0.241)
Observations	669	669	594	669	669	594	498	446
R^2	0.115	0.126	0.108	0.118	0.123	0.130	0.014	0.047
Log Likelihood	−920.34	−916.19	−806.57	−919.19	−917.58	−799.15	−690.08	−598.44

Constant and dummies representing missing address affiliation and unclassifiable subject fields not reported. Standard errors, reported in parentheses.
* Significance at 10% level.
** Significance at 5% level.
*** Significance at 1% level.

quicker at identifying false science. However, it may simply be the result of censoring, as any article recently published article could only appear in our data if it were retracted soon after publication, with many recently published articles still in the risk set for retraction in future years. To check for this alternative, we replicate the analyses of (4-1)–(4-3) using only the articles published between 1972 and 2000. A significant and negative coefficient on Calendar Year would be consistent with the finding that time-to-retraction decreased during this period. The results suggest some censoring, in that the magnitude of the coefficient on Publication Year is smaller in (4-4)–(4-6); however, the fact Publication Year enters negatively and significantly does confirm that time-to-retraction is declining over the final three decades of the 20th century.

4.4. Impact of retraction

To address our third research question – the impact of a retraction on subsequent research – we turn to the most novel aspect of our analysis. Prior studies on this topic have found that retracted articles to continue to receive citations and have concluded that scientists (wastefully) treat such articles as if their findings were valid even after they are acknowledged as false (Pfeifer and Snodgrass, 1990; Snodgrass and Pfeifer, 1992; Budd et al., 1998; Redman et al., 2008). However, they rely on descriptive techniques, such as comparing pre-retraction and post-retraction citation counts. In order to illuminate the causal impact of retraction on follow-on citations, we employ our control sample and a difference-in-differences analysis, thus allowing us to assess how retraction changes the rate at which retracted articles are cited relative to controls. We interpret follow-on citations as an indicator of the rate of knowledge accumulation within the sciences. We estimate the average impact of retraction on the forward citations of a research article by comparing the annual number of citations to a given article in the pre- vs. post-retraction periods, controlling for article age, calendar year of citation, and article-specific fixed effects. If we find a large effect from retraction, we can infer that the retraction system is providing new and valuable information to scientists.

We employ a conditional negative binomial model with age and year fixed effects for annual citations received by each article in our dataset. Our choice is guided by the fact that citation data in our sample (like other citation data) are skewed to the right and over-dispersed relative to a Poisson distribution. Our estimator identifies the average change in citations resulting from the change in the institutional or policy environment including article-specific fixed effects:

$$\text{CITES}_{i,t} = f(\gamma_i, \beta_t, \delta_t, \psi\text{POST-RETRACTION}_{i,t}; \varepsilon_{i,t}) \quad (3)$$

where γ_i is a fixed effect for each article, β_t is a year effect, δ_t reflects the age of the article, and POST-RETRACTION is a dummy variable equal to one for retracted articles only in the years after they have been retraction. (Thus, for an article published in 1996 but not retracted until 2002, POST-RETRACTION equals zero in years 1996–2002 and one in the years thereafter.) As article fixed effects identify the mean number of annual citations received by each article over its lifetime, the coefficient on POST-RETRACTION (ψ) reflects the decline in future citations induced by retractions. Our approach thus assesses the impact of retraction by calculating how the citation rate for a scientific publication *changes* following retraction events, accounting for fixed differences in the citation rate across articles and relative to the trend in citation rates for articles with similar characteristics.

We conduct conditional fixed effects negative binomial regression in each of our analyses. In addition to reporting the raw coefficients and standard errors, we report the incident rate ratio (IRR) associated with each coefficient. These are exponentiated versions of the negative binomial regression coefficients, which,

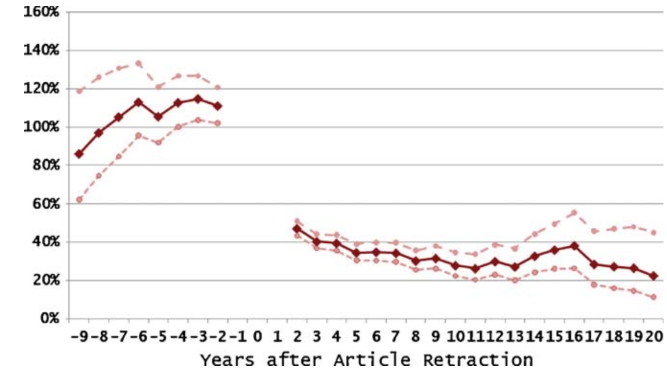

Fig. 3. Pre- and post-retraction effects on forward citations. Solid line reflects IRR from regression containing separate dummy variables for each year preceding and following Retraction, along with complete article, age, and calendar year fixed effects. Grey dotted lines represent 95% confidence intervals, based on robust standard errors, clustered by article.

unlike unadjusted coefficients, have a direct interpretation as elasticities. A coefficient of one implies that a one-unit change in an independent variable would have no effect on citations received whereas coefficients equal to 1.25 and 0.75 imply that a one unit change would be associated with a 25% boost and a 25% reduction in citations, respectively. For each of the models we compute bootstrapped standard errors, clustered by article.

