Asymmetric Information and R&D Disclosure: Evidence from Scientific Publications*

Stefano H. Baruffaldi[†a], Markus Simeth[‡b], and David Wehrheim[§c]

[a]University of Bath & Max Planck Institute for Innovation and Competition
[b]Copenhagen Business School
[c]IESE Business School

May 26, 2022

ABSTRACT

We examine how asymmetric information in financial markets affects firms' R&D disclosure policies, considering scientific publications as a disclosure channel. Difference-in-differences regressions around brokerage house mergers and closures (which increase information asymmetries through reductions in analyst coverage) indicate a quick and sustained increase in scientific publications from treated firms relative to the number of publications from control firms. The treatment effects are concentrated among firms with a large increase in information asymmetry, firms with greater financing constraints, firms run by managers with a short horizon, and firms with low proprietary costs. We do not find evidence of changes in R&D investments. Results from ordinary least squares regressions show that scientific publications by firms are positively associated with investor attention toward those firms. We complement these results with qualitative evidence from conference calls. Overall, our evidence lends support to the idea that R&D disclosure is motivated by the desire to communicate with investors but comes at the cost of rivals potentially obtaining valuable information. Our findings highlight the importance of financial analysts for R&D firms and financial markets for cumulative research.

KEYWORDS: financial analysts, financial signaling, information asymmetry, investor attention, patents, proprietary costs, R&D disclosure, scientific publications

---

*We thank Benjamin Balsmeier, Vicente Bermejo, Iván Blanco, Bruno Cassiman, Miguel Duro, Jose María Flores, Igor Kadach, Bastian Krieger, Johannes Luger, Ali Mohammadi, Maikel Pellens, Daniel Streitz, Neus Palomeras, Andrea Patacconi, Thomas Rønde, Valentina Tartari, Giovanni Valentini, and Moqi Xu, as well as workshop participants at the Barcelona GSE Summer Forum (2019), the University of Luxembourg (2019), Ecole Polytechnique Fédérale de Lausanne (2020), ZEW Mannheim (2021), the EU-SPRI conference (2021), the AEFIN (2021) conference, the EARIE (2021) conference, the EPIP (2021) conference, and the SEI faculty consortium at ESADE (2021) and seminar participants at CUNEF (2020), ESADE Entrepreneurship Institute (2020), IESE Business School (2020), UNU-MERIT (2021), and Norwich Business School (2021) for their helpful comments and suggestions. All errors are our own. The authors acknowledge financial support from the Independent Research Fund Denmark (Project 0133-00119B) and the Spanish Ministry of Science, Innovation and Universities (Project PGC2018-094418-B-I00). The authors are listed in alphabetical order.

†University of Bath, School of Management, United Kingdom; University of Bath and Max Planck Institute for Innovation and Competition, Munich, Germany; s.h.baruffaldi@bath.ac.uk.

‡Copenhagen Business School, Department of Strategy and Innovation, Denmark; ms.si@cbs.dk.

§IESE Business School, Strategic Management Department, Spain; dwehrheim@iese.edu.

1. Introduction

Understanding the consequences of information asymmetry in financial markets has been a topic of longstanding concern in accounting, economics, and finance (e.g., Merton, 1987; Diamond and Verrecchia, 1991). A central theme that emerges from these studies is that information asymmetry can lead to an increase in the cost of capital or even an inability to attract external capital.[1] There is considerable evidence that firms can reduce information asymmetry and thus their cost of capital by providing voluntary disclosures, such as management forecasts of earnings per share or details in annual reports and 10-K filings (e.g., Francis et al., 2008; Balakrishnan et al., 2014). However, this relationship is less clear in settings where proprietary costs are substantial because potential capital market benefits could be offset by information leakage to rivals (Bhattacharya and Ritter, 1983).

The tension between the costs and benefits of disclosure is of particular importance to firms that invest in research and development (R&D). Aboody and Lev (2001) emphasize important differences between R&D and other investments in terms of information asymmetry. First, R&D is firm specific, which makes it difficult for investors to derive inferences about the productivity and value of a firm's R&D by observing the R&D performance of other firms. Second, R&D is not traded in organized markets where investors can obtain information on R&D investments' productivity and value. Last, accounting principles require R&D to be immediately expensed and do not require financial reporting on the value or productivity of R&D investments. Hence, there should be substantial demand from financial markets for firms to disclose nonfinancial information about their R&D activities, in particular value-relevant information (Amir and Lev, 1996).

In this paper, we analyze the relation between information asymmetry and R&D disclosure and focus on a type of R&D disclosure that has received surprisingly little attention in the academic literature: publications in academic journals. Why should investors care about corporate publications in academic journals? A substantial proportion of firms' R&D outputs are published in academic journals, and large R&D-intensive companies such as Google, IBM and Merck are included in the prestigious Nature Index.[2,3] Scientific publications provide not only indirect information on R&D productivity but also direct information on firms' R&D capabilities thanks to the thorough peer-review process, which ensures that high academic standards are met, and well-known journal rankings (Dasgupta and David, 1994).[4] Moreover, it is typically basic research that is disclosed in the form of scientific publications (Arora et al., 2021), and the contribution of basic research to firm productivity and growth is much larger (by a factor of approximately 3) than that of other types of R&D (Griliches, 1986). Simeth and Cincera (2016) establish a positive relationship between scientific publications and firm value.

However, whether firms use scientific publications to actively communicate with investors is an open question. Some anecdotal evidence on this issue emerged in a recent policy workshop.[5] Jochen Maas, managing director of the R&D division at Sanofi-Aventis Germany, suggests that firms use scientific publications to certify and enhance their reputation, generate interest among investors and stimulate share prices. In addition, he points out that compared to alternatives such as patents, publications are a much faster way to build awareness of the firm among firm outsiders.[6] Importantly, there is also evidence that the decision to publish scientific papers is strategic. A number of case studies conclude that what most deters firms from publishing is the risk of knowledge leakage to rivals (Polidoro and Theeke, 2012). Accordingly, firms have formal clearance processes in place regarding what their scientists may or may not publish (Blumenthal et al., 1996).

We draw from a large literature exploring the causal effects of information asymmetry on various firm and capital market outcomes to provide systematic evidence that firms seek to communicate with investors through scientific publications and that such efforts lead to a lower cost of capital. Specifically, our tests exploit two natural experiments, namely, brokerage house mergers and closures, which generate plausibly exogenous reductions in analyst coverage. This approach was first advocated by Hong and Kacperczyk (2010), who document that such declines are associated with increases in the forecast error variance among analysts who continue to cover the stock. Using several proxies for information asymmetries, Kelly and Ljungqvist (2012) show that coverage terminations cause an increase in information asymmetry, a decrease in a firm's share price, and a reduction in retail investors' demand for the stock (see also Balakrishnan et al., 2014; Ellul and Panayides, 2018).

Our empirical approach uses 43 brokerage house mergers and closure events staggered over time from 2000 until 2010 and considers publicly traded U.S. firms that are active in R&D. We limit attention to R&D firms because research in accounting demonstrates that financial analysts are important information intermediaries for such firms (e.g., Barth et al., 2001; Palmon and Yezegel, 2012). Associated with these mergers are 760 firms that were covered in the year before the event by both merging houses or the closing house, our treatment sample. Using a difference-in-differences (DiD) approach, we compare changes in scientific publication rates of this treatment sample to a control sample of observationally similar firms unaffected by the brokerage house merger/closure, thus identifying the causal change in scientific publication rates from the loss of coverage.

We find that firms experiencing a decline in analyst coverage adjust scientific disclosures; namely, their publication rates increase. The increase in scientific publications is statistically significant and economically meaningful: our baseline analysis indicates a rise of between 12% and 15% in the number of publications due to shocks to information asymmetries. We limit concerns regarding systematic differences across firms, years, or events by including the respective fixed effects. We also account for differences in observable characteristics between the treatment and control samples by including control variables. Our results are robust to using a DiD matching estimator and continue to hold after we conduct a battery of additional tests. We examine the dynamic effects of broker mergers/closures and show that we do not violate the parallel trends assumption. Publication rates for the treatment sample increase mainly between the year before and the year after the shock, corroborating the interpretation that firms increase their scientific publications in an attempt to communicate with investors. This is further supported by evidence from conference calls.

The treatment effects that we identify are concentrated in the following subsamples: firms with a greater increase in information asymmetry (Kelly and Ljungqvist, 2012), firms with greater financial constraints (Derrien and Kecskés, 2013), firms run by managers with shorter horizons (Glaeser et al., 2020), and firms with lower proprietary costs (Glaeser, 2018). For instance, scientific publications increase by up to 21% in firms that are financially constrained, whereas the corresponding effect is insignificant for firms that are financially unconstrained. Similarly, we find that scientific publications increase by up to 25% for firms with low proprietary costs. Among the set of firms with high proprietary costs, the effect is economically small and statistically insignificant. The observed heterogeneity in the treatment effect is consistent with the existence of a trade-off between capital market benefits from increased R&D disclosure and the risk associated with revealing proprietary information to rivals (Bhattacharya and Ritter, 1983).

Next, we consider two important alternative perspectives. First, firms may simply communicate with investors through other channels, such as managerial earnings forecasts (as identified in Balakrishnan et al., 2014). However, the DiD tests do not show a significant change in managerial earnings forecast provision behavior in the treatment sample relative to the behavior of the control sample. Consistent with the arguments of Palmon and Yezegel (2012), among others, this result suggests that such alternative channels have limitations in reducing information asymmetries in the R&D context.

Second, we explore the alternative interpretation that analysts impose pressure on managers to meet short-term goals and that reductions in coverage encourage them to increase R&D investments, as suggested by He and Tian (2013). However, consistent with research in accounting (e.g., Barth et al., 2001; Palmon and Yezegel, 2012), the more recent literature in finance casts doubt on this interpretation (e.g., Clarke et al., 2015; Bellstam et al., 2021), at least in the R&D context, which is our focus. We observe no evidence of changes in R&D inputs, other R&D output measures, or the use of scientific research in R&D among the treatment sample. Nor do we see meaningful differences between the treatment and control samples in the incidence of hiring new scientists.

Finally, we explore one plausible channel for our results. We try to establish some direct evidence on whether scientific publications influence investors' attention and financial markets. Merton (1987) describes a channel whereby increased investor attention is associated with a lower cost of capital and higher share prices. We use three empirical proxies in ordinary least squares (OLS) regressions to analyze the relation between scientific publications and investor attention at the firm–month level. We find that scientific publications are associated with increased news articles, Google searches for company tickers, and news searching activity on Bloomberg terminals after we include firm–year fixed effects, month fixed effects, and controls for other stock characteristics related to investor attention.

Economic research highlights why scientific disclosure is important. Furman and Stern (2011) show that the deposit of individual cell lines amplifies the cumulative impact of individual scientific discoveries. Murray et al. (2016) show that increased openness in scientific research encourages entry by new researchers and exploration of more diverse research paths. Mukherjee and Stern (2009) examine the theoretical conditions supporting open science or secrecy, stressing that growing the stock of public knowledge requires a limit on the private returns obtained through secrecy. Given the ongoing concern about the apparent decline in publication rates among publicly traded firms in the U.S. and the associated slowdown in economic growth (Arora et al., 2021), it is important to understand firms' disclosure incentives. Our paper takes a step in this direction by exploring the impact of financial markets.

This study contributes to several strands of literature. First, our paper contributes to a growing empirical literature on the interaction between capital market benefits/proprietary costs and R&D disclosure choices (e.g., Dass et al., 2020; Glaeser et al., 2020). Notably, our paper is the first to examine disclosure choices in the context of scientific publications. Second, our paper adds to the literature on financial analysts. While earlier studies find that brokers and analysts provide information about a firm's R&D activities that matter to investors, some of the recent literature argues to the contrary that analysts ignore R&D. Our findings support the assertion of Palmon and Yezegel (2012, p. 624) that "*analysts covering R&D-intensive firms closely follow firms' research. They read scientific publications to assess the productivity and value of the firm's R&D investments*". Third, we contribute to the literature on financial markets and firms' R&D activities. Most studies in this literature draw inferences about changes in R&D activities from changes in observable R&D outcomes. Our findings suggest that this literature needs to account for potential changes in disclosure dynamics (see also Glaeser, 2018; Reeb and Zhao, 2020). Finally, we build on qualitative evidence on scientific publications by documenting large-scale evidence that firms publish scientific research, in part, to attract investor attention.

The rest of the paper is organized as follows. Section 2 develops the hypothesis to be tested and reviews the related literature. Section 3 provides qualitative evidence from conference calls. Section 4 describes the empirical setup and data. Section 5 establishes the empirical evidence on the link between information asymmetry and scientific publications. Section 6 examines alternative ways to corroborate our results. Section 7 establishes the empirical evidence on the link between scientific publications and investor attention. Section 8 concludes.

2. Theoretical development, context and related literature

2.1. Theoretical development

Information asymmetries among investors create trading frictions by inducing adverse selection, leading to lower liquidity (Glosten and Milgrom, 1985; Kyle, 1985). This illiquidity is priced by the market, increasing a firm's cost of capital (Brennan and Subrahmanyam, 1996; Amihud, 2002). Theoretical models propose that a firm's commitment to greater information disclosure can reduce information asymmetry and lower the cost of capital. For instance, Diamond (1985) shows that public disclosure reduces private information acquisition costs by some investors, thereby reducing information asymmetry among investors; Lambert et al. (2007) show that higher-quality or more precise firm-specific disclosures decrease the covariance of a firm's cash flow with the cash flows of other firms; and Merton (1987) shows that greater disclosure increases investors' awareness of a firm's existence and enlarges its investor base, which improves risk-sharing and reduces the cost of capital.

Although most empirical studies on the relation between disclosure and the cost of capital have focused on financial disclosure, the same mechanisms are likely to apply to nonfinancial disclosure. Indeed, past research has documented evidence on the disclosure of information related to corporate social responsibility, environmental initiatives, and customers (e.g., Ellis et al., 2012; Plumlee et al., 2015). Nevertheless, aside from R&D outputs such as new product/service introductions and downstream inventions (e.g., Dass et al., 2020; Hsu et al., 2021), little is known about the disclosure of scientific research. However, this is important because investors demand early-stage information on firms' R&D potential (Guo et al., 2004).

Scientific research is typically disclosed in the form of scientific publications, and there is a fair amount of research suggesting that scientific publications are value relevant (e.g., Simeth and Cincera, 2016; Arora et al., 2018). Of course, scientific publications can affect a firm's financial performance and value through channels other than those related to financial disclosure. For instance, some scientific publications may reduce information asymmetries by assisting investors in evaluating the characteristics of a firm's current and future sales revenues. Indeed, Azoulay (2002) demonstrates that prescription drug sales increase in response to scientific evidence regarding the safety and efficacy of new drug therapies published in peer-reviewed journals with good standing in their field. Furthermore, given the large number of head-to-head comparative studies published after a new drug's entry and the fact that these studies are important for physicians and patients to understand the benefits and risks of different treatments, this information is also relevant for assessing potential business-stealing effects.

Scientific publications may also enhance capital market participants' perceptions of the firm; i.e., firms may benefit from a positive association between academic recognition and firm capabilities (Audretsch and Stephan, 1996; Arora et al., 2018). When firms publish in academic journals, the scientific community (through the peer-review process) certifies that the research is consistent with academic standards. This enhances credibility, for instance, if a firm reports that a new drug was useful in treating a certain disease. Scientific publications also convey good news about the quality of the firm's R&D activities. A recent publication record, in particular, indicates that the firm can keep such activities close to the scientific frontier. It is therefore not uncommon for R&D-active firms to present details on scientific publications in their earnings conference calls when discussing the firm's accomplishments (see the discussion in Section 3). Furthermore, scientific publications may permit investors to infer information about successful partnerships with academic institutions.[7]

If investor attention is a scarce resource, then firms choosing to publish scientific articles may also benefit from increased investor attention. Studies suggest that increased investor attention to information events (e.g., earnings announcements, advertising, press coverage) is associated with improvements in liquidity (e.g., Grullon et al., 2004; Madsen and Niessner, 2019). Furthermore, publicly traded firms often manage investor attention strategically through, for instance, the use of investor relations (Bushee and Miller, 2012) or the timing of disclosure (deHaan et al., 2015). In the R&D context, Fitzgerald et al. (2021) propose that investors pay little attention to incremental R&D. The premise of their argument is that investors are not aware of such activities or are aware of the information but do not process it because it is familiar to them. However, Fitzgerald et al. (2021) also propose the opposite for explorative R&D; because individuals tend to place higher emphasis on novel and unique signals when selecting and processing information, investors devote more time to understanding the economic significance of ideas unfamiliar to them.

