Is the Patent System Sensitive to Information Quality?*

Janet Freilich† Soomi Kim‡

January 28, 2021

Abstract

We investigate whether the patent system is sensitive to information quality by observing how players in the patent system (applicants, examiners, and downstream readers) treat inaccurate information. We propose a novel approach to identify poor quality patents: patent-paper pairs where the paper has been retracted and the corresponding patent contains retracted information. We find that these patents are prosecuted, maintained, and cited at rates similar to control patents, despite containing inaccurate information. Insensitivity to information quality may lead to erroneous decisions during patent examination and has implications for patent quality, patent disclosure, and how patents facilitate knowledge flows.

*We thank Yonathan Arbel, Jonathan Ashtor, Pierre Azoulay, Christopher Cotropia, Rochelle Dreyfuss, Richard Freeman, Sepehr Shashahani, Michael Meurer, Fiona Murray, Nicholson Price, Scott Stern, David Schwartz, Sophie Sun, Andrew Toole, and Sam Zyontz for helpful feedback.

†Fordham Law School (jfreilich1@fordham.edu)

‡MIT Sloan of Management (skim20@mit.edu)

1 Introduction

The patent system is only as good as the information it produces. If patents contain poor quality information, examiners and readers cannot differentiate between correct and incorrect statements in patents. Patents may be erroneously granted to inventors who could not actually make the invention, and downstream patents may be erroneously rejected as preempted by poor-quality prior art.[1]

Consider the example of Theranos. By 2016, it was widely known that Theranos' vaunted technology—the ability to detect molecules in small amounts of blood—did not work. Yet in 2018, the U.S. Patent Office (USPTO) granted a Theranos patent claiming "a method of detecting an analyte in a . . . blood sample having a volume of less than about 500 μL" (U.S. Patent 10,156,579). The examiner did not question Theranos' claim. This reverberated downstream: in 2019, a different examiner cited a Theranos patent as evidence that a University of Arizona patent application claiming methods of detecting analytes in small drops of sweat (WO Patent App. 2018013579) was obvious, without acknowledging the public failure of Theranos' technology.

In this paper, we seek to understand whether the examples above are isolated incidents, or whether players in the patent system are broadly insensitive to the quality of information in patents. We make two contributions: a) we develop a new approach to measure poor information quality in the patent system and b) we combine quantitative empirical methods with qualitative analysis derived from manual review of patent documents to find that examiners and inventors generally do not react to poor information quality either at patent application or downstream stage.

Since there is no easy way to measure information quality in the patent system, it is challenging to study poor quality data in patents. Here, we propose a novel way of identifying poor quality data in patents: patent-paper pairs in which the paper has been retracted. Patent-paper pairs occur when the same piece of knowledge is disclosed in both a patent and a paper (Murray and Stern 2007). Because the paper has been retracted, we are confident that the retracted paper's corresponding patent—which we term an

[1]Examiners assess whether patent applications are novel and nonobvious (requirements for patentability) by searching for earlier published disclosures of the invention, called "prior art."

"unsupported patent"—makes claims unsupported by accurate data. Unlike papers, there is no mechanism to retract a patent,[2] so patents continue through the system even after the corresponding paper has been retracted. We identify the universe of all unsupported patents in the biomedical sciences, and then investigate, using matched controls, whether patents are treated differently once material in the patent is publicly acknowledged to be incorrect. Our results suggest that the patent system largely does not react to incorrect information contained in these unsupported patents, either during examination or downstream.

During prosecution of the unsupported patents themselves, we find that applicants overwhelmingly (95%) did not disclose the retraction to the USPTO, and, in two-thirds (66%) of families, continued prosecuting or maintaining patents after the corresponding retraction, despite clear knowledge of the questionable data contained in the patents. Examiners in turn almost always (93%) failed to discover that the application contained retracted material and did not reject unsupported patents at rates higher than controls.

We then turn to the downstream impact of unsupported patents. Confirming other studies of retractions in science (Furman et al. 2012; Azoulay et al. 2015; Azoulay et al. 2017; Jin et al. 2019; Lu et al. 2013), we find that on the paper side, citations to retracted papers drop significantly after retraction. In contrast, corresponding unsupported patents did not experience a change in citations after the retraction event relative to controls. Further, although some examiners cited unsupported patents as a justification for rejecting downstream patents as obvious or not novel, downstream applicants did not fight back. Only 0.6% challenged the rejection on the basis that the cited prior art contained retracted material.

We discuss possible explanations behind the patent system's insensitivity to information quality, including lack of knowledge and breakdown in institutional communication. Examiners rely on applicants to disclose information quality problems, but applicants here do not do so. Downstream examiners and applicants are also hampered in

[2]Patents can be invalidated or found unenforceable in litigation, but this is not the same as retraction. The outcomes in litigation mean that a patentee can no longer enforce her patent against others. However, the patent can nonetheless be cited as prior art against later applications, and a loss in litigation does not necessarily mean that information in the patent is wrong.

discovering poor-quality information. Because patents cannot be retracted, patents (unlike papers) have no clear marking that they contain mistakes.

In examining the quality of information in patents, this paper contributes to a large literature on the prevalence, causes, and effects of poor quality patents (Jaffe and Lerner 2004; Lanjouw and Shankerman 2004; Bessen and Meurer 2008; de Rassenfosse et al. 2016). While previous literature on poor quality patents has predominantly focused on patents that should never have been granted because they are either obvious or not novel, this paper joins a smaller literature that focuses on a different type of poor quality patents—patents with problematic data. These are patents that should never have been granted because the invention does not work or the patentee could not actually make the invention (Ouellette 2017; Freilich 2019; Freilich and Ouellette 2019; Freilich 2020). Poor quality patents are of great concern because they may be acquired by patent acquisition entities and asserted in a manner that taxes innovators (Jaffe and Lerner 2004; Government Accountability Office 2013; Scott Morton and Shapiro 2016), worsens patent thickets (Cohen 2004; Cockburn and MacGarvie 2009), and deters genuine innovators in the area covered by the patent (Government Accountability Office 2016).

Although this paper focuses specifically on the biomedical sciences and on the relatively few patents that incorporate retracted material, patents with poor quality information are ubiquitous (Freilich 2020). If the patent system is not sensitive to retracted information—where there is an easily accessible and reliable statement that the information is wrong—the patent system is likely also insensitive to other types of poor quality information. Such insensitivity leads to errors in the patent system, spreads poor quality information to the public, and damages the integrity of the patent system, which is supposed to disseminate information with the purpose of progressing science and provide incentives for innovation.

2 Institutional Context

The patent document contains a substantial amount of information about the invention. The quality of this information is vital to a functioning patent system at two stages: patent examination and downstream knowledge flows. Patent examiners must determine whether an

applicant invented something new and useful, and whether the applicant disclosed sufficient information about the invention to teach others how to make and use the technology (the "enablement" and "written description" requirements for patentability; 35 U.S.C. § 112). Examiners necessarily rely on the information provided in the application to assess the invention, whether it works, and whether it meets the requirements for patentability (Manual of Patent Examining Procedure § 2164). Examiners have access to little evidence about the invention beyond the words of the patent; thus, the quality of the information contained therein is of crucial importance (Freilich 2021).

If patent applications provide incorrect instructions on how to make an invention or falsely claim that a technology works, examiners who uncritically rely on information in the application will erroneously grant a patent. Because patent claims are always broader than the underlying data supporting the invention,[3] incorrect data can support claims that cover both a non-functional invention but also related inventions that (unbeknownst to the applicant) do work. Thus, examiner insensitivity to information quality may result in patentees being granted exclusive rights over useful technologies that the patentees did not and could not invent.

At the downstream stage, the information in a patent is perhaps even more important. First, this information becomes "prior art" to later patent applications and can be used to reject those downstream applications for obviousness or lack of novelty (Sampat 2010). Such a rejection is only correct if the information upon which it is based is correct.[4] Erroneous rejections may lead inventors to mistakenly narrow or abandon a meritorious patent application, dampening incentives for innovation. Properly balancing incentives between upstream and downstream innovators is a classic question in the design of the patent system (Scotchmer 1991; Williams 2013; Sampat and Williams 2019; Galasso and Schankerman 2014), and both erroneous grant and rejections are mistakes that damage the patent system.

Further, an important purpose of the patent system is to publicly disseminate knowledge that might otherwise be kept private (Ouellette 2012; Sampat 2018). This so-called "quid-pro-quo"—the patentee provides information in return for the patent

[3]For example, if a scientist discovers that a previously unknown molecule reduces tumor size in mice, she is likely to be able to get a patent that claims use of the molecule for any purpose whatsoever (Freilich 2020).

[4]More detail on the legal doctrines is provided in Appendix B.

right—was designed to fulfill the constitutional mandate to "promote the progress of science" (U.S. Constitution Art I, Sec. 8, Cl. 8). High information quality is key to the disclosure function of patents.

3 Data

In this section, we first describe our procedure to identify unsupported patents with problematic data—patent-paper pairs where the paper has been retracted. We then describe how we matched our unsupported patents to control patents from non-retracted patent-paper pairs. We also provide an overview of our main patent data sources.

1. **Retracted papers:** We began by retrieving all retracted papers from PubMed that were published between 2001 and July 2019. We specifically focused on papers that were indexed in Medline, which exclusively focuses on the life sciences, and matched these papers to data provided by Retraction Watch. Retraction Watch aims to compile a comprehensive dataset that captures retractions across all fields and systematically documents the reasons behind the retractions (Oransky and Marcus 2010). We excluded papers that are not original research articles, such as reviews, letters, and commentaries, leaving us with a set of 4,322 retracted papers. Finally, we only kept papers that specified that the retraction occurred because the information in the paper was incorrect; we excluded papers that were retracted due to reasons that do not cast doubt on the veracity of the retracted information (e.g. plagiarism, IRB problems).

2. **Identifying unsupported patents:** We identified all U.S. patent applications with (1) inventors who share the same name as the first or last author of the focal retracted publication and (2) filing dates within +/- 2 years of paper publication. We identified potential patent-paper pairs using a word similarity algorithm that calculated the number of words in common between the paper abstract and patent specification (described in more detail in Appendix A). To definitively identify patent-paper pairs, we manually reviewed potential pairs where the patent specification contained at least 90% of the words in the paper abstract. We then

manually reviewed all retraction notices to confirm that the patents incorporated the retracted material from their corresponding paper in their specification and included claims directed at least in part to the retracted material. From our sample of 4,322 retracted papers, 2% had a patent pair: we identified 107 patent-paper pairs (101 papers; 96 patents).