In Table 5, we report the results of estimates on the entire sample and on various subgroups of root article types. Each specification suggests that the impact of retraction is statistically significant and of a substantial magnitude. Our core results appears in (5-1): the coefficient on RETRACTED ARTICLE, POST-RETRACTION implies that annual citations of an article drop by 65% following retraction, controlling for article age, calendar year, and a fixed article citation effect. The effect of retraction on subsequent citation is robust across contexts. In (5-2)–(5-6), we examine the possibility that the discovery of false science may have different effects depending on the characteristics of the retracted paper. For example, one might expect that investigations of individuals at high status universities would be widely known within the community, such that a formal retraction is no longer “news” and has no impact on citations, since the shift in research away from the false knowledge had already taken place. However, we found no evidence for this scenario or others like it. Instead, retractions decline by a similar degree across each sub-sample of retracted articles. For example, the drop is similar for retractions from US-based authors (66.4%) compared to those outside the US (65.0%) and is similar for retractions from Top 25 university authors (60.5%). The effect of retraction does appear to be stronger in the most recent decade than in prior decades, although the large, statistically significant impact of retractions on future citations does not appear to be only induced by modern IT. The results in (5-7) and (5-8) suggest that papers retracted between 1972 and 1999 experienced a 63% decline in citations after retraction, while those retracted since 2000 experienced a 69% decline in citations.

While the coefficient on POST-RETRACTION in Table 5 indicates the *average* impact of retraction on future citations, it is also interesting to understand both pre- and post-retraction citation dynamics – i.e., whether the impact of the retraction occurs as a one-time change in the rate of follow-on research returning to the baseline over time, or whether the retraction induces ever-decreasing levels of citation. These dynamics are illuminating because, for example, a strong downward trend in citations in the years prior to retraction may indicate that the scientific community had been aware of the erroneous nature of retracted articles prior to their retraction. To address these issues Fig. 3 reports the results

Table 5
Principal analyses, average impact of retraction on follow-on citation, comparing across various sample sub-groups.

	(5-1) Sample = entire sample	(5-2) Sample = US authors	(5-3) Sample = non-US authors	(5-4) Sample = authors in Western Europe	(5-5) Sample = authors at US Top 25 universities	(5-6) Sample = papers retracted, 1972–1999	(5-7) Sample = papers retracted, 2000–2006	(5-8) Sample = articles in general interest journals
Retracted article, post-retraction	[0.348] −1.057 (0.010)***	[0.339] −1.081 (0.016)***	[0.358] −1.026 (0.023)***	[0.391] −0.939 (0.031)***	[0.411] −0.888 (0.035)***	[0.371] −0.993 (0.014)***	[0.322] −1.132 (0.017)***	[0.243] −1.413 (0.014)***
Papers	1680	611	591	378	128	960	720	303
Paper-year obs.	23,223	8545	6455	4238	2017	17,582	5641	3894
Log Likelihood	−31,270.76	−12,288.54	−8142.22	−5663.31	−2566.29	−22,660.40	−8560.75	−7369.46

Robust standard errors, clustered by article, are reported in parentheses. Coefficients for retracted article, window period included but not reported, where the window period includes the year of retraction and the year immediately preceding and following retraction.
*Significance at 10% level.
**Significance at 5% level.
*** Significance at 1% level.

of a regression in which we include separate dummy variables for each year preceding and following Retraction (along with complete article, age, and calendar year fixed effects).

In the years *prior* to retraction, there is no decline in citations, implying that retractions are not anticipated by the community. After retraction, however, the community's reaction is immediate and persistent: citations decrease by more than 50% in the second year after retraction and as much as 72% by the tenth year. Two key findings are worth emphasis: (a) there is no pre-retraction decline in citations and (b) retraction corresponds to an immediate decline whose magnitude grows for approximately two decades after citation. Taken together, these finding imply that retraction *causes* a substantial and long-lived decline in research building on the initially false publication.