Following this logic, there are several arguments for why scientific publications may be effective in attracting investor attention. First and foremost, the nature of scientific inquiry involves discovering new cause–effect relationships and technical principles, and novelty is an important criterion in the peer-review process (Dasgupta and David, 1994; Stephan, 1996). This creates significant barriers to publication in scientific journals, which in turn reduce the amount of information to be processed. Another important feature is that scientific quality is readily observable at the moment of publication (even for scientific laypeople) thanks to well-established journal rankings, making it easier for investors to compare R&D outputs across firms. Finally, the media also play a relevant role in the dissemination of scientific articles, as exciting research findings are frequently covered in the popular press, which further helps increase firm visibility among investors.[8]

Having established the benefits of scientific publications, we now move to the counterpart in the disclosure trade-off: proprietary costs. Early theoretical models argue that managers follow a full disclosure policy because, in the absence of disclosure, investors assume the worst regarding the firm's prospects and discount the value of the firm (Grossman, 1981). However, the existence of proprietary information extends the range of possible explanations for why managers withhold information: managers will only disclose proprietary information when the increase in firm value from disclosure exceeds the associated proprietary costs (Verrecchia, 1983). Proprietary costs can take several forms. For instance, knowledge about the initial development stage of a new technology could allow an existing rival to begin development, reducing a firm's advantage from the product's early market entry. However, when it is more difficult for a rival to steal a firm's ideas, the competitive costs of disclosure are reduced.

Entwistle (1999) interviewed senior executives of leading technology firms in Canada, and most interviewees reported being "*very concerned*" about revealing proprietary R&D information. Entwistle (1999) notes that "*this concern was most commonly expressed in terms of competitors using the disclosure to usurp the firm's advantage*" (p. 325). In terms of scientific publishing, revealing information about the firm's latest discoveries, at a minimum, provides useful information to rivals about the firm's research direction. At the maximum, it can facilitate rivals' attempts to imitate technologies, refine existing products, or make subsequent discoveries (Polidoro and Theeke, 2012).

A prominent example occurred at DuPont in the 1930s when the company lost its proprietary position on nylon to IG Farben due to a publication by Wallace Carothers, one of its leading scientists (Hounshell and Smith, 1988, p. 302). In 1931, Carothers published an article in the *Journal of the American Chemical Society* with a definitive statement that the compound caprolactam could not be polymerized. In 1937, while in negotiations on another matter, DuPont informed IG Farben that it had succeeded in developing a polyamide fiber, nylon 66. Once informed about nylon, IG Farben's scientists went over every piece of DuPont's research and found Carothers's statement on caprolactam, inspiring IG Farben to experiment with the compound for creating a synthetic nylon. By January 1938, IG Farben had succeeded in polymerizing it into nylon 6, an invention that turned out to be fatal for DuPont.

To protect firms against such competitive damages, most firms approach scientific publications strategically (Polidoro and Theeke, 2012). Upon hiring a scientist or engineer, almost all firms that invest in R&D require an employee to sign a preinvention assignment, requiring the employee to disclose and assign all inventions, improvements, or useful processes made during employment to the employer. Employment contracts also include nondisclosure clauses in which the scientist must agree not to disclose or divulge any confidential information or trade secrets to the public (Dworkin and Callahan, 1998). This last requirement is important because it allows the firm (not the scientist) to decide which pieces of information to make public and, if so, when to disclose them.

Firms have specific guidelines in place regarding what their scientists may or may not publish in academic articles.[9] Based on survey responses from senior executives at life science firms, Blumenthal et al. (1996) document publication delays and secrecy restrictions on information resulting from academic research. For instance, 58% of the respondents stated that their firm requires scientists to keep information confidential. Using survey responses from firms engaged in university licensing, Thursby and Thursby (2007) report that 90% of the university contracts included publication delay clauses. Taking a different perspective, Czarnitzki et al. (2014) asked academic scientists about any disclosure restrictions that they experienced in research funded by industry. Out of the respondents with industry sponsorship, 41% reported partial or full secrecy requirements on publications.

In sum, the costs associated with releasing publications include the fact that the information can help rivals and others that would use the information to the firm's disadvantage, whereas the benefits can include reduced information asymmetry regarding current and expected sales revenues, positive exposure via association with strong scientific capabilities, and, in response to visibility concerns, an increased capacity to attract attention from market participants. We hypothesize that a shift in the relative benefits of scientific publications due to shocks to information asymmetries increases the likelihood of firms choosing to publish scientific articles instead of keeping the information secret. Furthermore, scientific publications should enhance a firm's stock liquidity. Therefore, we expect that the relative benefits of scientific publications induce firms to increase their publication rates and result in these firms experiencing a reduction in information asymmetries among investors, which in turn improves stock liquidity.

2.2. Empirical context and related literature

To test our predictions above, we examine the effects of asymmetric information shocks among financial analysts. A long line of literature argues that analysts produce information that matters to market participants. Extensive evidence documents the beneficial and informative role played by analysts (e.g., Hong et al., 2000; Barth and Hutton, 2004). The literature also finds that analysts' reports impact stock prices by increasing investor awareness and demand for stocks (Womack, 1996; Loh and Stulz, 2011). By producing information about the firms that they cover, analysts also monitor these firms (Chen et al., 2015). Although analysts sometimes issue biased reports (Lin and McNichols, 1998), their incentives tend to be related to the accuracy of the information that they provide (Hong and Kubik, 2003).

There are some discrepancies in the literature with respect to the relation between analyst coverage and R&D activities. Barth et al. (2001) argue and find that analysts have more incentives to follow R&D-intensive firms than firms with lower or no R&D. This is because following firms with more R&D can yield more profitable investment recommendations and higher trading commissions. In addition, analysts expend greater effort to follow such firms. Similarly, Palmon and Yezegel (2012) argue that analysts possess the skills necessary to analyze R&D activities and find that analysts issue more informative recommendations for high-R&D firms (see also Barron et al., 2002). In contrast, He and Tian (2013) characterize analysts in a negative light, arguing that due to a lack of understanding, analysts ignore R&D when making stock recommendations. The authors show that firms with higher analyst coverage have lower R&D outputs. A common explanation for these discrepancies is that while studies such as Barth et al. (2001) focus on R&D inputs in terms of R&D expenditures, He and Tian (2013) focus on R&D outputs in terms of patents and patent citations.

However, the findings in Clarke et al. (2015) move the discussion toward a different explanation that is consistent with the argument of Barth et al. (2001). Specifically, the authors show that the negative relation between analyst coverage and patenting obtained by He and Tian (2013) is driven by firms with either zero patents or citations in the past, i.e., firms with little R&D. For firms with substantial R&D, this relation turns positive. The authors conclude that analysts have a rather good understanding of R&D activities and therefore play an important informative role (see also Guo et al., 2019). There is also evidence from analysts' reports on this matter. Bellstam et al. (2021) propose a new measure of corporate R&D derived from textual descriptions of firm activities in analysts' reports and find that this measure correlates strongly with patenting and R&D intensity among patenting firms, suggesting that analysts provide investors with valuable information on R&D.

The above discussion indicates that the specific context is critical when analyst coverage is considered as a proxy for information asymmetry. For this reason, we focus on R&D-active firms and assume that analysts covering such firms have sufficient knowledge to inform market participants about R&D activities. In addition to the evidence reported above, we conducted a number of interviews with investor relations professionals from firms involved in R&D, including the head of investor relations of a multinational pharmaceutical firm with revenues in excess of \$10 billion. All interview partners agreed that R&D-related knowledge is paramount for analysts following such firms. As one interview partner put it when asked whether analysts have sufficient training to understand scientific research, "*If you need to assess the future earnings of a company like ours, you need to understand the science*".

We differ from He and Tian (2013) in two other important aspects. First, instead of focusing on patenting, we focus on scientific publishing. Second, instead of interpreting scientific publications as a measure for R&D investment, we interpret the relation between publication rates and coverage as evidence that firms change their level of scientific disclosure in response to changes in the number of analysts following them. It is worth pointing out that ours is not the first paper to see communication possibilities in the relation between analyst coverage and R&D outcomes. Reeb and Zhao (2020, p. 157) remark that "*financial intermediaries potentially influence the disclosure of innovation rather than research and development success*", and there is a growing empirical literature whose findings cast doubt on the extent to which positive R&D outcomes are an unambiguous measure of firms' investments in R&D (e.g., Glaeser, 2018; Glaeser et al., 2020).

Our study is also related to Balakrishnan et al. (2014). These authors show that an increase in information asymmetry stemming from brokerage house mergers and closures leads to an increase in firms' voluntary disclosure. Similarly to our work, Balakrishnan et al. (2014) also maintain that for firms that increase their disclosures, there is an increase in the liquidity of their shares. However, in contrast to our work, Balakrishnan et al. (2014) focus on disclosure in the form of earnings guidance regarding earnings per share (EPS) numbers. Prior studies emphasize that such disclosures are less relevant for R&D-intensive firms (Amir and Lev, 1996; Palmon and Yezegel, 2012). In addition, the forward-looking information in such disclosures lends themselves to misrepresentation and bias (Rogers and Stocken, 2005; Glaeser et al., 2020). In contrast, we focus on the disclosure of R&D outcomes, motivating us to use scientific publications as a disclosure form.

3. Evidence from conference calls

Before turning to our systematic analysis, we provide evidence of the relevance of scientific publishing for R&D firms by looking at a conference call discussion between senior management and investors during times when a firm had difficulties raising external funds and investors were concerned about the firm's share price. Below, we quote from an earnings call of one of the firms included in our sample, Aastrom Biosciences, in August 2008:

<**Q – Scott Smith (Investor)**>: *Now can you answer a question sort of going back to the share price? And this goes back to promoting the company. Can you tell me why are you guys not promoting the company more regarding share price...? Why is Aastrom so quiet about this? Why not tell the world "Hey, we are doing this," and try to attract new investors and new money into the company? Considering that you have made cardiac the focus of Aastrom, wouldn't it makes sense to do this? It seems like Aastrom was extremely quiet throughout this period. And it really needs to be said that Aastrom needs to go the other direction. You need to get out there and promote this. You are telling us that everything is working and we believe you. But if you don't get the message out to the world, you are not going to attract new investors.*

<**A – George Dunbar (CEO)**>: *I think all I can say to put this in context is that we do not withhold any data that we have that we believe is relevant and meaningful. ... What really carries the day, and what is important, is meaningful clinical data with statistically relevant numbers behind it. Now that is something new for Aastrom. I have to tell you, Aastrom in the past, in my opinion, and others may disagree would me, were clearly trying to generate a lot of information to stay in front of things. It is not obvious to me that in doing so that that helped except possibly on a temporary basis. We are trying to build a strong foundation of meaningful value with very solid clinical data that will stand up not only to regulatory, but peer review for publications by journals and in academic publications. That is really the only way that this is going to gain traction. And in particular, the type of investors that we are interested in attracting... .*

Source: Aastrom Biosciences, 2008:Q4 Earnings Call, accessed from Thomson One.

As evidenced by this quote, it appears that scientific publications are an important communication device for R&D firms and could be powerful in attracting investor attention and financing. In the Online Appendix, Section A.1, we provide excerpts on the other potential benefits from scientific publishing (as discussed in Section 2.1). In addition, we provide excerpts on managerial preferences for scientific publishing. Managers frequently emphasize that they are cautious in revealing too much information through alternative disclosure channels prior to journal publication because that could compromise the peer-review process and preclude them from publishing. As one manager commented, "*We don't believe in science by press release; we believe in science by peer-reviewed journal publications. And so we really can't go too much into the data we're seeing right now*". This suggests that managers perceive scientific publications as a credible disclosure form. We now turn to this issue, using a statistical model.

4. Empirical setup and data

4.1. Identification strategy

The most straightforward way to examine how information asymmetry in financial markets affects firms' scientific publication behavior is to regress scientific publications on the number of analysts following. However, the estimates from such regressions are difficult to interpret because of endogeneity (omitted variables bias, reverse causality, etc.). For instance, if a positive relation between analyst following and scientific publications were uncovered, this could reflect the fact that analysts are attracted to firms that provide enhanced disclosure (as argued in Bushee and Miller, 2012). Indeed, as we show in the Online Appendix, Section A.3, the relation between analyst coverage and scientific publications (in a full panel) is positive and significant.

To overcome this problem, we consider a setting where there is an unexpected shock to the number of analysts. Specifically, we use brokerage house mergers and closures as a source of plausibly exogenous variation in firms' analyst coverage. The validity of the experiments relies on the assumption that broker mergers and closures lead to a reduction in analyst coverage for a given firm. Importantly, past research suggests that such coverage terminations are uncorrelated with unobservable firm characteristics (e.g., investment opportunities or managerial talent) and that the decisions to terminate coverage are not made by analysts themselves (e.g., Hong and Kacperczyk, 2010; Kelly and Ljungqvist, 2012).

We focus on broker disappearances between 2000 and 2010 and examine the effect of shocks to analyst coverage for the treated firms over the three fiscal years before $[-3,-1]$ and the three years after $[+1,+3]$ the brokerage merger/closure date.[10] We consider post-Regulation Fair Disclosure (Reg FD) broker disappearances to avoid complications related to the implementation of this important regulatory change for analysts (as in Balakrishnan et al., 2014). We stop with brokers that disappeared in 2010 to avoid truncation problems in certain outcome measures during the last years of our sample. Our treatment sample is a combination of firms affected by brokerage merger/closure events from Kelly and Ljungqvist (2012), who cover the period between 2000 and 2008, and firms affected by events from Fich et al. (2018), who provide data on brokerage closures and mergers that occurred between 2008 and 2014.

To identify firms whose coverage levels are affected by merger/closure events, we follow the approach put forth in He and Tian (2013) and Billett et al. (2017), among others. First, for each event, we define the event period as the six months around the month of broker disappearance; this accounts for the fact that mergers can span several days or even a couple of months. Next, we retrieve all firms covered by the brokers involved in the event in the 12 months before the event period $[-15, -3)$ and the analysts working for them. We assume that an analyst covers a firm if there is at least one earnings estimate in the Institutional Brokers' Estimate System (I/B/E/S) Detail History file for that firm in the pre-event period. Similarly, we assume that an analyst disappears if there is no earnings estimate from her in the I/B/E/S records in the 12 months after the event period $(+3, +15]$.

For brokerage closures, we retain firms for which the analyst disappears from I/B/E/S in the postevent period; using those analysts who issue no earnings estimates during this period ensures that analysts who transition to other brokerage houses do not continue to cover these firms. For brokerage mergers, we retain firms covered by both the acquirer and the target broker before the merger period and for which one of their analysts disappears; this ensures that the resulting loss in coverage is indeed due to the brokerage merger. Furthermore, we exclude from the sample those firms that are no longer covered by the acquirer in the period after the event; the reason for this restriction is that such terminations could be endogenous because the acquiring broker has chosen to stop covering the firm for reasons not observable to us.