The following example gives a sense of the closeness of the match between retracted material in the paper and the corresponding patent. As shown in Appendix Figure A.1, a paper discussing how vaccines can be preserved without refrigeration (Zhang et al. 2012) was retracted with a notice stating that "there were significant errors in the data analysis that formed the basis of Figs. 2 and 3 . . . and we are no longer confident in the results presented or the conclusions made" (Zhang et al. 2016). The corresponding patent application—filed two years after the paper retraction—contains figures identical to the retracted figures (U.S. App. 20170258889).

3. **Controls:** Following papers that have studied the impact of paper retractions in science (Furman et al. 2012; Azoulay et al. 2015; Jin et al. 2019), we sought matched controls. We found control, non-retracted patent-paper pairs, by matching on both paper and patent characteristics. We first began by identifying all non-retracted, original research papers in Medline that were published in the same year and journal as the retracted papers of our patent-paper pairs (n = 127,271 papers). We then gathered control papers with associated patents to identify control patent-paper pairs. We used the same word similarity algorithm described above and identified potential control pairs with overlap score greater than 90% (n = 11,225 pairs) and the same primary technology class as the unsupported patents, using the International Patent Classification (IPC) system at the class level (n = 4,550 pairs). We then manually reviewed potential control pairs to confirm that they are indeed pairs. Specifically, for each retracted patent-paper pair, we sorted the potential control pairs by their word overlap score and reviewed them in descending order of the score until we identified a true control pair. After this procedure, we were able to find control patent-paper pairs for 86 of our retracted patent-paper pairs. More details can be found in Appendix A.

4. **Identifying patent family members:** Because members of a patent family often contain identical or very similar specifications, our unit of analysis is a patent family. For our 86 unsupported and 86 control patents identified above, we sought all of their family members, including international members, where possible.[5] Our final sample consists of 86 unsupported patent families (558 individual patents) and 86 control patent families (528 individual patents).

5. **Patent data:** Filing year and inventor names were obtained from Reed Tech's Bulk Data Downloads. Data on prosecution dates and events were obtained from the USPTO's Patent Application Information Retrieval (PAIR) system. Forward citations, maintenance fee payment records, priority dates, and family members were obtained from Google Patents. Information on technology class was obtained from PatSnap. Prosecution histories were downloaded from the USPTO's Public PAIR system. To obtain a more granular understanding of the data, we also manually read all correspondences between the patent applicant and USPTO for both unsupported patents and downstream applications rejected over unsupported patents.

4 Empirical Strategy

To assess whether the patent system is sensitive to information quality, we studied players involved in several different stages of the patent system and their treatment of unsupported patents as compared to control patents.

We begin by investigating the reaction of applicants and examiners of unsupported patents:

$$\mathbf{Y}_i = \beta_0 \ + \beta_1 \mathit{Treated}_i + \mathbf{X}_i + \epsilon_i \tag{1}$$

$\mathbf{Y}_i$ is an indicator variable for a) whether the applicant continued to pay maintenance fees for or prosecute any patent in patent family i after retraction; and b) whether the examiner rejected any patent in patent family i for lack of either enablement or written description after retraction. $\mathit{Treated}_i$ is an indicator variable for whether the patent family is

[5] A patent family is a group of patents that relates to the same invention, related through priority claims. Members of a patent family include continuations, divisionals, and patents filed in other countries.

an unsupported or a control patent family. $\mathbf{X}_i$ is a set of controls, including retraction-year fixed effects, age-at-retraction fixed effects, and technology class fixed effects.[6]

For reactions in the downstream stage, we employ a staggered difference-in-differences approach to investigate whether citations to unsupported patent families decrease more post-retraction relative to the control patent families. We estimate the following regression equation:

$$Cites_{it} = \beta_0 \ + \beta_1 Post\_Retraction_{it} + \beta_2 Post\_Retraction_{it} \times Treated_i + \alpha_{it} + \delta_t + \gamma_i + \epsilon_{it} \quad (2)$$

$Cites_{it}$ is the number of citations patent family i receives in year t, $Post\_Retraction_{it}$ is an indicator that is zero before retraction and one after retraction, $Post\_Retraction_{it} \times Treated_i$ is an indicator that turns one after retraction for only unsupported patent families. β_1 controls for any leads and lags around the retraction event that are common to both unsupported and control patent families (Jaravel et al. 2018), while β_2 is our coefficient of interest and can be interpreted as the causal effect of retraction on patent citations. Standard errors are clustered at the patent family level.

δ_t are calendar year fixed effects that control for any calendar year shocks that impact citation rates of all patent families in a given year,[7] while γ_i are patent family fixed effects that control for patent family traits that could affect citations (e.g., technology class of the patent family). α_{it} fixed effects are indicator variables for patent family age that control for any lifecycle effects of a patent on its citations (for instance, newer patents may be cited more than older patents).[8] ϵ_{it} is the error term.

[6]Retraction year fixed effects consist of full set of seventeen indicator variables, from 2002 to 2019. Age-at-retraction fixed effects consist of eight indicator variables, with the age one indicator including all prior age indicators and the age eight indicator variable including all subsequent age indicators. Technology class fixed effects consist of six indicators.

[7]Calendar year fixed effects consist of twenty-one indicators, from 1999 to 2019, with the 1999 indicator including all prior calendar years.

[8]Patent family age was defined as the patent family's priority year minus the calendar year. Due to the imperfect measure of determining exact citation dates (see Appendix A for more details), the year of first citation precede the priority year for some patent families. Patent age fixed effects consist of fourteen age indicators, with the age zero indicator including all prior age indicators and the age thirteen indicator including all subsequent age indicators.

In order to understand the dynamic effects of retraction, we turn to Equation 3:

$$Cites_{it} = \beta_0 \ + \sum_{j=-n}^{N} \beta_1^j a_{it}^j + \sum_{j=-n}^{N} \beta_2^j a_{it}^j \times Treated_i + \alpha_{it} + \delta_t + \gamma_i + \epsilon_{it} \tag{3}$$

Equation 3 is a modified version of Equation 2 and includes separate indicator variables for each year before and after retraction, a_{it}^j, where the subscript j is the window of years we are interested in before and after the retraction year. For our main analyses, we looked at the window of 5 years around retraction.

In order to conduct our analyses, it is important to understand whether the retraction occurred before or after our outcomes of interest. For instance, to investigate whether an examiner reacts to retraction, the retraction must have first occurred before the patent arrives in the examiner's desk; if the retraction occurred after the patent had already been granted or rejected, then this patent should not be included in our sample for our analysis on examiner rejection/grant.

Appendix Figure A.2 provides a timeline of a patent application as it progresses through the patent system, as well as what fraction of our sample of patent families experienced a (real or inherited counterfactual) retraction at each stage of a patent life cycle. As detailed in Appendix A, we selected the appropriate subsample for each of our outcomes, such that the timing of the retraction could have impacted the outcome. This resulted in the following: the full sample of 172 patent families was used for our analysis on the impact of retraction on downstream citations, 126 patent families were used for our analysis on applicant's decision to prosecute/maintain the patent, and 100 patent families were used for our analysis on examiner's decision to reject or grant the patent.

5 Results

Table 1 reports the summary statistics of our final sample of retracted and control pairs; Panel A focuses on the patent side of the pairs, while Panel B focuses on the paper side. Although the retracted and control pairs were only matched on paper publication year,

paper journal,[9] and patent technology class, our sample is similar on other covariates, such as patent priority year. 86% of the patents are owned by academic institutions. None of the unsupported patents were involved in litigation, as compared to 3% of the control patents. As shown in Panel B, 60% of the retracted papers were retracted due to error or unreliable results, while 38% were retracted due to fraud or misconduct. As reported earlier, we excluded papers that were retracted due to reasons that do not cast doubt on the veracity of the retracted information (e.g. plagiarism, IRB problems).

5.1 How Do Applicants Treat Unsupported Patents?

At the outset, applicants overwhelmingly opt not to tell examiners about the retraction, despite patent law's duty of disclosure. Due to our sample size, we were able to manually review all correspondences (prosecution histories) between the examiner and applicant for unsupported patents. Only three applicants disclosed that their application contained retracted material. Two examiners promptly rejected the application while one applicant preemptively amended the claims to remove the retracted material (though the material remained in the disclosure), after which the application was granted.

Applicants of unsupported patents treated their patents somewhat differently than applicants of control patents, suggesting slight sensitivity to information quality among this group, although more than half of applicants continued to invest resources in unsupported patents and continue legal proceedings even after the corresponding paper has been retracted. For patent families that were "alive" at the time of the retraction (meaning that at least one family member was either being prosecuted or being maintained), 62% of applicants of unsupported patents continued to either prosecute patent applications or pay maintenance fees for granted patents,[10] as compared to 81% of control patent applicants. Columns 1 and 2 of Table 2 report estimations from Equation 1 and show that the probability of being maintained or prosecuted after retraction declines by 18% points.[11]

[9] Appendix Figure A.3 shows the distribution of the journal impact factor (JIF) of our retracted papers. More than 30% of our papers come from highly ranked journals such as *The New England Journal of Medicine*, *Nature*, *Science*, and *Cell*.

[10] Owners of granted patents must pay maintenance fees at regular intervals to avoid abandoning the patent.

[11] Appendix Table A.1 parallels the results in Table 2 but uses a logit specification and reports the average marginal effects; the results are similar.

Because applicants include at least one inventor who is also an author on the corresponding retracted paper, applicants are aware of the retraction, which may account for the decrease in prosecution and maintenance as compared to controls. More surprising is that over half of applicants continue supporting and maintaining unsupported patents—spending money to keep the unsupported patent alive despite the corresponding retraction.

5.2 How Do Examiners Treat Unsupported Patents?

Examiners appear overwhelmingly unaware that unsupported patent applications contain retracted material. After reviewing all examiner communications with applicants of unsupported patents, we found only four examiners who mentioned the retraction, each of whom immediately rejected the application.

However, examiners might reject an unsupported patent application because of the retraction without outright mentioning the retraction. If this was the case, examiners might reject the application either for lack of enablement (on the ground that retracted material cannot teach others how to make and use the invention) or lack of written description (on the ground that retracted material indicates that the inventor was not in possession of the invention). As shown in Columns 3 and 4 of Table 2, we find that examiners appear to be 17% points *less* likely to reject the unsupported patents. This suggests that examiners are not rejecting applications that contain retracted material and are likewise not policing the quality of information in patents.