One important concern about our findings is that the results are sensitive to potential differences in the nature of citations to the nearest neighbor controls and the retracted articles. In particular, spurious results could arise if there were systematic differences between the control group and the retracted articles in their temporal citation trajectories. For example, if nearest neighbor articles were more likely to be associated with "normal science" having relatively steady citation rates over time, whereas retracted articles were more likely to be associated with "hot science" with initially high rates of citation that declined over time, we may obtain results similar to those in Table 5, even if the event of retraction were to have an inconsequential impact on future citation rates. To address this possibility we implement two additional control approaches in Table 6.

In (6-1), we replicate model (5-1), excluding all nearest neighbor articles from our sample and, thus, estimating the impact of retraction based only on the treated (retracted) articles. In this formulation, the coefficient on RETRACTED ARTICLE, POST-RETRACTION represents the decrement to citations experienced by retracted articles relative to citations to those articles that will be retracted (but have not yet been retracted), controlling for other observables associated with the retracted articles, including. Although the impact of retraction is estimated to be only 45% rather than 65%, this regression also suggests that retracted articles experience a statistically significant and quantitatively meaningful decline in citations.

In the remaining columns of Table 6, we implement a matching approach to ensure a closer resemblance between the treated sample and the control sample. Specifically, we report the results of regressions in which the samples of included articles based on coarsened exact matching (CEM). The aim of this matching procedure is to create a set of control and treated articles whose citation trajectories would mirror each other were the treated articles not to have been retracted (Blackwell et al., 2009; Azoulay et al., 2011). Whereas the overall nearest neighbor control sample includes articles drawn from the same issue of the same journal as the retracted articles (matching on publication timing and journal), our CEM procedure enables us to match control and treated articles based on citations as well. Treated articles for which we cannot identify a match are dropped from the regression sample.

In columns (6-2)–(6-5), we report regressions based on three versions of the CEM procedure. In each version, we create three strata and run regressions that include only those control and retracted articles that can be matched on each of the three strata. Those strata include: (a) publication year, (b) journal or subject field (depending on the column), and (c) citations received (either citations received in the first year after publication or cumulative citations received by the age at which the retracted article is retracted). In order to be included in the CEM regressions, controls and retracted articles must match exactly on publication year and either journal or subject field and must both be in the same strata

Table 6
Robustness analyses, using sample of retracted articles only and sample based on coarsened exact matching.

	(6-1) Sample = Treated (retracted) articles only	(6-2) Sample = CEM, matched on journal, age, and cites in 1st year	(6-3) Sample = CEM, matched on field, age, and cites in 1st year	(6-4) Sample = CEM, matched on journal, age, and cumulative cites at time of retraction	(6-5) Sample = CEM, matched on field, age, and cumulative cites at time of retraction
Retracted article, post-retraction	[0.550] −0.599 (0.049)***	[0.365] −1.007 (0.038)***	[0.365] −1.008 (0.036)***	[0.330] −1.108 (0.047)***	[0.372] −0.989 (0.038)***
Papers	584	877	966	518	774
Paper-year obs.	8083	12,148	13,504	6838	10,329
Log Likelihood	−9492.52	−15,204.43	−16,907.98	−9236.33	−13,751.95

Robust standard errors, clustered by article, are reported in parentheses. Coefficients for retracted article, window period included but not reported, where the window period includes the year of retraction and the year immediately preceding and following retraction.
* Significance at 10% level.
** Significance at 5% level.
*** Significance at 1% level.

on citations received.[10] We replicate the regression reported in (5-1) for a variety of sample types in columns (6-2)–(6-5). The results of these analyses are consistent with those of Table 5, estimating that retraction decreases annual citation by between 63 and 67%.

4.5. Content analysis – pre- and post-retraction citations

Our statistical results suggest (a) that retraction is predicted by the number of citations that a paper receives in its initial publication year (as well as the elite status of the lead author's research university) and (b) that retraction serves as critical "news" to the scientific community, resulting in a 65% decrease in citations. These findings raise questions about the content of citations in both the pre- and post-retraction period. It is possible that early citations raise debate or question the validity of ultimately retracted research rather than build on that work. This would result in high early citation counts, but would yield a different interpretation – i.e., that it is not article prominence that draws scrutiny which eventually exposes false knowledge, but rather that early disputation about validity yields high citations for knowledge soon-to-be proven false. In the post-retraction period, a similar issue arises regarding whether ongoing citations actually build on the false knowledge, thus ignoring the retraction and wasting time and resources, or instead comment upon or provide countervailing results that acknowledge the research as false. The latter possibility would imply that our 65% estimate of the post-retraction decline forms a conservative estimate of the true impact of retraction on avoiding future false knowledge, as the actual decline in the use of the false knowledge is greater than is measured by the citation counts.