Our treatment sample comprises 760 firms corresponding to 43 broker disappearances between 2000 and 2010. Figure 1 (Panel A) depicts the distribution of broker disappearances by calendar year. As in previous studies (e.g., Derrien and Kecskés, 2013), we observe some clustering of broker disappearances in 2000 and 2001. Panel B depicts the distribution of treated firms by calendar year. Treated firms are even more strongly clustered in time: 390 firms are treated between 2000 and 2002, and the other 370 are treated between 2003 and 2010. To address the possibility that time series effects explain our results, we implement our identification strategy by using a DiD methodology. This allows us to contrast the changes in the variables of interest for treated firms before and after the shock with the changes in the variables of interest for the control firms. In our setting, the control firms are all stocks that have analysts following in the pre-event period but that are not covered by the two merging brokers or the closing broker.[11]

One remaining concern with our identification strategy is that the treated and control firms differ in terms of observable dimensions, which may affect the estimate on the coverage loss. For instance, our estimate might be driven by larger firms being covered by more brokers (and therefore being more likely to be treated) while also having higher scientific publication rates. It is thus important to control for such systematic differences in our empirical specification to further isolate the effect of the coverage shock. Following Irani and Oesch (2013, 2016) and Guo et al. (2019), we address this potential concern using two different approaches. First, our basic approach is to incorporate firm fixed effects and control variables into the DiD regression framework. Second, we implement a DiD matching estimator.

Panel A. Distribution of broker disappearances *Panel A. Distribution of treated firms*

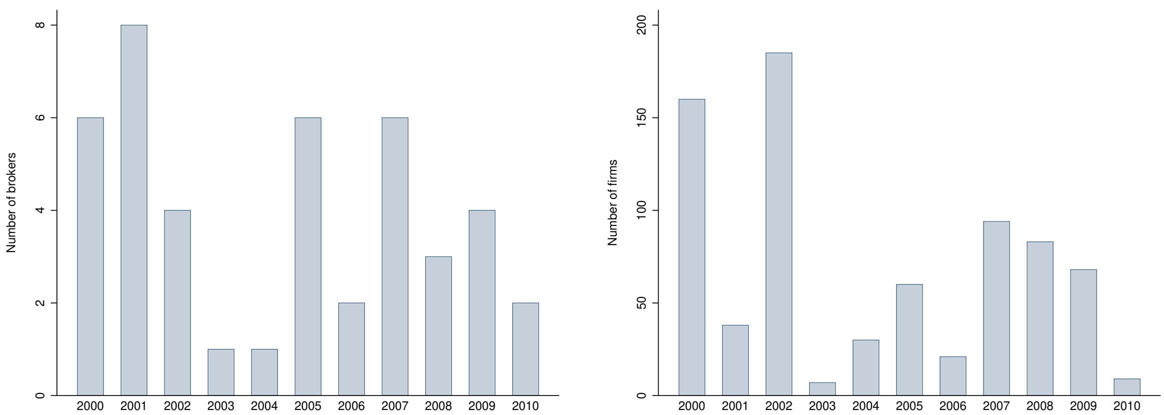

Fig. 1. Distribution of broker disappearances and treated firms by calendar year. This figure presents the distribution of brokerage house disappearances and treated firms by calendar year. *Panel A* depicts the distribution of brokerage house disappearances by calendar year. *Panel B* depicts the distribution of firms that lose an analyst due to brokerage house disappearances by calendar year. The sample comprises 760 treated firms that lost an analyst between the 2000 and 2010 mergers. These firms are U.S. public firms with at least one year of positive R&D expenditures and at least one patent over the period 1997-2014.

4.2. Sample construction

Our sample includes U.S. public firms for the period between 1997 and 2014.[12] We collect scientific publication information from Elsevier's Scopus database and analyst data from the I/B/E/S database. To calculate the control variables and the variables used in additional specifications, we add balance sheet data from Compustat and stock price data from the Center for Research in Security Prices (CRSP) database, institutional ownership data from Thomson's 13F database, text-based financial constraint measures from Gerard Hoberg's website, governance information from the Institutional Shareholder Services (ISS) database, and patent data from Noah Stoffman's website, among other data sources.

When constructing the sample of firm–year observations, we restrict the sample to observations with nonmissing (and positive values of) assets (item #6), sales (#12), and equity (#60) in the Compustat file (Billett et al., 2017). We eliminate financial and utility firms (firms with Standard Industrial Classification [SIC] codes 4900-4999 and 6000-6999) and firms not headquartered in the U.S. based on their current headquarters location (*LOC*). For reasons discussed in Section 2.2, we focus on firms that are active in R&D rather than sampling the entire Compustat universe. Since there is no common agreement on what defines R&D firms, we follow the recent literature on scientific publications (Arora et al., 2021), which requires firms to have at least one patent and at least one year of positive R&D spending during the sample period. For robustness, we replicate our analysis using several alternative definitions.

4.3. Scientific publications

We obtain scientific publication data from Scopus, which contains detailed information on approximately 71 million records from peer-reviewed journals, trade publications, book series, and conference proceedings from 1969 to 2020, and all citations made by these publications since 1970.[13] Each publication further includes information on the publication title, journal title, authors, and an affiliation field with the name and address of the publishing institute or the firm. We focus on "articles" and "conference proceedings" from the list of document types as the most relevant outlets for novel scientific results. To identify scientific publications of firms, we standardize the names in the affiliation field and match these names to all historical company names from the sample of CRSP/Compustat Merged Database firms. We present details on the extensive matching process in the Online Appendix, Section A.5.

The main corporate publication variable used in this paper is the firm's total number of scientific publications in peer-reviewed journals and conference proceedings in a given year (as in, e.g., Simeth and Cincera, 2016; Arora et al., 2021). Following the literature, we set the publication count to zero for firms without available publication information in the Scopus database and then use the natural logarithm of the publication count as the main publication measure in our analysis (LN_PUB). To avoid losing firm–year observations with zero publications, we add one to the actual values when we calculate the natural logarithm. Since we are interested in the decision to disclose scientific research and since the average delays from submission to publication in the natural sciences, engineering and biomedical research tend to be less than one year, our preferred specification relates the coverage shock in the current year to scientific publications over the same period.

4.4. *Control variables*

Our empirical setup enables us to add control variables to the specification. Incorporating these variables into our analysis mitigates concerns that observable differences between treated and control firms drive the estimates. When selecting control variables, we follow prior studies on the relation between analyst coverage and R&D outcomes that have developed a standard vector of firm and industry characteristics (e.g., He and Tian, 2013; Guo et al., 2019).

These variables are firm size, LN_TA, measured by the natural logarithm of total assets; investment in R&D, $RDTA$, measured by R&D expenditures scaled by total assets; firm age, LN_AGE, measured by the natural logarithm of number of years that the firm is listed on CRSP/Compustat; asset tangibility, $PPETA$, measured by net property, plant, and equipment scaled by total assets; investment in fixed assets, $CAPEXTA$, measured by capital expenditures scaled by total assets; profitability, ROA, measured by return on assets; leverage, LEV, measured by total debt to total assets; growth opportunities, Q, measured by Tobin's Q; financial constraints, $DELAYCON$, measured by the Hoberg and Maksimovic (2015) text-based index; institutional ownership, $INSTOWN$, measured by the fraction of institutional investors; product market competition, $HINDEX$, measured by the Herfindahl index based on sales; and $HINDEX^2$, the squared Herfindahl index.

Summary statistics for these variables for both the treatment and control samples are presented in Table 1. The Online Appendix, Section A.2, defines the variables used in this study and lists their sources.

4.5. *Model specification*

As discussed above, we implement our quasi-experiments using a DiD methodology. Specifically, we follow Irani and Oesch (2013, 2016) and Guo et al. (2019) in estimating the following model:

$$\begin{aligned} Y_{i,e,t} = \alpha + \beta_1 Post_{e,t} + \beta_2 Treated_{i,e} + \beta_3 Post_{e,t} \times Treated_{i,e} \\ + \gamma Z_{i,t} + \delta_t + \lambda_i + \theta_e + \varepsilon_{i,t} \end{aligned} \tag{1}$$

where i indexes the firm, e indexes the merger/closure event, t indexes the time, and $Y_{i,e,t}$ is the dependent variable, which is LN_PUB in our main specification. The variable $Post_{e,t}$ is a dummy equal to one if a firm–year observation is in the postevent period for event e, and

Table 1
Pre-event characteristics for treatment and control sample.

This table presents summary statistics for the treatment and control sample. *Panel A* reports summary statistics for the treatment sample. *Panel B* reports summary statistics for the control sample. The treatment sample is a combination of the brokerage merger and closure events from Kelly and Ljungqvist (2012) and Fich et al. (2018). Our sample includes 43 broker disappearances from 2000 to 2010. We follow the procedure put forth in He and Tian (2013) and Billett et al. (2017) to identify firms whose analyst coverage is reduced due to merger/closure events. The control sample is the remainder of firms in the CRSP/Compustat-merged universe with the required data and not covered by either the merging or closing brokers before the event. Variable definitions are provided in the Online Appendix, Section A.2.

	Panel A: Treatment sample						Panel B: Control sample					
Variable	Mean	25%	50%	75%	SD	N	Mean	25%	50%	75%	SD	N
LN_PUB	2.683	1.099	2.485	4.127	1.973	760	1.161	0.000	0.693	1.946	1.362	21,879
LN_AT	8.014	6.767	8.018	9.365	1.790	760	5.499	4.415	5.403	6.540	1.515	21,879
RDTA	0.074	0.018	0.057	0.107	0.074	760	0.093	0.017	0.053	0.122	0.118	21,879
LN_AGE	2.778	2.197	2.833	3.584	0.803	760	2.510	1.946	2.485	3.135	0.750	21,879
PPETA	0.205	0.085	0.160	0.270	0.162	760	0.187	0.076	0.150	0.262	0.144	21,879
CAPEXTA	0.048	0.022	0.037	0.063	0.035	760	0.043	0.018	0.032	0.054	0.039	21,879
ROA	0.125	0.085	0.135	0.195	0.128	760	0.052	0.008	0.104	0.163	0.202	21,879
LEV	0.165	0.013	0.143	0.269	0.153	760	0.141	0.000	0.089	0.244	0.155	21,879
Q	2.935	1.548	2.246	3.560	2.039	760	2.415	1.261	1.776	2.835	1.855	21,879
DELAYCON	-0.013	-0.064	0.000	0.009	0.073	760	-0.013	-0.080	-0.006	0.032	0.091	21,879
HINDEX	0.218	0.083	0.161	0.279	0.177	760	0.246	0.116	0.189	0.315	0.188	21,879
HINDEX^2	0.079	0.007	0.026	0.078	0.141	760	0.096	0.014	0.036	0.099	0.159	21,879
INSTOWN	0.685	0.566	0.708	0.823	0.178	760	0.544	0.342	0.569	0.755	0.256	21,879
COV	16.373	9.833	16.583	22.000	8.086	760	4.347	1.750	3.417	5.846	3.580	21,879
BASPRD	0.551	0.069	0.156	0.842	0.782	736	1.125	0.207	0.603	1.578	1.367	20,376

$Treated_{i,e}$ is a dummy equal to one if firm i is part of the treatment sample for that event. β_3 is the DiD coefficient, which captures the impact of coverage terminations on changes in scientific publications for the treated firms relative to the publications of control firms. The variables δ_t, λ_i, and θ_e correspond to year, firm, and merger/closure event fixed effects that account for time-invariant unobservable factors particular to a year, a firm, or a specific merger/closure event that may influence scientific publication behavior across units.[14] This specification enables us to incorporate control variables that account for other sources of differences across treated and control firms not picked up by the fixed effects. To this end, $Z_{i,t}$ is the set of control variables from Section 4.4. To control for serial correlation, we cluster standard errors at the firm level, as suggested by Irani and Oesch (2013, 2016).[15]

5. Empirical results

5.1. Average treatment effect

We start with the key idea of the experiment: on average, treated firms should lose one analyst relative to control firms. We test this by estimating Eq. (1) with the dependent variable replaced by analyst coverage (COV) and without including control variables. Given our experimental setup, we should observe a DiD coefficient close to minus one. Column 1 of Table 2 shows that this is indeed the case: the DiD coefficient is -1.029 and is significant at the 1% level. In terms of magnitude and significance, this result is consistent with related studies that utilize a similar research design (e.g., He and Tian, 2013; Irani and Oesch, 2016), despite the differences across samples.

Given this reduction in analyst coverage for the treatment sample, we also assess whether we observe an increase in information asymmetry among the investors in our data. The typical proxy used in the related literature is bid–ask spreads (e.g., Kelly and Ljungqvist, 2012; Balakrishnan et al., 2014). We report the results in column 2 of Table 2. The estimated DiD coefficient in the spreads equation ($BASPRD$) is 0.278 and is statistically significant at the 1% level. Hence, information asymmetry increases following coverage termination.

Next, we examine how this increase in information asymmetry translates to the scientific publication behavior of firms. Column 3 shows the results from estimating Eq. (1) where the dependent variable is the natural logarithm of one plus the number of scientific publications (LN_PUB). We obtain a DiD coefficient that is positive and significant at the 1% level. The point estimate in this column is 0.135, indicating that an exogenous increase in asymmetric information due to coverage shocks causes the firm to increase its scientific publications by approximately 14% relative to the number of publications of firms with no decrease in analyst coverage. Thus, the effect that we document is both statistically and economically meaningful. For robustness, column 4 presents the results based on the inverse hyperbolic sine transformation (IHS_PUB). If anything, the magnitude of the estimate becomes larger.

Our interpretation of this finding is that R&D firms react to an exogenous increase in information asymmetry by increasing their publication rates to communicate with investors. This rationale is consistent with economic models of financial signaling (e.g., Bhattacharya and Ritter, 1983). In such models, when managers face an increase in the cost of capital, they have a greater incentive to disclose their private information about firms' R&D outcomes to financial market participants, assuming such disclosure provides a credible signal about their firms' R&D prospects.

One concern with the above estimator is that the partial effect may capture systematic differences in characteristics between the treatment and control groups. To address these issues, we add the set of control variables employed by existing studies on the relation between analyst coverage and R&D. The results, shown in columns 5–8, indicate that our baseline estimate is robust to controlling for a large set of time-varying observable characteristics. Across all these specifications, the estimated partial effects remain significant at the 1% level and of a similar order of magnitude. The inclusion of additional control variables has a limited impact on the estimated treatment effect, suggesting that the coverage termination is plausibly exogenous and that the increase in scientific publications is not the result of omitted variable bias.

Table 2
Regressions of corporate publications on analyst coverage shocks.

This table presents the baseline difference-in-difference (DiD) results from the regression of corporate scientific publications on analyst coverage shocks. Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms assembled through the sample construction procedure described in Section 4.1. Variable definitions are provided in the Online Appendix, Section A.2. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Dep. Var.	COV (1)	BASPRD (2)	LN PUB (3)	IHS PUB (4)	COV (5)	BASPRD (6)	LN PUB (7)	IHS PUB (8)
TREATED × POST	-1.029***	0.278***	0.135***	0.146***	-0.993***	0.202***	0.116***	0.124***
	(0.229)	(0.038)	(0.032)	(0.036)	(0.211)	(0.035)	(0.031)	(0.035)
POST	-0.167***	0.053***	0.004	0.006	-0.077**	0.005	0.010	0.014
	(0.042)	(0.018)	(0.009)	(0.011)	(0.036)	(0.017)	(0.009)	(0.011)
TREATED	2.434***	-0.319***	-0.052	-0.052	2.059***	-0.152***	-0.060	-0.063
	(0.303)	(0.059)	(0.042)	(0.048)	(0.254)	(0.054)	(0.039)	(0.045)
LN_AT					2.302***	-0.553***	0.209***	0.251***
					(0.113)	(0.041)	(0.024)	(0.029)
RDTA					3.787***	-0.399	0.618***	0.770***
					(0.539)	(0.316)	(0.125)	(0.152)
LN_AGE					-0.294	0.024	-0.080	-0.090
					(0.220)	(0.086)	(0.053)	(0.064)
PPETA					1.365**	0.839***	0.131	0.165
					(0.662)	(0.205)	(0.144)	(0.177)
CAPEXTA					3.560***	-2.449***	0.190	0.223
					(1.093)	(0.438)	(0.251)	(0.308)
ROA					0.038	-0.844***	-0.038	-0.043
					(0.261)	(0.150)	(0.065)	(0.080)
LEV					-0.804**	0.841***	0.042	0.037
					(0.396)	(0.139)	(0.088)	(0.108)
Q					0.121***	-0.126***	-0.009	-0.011
					(0.024)	(0.011)	(0.006)	(0.007)
DELAYCON					0.991**	0.143	0.073	0.088
					(0.462)	(0.186)	(0.118)	(0.145)
HINDEX					0.018	-0.687	-0.585*	-0.683*
					(1.404)	(0.499)	(0.322)	(0.396)
HINDEX^2					-0.413	0.053	0.693**	0.821**
					(1.214)	(0.507)	(0.299)	(0.365)
INSTOWN					2.807***	-0.728***	-0.052	-0.043
					(0.307)	(0.113)	(0.073)	(0.088)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	122,429	114,536	122,429	122,429	122,429	114,536	122,429	122,429
Adjusted R^2	0.794	0.638	0.900	0.889	0.841	0.708	0.903	0.892

5.2. *Additional validation checks*

5.2.1. *Matching estimator*

Another potential concern is that if the treatment and control groups differ on observable characteristics, then they likely also differ on unobservable characteristics. If this is the case, including control variables in a linear regression framework might not adequately control for unobservable differences, especially if there are nonlinearities in the data (Irani and Oesch, 2013, 2016).