5.3 How Do Downstream Applicants and Examiners Treat Unsupported Patents?

Downstream Citations

In the scientific literature, retraction causes a sharp drop in citations to the retracted paper. To understand the downstream impact of unsupported patents, we ask how citations change after the corresponding retraction. Figure 1 shows raw mean annual citations to the patent-paper pairs analyzed in this study. Citations to retracted papers drop steeply

after retraction, while citations to the corresponding unsupported patents remain essentially unchanged.

We turn from raw descriptives to a difference-in-differences analysis. Table 3 reports the estimations of Equation 2.[12] In column 1, the outcome is the annual number of total citations, while columns 2 and 3 decompose the citation counts by whether the citation was added by the examiner or the applicant of the citing patent. Columns 4-6 report the logged specifications. Interestingly, in most of our specifications, the magnitude of the point estimates are positive. For instance, as shown in column 1, unsupported patent families experienced an increase of 0.22 total annual number of citations relative to controls, which is a 9% increase from a mean of 2.43.[13] When citations are logged transformed as in column 3, retraction was associated with a 4% increase in citations ($e^{0.041} - 1$). While our results are imprecisely estimated,[14] they suggest that downstream examiners and applicants do not appear to negatively react to retraction and continue to cite the unsupported patents. Alternative specifications using inverse hyperbolic sine transformation and Poisson model yield similar results (Appendix Table A.3). This insensitivity of the downstream patent players to retraction is in sharp contrast to the reaction by the paper publication system, where citations to retracted papers declined significantly by 60% relative to controls (Appendix Table A.4).

Figure 2 plots the event study graphs from estimating Equation 3. There are no noticeable pre-trends before retraction, and downstream examiners and applicants appear to not react to the retraction event.

Response to Rejections

In addition, downstream applicants have another venue to react to unsupported material. When examiners reject downstream applications as anticipated or obvious and cite to an unsupported patent as evidence that the invention was previous disclosed, this rejection is

[12]For our main analyses, our sample is unbalanced, as some patent families have fewer pre- or post-periods. Appendix A and Appendix Table A.2 detail a robustness check where we narrow our sample to a smaller but balanced sample. While the magnitudes change depending on the specifications, none of the coefficients are statistically significant.

[13]Appendix Figure A.4 shows the distribution of annual citations.

[14]For instance, in column 1, our point estimate is 0.22, with standard error of 0.66 and a 95% confidence interval of [-1.11, 1.51].

arguably incorrect. A retraction suggests that the inventor of the unsupported patent did not in fact make the invention, and thus that the downstream patent is novel. Further, a retraction may indicate that the scientific community believed that the invention in the unsupported patent did not work, and thus that the downstream patent is not obvious. There is therefore good reason for downstream applicants to argue that a rejection based on an unsupported patent is incorrect.

We read all communications between USPTO examiners and downstream applicants where an unsupported patent was cited in rejecting the downstream application. Only 3 out of 509 (0.6%) applicants responded to the office action with mention that the prior art contained retracted material. Although downstream applicants are highly incentivized to find and mention the upstream retraction, they do not do so. As with examiners, downstream applicants appear to be insensitive to the quality of information in cited patents.

6 Discussion

We show above that the players in the patent system are largely insensitive to information quality, leading to dissemination and use of incorrect information throughout the patent system. What accounts for this insensitivity? We suggest explanations below.

Applicants: Several factors may contribute to applicants' continued prosecution and maintenance of unsupported patents. First, prior research on retractions in scientific journals have found that many authors continue to believe in the veracity and utility of their work even after the retraction. After one inventor's application was rejected by a patent examiner because of the corresponding paper's rejection, the inventor responded that "[b]y definition, a retraction equates to the data having never been published... It is not a declaration that the data is incorrect" (U.S. Patent App. No. 20150110749). This atypical understanding of retraction and conviction about the invention may explain the choice to continue prosecuting the patent.

Second, high rates of continued prosecution may result from a breakdown in institutional communication. Inventors are often not intimately involved in patent prosecution. Rather, they delegate that duty—and decisions about the process—to their attorney and/or to an institutional party such as a technology transfer office (TTO). While

the inventor is well aware of the retraction, the attorney and TTO making decisions to contact the examiner and to continue prosecuting or maintaining the patent may not be aware of the retraction. For instance, Piero Anversa, a researcher at Harvard Medical School, published dozens of papers reporting his discovery of stem cells in the heart. But this discovery was entirely wrong, leading to the retraction of 46 Anversa papers from 2014 to 2019, a $10 million payment to the NIH after allegations of fraud in 2017, and a 2014 internal Harvard investigation that found falsified and fabricated data (Kolata 2018). Yet some of Anversa's patents were prosecuted and maintained through 2019—presumably because the institutions involved were not aware that the research was discredited. Further supporting the institutional communication breakdown hypothesis, one third of the papers in our sample were retracted after an institutional investigation—yet the corresponding patents were prosecuted and maintained at similar rates to those where an institutional investigation did not occur (50% vs. 56%).

Moreover, applicants' reluctance to abandon unsupported patents may reflect the patents' value. Some retractions may indicate a partially inoperable—but partially operable—technology. Patents based on entirely inoperable technology may still cover working, and valuable, technologies (Freilich 2020). Even patents that cover no operable technology can be monetized in nuisance litigation or provide value as part of a large patent portfolio (Hsu and Ziedonisn 2008). Perhaps reflecting the aforementioned possibilities, Theranos' patent portfolio retained value to investors even after its technology was entirely discredited (McKenna 2018). In our sample, two unsupported patents were sold after retraction, although the true number may be higher as we are not able to account for licenses and sales that are not reflected in the USPTO's assignments' database.

However, there is little evidence that the technological accuracy of the invention is linked to decisions to continue prosecuting or maintaining after the corresponding paper is retracted. We classified retracted papers based on whether the retraction notice retracted the entire paper or only a portion of the paper. Papers in the latter category might disclose operable discoveries. However, applicants on patents corresponding to partially retracted papers prosecuted and maintained their patents at rates comparable to completely retracted papers (56% vs. 52%).

Upstream examiners: Why do examiners grant patents containing unsupported material? Though we cannot exclude the possibility that examiners were aware of the retraction but did not feel a rejection was merited, we believe this is unlikely. The more plausible explanation is that examiners did not know about the retraction. Indeed, examiners are pressed for time (having fewer than 20 hours to review each application) (Merges 1999) and do not have resources to replicate experiments themselves. Examiners therefore rely on applicant disclosures to provide evidence of data quality (or lack thereof); since applicants do not generally disclose retractions, examiners would not discover them. Although examiners do independently search the field of the invention, they do so in the context of discovering prior art, and therefore truncate searches at the priority date of the application—usually before the retraction. The retraction notice would therefore not come up in an examiner search. Moreover, USPTO examiners search the non-patent literature using ProQuest, which does not update retracted articles with retraction notices, further hampering examiners' ability to discover the retraction.

Downstream examiners: Lack of knowledge is also likely why downstream examiners continue to cite unsupported patents. Unlike papers, unsupported patents have no visual notice indicating retraction, therefore providing no warning to citing examiners. Further, while examiners have at least a bachelor's degree in the scientific field in which they work, they may not be sufficiently familiar with the scientific literature to recognize that unsupported patents contain retracted material.

Downstream applicants: Downstream applicants whose patent application has been rejected over unsupported prior art do not raise the presence of retracted material in response to the rejection. Lack of awareness is again the likely explanation as unsupported patents are not marked as containing retracted material. While not all cited prior art is essential to the rejection, downstream applicants are likely unaware that the reference cited in rejecting their patent relies on retracted content. Further, downstream applicants themselves cite unsupported patents, yet another instance of insensitivity to information quality. Some of this unawareness may be attributed to poor communication between parties involved in patent applications. While the applicants themselves are deeply involved in the field of the invention and may (perhaps should) notice that the material in the unsupported patent has been retracted, as a functional matter, attorneys—not inventors—are often the ones

answering office actions, and attorneys would be less likely to recognize an unsupported patent, even if the retraction was high profile.

7 Conclusion

The patent system is largely insensitive to information quality. This has benefits: it is easier and faster to uncritically accept information provided by applicants. One could argue that most patents are low value and important patents that contain inaccurate data will be litigated; thus, battling bad data in the patent system ex-post may be the most efficient way to combat poor information quality since it would be too costly to examine every patent for accuracy. Furthermore, since the incidence of unsupported patents is low, perhaps one could accept these patents as an acceptable costs of having the patent system.

In response, we emphasize that unsupported patents—patents that contain retracted material—represent just a small fraction of the poor quality data in the patent system. Patents with poor quality information are ubiquitous (Freilich 2020). While we focused on the small universe of unsupported (biomedical) patents because retractions permit clear-cut classification of incorrect information in patents, future work could investigate other types of poor-quality information such as patent-paper pairs where the paper has received many negative citations that dispute the paper.

In addition, the patent system's failure to recognize inaccurate data ex-ante—before litigation—may undercut its ability to properly incentivize downstream innovation. Potential innovators may be deterred from innovation in the first place due to existing bad patents; in that case, litigation would not come into play at all. Patents granted on the basis of poor-quality information to patentees who could not actually develop the technology may therefore tax innovators who could make legitimate progress in the area covered by the patent because it is more difficult both to do downstream research and to get a patent as a subsequent entrant in a field (Roin 2008). Further, even patents that will be found invalid in litigation can be used in nuisance suits, a tactic exploited by so-called "patent trolls" (Cohen et al. 2019). The lack of information quality control by players in the patent system could therefore dampen incentives for innovation.

References

Alcácer, J. and M. Gittelman (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. *The Review of Economics and Statistics 88*(4), 774–779.

Azoulay, P., A. Bonatti, and J. Krieger (2017). The career effects of scandal: Evidence from scientific retractions. *Research Policy 46*(9), 1552–1569.

Azoulay, P., J. Furman, J. Krieger, and F. Murray (2015). Retractions. *The Review of Economics and Statistics 97*(5), 1118–1136.

Bacchiocchi, E. and F. Montobbio (2009). Knowledge diffusion from university and public research. A comparison between US, Japan and Europe using patent citations. *The Journal of Technology Transfer 34*(2), 169–181.