We investigated these questions by undertaking a content analysis of the papers citing a subset of twenty retracted articles in our sample. We considered ten papers retracted within 24 months of publication and ten papers retracted more than 24 months after publication. To analyze the aim and content of citations published in the pre-retraction period, we reviewed the detailed scientific content to discern whether the citations debated, disagreed with, or attempted to falsify the results of those ultimately retracted papers or whether they instead simply took the results as given and built upon then. Among papers retracted more than 24 months after publication, none experienced obvious debate among their first year citations. Consider the 1999 *Science* paper "Fusion-competent vaccines: broad neutralization of primary isolates of HIV," by LaCasse (1999), which was retracted in 2002. Although it received 29 citations in its year of publication, our reading of these citing articles suggests that none raise questions about its validity or debate its findings. Instead, they take the findings as given and either build upon the LaCasse et al. results or cite the paper to indicate the importance of this line of inquiry.

Among the ten papers were reviewed that were retracted fewer than 24 months after publication, only one experienced obvious debate in its first year after publication: A 1996 paper in *Science* by Arnold and co-authors ("Synergistic activation of estrogen receptor with combinations of environmental chemicals") was retracted a year later in 1997. This paper received 17 citations in 1996 and 91 citations in 1997. One of the 1996 citations (by M Joffe) raised questions about the original work, which prompted a defensive reply from JA McLachlan, the original paper's last author (and reprint author) (McLachlan et al., 1997). In January 1997, Ramamoorthy and co-authors published a paper in *Science* reporting their inability to replicate Arnold et al.'s results, prompting others in the literature to raise additional questions (Ramamoorthy et al., 1997).[11] In July 1997 (13 months after publication), McLachlan wrote a retraction to which all authors consented. While only indicative, our content analysis of pre-retraction citations suggests that many, but not all, build on the original research with only a few questioning its validity. Moreover, although the Arnold case constitutes an example in which a fraction of early citations is attributable to skepticism about the initial result, it does not appear that skeptical citations are sufficient in number to drive the overall correlation between early citations and retraction. This supports our interpretation that article importance (rather than notoriety) is correlated with retractions. It also supports the idea that pre-retraction citation trends reflect normal, follow-on research.

Turning to our analysis of post-retraction citations, we examined the content of articles citing already retracted articles to again assess the reason why articles are citing knowledge acknowledged to be false. Even for the twenty retracted articles in our subsample, this is a challenging task because of the technical nature of the papers and the number of post-retraction citations. We therefore classified post-retraction citations into only three categories: (a) those that explicitly acknowledged the false nature of

[10] In (6–2) and (6–3), we match based on the number citations received in the first year after publication, using six strata, corresponding approximately to the 0–50th, 50th–75th, 75th–95th, 95th–99th, and above the 99th percentiles. In (6–4) and (6–5), we match based on cumulative citations prior to retraction, with one strata corresponding to 0 citations (0-25th percentile), strata based on 1, 2, 3, 4, 5, and 10 citations, which fall between the 25th and 50th percentiles, 20 and 50 citations, which fall between the 50th and 80th percentiles, and 100 and 200, which comprise the 93rd and 98th percentiles (11 strata in total). Columns (6–2) and (6–4) involve matching based on journal, while (6–3) and (6–5) involve matching based on subject field.

[11] A number of articles described the debate and ongoing questions about Arnold et al.'s original findings, including Davidson and Yager (1997) and a March 1998 *Science* "Research News" report.

the retracted research, (b) those that provided evidence that the authors believed that the original article contained valid research contributions and, (c) those that cited to indicate the importance of the retracted paper's line of inquiry and that offer neither a positive nor negative interpretation of the article's specific findings. The Arnold et al. paper we described earlier provides a number of such examples. Danzo (1998) provides an example of a citation to Arnold that explicitly recognizes the original paper's retraction:

> "McLachlan's group [120] presented data showing synergistic activation of the oestrogen receptor by combinations of environmental xenobiotics. . . . The validity of the data presented by McLachlan was questioned by others who were unable to duplicate the published results [123]. Subsequently, the original paper was withdrawn by the authors [124]."