To address this concern, we implement a DiD matching estimator. We construct a control group of firms matched to the treated firms on a set of relevant ex ante characteristics measured in the year prior to the coverage shock. Our matching is similar in spirit to the approach of Irani and Oesch (2013, 2016), who match by firm size and performance. Since our purpose is to ensure that the treatment and control groups are similar in terms of the determinants of the scientific publications that we study, we match by total assets, R&D, and Q. We also match by bid–ask spreads because we hypothesize that information asymmetry affects publication rates. Last, we match on publication growth (i.e., the growth in the number of publications from years −3 to −1), which eliminates potential remaining differential trends in pre-event publication pattern (He and Tian, 2013).

We use a nearest-neighbor propensity score matching procedure, as these are often used in related studies (e.g., He and Tian, 2013; Irani and Oesch, 2016). In the first step, we run a logit regression of a dummy variable equal to one if a specific firm–year is classified as treated on our matching variables.[16] The sample used to estimate this regression consists of 570 treatment and 10,267 candidate control pre-event observations, which is the sample containing valid matching variables. The estimated coefficients from the logit regression are used to estimate the probabilities of treatment for each firm–year. These probabilities (or propensity scores) are then used to perform a nearest-neighbor match. We match with replacement, using a standard tolerance (0.005 caliper) and allowing for up to five unique matches per treated firm.

Table 3 presents the results. Columns 1–4 show summary statistics for the treatment and control groups after the matching procedure. The statistics indicate that for the variables that we match on, the differences between the treatment and control groups are small in terms of economic magnitudes and are statistically insignificant. Columns 5 and 6 show that the coverage shock continues to have a meaningful impact on the coverage and information asymmetry of treated firms relative to those of the matched control sample. The remaining columns indicate that the DiD matching estimator produces similar estimates of the average treatment effect, both in terms of economic magnitudes and statistical significance.[17]

Overall, these findings indicate that our baseline results from Table 2 are not driven by ex ante differences between the treatment and control groups. Specifically, we confirm that ex ante differences in firm size, R&D, growth opportunities, information asymmetry, or publication growth between the treatment and control groups do not drive our results. This provides further evidence in support of our empirical design.

Table 3
Regressions of corporate publications on analyst coverage shocks: Matching estimator.
This table presents the baseline difference-in-difference (DiD) regression results of corporate scientific publications on analyst coverage shocks by balancing the sample on pre-treatment covariates. The sample includes the treated and control firms assembled through the sample construction procedure described in Section 4.1 with valid matching variables in year -1. Treated firms are matched using a nearest-neighbor logit propensity score match with a 0.005 caliper and with matching of up to five control firms. Robust standard errors are clustered by firm (in parentheses). Variable definitions are provided in the Online Appendix, Section A.2. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Mean Treated (1)	Mean Control (2)	Difference Means (3)	Difference p-value (4)	COV (5)	BASPRD (6)	LN PUB (7)	IHS PUB (8)
TREATED × POST					-1.097**	0.183***	0.170***	0.209***
					(0.450)	(0.058)	(0.058)	(0.069)
POST					-0.403*	0.060	-0.082**	0.020
					(0.221)	(0.041)	(0.035)	(0.056)
TREATED					1.789***	-0.125	-0.107	-0.166
					(0.458)	(0.081)	(0.097)	(0.101)
Matching variables:								
LN_AT	7.013	7.012	0.001	0.993				
RDTA	0.087	0.095	-0.008	0.479				
Q	2.223	2.145	0.078	0.696				
BASPRD	0.604	0.720	-0.116	0.131				
PUB_GROWTH	0.287	0.232	0.055	0.633				
Firm fixed effects					Yes	Yes	Yes	Yes
Year fixed effects					Yes	Yes	Yes	Yes
Event fixed effects					Yes	Yes	Yes	Yes
Number of obs.					5,227	5,227	5,227	5,227
Adjusted R^2					0.819	0.689	0.848	0.838

5.2.2. Dynamic effects and parallel trends assumption

We examine whether the shock has a permanent effect on analyst coverage and the publication behavior of treated firms. In addition, we conduct falsification tests to determine whether the parallel trends assumption is violated.

Figure 2 (Panel A) depicts the dynamic effects of the shock on analyst coverage. On the yaxis, the graph shows the number of analysts following a firm; the x-axis shows the time relative to the shock (ranging from three years prior to the shock until three years after). The vertical lines in the figure correspond to the 90% confidence intervals of the coefficient estimates.[18] The results indicate that there are no pre-event trends in the data and that coverage decreases by roughly one analyst between year −1 and year +1. Moreover, we see no postevent trends in the data and can observe that the effect of the shock on analyst coverage is permanent. This is consistent with Hong and Kacperczyk (2010), Derrien and Kecskés (2013), and others, who also find a permanent reduction in coverage.

Panel B depicts the dynamic effects of the shock on scientific publications. We see that the pre-event trend is similar for the treatment and control firms. Specifically, there is no statistically significant difference in publication rates between the treatment and control firms over the years from −3 to −1. Therefore, we fail to invalidate the parallel trend assumption, suggesting that it is not violated in our setup. We also see a permanent increase in scientific publications in the period after the shock. Especially strong is the increase between year −1 and year +1. This is intuitive under a disclosure explanation because the impact of a change in R&D disclosure behavior (in the form of scientific publications) can quickly become visible.

Panel A. Difference (treated - control) in coverage *Panel B. Difference (treated - control) in publications*

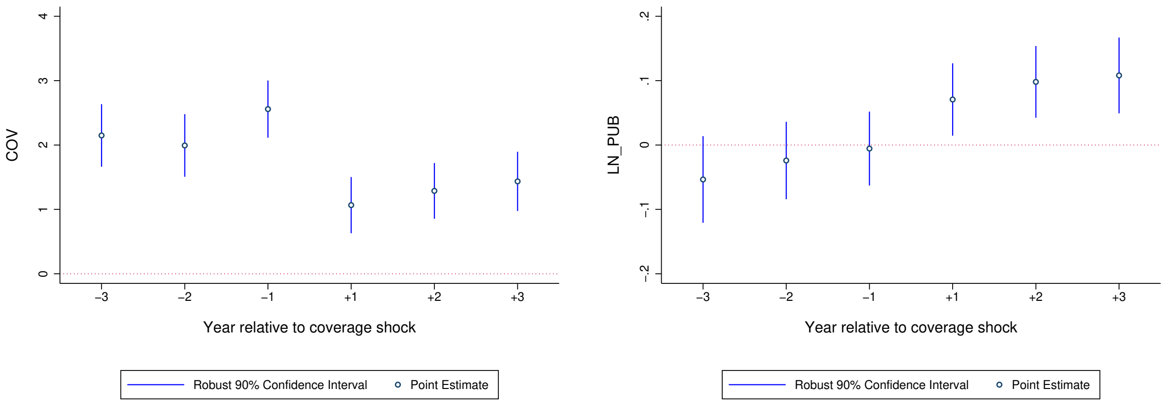

Fig. 2. Dynamic effects of the analyst shock on coverage and scientific publications. This figure shows a visual difference-in-differences examination of the effect of the analyst shock on coverage and scientific publications for the treatment sample relative to the control sample, from three years before the shock to three years after (for similar graphs, see, e.g., Billett et al. 2017). *Panel A* depicts the dynamic effects of the shock on coverage. *Panel B* depicts the dynamic effects of the shock on publications. The vertical lines in the figure correspond to the 90% confidence intervals of the coefficient estimates.

5.2.3. *Merger/closure characteristics*

Next, we confirm that our baseline results are not driven by broker disappearances in specific years or by the small number of mergers/closures that cause a large number of firms to be treated. Moreover, we show that our results are not driven by either broker mergers or broker closures alone and that they remain robust to the exclusion of broker disappearances not included in the list provided by Kelly and Ljungqvist (2012). To this end, we perform four different analyses.

We tabulate the results in the Online Appendix, Table AT3. First, we estimate Eq. (1) only for the group of brokers that disappeared in 2000, 2001, 2002, 2008, and 2009, contrasting these results with those for the group of brokers that disappeared in the other years. The DiD coefficient remains positive and significant in all subperiods. Second, we repeat our analysis separately for the group of the top 10 brokers ranked by the number of firms that lose an analyst; this group collectively accounts for 73% of the treated firms. The DiD estimate is positive and significant in both subsamples, suggesting that our results are not driven by a small number of brokers accounting for a large number of coverage terminations. Third, we repeat our analysis for mergers and closures separately and find that our results are similar across both groups. In the fourth test, we exclude broker disappearances not included in the list in Kelly and Ljungqvist (2012). The results are stronger when we restrict our sample to this list.

5.3. *Additional robustness checks and extension*

We conduct a rich set of additional tests to check the robustness of our baseline results in terms of magnitudes and statistical significance. We also consider additional publication measures to strengthen the financial signaling interpretation. The results are tabulated in the Online Appendix, Tables AT4–AT8.

First, we rerun our DiD specification on a balanced panel to check for attrition bias. This attempt delivers similar results. Our results are also robust to including industry-by-year fixed effects, which allow us to control for time-varying differences in scientific publishing across different industries. Throughout the paper, we use publication information from Scopus. While Scopus provides higher coverage in natural science, engineering and biomedical research than the Web of Science (WoS), it may suffer from other limitations. Therefore, we replicate our baseline results with data from the WoS and show that our results remain similar.

Second, we test whether our results suffer from serial correlation issues. To this end, we experiment with clustering standard errors at either the event level or the event–firm level. Using different clustering schemes makes little difference. If anything, by clustering standard errors at the firm level throughout the paper, we overstate the standard deviation of the DiD estimator. We also repeat our DiD analysis by collapsing the firm–year observations by broker into pre- and post-periods. This experiment also yields similar results.

Third, our baseline uses a three-year window before and after the coverage shock with a 12-month disappearance period. To explore the sensitivity of our results to this choice, we first move the pretreatment interval further backward from the event year by either one or two years while keeping the posttreatment window constant. We also move the posttreatment window further outward by either one and two years while keeping the pretreatment window constant. The DiD estimate remains robust in all these cases. Furthermore, we explore the sensitivity of the publication results to the selection of the three-year measurement window. We obtain similar results if we use a two-year, four-year or five-year window.

Fourth, we check whether our results are robust to using alternative definitions of R&D firms. As discussed in Section 2.2, we limit our attention to R&D firms because the empirical evidence consistently indicates that analyst coverage is beneficial and informative for such firms. We define such firms as those with at least one patent and at least one year of positive R&D during our sample period, following Arora et al. (2021). However, there might be concerns that R&D reporting in Compustat is incomplete. To this end, we relax this requirement and obtain robust results. We also show that our results remain unchanged if we require sample firms to have at least one scientific publication.

Finally, we use a series of additional publication metrics to explore the idea that increased information asymmetry due to the coverage shock not only increases the level of scientific publications but also causes a change in their composition. First, we examine two measures of publication quality based on the number of citations received in subsequent years and the journal impact factor. Second, we look at links with university-based scientists. Audretsch and Stephan (1996) argue that such links lend more prestige to firms. The results suggest that the treatment effects are stronger when we consider quality and/or affiliations with universities.

5.4. Addressing the reuse of natural experiments

Heath et al. (2021) show that repeated use of a natural experiment may increase the likelihood of false discoveries because of multiple hypothesis testing. This is relevant to us because both brokerage house mergers and closures have been used in several other studies focusing on many different dependent variables. As a first stage of the research design, the authors recommend validating a significant relation between the experiment and the main explanatory variable. As we show above, we do find a significant first-stage relationship—, i.e., an increase in information asymmetry, for the treatment group, even when we include control variables in our linear regression framework or use a matching estimator.

The authors also recommend assessing the validity of the exclusion restriction against existing findings in the literature. To address this concern, we note that prior studies on corporate policies using this experiment do *not* focus on R&D-active firms, which we study here. Indeed, virtually all studies are based on the entire Compustat universe, which implies, by construction, a focus on non-R&D-active firms. For instance, Chen et al. (2015) and Irani and Oesch (2016) focus on corporate governance and financial reporting variables, respectively. However, both dimensions, like many others, exhibit substantial structural differences between R&D and non-R&D firms (e.g., Aboody and Lev, 2001; Coles et al., 2008). Perhaps more relevant for our study, Derrien and Kecskés (2013) document that capital and acquisition expenditures and R&D *decrease* in response to the shocks, as do financing and cash holdings. Such adverse consequences for financing and investment make, if anything, communication with investors more important. Therefore, our finding of an increase in publication rates (which also unfolds rapidly after brokerage house events) is consistent with this channel.

In contrast, He and Tian (2013) show that brokerage merger/closure shocks *increase* investments in R&D, as measured by patent and patent citation counts. Under the channel that they identify, treated firms face less pressure to meet short-term earnings targets imposed by analysts, which in turn encourages managers to make long-term investments. While this channel is plausibly operative for firms with little or no R&D, Clarke et al. (2015) and Guo et al. (2019) show that it is not relevant for R&D firms, which, again, are our focus. Prior evidence still supports the view that such shocks, irrespective of the sample, lead to an increase in asymmetric information and the cost of capital (Kelly and Ljungqvist, 2012; Balakrishnan et al., 2014; Ellul and Panayides, 2018). Consequently, we interpret an increase in publication rates as reflecting managerial attempts to combat the adverse consequences of increased information asymmetry. We provide evidence on this alternative channel in Section 6.

5.5. *Heterogeneity in the treatment effect*

We investigate several cross-sectional implications of our main results. Specifically, we test whether our treatment effect is more pronounced when asymmetric information concerns are high, when financial constraints are present, when managers have shorter horizons, and when proprietary costs are elevated. We present triple differences, wherein we compare the DiD results for two subsamples created by sampling on each of the above four general constructs.

5.5.1. *Conditioning on changes in information asymmetry*

We hypothesize that the analyst shock impacts publication policies by causing an increase in information asymmetry. If this is the case, then increases in publication rates should be largest for firms for which information asymmetry increases the most as a result of the shock.

To test this conjecture, we condition on changes in the following variables: analyst coverage, bid–ask spreads, the Amihud (2002) illiquidity measure, and the ratio of zero and missing returns days to the total number of days. We compute these variables following Kelly and Ljungqvist (2012). We classify firms in the top tercile of the change in the information asymmetry proxy as having a large change and firms in the bottom tercile as having a small change. Table 2 shows that coverage decreases and the bid–ask spread increases due to the shock. We confirm that there is also an economically and statistically significant increase in information asymmetry for treated firms in the other two proxies.

Table 4, Panel A, presents the results. The effect of the shock on scientific publications is larger for firms with a larger increase in information asymmetry. Using analyst coverage as an example, the treatment effect is large in magnitude (0.200) and significant at the 1% level for firms with a large decrease in analyst coverage. In contrast, the estimated treatment effect is smaller (0.037) and insignificant for firms with a small decrease in coverage. Overall, the results suggest that the effect of the shock is indeed larger for firms with a larger change in information asymmetry.