Bessen, J. and M. Meurer (2008). *Patent Failure: How Judges, Bureaucrats, and Lawyers Put Innovators at Risk*. Princeton University Press.

Cockburn, I. and M. MacGarvie (2009). Patents, thickets and the financing of early-stage firms: Evidence from the software industry. *Journal of Economics & Management Strategy 18*(3), 729–773.

Cohen, L., U. Gurun, and S. D. Kominers (2019). Patent trolls: Evidence from targeted firms. *Management Science 65*(12), 5461–5486.

Cohen, W. (2004). Patent and appropriation: Concerns and evidence. *The Journal of Technology Transfer 30*(1-2), 57–71.

de Rassenfosse, G., W. Griffiths, A. Jaffe, and E. Webster (2016). Low-quality patent in the eye of the beholder: Evidence from multiple examiners. *NBER Working Paper No. 22244*.

Fehder, D. C., F. Murray, and S. Stern (2014). Intellectual property rights and the evolution of scientific journals as knowledge platforms. *International Journal of Industrial Organization 36*, 83–94.

Freilich, J. (2019). Prophetic patents. *U.C. Davis Law Review 53*, 663–731.

Freilich, J. (2020). The replicability crisis in patent law. *Indiana Law Journal 95*, 431–483.

Freilich, J. (2021). Matching and digging: Evidentiary analysis at the patent office. *Fordham Law Review* (Forthcoming).

Freilich, J. and L. L. Ouellette (2019). Science fiction: Fictitious experiments in patents. *Science 364*(6445), 1036–1037.

Furman, J., K. Jensen, and F. Murray (2012). Governing knowledge in the scientific community: Exploring the role of retractions in biomedicine. *Research Policy 41*(2), 276–290.

Galasso, A. and M. Schankerman (2014). Patents and cumulative innovation: Causal evidence from the courts. *The Quarterly Journal of Economics 130*(1), 317–369.

Government Accountability Office (2013). Intellectual property: Assessing factors that affect patent infringement litigation could help improve patent quality. *GAO-13-465*.

Government Accountability Office (2016). Intellectual property: Patent office should define quality, reassess incentives, and improve clarity. *GAO-16-490*.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). The NBER patent citation data file: Lessons, insights and methodological tools. *NBER Working Paper No. 8498*.

Hsu, D. and R. Ziedonisn (2008). Patents as quality signals for entrepreneurial ventures. *Academy of Management Proceedings*.

Huang, K. G. and F. E. Murray (2009). Does patent strategy shape the long-run supply of public knowledge? evidence from human genetics. *The Academy of Management Journal 52*(6), 1193–1221.

Jaffe, A. and J. L. Lerner (2004). *Innovation and Its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What to do About It*. Princeton University Press.

Jaffe, A. and M. Trajtenberg (1999). International knowledge flows: Evidence from patent citations. *Economics of Innovation and New Technology 8*(1-2), 105–136.

Jaravel, X., N. Petkova, and A. Bell (2018). Team-specific capital and innovation. *American Economic Review 108*(4-5), 1034–73.

Jin, G. Z., B. Jones, S. F. Lu, and B. Uzzi (2019). The reverse Matthew Effect: Consequences of retraciton in scientific teams. *The Review of Economics and Statistics 101*(3), 492–506.

Kolata, G. (2018). He promised to restore damaged hearts. Harvard says his lab fabricated research. *New York Times*.

Lanjouw, J. and M. Shankerman (2004). Patent quality indicators and research productivity: Measuring innovation with multiple indicators. *The Economic Journal 114*(495), 441–465.

Lu, S. F., G. Z. Jin, B. Uzzi, and B. Jones (2013). The retraction penalty: Evidence from the web of science. *Nature Scientific Reports 3*, 1–3.

McKenna, F. (2018). The last days of Theranos – the financials were as overhyped as the blood tests. *MarketWatch*.

Meyer, M. (2006). Are patenting scientists the better scholars?: An exploratory comparison of inventor-authors with their non-inventing peers in nano-science and technology. *Research Policy 35*(10), 1646 – 1662.

Murray, F. and S. Stern (2007). Do formal intellectual property rights hinder the free flow of scientific knowledge?: An empirical test of the anti-commons hypothesis. *Journal of Economic Behavior & Organization 63*(4), 648–687.

Nicholas, T. (2008). Does innovation cause stock market runups? evidence from the great crash. *American Economic Review 98*(4), 1370–96.

Oransky, I. and A. Marcus (2010). Why write a blog about retractions.

Ouellette, L. L. (2012). Do patents disclose useful information? *Harvard Journal of Law and Technology 25*(2), 545–608.

Ouellette, L. L. (2017). Who reads patents? *Nature Biotechnology 35*(5), 421–424.

Roin, B. (2008). Unpatentable drugs and the standards of patentability. *Texas Law Review 87*, 503–570.

Sampat, B. (2010). When do applicants search for prior art. *The Journal of Law and Economics 53*(2), 399–416.

Sampat, B. (2018). A survey of empirical evidence on patents and innovation. *NBER Working Paper No. 25383*.

Sampat, B. and H. L. Williams (2019, January). How do patents affect follow-on innovation? evidence from the human genome. *American Economic Review 109*(1), 203–36.

Scotchmer, S. (1991). Standing on the shoulders of giants: Cumulative research and the patent law. *Journal of Economic Perspectives 5*(1), 29–41.

Scott Morton, F. and C. Shapiro (2016). Patent assertions: Are we any closer to aligning reward to contribution? *Innovation Policy and the Economy 16*, 89–133.

Thompson, N. C., A. A. Ziedonis, and D. C. Mowery (2018). University licensing and the flow of scientific knowledge. *Research Policy 47*(6), 1060 – 1069.

Williams, H. L. (2013). Intellectual property rights and innovation: Evidence from the human genome. *Journal of Political Economy 121*(1), 1–27.

Zhang, J., E. Pritchard, X. Hu, T. Valentin, B. Panilaitis, F. G. Omenetto, and D. L. Kaplan (2012). Stabilization of vaccines and antibiotics in silk and eliminating the cold chain. *Proceedings of the National Academy of Sciences 109*(30), 11981–11986.

Zhang, J., E. Pritchard, X. Hu, T. Valentin, B. Panilaitis, F. G. Omenetto, and D. L. Kaplan (2016). Retraction for zhang et al., stabilization of vaccines and antibiotics in silk and eliminating the cold chain. *Proceedings of the National Academy of Sciences 113*(26), E3810–E3810.

Patents

Gibbons, Ian. 2018. "Methods for the Detection of Analytes in Small-Volume Blood Samples." US10156579.

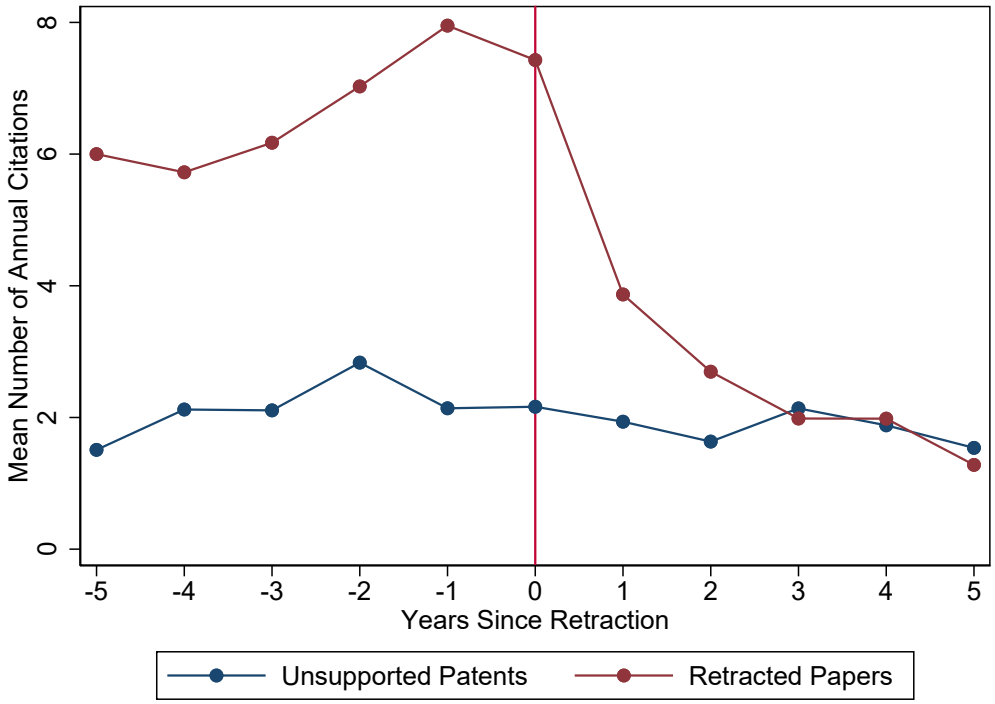
Kaplan, David and Fiorenzo Omenetto. 2017. "Compositions and Methods for Stabilization of Active Agents." US2017025889.

Katchman, Benjamin. 2017. "Sweat as a Biofluid for Analysis and Disease Identification" WO2018013579A1.

Losick, Richard. 2013. "Methods and Compositions for Treating Biofilms." US20140056951.

Vacanti, Charles. 2013. "Generating Pluripotent Cells de Novo." US20150110749.