Miller and Sharpe (1998) also cite Arnold et al. (1996) in full awareness of its retracted status:

> "This possibility was fuelled by the report (Arnold et al., 1996) that two very weak environmental oestrogens (dieldrin and endosulfan) could induce a 1000-fold greater oestrogenic effect in combination than on their own, in an in vitro screening system based on yeast transfected with ERa. A number of other laboratories have subsequently been unable to confirm this synergism (Ashby et al., 1997; Ramamoorthy et al., 1997) and the original findings have recently been retracted by the authors." (McLachlan, 1997)

By contrast, we found a number of papers proceeding on the basis that the findings in Arnold et al. (1996) were valid. A full ten years after retraction Evans and N'ipper (2007) cite Arnold et al. (1996) stating:

> "Previous studies have also found endocrine disrupting chemicals, i.e. polychlorinated biphenyls and certain pesticides, when released in combination into the environment have the potential to act synergistically." (Arnold et al., 1996)

Other papers do not take a position regarding the validity of Arnold et al. (1996), but cite the paper in order to note that the issue it was investigating was a research area worthy of attention. For example, Campbell et al. (2006) write:

> "This paper focuses on estrogenic EDCs, we will designate as e-EDCs, that are either hormonal estrogens or chemicals which mimic or induce estrogen-like response in an organism. . . . This broad class of chemicals includes both natural and synthetic estrogens (e.g. xenoestrogens and pseudoestrogens). Specific examples of e-EDCs include: pesticides like atrazine, dieldrin, and toxaphene ([Arnold et al., 1996a], [Ramamoorthy et al., 1997] and [Hayes et al., 2002]), surfactants such a alkyphenol-ethoxalates, . . . natural hormones and pharmaceutical estrogens, . . . phytoestrogens including isoflavonoides and coumestrol, . . . as well as other industrial compounds like bisphenol . . ."

While it is beyond the scope of our analysis to classify each post-retraction citation for each article into one of the three types we identify above, our subset of the data suggests that more than 50% of post-retraction citations came either from articles that explicitly acknowledge the retraction or that cited the retracted article purely for the purpose of identifying a potential fruitful area for research. Thus, we conclude from our qualitative review that fewer than 50% of post-retraction citations built unknowingly on false knowledge.[12] In other words, our estimated citation decline underestimates the value of retractions in preventing wasteful research along false paths.

5. Discussion

Our study supports both optimistic and pessimistic assessments of retraction as a governance mechanism alerting the scientific community to false knowledge. First, we find that retractions are *not* systematically linked to specific institutions, countries, or co-authoring arrangements. Instead, consistent with the model of Lacetera and Zirulia (2011), the strongest predictors of a retraction are measures of a manuscript's prominence, particularly high early citations and the ranking of the corresponding author's institution. One implication of this finding is that "high profile" false science is unlikely to go undetected, perhaps because of the additional scrutiny that such manuscripts attract. In contrast, less prominent science is far less likely to be acknowledged as false. This suggests either that the scientific community holds attention-grabbing papers to a higher standard or (and, possibly, and) that within the vast under-scrutinized literature much false science remains unacknowledged. This result contrasts with the finding that the lag between publication and retraction is uncorrelated with article, author, or institution characteristics.

Of course, this analysis begs the question of how much false science is actually being produced (Cokol et al., 2007; Fanelli, 2009). Like studies of criminal behavior based on reported crimes, our analysis of false science is necessarily limited by our ability to detect violations; we can only analyze articles retracted from the literature (in the observable period), thus obscuring the background rate at which false science is actually produced. As is the case with all governance systems intended to filter good and bad information, decision-makers must determine the appropriate level of the barrier to initial production (in this case, the barrier to initial publication). On the one hand, high barriers to publication reduce the likelihood that published findings will be subsequently determined to be false. On the other hand, high barriers limit the amount of knowledge published and the increased scrutiny of submitted research may delay publication, perhaps rendering the system of knowledge production less efficient. Our results, however, do suggest that particularly pernicious violations of scientific integrity (i.e., those associated with prominent research areas and committed by high-profile individuals) are discovered.

Whatever the precise strength of the filter on false science, the strongest and most robust finding in our analysis is that when false knowledge is identified and signaled to the community via a retraction, the signal is critical and leads to an immediate and long-lived decline in citations. Particularly in light of our qualitative evidence, which suggests that many post-retraction citations do not build on false knowledge, this provides compelling evidence that the system of retractions is an important mode of governance, which alerts the scientific community to the presence of false knowledge, helping investigators of all varieties to avoid follow-on research predicated on false science, potentially saving tens of millions of dollars per year. This is true regardless of observable characteristics: retraction invariably causes a swift, substantial and long-lived (though rarely complete) decline in follow-on citations.