5.5.2. *Conditioning on financial constraints*

Derrien and Kecskés (2013) show that an increase in information asymmetry due to an analyst shock leads to a reduction in external financing and investment, especially for firms that are financially constrained. For financially unconstrained firms, the decrease in analyst coverage is largely irrelevant, as they have sufficient internal capital. Therefore, we expect more pronounced treatment effects for firms that are financially constrained.

To test this, we create subsamples of firms based on whether they are financially constrained. The extant literature offers numerous proxies for financial constraints. We begin with the delaycon measure from Hoberg and Maksimovic (2015). We consider unconstrained (constrained) firms to represent those with scores in the bottom (top) tercile. Billett et al. (2017) use two indirect proxies (having a credit rating or paying dividends) to classify firms as unconstrained; we use both. Another proxy that we consider is firm age. Younger firms tend to be characterized by a higher degree of information asymmetry and high growth opportunities. Older and more established firms tend to have a lower degree of informational asymmetry with outside investors and fewer growth opportunities. We classify firms in the bottom (top) tercile of the age distribution as constrained (unconstrained).

Table 4, Panel B, presents the results. The effect of the analyst shock on scientific publications is larger for firms that are financially constrained. When we use the measure from Hoberg and Maksimovic (2015), for example, the estimated marginal effect of the shock for financially constrained firms is positive and significant. The DiD coefficient is larger in magnitude (0.209) than the corresponding average treatment effect for the full sample (see Table 2, column 7) and significant at the 1% level. In contrast, the estimated treatment effect is smaller (0.087) and insignificant for financially unconstrained firms. Overall, the results suggest that the effect of the shock is larger for financially constrained firms.

5.5.3. Conditioning on manager horizon

We examine whether changes in scientific disclosure behavior are motivated by managerial incentives. Glaeser et al. (2020) argue and find that when investors believe that managers' horizons are short, i.e., when managers' incentives are more closely aligned with maximizing current share prices, they can be expected to be more likely to disclose R&D outcomes. The authors also show that investors discount the value of R&D whose outcome is undisclosed to a greater extent when the manager's horizon is short. The intuition is that when investors believe that the manager seeks to maximize short-term stock prices, they are more likely to interpret nondisclosure as withholding of bad news. This should increase the pressure on managers to disclose.

To test this idea, we condition upon proxies for managerial entrenchment and institutional ownership. Prior literature finds that less entrenched managers and those in firms with significant ownership by short-term investors reduce investments in R&D, consistent with such managers having short horizons (e.g., Bushee, 1998; Keum, 2021). We use two proxies for managerial entrenchment: the governance index and the entrenchment index. Lower values correspond to a stronger market for corporate control (i.e., facilitating takeovers), thus leading to less entrenched managers. We consider short-horizon (long-horizon) managers to be those with scores below (above) the median. We further consider whether the CEO is also the chair of the board. In firms with such CEO duality, the CEO is more entrenched (Irani and Oesch, 2013). The level of ownership by short-term-oriented institutional investors offers additional variation to conduct another test. We measure this indicator using the fraction of a firm's outstanding shares held by transient investors (Bushee, 1998). We classify managers in the top (bottom) tercile as having a short (long) horizon.

Table 4, Panel C, presents the results. The effect of an analyst shock on scientific publications is larger when a manager's horizon is short. Using the governance index as an example, the estimated treatment effect is larger in magnitude (0.204) in the presence of strong shareholder rights, i.e., when the manager is less entrenched, and significant at the 1% level. Among the set of firms with weak shareholder rights and more entrenched managers, the estimated treatment effect is smaller (0.056) and insignificant. Similarly, we see that the treatment effect is stronger when ownership by short-term investors is high (0.175) than when ownership by short-term investors is low (0.040). Overall, the results suggest that the effect of the shock is larger when managers have short horizons and thus are under more pressure to disclose.

5.5.4. Conditioning on proprietary costs

We also relate firms' publication responses to proprietary costs of disclosure. Guo et al. (2004) characterize such costs in terms of the propensity of firms to patent. The authors show that firms are less reluctant to disclose proprietary information when patent protection is available. This is consistent with evidence (e.g., Galasso and Schankerman, 2014) that patent rights often block downstream innovations and thus prevent imitation by rivals.[19] Following Hall and Ziedonis (2001), we proxy for the propensity of firms to patent using the ratio of patents to R&D expenditures. Firms in the top tercile of the propensity to patent are classified as having low disclosure costs, and firms in the bottom tercile are classified as having high disclosure costs.

The literature suggests that trade secrets and noncompete agreements increase the proprietary costs of disclosure because these agreements reduce information leakage through employee movements. For instance, Aobdia (2018) finds a negative relation between a state's propensity to enforce noncompete agreements and disclosure activities (e.g., management forecasts) of firms headquartered in this state. Li et al. (2018) examine the adoption of the inevitable disclosure doctrine (IDD) by state courts and find that firms respond to IDD adoption by reducing the level of disclosure regarding their customers' identities. Glaeser (2018) examines the passage of the Uniform Trade Secrets Act (UTSA) by states and finds, based on 10-Ks, a reduction in the disclosure of proprietary information related to trade secrets.

Following these studies, we use three proxies for firms' incentives to use trade secrecy and noncompete agreements. The first proxy is a dummy variable for whether a state has rejected (adopted) the IDD. We use information on states' IDD rejections and adoptions from Li et al. (2018). A headquarters location in a state that has rejected (adopted) the IDD suggests low (high) disclosure costs. The second proxy is the effective UTSA index developed by Png (2017), which measures the strength of the legal protection of trade secrets under common law in each state over time up to the enactment of a statute and the increase in legal protection due to the enactment of the statute. The third proxy is the noncompete enforcement variable developed by Garmaise (2011). Firms in the bottom (top) half of these indexes are classified as having low (high) disclosure costs.

Table 4, Panel D, presents the results. The effect of an analyst shock on scientific publications is larger when firms have low disclosure costs. If we consider the propensity of firms to patent, for example, the estimated treatment effect is larger in magnitude (0.249) when firms rely more heavily on patents to secure their returns to R&D and is significant at the 1% level. Among the set of firms that do not rely heavily on patents, the estimated treatment effect is smaller (0.032) and insignificant. Similarly, we observe stronger treatment effects for firms headquartered in states not enforcing noncompete agreements (0.173) than for firms headquartered in states enforcing such agreements (0.052). Overall, the results suggest that the effect of the shock is larger when firms have low disclosure costs. This is consistent with the idea that firms trade off disclosure costs against the capital market benefits when making decisions about whether to publish scientific research outcomes or keep them secret.

Table 4
Heterogeneity tests.

Panel A presents DiD regression results conditioning on changes in information asymmetry between the year after an analyst shock (year +1) and the year before (year −1). Firms in the top (bottom) tercile of the change in information asymmetries are classified as having a large (small) change. To measure information asymmetry, we use analyst coverage (column 1), bid–ask spread (column 2), Amihud's illiquidity measure (column 3), and return ratio (column 4). *Panel B* presents the results conditioning upon firms' pre-event (year −1) levels of financial constraints. For the continuous measures, constrained and unconstrained firms are divided based on the median values. In column (1), we use the delaycon measure (Hoberg and Maksimovic, 2015); in column 2, we consider whether firms have a credit rating; in column 3, we consider whether firms pay dividends; and in column 4, we differentiate by firm age, with young firms classified as constrained. *Panel C* reports the results depending on pre-event (year −1) proxies for managerial horizon. The distinction between short and long horizons is computed based on the top and bottom terciles of the continuous measures. We use the G-index (column 1), E-index (column 2), CEO duality (column 3), and the share of transient institutional ownership (column 4). *Panel D* shows the results depending on pre-event (year −1) disclosure costs. We consider firms' patenting intensity (column 1) and three proxies for firms' opportunities to rely on trade secrecy as an appropriability instrument (columns 2-4). The sample includes the treated and control firms assembled through the sample construction procedure described in Section 4.1. See Table 2 for control variables. Variable definitions are provided in the Online Appendix, Section A.2. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Clustered standard errors in parentheses.

Panel A: Conditioning on an increase in information asymmetry

Measure Dep. var: LN_PUB	Relative Decrease in COV (1)	Relative Increase in BASPRD (1)	Relative Increase in AMIHUD (3)	Relative Increase in RETR (4)
TREATED × POST × 1(Large change)	0.200***	0.190***	0.223***	0.162***
	(0.057)	(0.047)	(0.050)	(0.046)
TREATED × POST × 1(Small change)	0.037	0.050	0.042	0.050
	(0.061)	(0.057)	(0.047)	(0.059)
Difference	F=4.02	F=4.21	F=7.22	F=2.55
	(p=0.05)	(p=0.04)	(p=0.01)	(p=0.11)
Controls included	Yes	Yes	Yes	Yes
Firm, year, and event fixed effects	Yes	Yes	Yes	Yes
Number of obs.	76,506	71,964	71,765	77,949
Adjusted R^2	0.895	0.907	0.905	0.899

Panel B: Conditioning on pre-event financial constraints

Measure Dep. var: LN_PUB	DELAYCON (1)	CREDIT RATING (1)	PAYOUT (3)	AGE (4)
TREATED × POST × 1(Financially constrained)	0.209***	0.195***	0.189***	0.202***
	(0.048)	(0.042)	(0.047)	(0.068)
TREATED × POST × 1(Financially unconstrained)	0.087	0.036	0.037	0.009
	(0.056)	(0.044)	(0.037)	(0.038)
Difference	F=2.77	F=7.53	F=6.57	F=6.13
	(p=0.10)	(p=0.01)	(p=0.01)	(p=0.01)
Controls included	Yes	Yes	Yes	Yes
Firm, year, and event fixed effects	Yes	Yes	Yes	Yes
Number of obs.	80,247	121,515	121,515	85,494
Adjusted R^2	0.906	0.903	0.903	0.910

(Continued)

(Continued)

Panel C: Conditioning on pre-event manager horizon

Measure Dep. var: LN_PUB	GINDEX (1)	EINDEX (1)	DUAL CEO (3)	INSTOWN TRA (4)
TREATED × POST × 1(Short-horizon managers)	0.204***	0.189***	0.218***	0.175***
	(0.047)	(0.048)	(0.049)	(0.050)
TREATED × POST × 1(Long-horizon managers)	0.056	0.057	0.074*	0.040
	(0.051)	(0.046)	(0.041)	(0.061)
Difference	F=4.85	F=4.32	F=5.58	F=2.95
	(p=0.03)	(p=0.04)	(p=0.02)	(p=0.09)
Controls included	Yes	Yes	Yes	Yes
Firm, year, and event fixed effects	Yes	Yes	Yes	Yes
Number of obs.	56,022	51,492	61,704	79,829
Adjusted R^2	0.913	0.915	0.909	0.897

Panel D: Conditioning on pre-event disclosure costs

Measure Dep. var: LN_PUB	PAT INT (1)	IDD (2)	UTSA (3)	NON COMP (4)
TREATED × POST × 1(Low disclosure costs)	0.249***	0.159***	0.146***	0.173***
	(0.058)	(0.051)	(0.035)	(0.046)
TREATED × POST × 1(High disclosure costs)	0.032	-0.036	-0.035	0.052
	(0.053)	(0.103)	(0.085)	(0.043)
Difference	F=7.98	F=2.88	F=3.93	F=3.68
	(p=0.00)	(p=0.09)	(p=0.05)	(p=0.05)
Controls included	Yes	Yes	Yes	Yes
Firm, year, and event fixed effects	Yes	Yes	Yes	Yes
Number of obs.	51,923	67,658	102,581	102,527
Adjusted R^2	0.902	0.903	0.904	0.904

6. Alternative perspectives

This section considers alternative ways to corroborate our results and interpretations. In Section 6.1, we explore whether increased information asymmetries due to a coverage shock are mitigated through other sources of information. In particular, we ask whether R&D firms also offer management earnings guidance more often. In Section 6.2, we consider an alternative interpretation of our main results. Instead of the relation between asymmetric information and R&D disclosure driving the results, perhaps the analyst loss event reduces short-term pressure on managers to meet earnings targets.

6.1. *Other sources of information*

Managers can potentially reduce information asymmetry by communicating information about their R&D activities through disclosure forms other than scientific publications. Balakrishnan et al. (2014) find that shocks to information asymmetry encourage managers to provide earnings guidance more often. Under the assumption that investors can extract useful R&D-related information from earnings guidance, we should also observe greater provision of earnings guidance driven by the shock. However, Palmon and Yezegel (2012) and Glaeser et al. (2020), among others, argue that such disclosure forms are subject to severe weaknesses that are most prominent for firms involved in R&D.

We procure management earnings guidance from the I/B/E/S Guidance file, which includes information previously available in the Company Issued Guidance file and information from the defunct First Call database. We include forecasts of both annual and quarterly EPS and drop observations with missing earnings announcement dates or with guidance dates occurring on or after the actual earnings announcement date. To ensure the highest degree of precision, we restrict our analysis to guides, i.e., firms included in the I/B/E/S Guidance database that provide some earnings guidance (as in the approach in Balakrishnan et al., 2014).

The results are presented in Table 5, column 1. We estimate Eq. (1) with the dependent variable replaced by the natural logarithm of the amount of management earnings guidance made during the fiscal year. The estimated treatment effect indicates that our sample firms do not provide earnings guidance more often. The point estimate is small and statistically insignificant. This is consistent with the view that for R&D firms, provision of management earnings guidance is unlikely to mitigate the consequences of an increase in information asymmetry. We conclude that such disclosure forms are not an effective substitute for scientific publications.

6.2. *The dark side argument*

Up to this point in the paper, we have interpreted our results as consistent with a positive relation between information asymmetry and R&D disclosure. An alternative interpretation may be that shocks to analyst coverage reduce pressure on managers to meet short-term goals, which encourages them to increase long-term investments, as suggested by He and Tian (2013). Given that both interpretations are possible under the main evidence, we now examine an independent implication.

In particular, if our results thus far are due to less pressure from analysts that in turn encourages investments in R&D, then we should also observe an increase in R&D inputs and/or other R&D-related outputs caused by the shock. We report the results from this exercise in the remaining columns of Table 5. Column 2 reports the regression results from estimating Eq. (1) with the dependent variable replaced by R&D expenditure and shows an insignificant DiD estimator. In column 3, the dependent variable is replaced by citation-weighted patent counts, the most common measure of R&D investments. The treatment coefficient remains small and insignificant. These findings do not support the investment explanation for our main results or, more broadly, the claim that analyst coverage has a dark side.

However, a specific concern for us is whether scientific research–related investments increase due to coverage shocks. While the above evidence suggests that R&D investments do not change, we recognize that our measures do not account for the composition of R&D activities. To address this issue, we consider the use of scientific research in patents. Following Marx and Fuegi (2020), we consider patents to be science-based if they contain at least one citation of scientific research on their front page. We calculate two variables: the fraction of science-based patents (column 4) and the fraction of science-based patents with external references (column 5). Note that the construction of these variables constrains the relevant sample to firm–year observations with at least one patent. In both cases, we observe a pattern of the treatment effect that is very similar to that of the previous results: the estimated DiD coefficient is small and statistically insignificant.

We also check whether the coverage stock causes the hiring of new scientists, which would be another indication of increased investments in scientific research. It does not. To show this, we leverage Scopus's unique author identifier. The unique identifiers are based on sophisticated disambiguation routines that separate similar scientists based on various characteristics. We define a new hire as a scientist with a single publication at a sample firm in a given year and at least one publication reflecting a different affiliation (e.g., with another firm, university or institute) before that year. We then aggregate the sum of all new hires at a sample firm in a given year and use the natural logarithm of (one plus) this raw measure of hiring in our analysis. The last two columns of Table 5 show the results. Column 6 reports estimates based on the full sample, while the estimates in column 7 use firms that experience an increase in scientific publications following the shock. In both specifications, we obtain weak and insignificant treatment effects.

Table 5
Test of alternative perspectives.