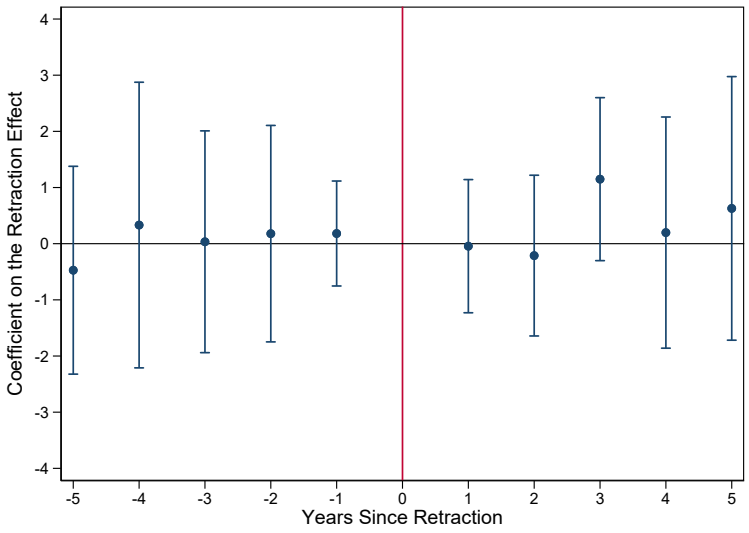
Figure 1: Unsupported Patents vs. Retracted Papers: Mean Annual Citations Received



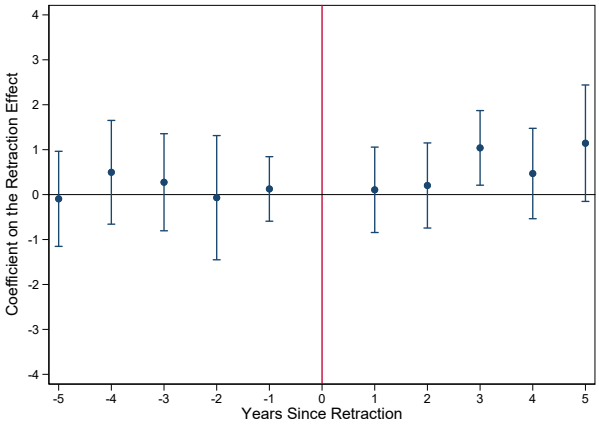
Notes: This figure plots the raw mean number of annual citations received by the 86 unsupported patent families of our main sample and their corresponding retracted papers -/+5 years since retraction.

Figure 2: Event Study: Impact of Retraction on Downstream Citations

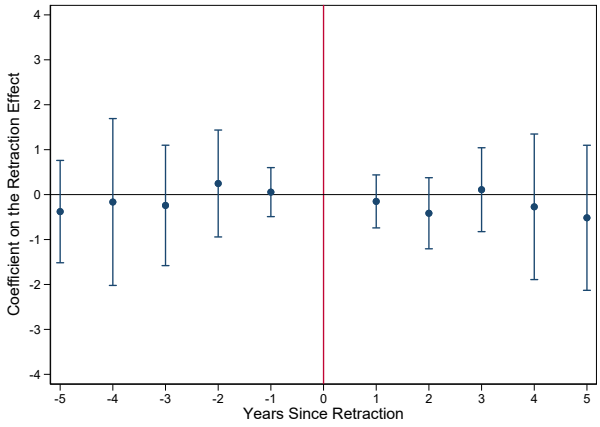
(A) Total Citations



(B) Examiner-Added Citations



(C) Applicant-Added Citations



Notes: These figures plot the coefficients from the estimation of Equation 3 and 95% confidence intervals, and show the impact of retraction on downstream citations. The outcome variable is the number of annual citations received by the patent family -/+5 years since retraction; Panel A plots the total citations, while Panels B and C decompose the citation count into examiner-added and applicant-added citations. The unit of analysis is a patent family X year, and the sample includes 1,540 patent family-years.

Table 1: Summary Statistics

(A) Patents

	(1) Unsupported	(2) Controls
Priority year of the family	2007.06	2007.59
Retraction year (real or counterfactual)	2013.26	2013.26
Family age at retraction year	6.20	5.66
Family size at retraction year	6.21	5.34
Academic assignee	86%	86%
Technology class		
A61: Medical	56%	56%
C12: Biochemistry, microbiology, etc.	24%	24%
C07: Organic chemistry	10%	10%
G01: Measuring, testing	5%	5%
A01: Agriculture	3%	3%
G06: Computing, calculating, counting	1%	1%
Maintained or prosecuted by applicant after retraction	62%	81%
Rejected by examiner after retraction	20%	38%
N of annual citations		
5 years before retraction	1.99	3.31
5 years after retraction	1.59	3.02
Litigated	0%	3%
N of patent families	86	86
N of patents	558	528

(B) Papers

	(1) Retracted	(2) Controls
Publication year	2008.93	2008.81
Retraction year (real or counterfactual)	2013.35	2013.26
Age at retraction	4.42	4.44
Retraction reason		
Error; unreliable results; contaminated materials	60%	-
Fabrication; fraud; misconduct	38%	-
Unknown	2%	-
Duplication; plagiarism	0%	-
N of annual citations		
5 years before retraction	7.10	13.44
5 years after retraction	2.49	20.08
N of papers	84	86

Notes: This table reports the mean summary statistics of our main sample: the universe of retracted patent-paper pairs in the biomedical sciences (as indexed in Medline from 2001 to July 2019) and their control, none-retracted patent-paper pairs. The controls were exactly matched on the publication year of the paper, journal of the paper, and the primary technology class of the patent and inherited the counterfactual retraction date from their retracted counterparts. Two of the retracted papers were associated with two patents. All of the patent summary statistics are based on the full sample of 172 patent families, except for the statistics on maintenance/prosecution by the applicant and examiner rejection, which were based on subsamples of 126 and 100 families, respectively, that were used for the analyses on Table 2.

Table 2: Impact of Retraction on Prosecution, Maintenance, and Rejection

VARIABLES	(1) Applicant Action	(2) Applicant Action	(3) Examiner Rejection	(4) Examiner Rejection
Retracted	-0.190	-0.175	-0.180	-0.167
	(0.079)	(0.081)	(0.090)	(0.096)
Retraction-year FE	NO	YES	NO	YES
Age-at-retraction FE	NO	YES	NO	YES
IPC class FE	NO	YES	NO	YES
N of patent families	126	126	100	100

Notes: The table reports estimation from Equation B. Linear probability model was used, and the unit of analysis is a patent family. Columns 1-2 show applicants' reaction to retraction and report whether the applicant continued to pay maintenance fees for or prosecute any patent in the family after retraction. Columns 3-4 show examiners' reaction to retraction and whether the examiner rejected any patent in the family for lack of either enablement or written description after retraction. We selected the appropriate subsamples for each of our analyses, such that the timing of the retraction could have impacted the outcome (see Appendix Figure A.2 for more details); 126 patent families were used for our analysis on applicant action, and 100 patent families were used for our analysis on examiner rejection. Robust standard errors are in parentheses.

Table 3: Impact of Retraction on Downstream Citations

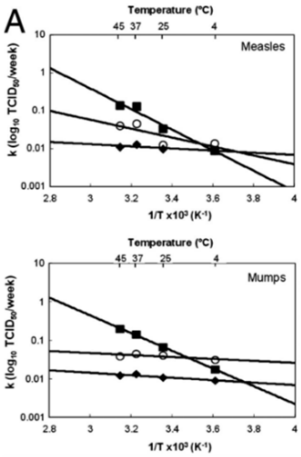
VARIABLES	(1) Total	(2) Examiner	(3) Applicant	(4) Log(1+Total)	(5) Log(1+Examiner)	(6) Log(1+Applicant)
Treat X Post-Retraction	0.215	0.415	-0.200	0.041	0.029	0.026
	(0.659)	(0.380)	(0.468)	(0.087)	(0.073)	(0.079)
Post-Retraction	-0.633	-0.736	0.103	-0.114	-0.127	0.005
	(0.403)	(0.328)	(0.320)	(0.067)	(0.065)	(0.059)
Calendar-year FE	YES	YES	YES	YES	YES	YES
Patent family age FE	YES	YES	YES	YES	YES	YES
Patent family FE	YES	YES	YES	YES	YES	YES
N of patent families	172	172	172	172	172	172
N of patent family-years	1540	1540	1540	1540	1540	1540

Notes: This table reports results from the estimation of Equation 2 and shows the impact of retraction on downstream citations. The unit of analysis is a patent family X year, and the full sample of 172 families was included. In column 1, the outcome variable is the number of total annual citations received by the patent family -/+5 years since retraction, while columns 2 and 3 decompose the citation counts by whether the citation was added by the examiner or the applicant. Columns 4-6 report log transformed citation outcomes. Standard errors are clustered at the patent family level.

Appendix Figure A.1: Retracted Paper-Unsupported Patent Pair Example

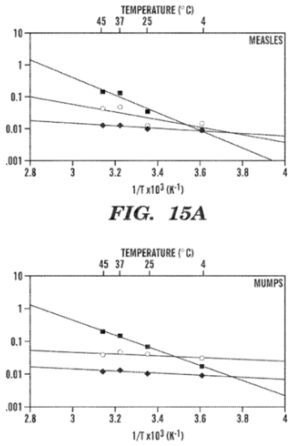
"It has come to our attention that there were significant errors in the data analysis that formed the basis of Figs. 2 and 3…and we are no longer confident in the results presented or the conclusions made…"

Figure 3(a) (partial)



Zhang et al., Stabilization of vaccines and antibiotics in silk and eliminating the cold chain, *PNAS* (109:11981-11986) 2012