Broadly speaking, the importance of understanding the governance of knowledge production is growing as the scientific enterprise expands in scale and scope (Wuchty et al., 2008; Van Noorden, 2011), thus rendering informal networks of trusted peers

[12] It may be interesting to note that we did not compare the likelihood of "research line citations" (i.e., those that cited to indicate the importance of the retracted paper's line of inquiry and that offer neither a positive nor negative interpretation of the article's specific findings) in the control and retracted samples; thus, we cannot say whether retraction shifted the balance of citation types. This may be an interesting area for future research.

incapable of diffusing information about false publications to the entire relevant scientific community (Crane, 1969). It also becomes increasingly salient for those who generate and use knowledge in the emerging range of knowledge-based communities whose distributed nature, limited social interactions and recent formation preclude the use of informal relationships as the key filter to identify and share information regarding false findings. The need for stronger institutional monitoring of false science may be even greater due to recent funding choices made by agencies in the United States and the European Union, which increasingly promote the establishment of larger, more distributed research teams. This trend towards bigger and broader knowledge production (described, for example, by Giles, 2005) places an even greater responsibility on the existing system of retraction as the key bulwark against false science.

Acknowledgements

We thank Pierre Azoulay, Scott Stern, Nico Lacetera, Dietmar Harhoff, and participants in numerous seminars for comments and suggestions. Lisa Bassett, Anne-Marie Crain, Michaël Bikard, Devin Fensterheim, Robyn Fialkow, Jacob Magid, and Lexie Somers provided exceptional research assistance. All errors are our own. Financial support for this research was provided by the National Science Foundation, under grant #0738394.

References

Arnold, S.F., Klotz, D.M., Collins, B.M., Vonier, P.M., Guillette Jr., L.J., McLachlan, J.A., 1996. Synergistic activation of estrogen receptor with combinations of environmental chemicals. Science 272 (5267), 1489–1492.

Atlas, M.C., 2004. Retraction policies of high-impact biomedical journals. Journal of the Medical Library Association 92 (2), 242–250.

Azoulay, P., Graff Zivin, J., Sampat, B., 2011. The diffusion of scientific knowledge across time and space: evidence from professional transition for the superstars of medicine (with Joshua Graff Zivin and Bhaven Sampat). In: Lerner, J., Stern, S. (Eds.), Proceedings of the NBER Rate & Direction of Inventive Activity Conference.

Babbage, C., 1830. Reflections on the Decline of Science in England. B. Fellowes and J. Booth, London, UK.

Bailey, C.D., Hasselback, J.R., Karcher, J.N., 2001. Research misconduct in accounting literature: a survey of the most prolific researchers' actions and beliefs. Abacus 37 (1), 26–54.

Benos, D.J., et al., 2005. Ethics and scientific publication. Advances in Physiology Education 29, 59–74.

Biagoli, M., 2000. Rights or rewards? Changing contexts and definitions of scientific authorship. Journal of College and University Law 27 (1), 83–108.

Blackwell, M., Iacus, S., King, G., Porro, G., 2009. CEM: coarsened exact matching in Stata. The Stata Journal 9, 524–546.

Broad, W., Wade, N., 1983. Betrayers of the Truth. Simon & Schuster, New York.

Budd, J.M., Sievert, M., Schultz, T.R., 1998. Phenomena of retraction reasons for retraction and citations to the publications. Journal of the American Medical Association 280 (3), 296–297.

Budd, J.M., Sievert, M., Schultz, T.R., Scoville, C., 1999. Effects of article retraction on citation and practice in medicine. Bulletin of the Medical Library Association 87 (4), 437–443.

Campbell, C.G., Borglin, S.E., Green, F.B., Grayson, A., Wozei, E., Stringfellow, W.T., 2006. Biologically directed environmental monitoring, fate, and transport of estrogenic endocrine disrupting compounds in water: a review. Chemosphere 65 (8), 1265–1280.

Chubin, D.E., 1985. Misconduct in research: an issue of science policy and practice. Minerva 23 (2), 175–202.

Cookson, C., 2009. The real lab rats. Financial Times May 23.

Crane, D., 1969. Social structure in a group of scientists: a test of the "invisible college" hypothesis. American Sociological Review 34 (3), 335–352.

Cokol, M.I.I, Rodriguez-Esteban, R., Rzhetsky, A., 2007. How many scientific papers should be retracted? EMBO Reports 8 (5), 421.

Couzin, J., 2006. Breakdown of the year: scientific fraud. Science 314 (5807), 1853.

Couzin, J., Unger, K., 2006. Cleaning up the paper trail. Science 312 (5770), 38–43.