This table presents the difference-in-difference (DiD) results of the regression of management earnings guidance, R&D expenditures, patenting, and the hiring of new scientists on analyst coverage shocks. The sample includes the treated and control firms assembled through the sample construction procedure described in Section 4.1. Robust standard errors are clustered by firm (in parentheses). Variable definitions are provided in the Online Appendix, Section A.2. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Dep. Var.	LN GUID	LN RD	LN PAT	PAT SCI	PAT SCI EXT	LN NEW HIRES	LN NEW HIRES
Sample	Guiding Firms	Full	Full	Patenting Firms_t	Patenting Firms_t	Full	$\triangle\ \text{PUBS}_{pre \rightarrow post} > 0$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
TREATED × POST	0.003	0.027	-0.007	-0.006	-0.003	0.006	0.042
	(0.033)	(0.032)	(0.066)	(0.009)	(0.009)	(0.029)	(0.043)
POST	0.011	-0.002	-0.130***	-0.007	0.002	0.004	0.123***
	(0.011)	(0.007)	(0.024)	(0.004)	(0.004)	(0.008)	(0.023)
TREATED	0.019	0.044	-0.086	0.008	0.004	0.011	0.025
	(0.036)	(0.051)	(0.089)	(0.011)	(0.012)	(0.041)	(0.078)
Controls included	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	53,127	114,020	122,429	82,335	82,335	122,429	40,885
Adjusted R^2	0.605	0.934	0.751	0.737	0.636	0.771	0.775

7. A plausible mechanism: Investor attention

Our results provide robust evidence that shocks to information asymmetries change scientific disclosure practices. The qualitative evidence from Section 3 suggests that one possible explanation for this result is that scientific publications are effective in increasing investors' attention to a firm. Merton (1987) develops a model that incorporates limited investor attention and analyzes the capital market equilibrium in this setting, including the implications for asset prices. He shows that firms that are relatively more familiar to investors exhibit a lower cost of capital and higher share prices (e.g., see also Grullon et al., 2004; Madsen and Niessner, 2019). Therefore, increased investor attention is one plausible channel through which scientific publications could mitigate the consequences of the shock.

To provide systematic evidence on this channel, we link several proxies for investor attention to scientific publications. Following Da et al. (2011) and Madsen and Niessner (2019), among others, we measure inventor attention using the monthly Google Search Volume Index (GSVI). LN_GSVI is calculated as the natural logarithm of the GSVI. We also use a passive attention measure based on media coverage in newspapers (e.g., deHaan et al., 2015). We obtain data on news coverage from RavenPack. LN_NEWS is the natural logarithm of (one plus) the number of news articles published in a month about a firm. Searching and news reading activity on Bloomberg terminals is another measure (e.g., Ben-Rephael et al., 2017). Each month, for each stock, we calculate the ratio of days with abnormal attention (i.e., when the Bloomberg daily maximum attention score is 3 or 4) ($AIAR$).

Compared to individual investors who search for information on Google, those searching for information on Bloomberg terminals are more likely to be institutional investors (e.g., Ben-Rephael et al., 2017). LN_GSVI and $AIAR$ are positively correlated in our sample, but the level of the correlation is rather low (0.154), consistent with the finding of Ben-Rephael et al. (2017), and suggests that the two measures capture different aspects of investor attention. The intuition behind LN_NEWS is that the number of published newspaper articles is correlated with both the breadth of scientific news dissemination (i.e., the existence of more articles means that more investors learn about scientific activities) and the depth of news dissemination (i.e., investors can acquire more scientific information when more articles are available) (Bushee et al., 2010). We confirm a positive correlation between LN_NEWS and LN_GSVI (0.149) and $AIAR$ (0.603).

We use the following OLS regressions to examine the relation between investor attention and scientific publications:

$$Y_{i,m} = \alpha + \beta LN\_PUB_{i,m} + \gamma Z_{i,m} + \delta_m + \eta_{i,y} + \varepsilon_{i,m} \tag{2}$$

where i indexes the firm, m indexes the month, and $Y_{i,m}$ is the dependent variable, which is the Google search volume (LN_GSVI), the level of news coverage (LN_NEWS), or the attention measure from Bloomberg ($AIAR$). Our main variable of interest is the natural logarithm of (one plus) the monthly count of scientific publications (LN_PUB). Z is the vector of control variables, which includes several stock characteristics associated with investor attention. We include the natural logarithm of firm market capitalization (LN_MCAP), the natural logarithm of the number of months since CRSP first reported return data for the firm (LN_AGE), and the natural logarithm of (one plus) the number of analysts following (LN_COV) (Da et al., 2011). Based on Grullon et al. (2004), we control for share turnover (LN_TURN), measured by the monthly average of the share volume divided by shares outstanding; stock returns (RET), measured by the monthly average of daily stock returns; and return volatility (SD_RET), measured by the monthly standard deviation of daily returns. Our regressions also include dummy variables for whether a stock is included in the S&P 500 index and for whether a stock is listed on the NASDAQ. Following Madsen and Niessner (2019), we further include firm–year fixed effects ($\eta_{i,y}$) and month fixed effects (δ_m) to control for differences across firms and months. Standard errors are clustered at the firm–year level.[20]

Columns 1 through 6 of Table 6 present the results with and without control variables. In all six columns, we obtain a positive and statistically significant relationship between the number of scientific publications and investor attention. In terms of economic significance, a one standard deviation increase in the number of scientific publications (1) is associated with a 1% increase in Google searches for company tickers (column 2) and in the number of news articles (column 4). For context, Madsen and Niessner (2019) find that Google searches increase by approximately the same magnitude when firms announce new products in newspapers. Similarly, a one standard deviation increase in the number of scientific publications is associated with a 3% increase in news reading and news searches on Bloomberg terminals relative to the sample average (0.181). Overall, the three attention proxies indicate increased investor attention to scientific publications.

We also examine whether this increased attention has an effect on stock liquidity. To do so, we estimate Eq. (2) with the dependent variable replaced by the bid–ask spread. We calculate this measure as the monthly average of daily bid–ask spreads (multiplied by 100). Columns 7 and 8 of Table 6 present the results. The coefficients on scientific publications are negative and statistically significant. We obtain a 2% decrease in the bid–ask spread relative to the sample average (0.450) with an increase in scientific publications by one standard deviation (column 8). Overall, these results point to the plausibility of a channel whereby scientific publications are effective in increasing investor attention and thereby improving liquidity, as theory would predict.

Table 6
Regressions of investor attention on scientific publications.

This table presents ordinary least squares (OLS) regression results of investor attention on the number of scientific publications. The unit of observation is the firm–month. Our initial sample includes U.S. public firms with at least one year of positive R&D expenditures and at least one patent over the period from 2004 to 2014. Our sample period begins in 2004 because Google Trends provides data on search term frequency dating back to January 2004 (https://trends.google.com/trends). Bloomberg's historical attention measure begins on 2/17/2010. Historical data are missing for the periods between 12/6/2010-1/7/2011 and 8/17/2011-11/2/2011 (Ben-Rephael et al., 2017). For this reason, the sample size is reduced when we use the Bloomberg measure. In all specifications, we include firm–year fixed effects and month fixed effects. Robust standard errors (in parentheses) are clustered by firm–year. Variable definitions are provided in the Online Appendix, Section A.2. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable	LN GSVI (1)	LN GSVI (2)	LN NEWS (3)	LN NEWS (4)	AIAR (5)	AIAR (6)	BA SPRD (7)	BA SPRD (8)
LN_PUB	0.010***	0.010***	0.019***	0.011***	0.006***	0.005***	-0.006***	-0.007***
	(0.003)	(0.003)	(0.005)	(0.004)	(0.001)	(0.001)	(0.002)	(0.002)
LN_MCAP		0.138***		0.185***		0.037***		-0.429***
		(0.014)		(0.017)		(0.007)		(0.025)
LN_AGE		-0.034		-0.021		0.019		0.120
		(0.047)		(0.069)		(0.026)		(0.078)
LN_COV		0.010***		0.274***		0.013***		0.001
		(0.002)		(0.003)		(0.001)		(0.002)
LN_TURN		0.098***		0.323***		0.032***		-0.238***
		(0.006)		(0.008)		(0.003)		(0.014)
RET		-1.254***		-1.077**		-0.446**		4.560***
		(0.348)		(0.479)		(0.177)		(0.458)
SD_RET		2.740***		7.950***		1.183***		6.483***
		(0.214)		(0.262)		(0.121)		(0.408)
NASDAQ		-0.057		-0.100		-0.010		0.477**
		(0.108)		(0.091)		(0.039)		(0.212)
SP500		-0.026		0.060		0.030*		0.033
		(0.040)		(0.057)		(0.015)		(0.055)
Firm × Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	69,027	69,027	69,027	69,027	19,778	19,778	69,027	69,027
Adjusted R^2	0.820	0.826	0.728	0.797	0.806	0.817	0.869	0.886

8. Conclusion

We examine how information asymmetry in financial markets affects firms' scientific disclosure policies. We use two natural experiments to identify changes in analyst coverage that are exogenous to firm policies: broker closures and broker mergers. Using a DiD approach, we find that a reduction in analyst coverage leads to an increase in the number of scientific publications. Moreover, the results are stronger with respect to firms for which the decrease in analyst coverage is more costly: those with a larger increase in information asymmetry, those that are financially constrained, and those run by managers with short horizons. Similarly, our results are stronger for firms with low disclosure costs. Attempts to attract investors to offset the consequences of an increase in information asymmetry appear to be a plausible underlying economic mechanism.

Our results have implications for the literature on financial analysts. If firms respond to a loss of coverage by increasing their R&D disclosure, then it must be the case that brokers and analysts produce research that is relevant to investors and firms. In addition, we show that this loss does not appear to change investments in R&D. Hence, it is inappropriate to conclude that analyst coverage has a dark side, at least for R&D firms. Our findings suggest that future researchers should pay closer attention to the context when testing theories about the effects of analyst coverage on corporate policies.

We also contribute to the literature on corporate R&D by showing that information asymmetry in financial markets causes changes in R&D disclosure. This finding suggests that the large literature that infers changes in R&D investments based on changes in observable R&D outcomes is incomplete, as the decision to disclose such outcomes is deliberate. Managers choose to disclose R&D outcomes when the benefits of disclosure outweigh the costs. Our work represents one of the few empirical studies on the role of scientific publications in financial markets. We also examine additional metrics based on patenting and patent citations. Our findings suggest that future researchers should use multiple metrics when drawing inferences about the causal relation between financial markets and corporate R&D.

Notes

[1]We refer to information asymmetry as differences in information among investors, including differences both in knowledge about the firm across investors and in the fraction of investors who know about the firm.

[2]In 2015, the average number of scientific publications by publicly traded firms in the U.S. was approximately 19. In absolute terms, these firms have published more than 27,000 articles in academic journals—a substantial number (Arora et al., 2021).

[3]See Nature Research, Nature Index, 2021, Annual tables. Available at https://www.natureindex.com/annual-tables/2021 (last accessed October 8, 2021).

[4]As noted by Amir and Lev (1996), this information is generally not available in financial statements.

[5]See Leibniz Centre for European Economic Research, ZEW/ISI Workshop on the Exchange Between Research, Businesses and, Policymaking, 2020, Panel 2: What motivates firms to publish research articles? Available at https://www.zew.de/en/zew/news/why-do-firms-publish-their-research (last accessed October 14, 2021).

[6]This is especially true for publications in hard sciences, which are the focus of this paper. For instance, the median time from submission to final editorial acceptance for all Nature journals ranges between 59 and 284 days. Statistics are available at https://www.nature.c m/nature-portfolio/about/journal-metrics (last accessed February 1, 2022).

[7]In the context of U.S. biotechnology firms, Audretsch and Stephan (1996) show that university-based scientists are extensively involved in the research agendas of firms.

[8]Scientific publications by firms can generate substantial media coverage, as illustrated by Google's attempt to demonstrate quantum supremacy in a Nature article titled *"Quantum supremacy using a programmable superconducting processor"*. The article (see https://doi.org/10.1038/s41586-019-1666-5) was submitted on July 22, 2019, accepted on September 2019, and published on October 23, 2019. On the day of publication, Google released a press statement, and all major news outlets around the globe featured the article and commentary on Google's experiment, including The New York Times, CNN and Newsweek. Overall, according to the PlumX metrics database, the article was mentioned 460 times in news outlets by December 2, 2021 (see https://plu.mx/w/a/-p-0alZf8LQIe_tAU10XTyzYXw6jGEF4v9baor5zhwk).

[9]Hounshell and Smith (1988) note that a representative clearance process for scientific publications involves the research director, the patent division and other interested divisions. Other divisions can have veto power over the publication of work done by another division if a paper is perceived as relating to that division's business.

[10]The choice of a three-year window before and after the coverage loss follows the precedent in related literature. However, our results are similar irrespective of whether we use a two-, four-, or five-year window, as shown in the Online Appendix, Table AT6.

[11]To address the concern that firm–year observations can overlap across events, we restrict the control sample to firms not included in any treatment cohort during the relevant pretreatment or posttreatment time window. This design implies that certain firms can serve as treatment firms in one period and control firms in another, although never within the seven-year window surrounding treatment. To address the concern about treatment effect heterogeneity raised by de Chaisemartin and D'Haultfœuille (2020) and Goodman-Bacon (2021), we perform the Goodman-Bacon decomposition and analyze the weights underlying our staggered DiD regressions. The

decomposition results suggest that our regressions put 97.1% of the weight on the "non-problematic" comparison between treated and never treated observations.

[12]Our identification strategy employs broker disappearances between 2000 and 2010, and we examine the three fiscal years before and after each event. Following the literature, we also construct a 12-month "disappearance period" symmetrically around the identified events to avoid overlaps in years −1 and +1. For these reasons, the baseline sample covers the period between 1997 and 2014.

[13]Elsevier's Scopus is also the preferred source used by the National Science Board to track scientific research trends in the U.S. and for international comparisons. See https://ncses.nsf.gov/pubs/nsb20206.

[14]Event fixed effects refer to 43 dummy variables, one for each specific merger/closure event included in our sample.

[15]In robustness tests, we also cluster standard errors at the firm–event level or the event level or collapse the time series information into two effective periods (before and after the event).

[16]We obtain similar results when we use a probit regression to predict propensity scores.

[17]We obtain similar results if we match by publication levels instead of publication growth rates or by analyst coverage instead of bid–ask spreads. For instance, repeating the specification in Table 3, column 7, but using publication levels as a matching variable instead of publication growth leads to a DiD coefficient (standard error) of 0.160 (0.044) for a sample of 8,225 observations. Substituting bid–ask spreads for analyst coverage leads to a DiD coefficient (standard error) of 0.192 (0.074) for 2,496 observations. We also conduct many other experiments, such as matching on fewer variables (total assets and Q) or using alternative matching procedures (coarsened exact matching or entropy balanced matching). The DiD matching estimator is robust across all these experiments.

[18]The graphs plot the point estimates and 90% confidence intervals of the estimates on β_τ from the following regression:

$$Y_{i,e,t} = \alpha + \sum_{\tau=-3}^{3} \beta_\tau \; Treated_{i,e} \times L_{t-\tau} + \delta_t + \lambda_i + \theta_e + \varepsilon_{i,t}$$

where $Y_{i,e,t}$, the dependent variable, is either COV, firm i's number of analysts following in a given year, or LN_PUB, the natural logarithm of (one plus) firm i's number of scientific publications in a given year. δ_t, λ_i, and θ_e are dummies that correspond to year, firm, and merger/closure event fixed effects, respectively. $Treated_i \times L_{t-\tau}$ is the interaction term between a variable indicating the year relative to the analyst shock (ranging from three years prior to the shock until three years after) and a dummy equal to one if firm i is part of the treatment sample for event e. The sample includes the treated and control firms assembled through the sample construction procedure described in Section 4.1.

[19]We do not claim that patents do not reveal proprietary information (e.g., as shown by Kim and Valentine, 2021) but that successful imitation by rivals becomes more costly when patent protection is available.

[20]The results are robust to clustering standard errors at the firm level or the firm and month level (two-way clustering).

References

Aboody, D. and B. Lev (2001). Information asymmetry, R&D, and insider gains. *The Journal of Finance 55*, 2746–2766.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31–56.