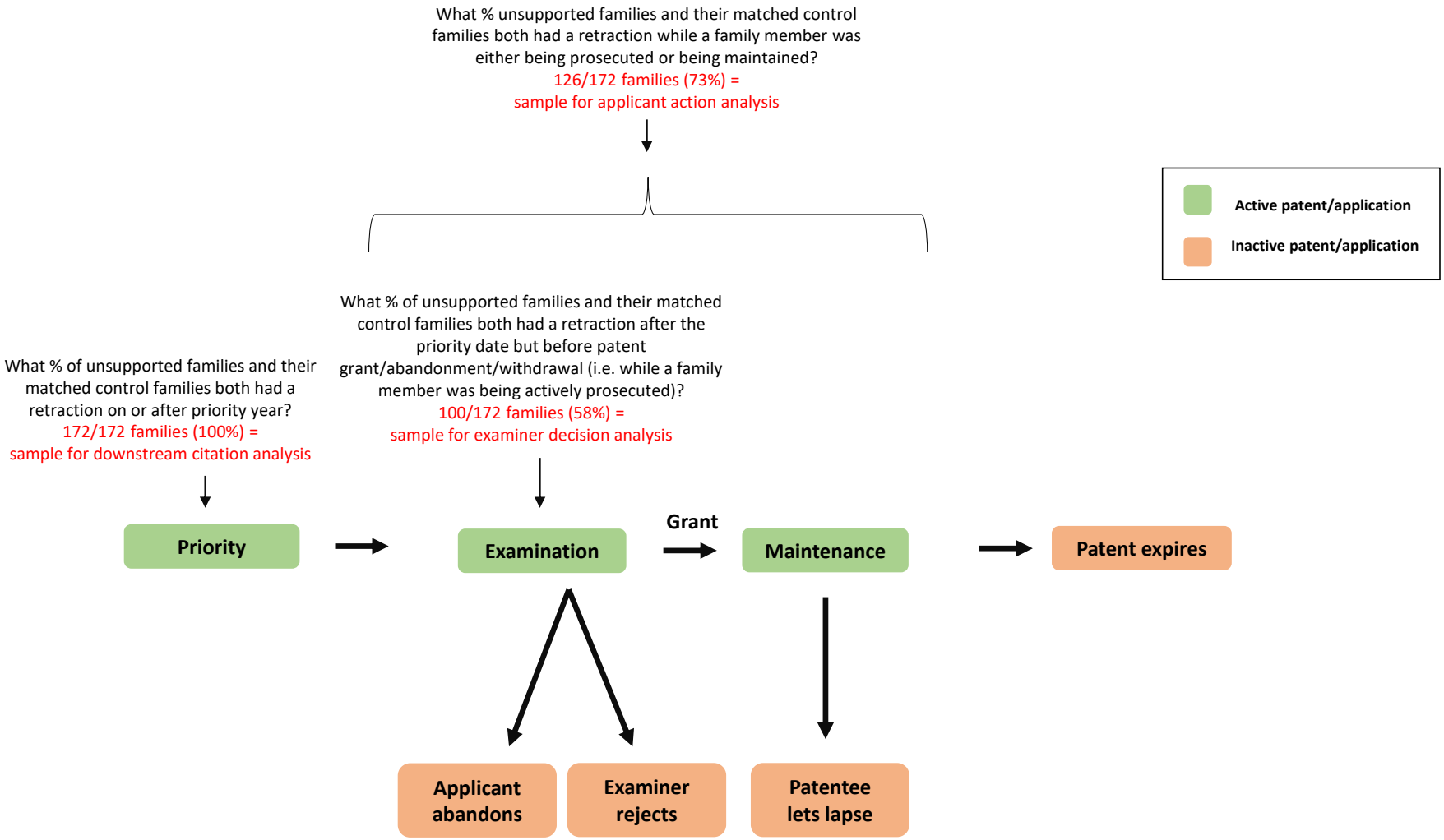
Figure 15



U.S. Patent Application No. 20180360947 (filed by Tufts university 2 years after paper retraction)

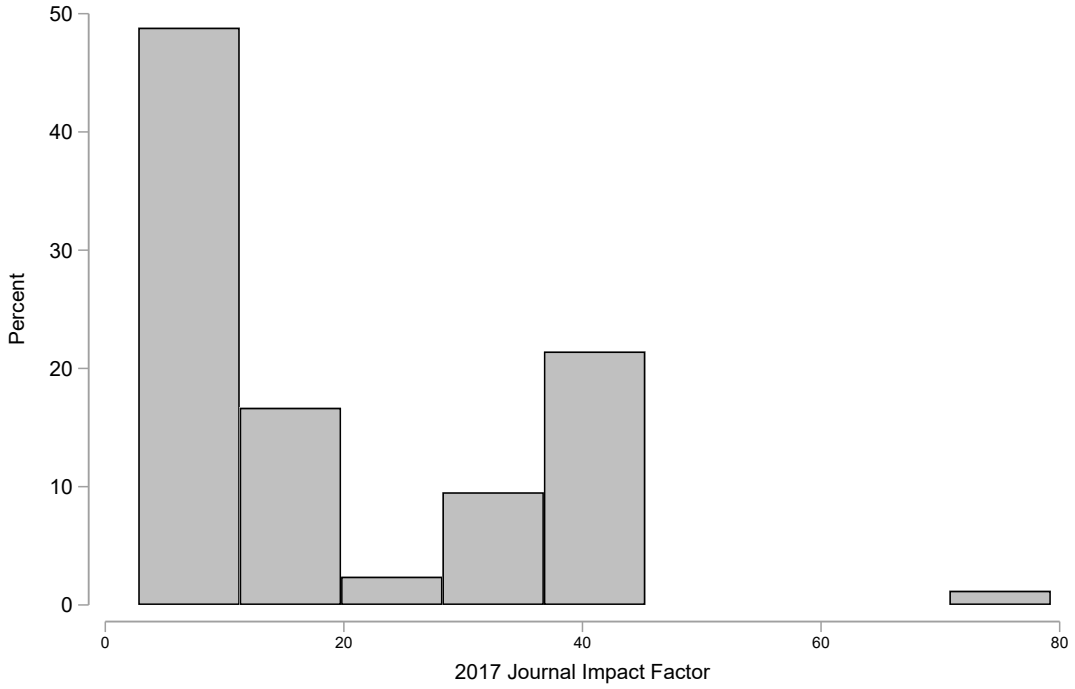
Notes: This example of a retracted paper-unsupported patent pair gives a sense of the closeness of the match between retracted material in the paper and the corresponding patent. A paper discussing how vaccines can be preserved without refrigeration (Zhang et al. 2012) was retracted with a notice stating that "there were significant errors in the data analysis that formed the basis of Figs. 2 and 3...and we are no longer confident in the results presented or the conclusions made" (Zhang et al. 2016). The corresponding patent application – filed two years after the paper retraction – contains figures identical to the retracted figures (U.S. App. 20170258889).

Appendix Figure A.2: Timing of Retraction and Sample Selection



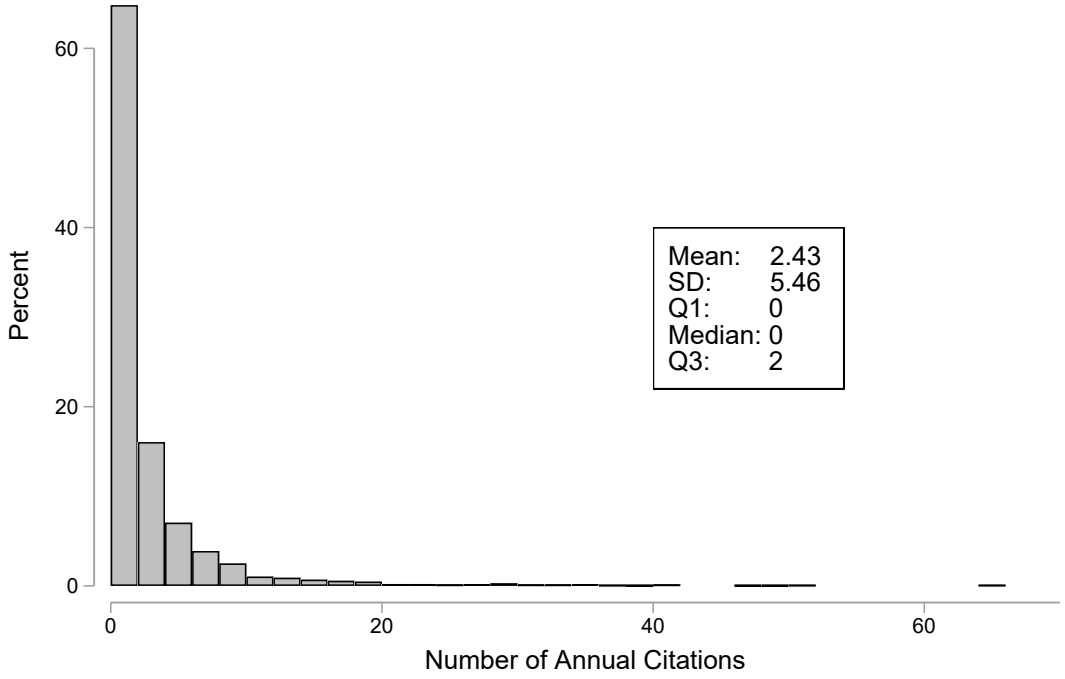
Notes: This figure provides a timeline of a patent application as it progresses through the patent system, as well as what fraction of our sample of 172 patent families (86 unsupported and 86 controls) experienced a (real or counterfactual) retraction at each stage of a patent life cycle. We selected the appropriate subsamples for each of our analyses on applicant actions (Table 2), examiner decisions (Table 2), and downstream citations (Table 3), such that the timing of the retraction could have impacted the outcome. Note that 72/86 (83%) of the unsupported families had a retraction while a family member was either being prosecuted or being maintained, while 61/86 (71%) of the unsupported families had a retraction after the priority date but before patent grant/abandonment/withdrawal. The subsamples in the above figure are smaller since we further restricted the sample to unsupported families whose matched control families also had relevant retraction timing. See Appendix A for details.

Appendix Figure A.3: Distribution of the Journal Impact Factor of Retracted Publications



Notes: This figure plots the distribution of the journal impact factor (measured in 2017) of the 84 retracted papers in our main sample. The journal impact factor was obtained from Clarivate Analytics. More than 30% of the papers come from highly ranked journals such as *The New England Journal of Medicine* (JIF 79), *Nature* (JIF 42), *Science* (JIF 41), and *Cell* (JIF 31).

Appendix Figure A.4: Distribution of Patent Citations



Notes: This figure plots the distribution of annual citations received by the 172 patent families in our main sample (86 unsupported and 86 control families).

Appendix Table A.1: Impact of Retraction on Prosecution, Maintenance, and Rejection - Logit

VARIABLES	(1) Application Action	(2) Application Action	(3) Examiner Rejection	(4) Examiner Rejection
Retracted	-0.187	-0.233	-0.177	-0.227
	(0.075)	(0.086)	(0.085)	(0.097)
Retraction-year FE	NO	YES	NO	YES
Age-at-retraction FE	NO	YES	NO	YES
IPC class FE	NO	YES	NO	YES
N of patent families	126	94	100	69

Notes: The table reports estimation from Equation B, modified for logit specification, and parallels results from Table 2. Average marginal effects are reported, and the unit of analysis is a patent family. Columns 1-2 show applicants' reaction to retraction and report whether the applicant continued to pay maintenance fees for or prosecute any patent in the family after retraction. Columns 3-4 show examiners' reaction to retraction and whether the examiner rejected any patent in the family for lack of either enablement or written description after retraction. Addition of fixed effects into the model in columns 2 and 4 led to some patent families dropping out of the regression due to perfect prediction. Robust standard errors are in parentheses.