Culliton, B.J., 1983. Coping with fraud: the Darsee case. Science 220 (4592), 31–35.

Danzo, B.J., 1998. The effects of environmental hormones on reproduction. Cellular and Molecular Life Sciences 54 (11), 1249–1264.

Dasgupta, P., David, P.A., 1994. Towards a new economics of science. Research Policy 23, 487–521.

David, P., Pozzi, A., 2007. Empirical realities of scientific misconduct in publicly funded research. Working paper. Presented at the ESF-ORI First World Conference on Research Integrity: Fostering Responsible Research, Lisbon, 16–19 September, 2007.

Davidson, N.E., Yager, J.D., 1997. Pesticides and breast cancer: fact or fad? Journal of the National Cancer Institute 89 (23), 1743.

Deer, B., 2004. Revealed: MMR research scandal. Sunday Times (of London) February 22, Accessed from www.timesonline.co.uk/tol/life_and_style/health/article1027636.ece.

Deer, B., 2009. MMR doctor Andrew Wakefield fixed data on autism. Sunday Times (of London) February 8, Accessed from http://www.timesonline.co.uk/tol/life_and_style/health/article5683671.ece.

Doran, J., 2007. Wikipedia chief promises change after 'expert' exposed as fraud. Sunday Times (of London) March 6, Accessed from http://technology.timesonline.co.uk/tol/news/tech_and_web/article1480012.ece.

Evans, A.D., N'ipper, M., 2007. Toxicity of phenanthrene and lindane mixtures to marine invertebrates. Environmental Toxicology 22 (5), 495–501.

Fanelli, D., 2009. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS ONE 4 (5), e5738.

Ferraro, F., O'Mahony, S. Managing the boundaries of a "open" project. In: J. Padgett, W. Powell (Eds.), The Emergence of Organizations and Markets (Chapter 18), forthcoming.

Fisman, R., Miguel, E., 2007. Corruption, norms, and legal enforcement: evidence from diplomatic parking tickets. Journal of Political Economy 115 (6), 1020–1048.

Franklin, A., Edwards, A.W.F., Fairbanks, D., Hartl, D., Seidenfeld, T., 2008. Ending the Mendel-Fisher Controversy. University of Pittsburgh Press.

Furman, J.L., Stern, S., 2011. Climbing atop the shoulders of giants: the impact of institutions on cumulative research. American Economic Review 101 (5), 1933–1963.

Garfield, E., 1989. Journal editors owe readers timely action on retractions. The Scientist 3, 332–333.

Garfield, E., 1988. Why the scientist welcomes corrections. The Scientist 2 (14), 12.

Giles, J., 2005. Internet encyclopaedias go head to head. Nature 438, 900–901.

Glaeser, E., 2006. Researcher Incentives and Empirical Methods, NBER Technical Working Paper #329.

Godlee, F., 2011. The fraud behind the MMR scare. British Medical Journal 342, d22.

Hamermesh, D.S., 1997. Some thoughts on replications and reviews. Labour Economics 4, 107–109.

Hamermesh, D.S., 2007. Replication in Economics. Forschungsinstitut zur Zukunft der Arbeit (Institute for the Study of Labor), IZA discussion paper, #2760.

Hargadon, A.B., Bechky, B.A., 2006. When collections of creatives become creative collectives: a field study of problem solving at work. Organization Science 17, 484–500.

Huckman, R., Pisano, G., 2006. The firm specificity of individual performance: evidence from cardiac surgery. Management Science 52 (4), 473–488.

Kennedy, D., 2006. Editorial retraction. Science 311 (5759), 335.

Kevles, D., 1998. The Baltimore Case: A Trial of Politics, Science, and Character. W.W. Norton & Co., New York.

Kowalczyk, L., 2009. Doctor accused of faking studies: his work dealt with pain drugs. Boston Globe March 11.

Kotzin, S., Schuyler, P.L., 1989. NLM's practices for handling errata and retractions. Bulletin of the Medical Libraries Association 77, 337–342.

LaCasse, R.A., 1999. Fusion-competent vaccines: broad neutralization of primary isolates of HIV. Science 283 (5400), 357–362.

LaFollette, M.C., 1992. Stealing Into Print: Fraud, Plagiarism, and Misconduct in Scientific Publishing. University of California Press, Berkeley, CA.

Lacetera, N., Zirulia, L., 2011. The economics of scientific misconduct. Journal of Law Economics and Organization 27 (3), 568–603.