Amir, E. and B. Lev (1996). Value-relevance of nonfinancial information: The wireless communications industry. *Journal of Accounting and Economics 22*(1), 3–30.

Aobdia, D. (2018). Employee mobility, noncompete agreements, product-market competition, and company disclosure. *Review of Accounting Studies 23*(1), 296–346.

Arora, A., S. Belenzon, and A. Patacconi (2018). The decline of science in corporate R&D. *Strategic Management Journal 39*(1), 3–32.

Arora, A., S. Belenzon, and L. Sheer (2021, March). Knowledge spillovers and corporate investment in scientific research. *American Economic Review 111*(3), 871–98.

Audretsch, D. B. and P. E. Stephan (1996). Company-scientist locational links: The case of biotechnology. *The American Economic Review 86*(3), 641–652.

Azoulay, P. (2002). Do pharmaceutical sales respond to scientific evidence? *Journal of Economics & Management Strategy 11*(4), 551–594.

Balakrishnan, K., M. B. Billings, B. Kelly, and A. Ljungqvist (2014). Shaping liquidity: On the causal effects of voluntary disclosure. *The Journal of Finance 69*(5), 2237–2278.

Barron, O. E., D. Byard, C. Kile, and E. J. Riedl (2002). High-technology intangibles and analysts' forecasts. *Journal of Accounting Research 40*(2), 289–312.

Barth, M. E. and A. P. Hutton (2004). Analyst earnings forecast revisions and the pricing of accruals. *Review of Accounting Studies 9*(1), 59–96.

Barth, M. E., R. Kasznik, and M. F. McNichols (2001). Analyst coverage and intangible assets. *Journal of Accounting Research 39*(1), 1–34.

Bellstam, G., S. Bhagat, and J. A. Cookson (2021). A text-based analysis of corporate innovation. *Management Science 67*(7), 4004–4031.

Ben-Rephael, A., Z. Da, and R. D. Israelsen (2017). It depends on where you search: Institutional investor attention and underreaction to news. *The Review of Financial Studies 30*(9), 3009–3047.

Bhattacharya, S. and J. R. Ritter (1983). Innovation and communication: Signalling with partial disclosure. *The Review of Economic Studies 50*(2), 331–346.

Billett, M. T., J. A. Garfinkel, and M. Yu (2017). The effect of asymmetric information on product market outcomes. *Journal of Financial Economics 123*(2), 357–376.

Blumenthal, D., N. Causino, E. Campbell, and K. S. Louis (1996). Relationships between academic institutions and industry in the life sciences - An industry survey. *New England Journal of Medicine 334*(6), 368–374.

Brennan, M. J. and A. Subrahmanyam (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics 41*(3), 441–464.

Bushee, B. J. (1998). The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review 73*(3), 305–333.

Bushee, B. J., J. E. Core, W. Guay, and S. J. W. Hamm (2010). The role of the business press as an information intermediary. *Journal of Accounting Research 48*(1), 1–19.

Bushee, B. J. and G. S. Miller (2012). Investor relations, firm visibility, and investor following. *The Accounting Review 87*(3), 867–897.

Chen, T., J. Harford, and C. Lin (2015). Do analysts matter for governance? Evidence from natural experiments. *Journal of Financial Economics 115*(2), 383–410.

Clarke, J., N. Dass, and A. Patel (2015). When do analysts impede innovation? Unpublished working paper, Georgia Institute of Technology and Wake Forest University.

Coles, J. L., N. D. Daniel, and L. Naveen (2008). Boards: Does one size fit all? *Journal of Financial Economics 87*(2), 329–356.

Czarnitzki, D., C. Grimpe, and A. A. Toole (2014). Delay and secrecy: Does industry sponsorship jeopardize disclosure of academic research? *Industrial and Corporate Change 24*(1), 251–279.

Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. *The Journal of Finance 66*(5), 1461–1499.

Dasgupta, P. and P. A. David (1994). Toward a new economics of science. *Research Policy 23*(5), 487–521.

Dass, N., V. Nanda, H. D. Park, and S. C. Xiao (2020). Intellectual property protection and financial markets: Patenting versus secrecy. *Review of Finance 25*(3), 669–711.

de Chaisemartin, C. and X. D'Haultfœuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–2996.

deHaan, E., T. Shevlin, and J. Thornock (2015). Market (in)attention and the strategic scheduling and timing of earnings announcements. *Journal of Accounting and Economics 60*(1), 36–55.

Derrien, F. and A. Kecskés (2013). The real effects of financial shocks: Evidence from exogenous changes in analyst coverage. *The Journal of Finance 68*(4), 1407–1440.

Diamond, D. W. (1985). Optimal release of information by firms. *The Journal of Finance 40*(4), 1071–1094.

Diamond, D. W. and R. E. Verrecchia (1991). Disclosure, liquidity, and the cost of capital. *The Journal of Finance 46*(4), 1325–1359.

Dworkin, T. M. and E. S. Callahan (1998). Buying silence. *American Business Law Journal 36*(1), 151–191.

Ellis, J. A., C. E. Fee, and S. E. Thomas (2012). Proprietary costs and the disclosure of information about customers. *Journal of Accounting Research 50*(3), 685–727.

Ellul, A. and M. Panayides (2018). Do financial analysts restrain insiders' informational advantage? *Journal of Financial and Quantitative Analysis 53*(1), 203–241.

Entwistle, G. M. (1999). Exploring the R&D disclosure environment. *Accounting Horizons 13*(4), 323–342.

Fich, E. M., J. L. Juergens, and M. S. Officer (2018). Analyst coverage and acquisition returns: Evidence from natural experiments. Unpublished Working Paper, Drexel University, U.S. Securities and Exchange Commission and Loyola Marymount University.

Fitzgerald, T., B. Balsmeier, L. Fleming, and G. Manso (2021). Innovation search strategy and predictable returns. *Management Science 67*(2), 1109–1137.

Francis, J., D. Nanda, and P. Olsson (2008). Voluntary disclosure, earnings quality, and cost of capital. *Journal of Accounting Research 46*(1), 53–99.

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

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

Garmaise, M. J. (2011). Ties that truly bind: Noncompetition agreements, executive compensation, and firm investment. *Journal of Law, Economics, & Organization 27*(2), 376–425.

Glaeser, S. (2018). The effects of proprietary information on corporate disclosure and transparency: Evidence from trade secrets. *Journal of Accounting and Economics 66*(1), 163 – 193.

Glaeser, S., J. Michels, and R. Verrecchia (2020). Discretionary disclosure and manager horizon: Evidence from patenting. *Review of Accounting Studies 25*(2), 597–635.

Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics 14*(1), 71–100.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277.

Griliches, Z. (1986). Productivity, R&D, and basic research at the firm level in the 1970's. *The American Economic Review 76*(1), 141–154.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *The Journal of Law & Economics 24*(3), 461–483.

Grullon, G., G. Kanatas, and J. P. Weston (2004). Advertising, breadth of ownership, and liquidity. *The Review of Financial Studies 17*(2), 439–461.

Guo, B., D. Pérez-Castrillo, and A. Toldrà-Simats (2019). Firms' innovation strategy under the shadow of analyst coverage. *Journal of Financial Economics 131*(2), 456–483.

Guo, R.-J., B. Lev, and N. Zhou (2004). Competitive costs of disclosure by biotech IPOs. *Journal of Accounting Research 42*(2), 319–355.

Hall, B. H. and R. H. Ziedonis (2001). The patent paradox revisited: An empirical study of patenting in the U.S. semiconductor industry, 1979-1995. *The RAND Journal of Economics 32*(1), 101–128.

He, J. J. and X. Tian (2013). The dark side of analyst coverage: The case of innovation. *Journal of Financial Economics 109*(3), 856–878.

Heath, D., M. C. Ringgenberg, M. Samadi, and I. M. Werner (2021). Reusing natural experiments. Unpublished working paper, University of Utah, Federal Reserve Board of Governors, The Ohio State University, and CEPR.

Hoberg, G. and V. Maksimovic (2015). Redefining financial constraints: A text-based analysis. *The Review of Financial Studies 28*(5), 1312–1352.

Hong, H. and M. Kacperczyk (2010). Competition and bias. *The Quarterly Journal of Economics 125*(4), 1683–1725.

Hong, H. and J. D. Kubik (2003). Analyzing the analysts: Career concerns and biased earnings forecasts. *The Journal of Finance 58*(1), 313–351.

Hong, H., T. Lim, and J. C. Stein (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance* *55*(1), 265–295.

Hounshell, D. and J. K. Smith (1988). *Science and corporate strategy: Du Pont R&D, 1902-1980.* Cambridge University Press.

Hsu, P.-H., D. Li, Q. Li, S. H. Teoh, and K. Tseng (2021). Valuation of new trademarks. Management Science, forthcoming.

Irani, R. M. and D. Oesch (2013). Monitoring and corporate disclosure: Evidence from a natural experiment. *Journal of Financial Economics* *109*(2), 398–418.

Irani, R. M. and D. Oesch (2016). Analyst coverage and real earnings management: Quasi-experimental evidence. *Journal of Financial and Quantitative Analysis* *51*(2), 589–627.

Kelly, B. and A. Ljungqvist (2012). Testing asymmetric-information asset pricing models. *The Review of Financial Studies* *25*(5), 1366–1413.

Keum, D. D. (2021). Innovation, short-termism, and the cost of strong corporate governance. *Strategic Management Journal* *42*(1), 3–29.

Kim, J. and K. Valentine (2021). The innovation consequences of mandatory patent disclosures. *Journal of Accounting and Economics* *71*(2), 101381.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics* *132*(2), 665–712.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica* *53*(6), 1315–1335.

Lambert, R., C. Leuz, and R. E. Verrecchia (2007). Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research* *45*(2), 385–420.

Li, Y., Y. Lin, and L. Zhang (2018). Trade secrets law and corporate disclosure: Causal evidence on the proprietary cost hypothesis. *Journal of Accounting Research* *56*(1), 265–308.

Lin, H.-W. and M. F. McNichols (1998). Underwriting relationships, analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics* *25*(1), 101–127.

Loh, R. K. and R. M. Stulz (2011). When are analyst recommendation changes influential? *The Review of Financial Studies* *24*(2), 593–627.

Madsen, J. and M. Niessner (2019). Is investor attention for sale? The role of advertising in financial markets. *Journal of Accounting Research* *57*(3), 763–795.

Marx, M. and A. Fuegi (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal* *41*(9), 1572–1594.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance* *42*(3), 483–510.

Mukherjee, A. and S. Stern (2009). Disclosure or secrecy? The dynamics of open science. *International Journal of Industrial Organization* *27*(3), 449–462.

Murray, F., P. Aghion, M. Dewatripont, J. Kolev, and S. Stern (2016). Of mice and academics: Examining the effect of openness on innovation. *American Economic Journal: Economic Policy* *8*(1), 212–52.

Palmon, D. and A. Yezegel (2012). R&D intensity and the value of analysts' recommendations. *Contemporary Accounting Research* *29*(2), 621–654.

Plumlee, M., D. Brown, R. M. Hayes, and R. S. Marshall (2015). Voluntary environmental disclosure quality and firm value: Further evidence. *Journal of Accounting and Public Policy* *34*(4), 336–361.

Png, I. P. L. (2017). Law and innovation: Evidence from state trade secrets laws. *The Review of Economics and Statistics* *99*(1), 167–179.

Polidoro, F. and M. Theeke (2012). Getting competition down to a science: The effects of technological competition on firms' scientific publications. *Organization Science* *23*(4), 1135–1153.

Reeb, D. M. and W. Zhao (2020). Patents do not measure innovation success. *Critical Finance Review* *9*(1-2), 157–199.

Rogers, J. L. and P. C. Stocken (2005). Credibility of management forecasts. *The Accounting Review* *80*(4), 1233–1260.

Simeth, M. and M. Cincera (2016). Corporate science, innovation, and firm value. *Management Science* *62*(7), 1970–1981.

Stephan, P. E. (1996). The economics of science. *Journal of Economic Literature* *34*(3), 1199–1235.

Thursby, J. G. and M. C. Thursby (2007). University licensing. *Oxford Review of Economic Policy* *23*(4), 620–639.

Verrecchia, R. E. (1983). Discretionary disclosure. *Journal of Accounting and Economics* *5*, 179–194.

Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value? *The Journal of Finance* *51*(1), 137–167.

Online Appendix

This Online Appendix provides supplemental information and robustness tests that accompany the main article. In Section A.1, we provide additional passages taken from conference calls that show qualitative evidence on the potential benefits of scientific publications. In Section A.2, we provide variable definitions. In Section A.3, we show standard ordinary least squares (OLS) regressions of the number of scientific publications on analysts' coverage. In Section A.4, we show additional DiD robustness results in tables that do not appear in the article. Finally, in Section A.5, we provide details on the matching procedure between records from the North American Center for Research in Security Prices (CRSP)/Compustat Merged Database and scientific publication data from Elsevier's Scopus database.

A.1. Additional conference call transcripts

In this section, we provide additional passages from transcripts of conference calls discussing several aspects of scientific publishing.

Passages related to the benefits of scientific publishing

<**Thomas Patton (CEO)**>: *As a reminder, during October, there were eight key scientific publications presented at the annual meeting of the American Society of Anesthesiology and the associated SNACC [Society for Neuroscience in Anesthesiology and Critical Care] meeting on either the accuracy of FORE-SIGHT or the use of FORE-SIGHT in various clinical situations. I discussed these abstracts in some detail last quarter, but I will remind you that one paper which we commissioned directly compared FORE-SIGHT with four competitive oximetry devices. In each case, the precision of the oximeter was reported as a standard deviation from an invasive reference. And FORE-SIGHT was the lowest amongst all monitors tested, with a standard deviation of just plus or minus 3%. This compares with a standard deviation of plus or minus 9.7% for the leading competitor. The number of scientific publications regarding the use of FORE-SIGHT oximetry now stands at 175. There were 49 new publications issued in 2011, and we have added another 19 publications to date in 2012. Expanding the body of scientific knowledge regarding FORE-SIGHT amongst clinicians is an important part of our strategy to drive clinical education and expanded product use, and we are making good progress in that initiative... .*

Source: CAS, 2011:Q4 Earnings Call, accessed from Thomson One.

<**Michael Becker (CEO)**>: *Now I would like to turn to provide an update regarding some of our future potential growth drivers. ... Our optimism about Combidex was reinforced by the publication in June of 2003 of a study from the New England Journal of Medicine showing that Combidex aids in the non-invasive diagnosis of otherwise undetectable lymph node metastasis in patients with prostate cancer... .*

Source: Cytogen, 2003:Q3 Earnings Call, accessed from Thomson One.

<**Stan Crooke (CEO)**>: *In December, we had two papers published in the New England Journal of Medicine. The first was on APOCIIIRx and the second was on Factor XIRx. These were both relatively small-ish Phase II programs and so we were very excited to have these publications and it says two things to me that I want to share with you. I think it says that the technology is now fully accepted and it says we're doing important work with the technology... .*

Source: Isis, 2015:Q1 Healthcare Conference, accessed from Thomson One.

<**Kerry Gray (CEO)**>: *I'd now like to briefly update you on publications and certain things we're looking at in the scientific area. Since the last call, we have been in contact with a number of academic institutions that have expressed interest in assisting the Company to determine the scientific basis behind a number of the observed benefits of Altrazeal. We are constantly asked by not only our physicians, but also our strategic partners and potential partners, as to why patients experienced almost immediate pain relief, healing is accelerated, and there is seen improvement in wound exudate and its control. To date, we have only been able to offer hypotheses on these subjects. While it is not crucial to the commercial success of Altrazeal, it would certainly add further credibility to Altrazeal if we had scientific data supporting a mechanism of action. ... Scientific publication on the basic science will not only add additional credibility, but it will also provide further marketing support.*

Source: ULURU, 2014:Q1 Earnings Call, accessed from Thomson One.