Appendix Table A.2: Impact of Retraction on Downstream Citations - Balanced Sample

VARIABLES	(1) Total	(2) Examiner	(3) Applicant	(4) Log(1+Total)	(5) Log(1+Examiner)	(6) Log(1+Applicant)
Treat X Post-Retraction	0.677	0.543	0.134	0.054	-0.052	0.132
	(0.766)	(0.520)	(0.547)	(0.104)	(0.098)	(0.093)
Post-Retraction	-0.037	-0.470	0.434	-0.049	-0.068	-0.001
	(0.604)	(0.487)	(0.523)	(0.100)	(0.101)	(0.084)
Calendar-year FE	YES	YES	YES	YES	YES	YES
Patent family age FE	YES	YES	YES	YES	YES	YES
Patent family FE	YES	YES	YES	YES	YES	YES
N of patent families	96	96	96	96	96	96
N of patent family-years	672	672	672	672	672	672

Notes: This table parallels the results from Table 3, but using a balanced sample and examining -/+3 year window around the retraction event. We selected unsupported patents that have a full citation history of -/+3 years around the retraction event and whose control patents also have a full citation history of -/+3 years around the retraction event. This led to a sample of 48 unsupported and 48 control patent families. The unit of analysis is a patent family X year.

Appendix Table A.3: Impact of Retraction on Downstream Citations - Alternative Specifications

VARIABLES	(1) IHS(Total)	(2) IHS(Examiner)	(3) IHS(Applicant)	(4) Poisson(Total)	(5) Poisson(Examiner)	(6) Poisson(Applicant)
Treat X Post-Retraction	0.051	0.034	0.037	0.074	0.164	-0.097
	(0.109)	(0.093)	(0.099)	(0.166)	(0.188)	(0.233)
Post-Retraction	-0.141	-0.157	0.006	-0.143	-0.239	-0.027
	(0.084)	(0.083)	(0.075)	(0.123)	(0.148)	(0.186)
Calendar-year FE	YES	YES	YES	YES	YES	YES
Patent family age FE	YES	YES	YES	YES	YES	YES
Patent family FE	YES	YES	YES	YES	YES	YES
N of patent families	172	172	172	161	161	109
N of patent family-years	1540	1540	1540	1460	1460	1018

Notes: This table parallels the results from Table 3, but using alternative specifications. In columns 1-3, inverse hyperbolic sine transformation of the number of citations was used as the outcome. In columns 3-5, instead of OLS, quasi-maximum likelihood estimates based on conditional fixed effects Poisson model was used. For the Poisson specifications, some patent families never received a citation in our time period and hence dropped out of the regressions. Standard errors are clustered at the paper level for the OLS specification, and robust standard errors are reported for the Poisson specification.

Appendix Table A.4: Impact of Retraction on Downstream Citations - Papers

VARIABLES	(1) Citations	(2) Log(1+Citations)	(3) IHS(Citations)	(4) Poisson(Citations)
Treat X Post-Retraction	-10.896	-0.917	-1.103	-1.524
	(3.535)	(0.092)	(0.108)	(0.145)
Post-Retraction	2.492	0.236	0.287	0.147
	(1.650)	(0.065)	(0.078)	(0.053)
Calendar-year FE	YES	YES	YES	YES
Paper age FE	YES	YES	YES	YES
Paper FE	YES	YES	YES	YES
N of papers	170	170	170	168
N of paper-years	1407	1407	1407	1390

Notes: This table parallels the results from Table 3, modified for papers. The unit of analysis is a paper X year. Standard errors are clustered at the paper level for the OLS specifications, and robust standard errors are reported for the Poisson specification. Calendar year fixed effects consist of full set of nineteen indicator variables from 2001 to 2019; age fixed effects consist of twelve indicators, with the age eleven indicator including all subsequent age indicators.

A Data Appendix

A.1 Sample Construction

Retracted patent-paper pairs

Previous research on patent-paper pairs has, broadly speaking, used two general definitions of such pairs: (1) two documents that share temporal proximity and (most of) the same authors (Meyer 2006; Thompson et al. 2018); or (2) two documents that share temporal proximity, the same authors, and have significant overlap in content (Fehder et al. 2014; Huang and Murray 2009; Murray and Stern 2007). We use the second approach – seeking patent-paper pairs with significant overlap in content – because we wish to study patents that contain the same information as retracted papers.

We identified all retracted papers from PubMed, and specifically, those indexed in Medline, limiting our study to the life sciences. We attempted to identify retracted papers indexed in the Web of Science, which would allow us to expand beyond life sciences, but found only one additional patent-paper pair that fits our definition of pairs. We therefore chose to exclude the Web of Science result to focus on the life sciences. We hypothesize that the Web of Science did not yield many patent-paper pairs because (1) there are fewer retractions outside of the life sciences; (2) many non-life sciences retractions are retractions of conference proceedings with missing abstracts so we could not evaluate their information content; and (3) many non-life sciences retractions came from authors outside of the United States, and our methodology specifically sought US patents.[15]

In addition to Medline, we relied on data from Retraction Watch, which aims to compile a comprehensive dataset that captures retractions across all fields and systematically documents the reasons behind the retractions.

We identified patent-paper pairs as follows:

1. Identifying retracted papers: We retrieved all retracted papers from Medline that were published between 2001 and July 2019. We excluded papers that are not original research articles, such as reviews, letters, and commentaries. We matched the papers to data provided by Retraction Watch in May 2019 in order to obtain data on reason for reaction. We only kept papers that specified that the retraction occurred because information in the paper was incorrect (or did not provide a reason for retraction). We excluded papers

[15]Note that non-US inventors can file US patents, but they are presumably less likely to do so as compared to US inventors.

retracted due to reasons that cast no doubt on the veracity of the retracted information (e.g. plagiarism, IRB problems).

2. Identifying patent pairs:

   (a) From the set of retracted papers identified in Step 1, we identified all US patent applications with (i) inventors who share the same name (first and last) as the first or last author of the focal retracted publication and (ii) filing dates within +/- 2 years of paper publication. We obtained bibliographic patent data from Reed Tech (`https://patents.reedtech.com/parbbib.php`), which provides bulk files with author information for all applications filed each week.

   (b) We ran a word similarity algorithm that calculated the number of words in common between the paper abstract and the patent specification. The algorithm stems words in the paper abstract and patent specification. The algorithm then takes each word in the abstract and seeks that word in the specification. Finally, the algorithm calculates the percentage of stemmed words in the abstract that are also in the specification.

   (c) For potential pairs that had >90% overlap between words in the paper abstract and the patent specification, we manually reviewed the patent and the paper to verify that the potential pairs identified by the algorithm are indeed true patent-paper pairs. Our manual review incorporated two steps: (i) Review to determine if the potential match is a true patent-paper pair by making sure that the patent and the paper contain the same information in the text or in the figures. (ii) Review to determine if the retracted material (as specified by the retraction notice) from the paper is present in the patent. For retraction notices that did not specify the particular part of the paper that was retracted, we assumed that the entire paper was retracted.

3. Some of our unsupported patents belonged to the same patent family. Since we conducted all of our analyses at the family level, we deduplicated these patents from our sample. We chose the pair whose paper was retracted first, and if there were still ties, we chose the paper whose paper was published first.

After these steps, we identified 107 patent-paper pairs at the family level (101 papers; 96 patent families).

Control patent-paper pairs

Following papers that have studied the impact of paper retractions in science (Furman et al. 2012; Azoulay et al. 2015; Jin et al. 2019), we sought matched controls. We found control, non-retracted patent-paper pairs by matching on both paper and patent characteristics.

1. Identifying control papers: we identified all non-retracted, original research papers in Medline that were published in the same year and journal as the retracted papers of our patent-paper pairs (n = 142,579 papers)

2. Identifying control patent-paper pairs: we determined which control papers had associated patents using the same word similarity algorithm described earlier and identified potential control pairs with word overlap score >0.9 (n = 11,225 pairs). We further narrowed down this pool by focusing on potential control pairs with the same primary technology class as the unsupported patents, using the International Patent Classification (IPC) system at the class level (n = 4,550 pairs).

   Note that while we would have liked to have matched controls on additional covariates of interest, the pool of potential controls for each unsupported patent is highly skewed. Some unsupported patents have very few potential controls left after matching on paper publication year, paper journal, and IPC class and thus we were unable to match on additional covariates.

3. After matching on paper publication year, journal, and IPC class, we then manually reviewed the pool of potential control pairs to confirm that these potential pairs were indeed pairs. Specifically, for each retracted patent-paper pair, we sorted the potential control pairs by the word overlap score measuring correspondence between patent and paper and manually reviewed them in descending order of the score until we identified a true control pair. By sorting the potential control pairs by their word overlap score, we prioritize the manual review of potential control pairs in which the patent closely copies the language from the paper or if the patent and the paper use common, generic language, but we believe this should not affect the treatment assignment (retraction) or the outcome (e.g., citations that the patent receives).[16]

[16]We are facing a measurement error problem since the world similarity algorithm is not perfect. We have a pool of potential control patent-paper pairs that were algorithmically identified—only some of them are true patent-paper pairs, while others are false pairs. We argue that this measurement error is random—being a true vs. false pair in this pool of potential pairs is not a confounder and does not affect the treatment assignment (retraction) or the outcome (e.g., citations that the patent receives). Although patent-paper pairs may receive more citations than non-pairs, all of our potential patent-paper pairs "look like" patent-paper pairs, so even if some of them might not actually be true pairs, they will likely still be cited highly.

4. Some unsupported patents have the same journal, publication year, and IPC class (and hence have the same control pool). We randomly assigned one control pair to each such retracted pair, so the control pair can inherit one counterfactual retraction year.

5. Finally, some patents are associated with multiple papers. For these cases, we chose the pair whose paper was retracted first. If there were still patents left associated with multiple papers, then we chose the pair whose paper was published first.

After the 1-to-1 matching procedure, we were able to find control patent-paper pairs for 86 of our retracted patent-paper pairs, leaving us with a sample of 86 retracted pairs (86 unsupported patents and 84 retracted papers; two of the retracted papers were associated with two patents) and 86 control pairs (86 control patents and 86 non-retracted papers).

Identifying patent family members

Because members of a patent family often contain identical or very similar specifications, our unit of analysis is a patent family. For our 86 unsupported and 86 control patents identified above, we sought all of their family members. We defined related applications as both applications filed in other countries (for example, a U.S. patent may have a Japanese counterpart) and parent and/or child applications, including divisionals, continuations, and continuations-in-part. We obtained family information from Google Patents. Our information is current as of March 2020.

Timing of retraction and sample selection

In order to conduct our analyses, it is important to understand whether the retraction occurred before or after our outcomes of interest. For instance, to investigate whether an examiner reacts to retraction, the retraction must have first occurred before the examiner makes a final decision to grant or reject the patent;[17] if the retraction occurred after the patent had already been granted or rejected, then this patent should not be included in our sample for our analysis on examiner rejection/grant.