List, J.A., Bailey, C.D., Euzent, P.J., Marti, T.L., 2001. Academic economists behaving badly? A survey on three areas of unethical behavior. Economic Inquiry 39 (1), 162–170.

McLachlan, J.A., Arnold, S.F., Klotz, D.M., Collins, B.M., Vonier, P.M., Guillette, L.J., 1997. Potency of combined estrogenic pesticides – response. Science 275 (5298), 405–406.

Marshall, E., 2000. Scientific misconduct – how prevalent is fraud? That's a million-dollar question. Science 290, 1662–1663.

Martin, B.R., et al., 2007. Keeping plagiarism at bay – a salutary tale. Research Policy 36 (7), 905–911.

Mallove, E.F., 1991. Fire From Ice: Searching for the Truth Behind the Cold Fusion Furor. John Wiley & Sons, New York, NY.

Miller, W.R., Sharpe, R.M., 1998. Environmental oestrogens and human reproductive cancers. Endocrine-Related Cancer 5 (2), 69–96.

Merton, R.K., 1957. Social Theory and Social Structure. The Free Press, New York, NY.

Merton, R.K., 1973. The Sociology of Science. The University of Chicago Press, Chicago, IL.

Miller, D.J., Hersen, M., 1992. Research Fraud in the Behavioral and Biomedical Sciences. John Wiley & Sons, New York, NY.

Mokyr, J., 2002. The Gifts of Athena: Historical Origins of the Knowledge Economy. Princeton University Press, Princeton, NJ.

Nash, L.K., 1956. The origin of Dalton's chemical atomic theory. Isis 47 (2), 101–116.

National Library of Medicine, 2009. Fact sheet: errata, retractions, partial retractions, corrected and republished articles, duplicate publications, comments (including author replies), updates, patient summaries, and republished (reprinted) articles policy for MEDLINE, 2008. Published online at http://www.nlm.nih.gov/pubs/factsheets/errata.html (accessed 29.01.09).

Office of Science, Technology Policy, 2000. Federal research misconduct policy. Federal Register 65 (December 6 (235)), 76260–76264.

Pfeifer, M.P., Snodgrass, G.L., 1990. The continued use of the continued use of retracted, invalid scientific literature. Journal of the American Medical Association 263 (10), 1420–1423.
Pool, R., 1990. Cold fusion at Texas A&M: problems, but no fraud. Science 250 (4987), 1507–1508.
Potter, W., 2003. Student sentenced for faking theft. The Chronicle of Higher Education 50.3 (September 12).
Redman, B.K., Yarandi, H.N., Merz, J.F., 2008. Empirical developments in retraction. Journal of Medical Ethics 34, 807–809.
Ramamoorthy, K., Wang, F., Chen, I.C., Safe, S., Norris, J.D., McDonnell, D.P., Gaido, K.W., Boccinfuso, W.P., Korach, K.S., 1997. Potency of combined estrogenic pesticides. Science 275 (5298), 405.
Resnick, D.B., 2003. From Baltimore to Bell labs: reflections on two decades of debate about scientific misconduct. Accountability in Research 10 (2), 123–135.
Reich, E.S., 2009. Plastic Fantastic: How the Biggest Fraud in Physics Shook the Scientific World. Palgrave-Macmillan, New York, NY.
Sox, H.C., Rennie, D., 2006. Research misconduct, retraction, and cleansing the medical literature: lessons from the Poehlman case. Annals of Internal Medicine 144 (8), 609–613.
Snodgrass, G.L, Pfeifer, M.P., 1992. The characteristics of medical retraction notices. Bulletin of the Medical Library Association 80 (4), 328–334.
Steneck, N.H., 2006. Fostering integrity in research: definitions, current knowledge, and future directions. Science and Engineering Ethics 12, 53–74.
Stephan, P.E., 1996. The economics of science. Journal of Economic Literature 34 (3), 1199–1235.
Stern, C., Sherwood, E.R., 1966. The Origin of Genetics. A Mendel Source Book. W. H. Freeman & Co., San Francisco, CA.
Van Noorden, R., 2011. Science publishing: the trouble with retractions. Science 478, 26–28.
Weiner, J.S., 1955. The Piltdown Forgery. Oxford University Press, New York.
Westfall, R.S., 1973. Newton and the fudge factor. Science 179 (4075), 751–758.
Wuchty, S., Jones, B.F., Uzzi, B., 2008. The increasing dominance of teams in production of knowledge. Science 316, 1036–1039.
Zuckerman, H., Merton, R.K., 1971. Patterns of evaluation in science: institutionalisation, structure and functions of the referee system. Minerva 9, 66–100.