Appendix Figure A.2 provides a timeline of a patent application as it progresses through the patent system, as well as what fraction of our sample of patent families experienced a (real or inherited counterfactual) retraction at each stage of a patent life cycle. A patent application

[17]Note here that we use the term "final" rejection colloquially to mean a rejection after which the applicant stops pursuing the application. "Final rejection" is also a term of art used by PTO examiners to describe certain types of rejections but it is possible under some circumstances for the applicant to continue pursuing the application even after such a rejection.

undergoes the following stages: (i) the applicant files the first patent application in the family (the priority stage); (ii) provided the applicant does not abandon the application beforehand, the application arrives in the examiner's desk, who either rejects or grants the patent; (iii) if the patent is granted, the applicant pays maintenance fees to keep the granted patent active or chooses to let the patent lapse; and (iv) finally, typically after twenty years, the terms of the patent expire.

We selected the appropriate subsample for each of our outcomes, such that the timing of the retraction could have impacted the outcome. Among our sample of 172 patent families, all of them had a retraction on or after the priority year, so the full sample was used for our analysis on the impact of retraction on downstream citations. 63 of our unsupported patent families and their matched control families both had a retraction while a family member was either being prosecuted or maintained, so this sample of 126 families was used for our analysis on applicant's decision to prosecute/maintain the patent. 50 of our unsupported patent families and their matched control families had a retraction after the priority date but before patent grant or abandonment (i.e. during prosecution), so this sample of 100 families was used for our analysis on the examiner's decision to reject or grant the patent.

Unbalanced vs. balanced samples for downstream citation analysis

Our main analysis on downstream citations examines a period of +/- 5 years around the retraction event. This sample is unbalanced, as there are some patent families that experienced retraction early, leaving them with fewer years of pre-retraction citation data, while some patent families were retracted recently, leaving them with fewer years of post-retraction citation data. In order to retain as much of our sample as possible, our main analysis is based on the unbalanced sample.

As a robustness check, we narrow our sample to a smaller but balanced sample. We selected unsupported patents that themselves have full citation history of +/- 3 years around the retraction event and whose control patents also have full history of +/- 3 years around the retraction event.[18] This led to a sample of 48 unsupported and 48 control pairs. As shown in Appendix Table A.2, while the magnitudes change depending on the specifications, none of the coefficients are statistically significant and the results remain similar.

[18]We chose a smaller window of +/-3 years of the retraction event instead of the +/-5 years window because the +/-5 years window yielded just 23 unsupported and 23 control pairs.

A.2 Outcomes Data

Prosecution and maintenance fee payment

Prosecution: We obtained data on whether a US patent application was prosecuted after the retraction event. Data was collected manually from the USPTO's Public Patent Application Information Retrieval system (PAIR). We considered an application to have been prosecuted after retraction if the prosecution file history contained a filing that required an affirmative action from the applicant, as opposed to the examiner, and that filing occurred after the retraction date. For example, filing a response to an office action, requesting an interview, paying an issue fee, or filing an IDS all require affirmative action from the applicant and were counted if they occurred after the retraction date. By contrast, the examiner issuing a rejection, filing a search report, or granting a patent were not counted because the action was not initiated by the applicant. Our information is current as of August 2020 and applies to US patents only.

Maintenance fee payment: We collected data on whether unsupported and control patents were maintained by the applicant after the retraction event. We obtained information on maintenance fee payment and dates of payment from Google Patents, and included that information in our analysis for every country for which the information was available. Our information is current as of August 2020.

Examiner decision

We obtained data on whether US patent applications were rejected by examiners for either lack of enablement or lack of written description, which might indicate that the examiner was skeptical of the validity of the data. Rejection data was aggregated by the USPTO's Office of the Chief Economist.[19] A limitation of this data is that it is only current through 2017, and does not include rejections that occurred after 2017. We searched the rejection data for entries containing the relevant application number and the text "112,a", which indicates a rejection under 35 U.S.C. 112(a), the statutory section directed to enablement and written description.[20]

[19] Available here: `https://www.uspto.gov/learning-and-resources/electronic-data-products/office-action-research-dataset-patents`.

[20] 35 U.S.C. 112(a): "The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor or joint inventor of carrying out the invention."

Downstream citations

We obtain citation data for all family members from Google Patents, which tracks citations from patents filed in 22 jurisdictions.

We then sought to determine the citation date for each citing document. Generally, studies that use patent citation dates assume that the citation date is (1) the priority date (Bacchiocchi and Montobbio 2009); (2) the year of patent filing (Alcácer and Gittelman 2006); or, most commonly (3) the patent grant date (Hall et al. 2001; Jaffe and Trajtenberg 1999; Nicholas 2008). Because we seek to understand how citation patterns to patents change when the corresponding paper is retracted, all of these strategies are imperfect measures. Patent grant date is certainly later than both applicant-added and examiner-added citations, and, moreover, we use patent applications, not all of which have been granted. Priority date might be the correct citation date for some applicant-added citations, but is certainly erroneously early for examiner-added citations. We chose to use filing date to approximate citation dates because, although it is also likely early for examiner-added citations, it is more accurate than either the patent grant date or the priority date. Note that due to the imperfect measure of determining exact citation dates, the year of first citation can precede the priority year for some of our patent families.

For our analysis, when we count the number of citations to our focal unsupported or control patent, we count the number of total citations to the focal patent's family in a given year. For instance, if a downstream patent cites to two members of our unsupported patent's family, two citations are counted. Similarly, if a downstream patent cites to an unsupported patent and a family member of that downstream patent also cites to the unsupported patent, two citations are counted. Citation data was collected in March 2020.

A.3 Other Data

Partial or full paper retraction

In order to determine whether parts of a retracted paper remained good science even after the retraction, we read the text of each retraction notice and classified it as partially or entirely retracting the paper. Partial retractions suggest that some of the results reported in the paper are

valid, despite the retraction.[21] Total retractions suggest that none of the results reported in the paper are valid.

Citation data for papers

Downstream citation data for the papers of our patent-paper pairs was exported from PubMed in November 2019 and August 2020. Although this citation data only contains downstream citations from other PubMed articles and may miss citations from non-PubMed articles, since our study focuses on the life sciences, we expect this citation data to capture most, if not all, of the downstream citations that our papers have received.

Patent owners

In order to distinguish academic and industry patent applicants, we sought information on the institutional affiliation of the inventor, defined as the first-listed assignee (information from Google Patents). We then classified the assignee as either a non-profit or for-profit institution. Assignees who were individuals were classified as non-profits.

Patent technology class

We used the International Patent Classification (IPC) system at the class level to identify a patent's technology class. Data on IPC class was obtained from PatSnap in October and November 2019. A patent can be associated with multiple IPC classes, and we used the primary IPC class of the patent, as identified by PatSnap.

Litigation

Litigation data was obtained from Google Patents, which sources its data from Darts-ip's Global Patent Litigation Dataset. Litigation includes court trials but does not include administrative proceedings.

[21] For example: "During investigation of engineered point mutants of the G2A receptor, we were unable to repeat these radiolabeled ligand-binding studies following similar protocols. Alternative protocols... also failed to establish direct G2A binding. This calls into question the major conclusion that LPC and SPC are direct ligands for G2A....[but] Data generated by Dr. Kabarowski demonstrating cellular migration dependent on LPC addition and G2A receptor expression have been reproduced and extended in independent work. We believe these data to be accurate and reproducible and therefore conclude that G2A is an effector of LPC action in certain cell types." Owen N. Witte, et al., Retraction, Science 307(5707): 206 (2005).

B Legal Appendix

This appendix elaborates on several legal doctrines that are mentioned in the article:

Duty of disclosure

Applicants for US patents owe a duty of disclosure, candor, and good faith to the USPTO (37 CFR 1.56). This duty applies not just to inventors, but to everyone involved in filing or prosecuting the application, including attorneys, assignees, and "every other person who is substantively involved in the preparation or prosecution of the application." (MPEP 2001.01). Specifically, the duty requires disclosure of "all information known to that individual to be material to patentability." (MPEP 2001.01).

Would knowingly prosecuting a patent application that contains retracted material, and failing to disclose that fact to the PTO, be in violation of the duty of disclosure? The answer will depend on the precise facts of the retraction and patent in question, in particular how important the retracted material was to patent grant, but it is likely that such a failing would often violate the duty of disclosure. Although we were unable to find cases directly addressing the question of retracted information, several cases have found that including fabricated data[22] in the patent specification violates the duty of disclosure. Techno Corp. v. Kenko USA, Inc., 515 F.Supp.2d 1086 (N.D. Cal. 2007) (finding inequitable conduct when the patent specification included fabricated human clinical trials); Hoffmann-La Roche, Inc. v. Promega Corp., 323 F.3d 1354 (Fed. Cir. 2003) (holding that describing a prophetic example in the past tense (indicating that it had been conducted, when that was not the case) constituted intent to deceive.).

Circumstances under which an examiner's rejection based on prior art that contains retracted material would be erroneous:

Patent applications can be rejected for (1) lack of novelty or (2) obviousness. Examiners issuing either rejection will cite to specific prior art, but the impact of retracted material in that prior art is different for the two rejections.

1. Lack of novelty: Patents are rejected for lack of novelty (35 U.S.C. 102) if the invention has previously been disclosed in the prior art. Examiners may only make this rejection is the prior art is enabled, meaning that the prior art discloses the invention in sufficient detail that

[22]Note that not all unsupported patents contain *fabricated* data—many contain retracted material that results from errors, not fraud.

others in the field could make and use the invention. In other words, a rejection based on lack of novelty is not correct if the invention described by the prior art does not work[23]—as would often be the case with retracted material.

2. Obviousness: Patents are rejected for obviousness (35 U.S.C. 103) if the invention is obvious in light of one or more prior art references. Unlike rejections for lack of novelty, the prior art in this case does not need to be enabled. Thus, as a theoretical matter, retracted material could properly be used as prior art for obviousness rejections. In practice however, it would often not be proper for an examiner to cite retracted material as part of an obviousness rejection. This is because the fact of retraction demonstrates that the retracting scientist could not actually make the invention, which in turn suggests that it is not obvious to scientists in the field how to make the invention. Thus, using retracted material as prior art will often support a finding of nonobviousness more than a finding of obviousness.

[23]And other scientists could not figure out how to make it work from the description provided in combination with their knowledge of the field.