<**Parker Petit (CEO)**>: *It is one of the markets that we'll be pursuing in 2015. We have a publication that's being worked on right now, we like to step into markets where we've gotten some clinical data and have gotten it through a publication, peer-reviewed publication process, then we feel comfortable with the talking about the process and procedure and what we recommend the physicians to do about it. So that's well under way and we'll keep everybody informed as it plays out. But it's an exciting opportunity for us. But we have more data to develop and get published... .*

Source: MiMedx, 2014:Q2 Earnings Call, accessed from Thomson One.

<**Don Kania (CEO)**>: *In a growing number of transactions we are displacing more traditional technologies such as x-ray crystallography. We are excited by the increased number of publications and recognition of cryo-EM in the leading scientific journals. We are especially proud that in January Nature Methods named cryo-EM as Method of the Year. And 2015 saw numerous papers highlighting the ability to achieve near atomic resolution with all our cryo-EM solutions... .*

Source: FEI, 2015:Q4 Earnings Call, accessed from Thomson One.

<**Jorge Garces (VP)**>: *Our plans have been to display some of the data at AACC, but already at [CDS] and some conferences we have a number of customers that have presented abstracts and posters, so that will continue as well as publications. We want to publish the data in peer review journals—that way it has a lot more bang for the buck than just presenting it at a meeting... .*

Source: Hologic, 2015:Q2 Acquisition Announcement, accessed from Thomson One.

<**Gregory Moore (VP)**>: *This is an amazing group of some of the world's leading engineers, leading scientists and researchers that go all the way from neuroscience to communications technology. Microsoft has really put some of the leading work out there over the last decade in open source in the AI and machine learning community and is an active publisher in this space but also in deep scientific areas like synthetic biology and is partnering with the best universities in the world, Cambridge and Oxford... .*

Source: Microsoft, 2020:Q2 RBC Capital Markets, accessed from Thomson One.

Passages related to preferences for scientific publishing

<**Christopher Anzalone (CEO)**>: *We don't believe in science by press release; we believe in science by peer-reviewed journal publications. And so we really can't go too much into the data we're seeing right now. What we anticipate is that once the Phase 1 is done, we will write that up and present it at a meeting or publish it. And if we start talking about these data as they come out, we lose the ability to do that. And therefore, I think that data loses a lot of its power. So we really can't speak too much to what we're seeing right now... .*

Source: Arrowhead, 2010:Q4 Earnings Call, accessed from Thomson One.

<**Christopher Anzalone (CEO)**>: *There is always that tension between what you can do and not lose your ability to present it at big international meetings like AASLD [American Association for*

the Study of Liver Diseases] and EASL [European Association for the Study of the Liver] —cannot be prevented from publishing while giving investors enough to really understand the basics of what is in the data. So that is a tension that all of us in the industry face. And we will do our best to give as much information as we can, you know, without giving so much that we are precluded from scientific publication and presentation of the data. There is always that tension. I'm sure you are well aware of that... .

Source: Arrowhead, 2015:Q2 Earnings Call, accessed from Thomson One.

<**Peter Kim (VP)**>: *With regard to the press release, it certainly is our preference, as you know, to present our data at scientific conferences in a peer-reviewed setting and not be a press release. However, as we indicated in this press release, we have already submitted abstracts to the AASLD, and we are going to be submitting additional abstracts this week with other data from our Phase III studies. And given the interest in these data, we believe it was important to share the top-line data at this time, and that is what we are doing. We are sharing just the top-line data. But again, out of respect for the peer review process that really is our preferred avenue for disclosing data at medical meetings and publications, we are not going to be sharing additional data here... .*

Source: Merck, 2010:Q3 Conference Call on Boceprevir, accessed from Thomson One.

<**Elias Zerhouni (VP)**>: *Yes, I'll take the dengue question. This is Elias. In terms of the specific results, I don't want to really prevent the publication in a top-notch journal, which journals are really anxious to publish this before the end of this quarter, so the results will be out there. But it's trending all in the right direction. Hospitalizations, definitely down... .*

Source: Sanofi, 2014:Q1 Earnings Call, accessed from Thomson One.

<**Mads Krogsgaard Thomsen (VP)**>: *As an integral part of the process Novo Nordisk will submit a briefing book to the FDA [Food and Drug Administration] and the Advisory Committee summarizing the data that will form the basis for the liraglutide benefit risk assessment at the hearing on April 2. The briefing book will be made public at the FDA's website shortly before the meeting. Until then, Novo Nordisk will only elaborate on data that has published in scientific journals or at scientific conferences, as is customary when a company is in a regulatory approval process... .*

Source: Novo Nordisk, 2008:Q4 Earnings Call, accessed from Thomson One.

<**Arvin Sood (VP)**>: *Recently we also conducted a rather large study. This was 1,100 patients comparing denosumab to alendronate, or Fosamax, which is pretty much the gold standard in the bisphosphonate class. Once again, we met all the primary and secondary end points. We did not communicate a lot of details from this particular study because we didn't want to compromise the scientific publication. We are going to be presenting the study in complete detail at the upcoming ECTS meeting on May 28, and ECTS just stands for European Calcified Tissue Society meeting. And we are going to conduct a conference call, by the way, in conjunction with that, so those of you who might be interested will be able to participate in that.*

Source: Amgen, 2008:Q2 Growth Stock Conference, accessed from Thomson One.

A.2. Variable definitions

In this section, we provide definitions for the variables used in our empirical analysis and list their sources.

- AIAR is the Ben-Rephael et al. (2017) abnormal institutional attention ratio, calculated as the number of days when the Bloomberg attention score is 3 or 4 divided by the total number of days (Bloomberg).

- AMIHUD is the Amihud's (2002) illiquidity measure, calculated as the average daily absolute value of stock returns divided by the dollar value of trading volume during the fiscal year (CRSP).
- BASPRD is the bid–ask spread, calculated as the average daily ask price minus the bid price divided by the average of the ask price and bid price (CRSP).
- CAPEXTA is capital expenditures (#128) divided by the book value of total assets (#6) (Compustat).
- COV is analyst coverage, defined as the number of analysts covering the firm (I/B/E/S).
- CREDIT_RATING equals one if the firm has a Standard & Poor's (S&P) Domestic Long-Term Issuer Credit Rating (splticrm) from AAA to C (S&P credit ratings).
- DELAYCON is financial constraints calculated based on a text-based index of financial constraints from Hoberg and Maksimovic (2015) (Gerard Hoberg's website).
- DUAL_CEO denotes whether the CEO is also the chair of the board of the company (ISS).
- EINDEX is the entrenchment index, which is based on the six shareholder rights (ISS).
- GINDEX is the governance index, which is based on the 24 shareholder rights (ISS).
- HINDEX is the Herfindahl index of the sales of the four-digit SIC industry to which a firm belongs. HINDEX^2 is the square of HINDEX (Compustat).
- IDD is the state-level adoption of the inevitable disclosure doctrine (Compustat, Li et al. (2018)).
- IHS_PUB is the inverse hyperbolic sine transformation of the number of scientific publications (Scopus).
- INSTOWN is institutional ownership, calculated as the arithmetic mean of the four quarterly institutional holdings reported through form 13F (TR 13F).
- INSTOWN_TRA is the percentage of shares outstanding owned by transient institutional investors, as defined in (Bushee, 1998) (TR 13F, CRSP, Brian Bushee's website).
- LEV is leverage, defined as the book value of debt (#9+#34) divided by the book value of total assets (#6) (Compustat).
- LN_AGE is the log of the number of years in which the firm appears in Compustat.
- LN_AT is the log of the book value of total assets (#6) (Compustat).
- LN_GSVI is the log of the monthly Google Search Volume Index based on the stock ticker (Google Trends).
- LN_GUID is the log of (one plus) the number of managerial earnings forecasts (I/B/E/S).
- LN_MCAP is the log of a stock's monthly market capitalization (CRSP).

- LN_NEW_HIRES is the log of (one plus) the number of new scientists publishing with a firm (Scopus).
- LN_NEWS is the log of (one plus) the number of news articles published on the Dow Jones newswire during the months (Ravenpack).
- LN_PAT is the log of (one plus) the number of patents filed weighted by the number of forward citations (Noah Stoffman's website).
- LN_PUB is the log of (one plus) the number of scientific publications (Scopus).
- LN_TURN is share turnover, calculated as the log of the monthly average of daily shares traded divided by the number of shares outstanding (CRSP).
- NASDAQ is a dummy variable that takes the value of one if a firm is listed on the NASDAQ and zero otherwise (CRSP).
- NON_COMP is the state-level enforcement of noncompete agreements (Compustat, Garmaise (2011)).
- PAT_INT is the patent to R&D expenditure ratio (Noah Stoffman's website, Compustat).
- PAT_SCI is the ratio of science-based patents to a firm's total number of patents, where science-based patents are defined as patents with at least one reference to scientific research on their front page (Noah Stoffman's website, Marx and Fuegi (2020)).
- PAT_SCI_EXT is the ratio of science-based patents with external references to a firm's total number of patents (Noah Stoffman's website, Marx and Fuegi (2020)).
- PAYOUT equals one if the firm has a positive dividend payout (#21) (Compustat).
- PPETA is net property, plant & equipment (#8) divided by the book value of total assets (#6) (Compustat).
- PUB_GROWTH is the growth in scientific publications from year -3 to year -1 (Scopus).
- Q is the market-to-book ratio, calculated as the market value of equity (#199 $\times$ #25) plus the book value of total assets (#6) minus the book value of equity (#60) minus balance sheet deferred taxes (#74, set to zero if missing) divided by the book value of total assets (#6) (Compustat).
- RDTA is R&D expenditure (#46) divided by the book value of total assets (#6), set to zero if missing (Compustat).
- RET is stock returns, calculated as the monthly average of daily returns (CRSP).
- RETR is the return ratio, calculated as the number of days with zero or missing returns divided by the total number of trading days (CRSP).
- ROA is the return on assets, defined as operating income before depreciation (#13) divided by the book value of total assets (#6) (Compustat).

- SD_RET is return volatility, calculated as the monthly standard deviation of daily returns (CRSP).
- S&P500 is a dummy variable that takes the value of one if a firm is a member of Standard & Poor's 500 index and zero otherwise (CRSP).
- UTSA is the state-level enactment of the Uniform Trade Secrets Act (Compustat, Png (2017)).

A.3. Results from OLS estimation

In this section, we report results on the association between analyst coverage and scientific publications from OLS regression models. We estimate the following model:

$$LN\_PUB_{i,t} = \alpha + \beta LN\_COV_{i,t} + \gamma Z_{i,t} + \delta_t + \lambda_i + \varepsilon_{i,t} \tag{1}$$

where i indexes firms and t indexes time. The dependent variable, LN_PUB, is the natural logarithm of (one plus) the total number of scientific publications for firm i. The analyst coverage measure, LN_COV, is measured for firm i over its fiscal year t as the logarithmic transformation of analyst coverage. The vector $Z_{i,t}$ contains firm and industry characteristics that could affect a firm's publication behavior, detailed in Section 4.4 in the main article. δ_t and λ_i correspond to year and firm fixed effects, respectively; standard errors are robust to heteroskedasticity and are clustered at the firm level.

Table AT1 presents descriptive statistics for this sample. Table AT2 presents the results from estimating Eq. (1). The first three columns regress the number of scientific publications only on the coverage variable. We begin without any fixed effects. The next two columns gradually add firm and year fixed effects. The final three columns gradually add the set of control variables. Under the hypothesis that information asymmetries encourage managers to increase scientific publication rates (because publications are value relevant), we expect the coefficient estimate on LN_COV to be negative (because the number of analysts following is inversely related to information asymmetries).

However, across all the columns of Table AT2, the coefficient on LN_COV is positive and significant; hence, higher coverage is associated with more scientific publications. This is inconsistent with our hypothesis and likely occurs due the endogenous relationship between analyst coverage and scientific publications. For instance, it is possible that larger increases in scientific publications are seen only among larger firms because those firms are less concerned about proprietary costs. Alternatively, analysts may have more interest in covering or willingness to cover firms that adopt more forthcoming disclosure practices, as Bushee and Miller (2012) propose, with this mechanism also extending to scientific publications. Whatever the case, these are panel regressions and are difficult to interpret due to the well-known difficulties in controlling for endogeneity in estimations of the relationship between information asymmetries in capital markets and corporate policies.

A.4. Additional robustness tests

In this section, we report the results of additional DiD regressions. The results are discussed in Sections 5.2.3 and 5.3 of the article.

- Table AT3 shows the difference-in-difference regression results when we condition on various brokerage house merger/closure characteristics.
- Table AT4 shows the difference-in-difference regression results from miscellaneous robustness tests.
- Table AT5 shows the difference-in-difference regression when we collapse the panel at the firm–event level.
- Table AT6 shows the difference-in-difference regression results when we use alternative time windows.
- Table AT7 shows the difference-in-difference regression results when we use an alternative definition of R&D-active firms.
- Table AT8 shows the difference-in-difference regression results when we use additional publication measures.

A.5. Matching scientific publications to Compustat

In this section, we describe our approach to matching firms in the Compustat universe to scientific publication data from Scopus for the sample period between 1997 and 2014. The process is divided into five different steps.

Step 1: Selecting the sample and accounting for name changes

We begin with records from Compustat and select all firms with positive assets (Compustat Annual Item #6), sales (#12), and equity (#60) and require firms to have positive R&D expenses (#46) for at least one year during our sample period. We exclude financial and utility firms (with Standard Industrial Classification [SIC] codes 4900-4999 and 6000-6999) as well as firms not headquartered in the U.S. based on their current headquarters location (LOC). After matching the remaining firms to the updated KPSS patent dataset (version released on June 8, 2021) created by Kogan et al. (2017), we further restrict our sample to Compustat firms with at least one patent during our sample period. Before matching those firms to the scientific publication data in Scopus, one important issue to consider is that firms listed in Compustat appear under their most recent name ($CONM$), while scientific publication records contain the firm name at the time of their publication. Without correcting for this, we may undercount scientific publications from earlier sample years. A prominent example is Google. Google is listed in the Compustat database under the name "Alphabet" but has published in the past (and is still publishing) under the name "Google". Following Arora et al. (2021), we complement the most recent firm names listed in Compustat with the historical firm names listed in the

CRSP Monthly Stock File. We find that our sample contains approximately 30% of Compustat records with more than one related name.

Step 2: Cleaning and standardizing firm names

Firm names available in the Compustat database contain legal identifiers and name components that have little meaning for identifying false positives while possibly reducing the recall rate when querying the Scopus database. We therefore develop and execute an extensive cleaning script to remove all these name components. This step is particularly important, as we download publications using an online source. The Scopus application programming interface (API) provides access to a rather flexible search engine that allows partial matches, word permutations, Boolean search queries, and the use of wild cards. However, it does not allow fuzzy matching of words. This implies that names have to be carefully precleaned to ensure that recall is complete. Common legal extension terms include "CO", "CORP", "CORPORATION", "INC", "HLDG", and "HOLDINGS", among many others (including various combinations thereof). For instance, the raw name "LILLY (ELI) & CO" is cleaned toward "ELI LILLY" (and the permutation thereof). We further harmonize various abbreviations of firm name extensions, as displayed below. We follow a hybrid approach for these name terms since they possibly reduce the recall but also have the potential to decrease noise in the querying of firm names through the Scopus API. As a general rule, these extensions have been removed, but if pretests and manual inspections indicate problematic remaining names creating enormous noise, the extensions are retained and harmonized to limit false positives. Examples include "COMM", "ELECTR", "INSTR", "LAB", "PHARM", "SYS", and "TECH", among others. Importantly, beyond generic corrections, we screen all firm names and make various manual corrections of idiosyncratic firm names. Moreover, problematic names are pretested in the standard Scopus API to identify whether they create false positives or false negatives.

Step 3: Querying the standardized firm names through the Scopus API

Once the firm names are prepared for the publication download, we run a query with the Scopus API for all possible publications by our sample firms and add additional information to improve the matching precision. First, we download all publications that match any firm name in the affiliation name field. Each author of a scientific publication has an associated affiliation, and this field allows us to search for publications by institutions such as firms or universities. In the Scopus query syntax, it is possible to restrict the search to the part of the affiliation field referring to the name of the affiliated institution (i.e., the address can be excluded) where the name of a firm would be reported (Scopus query command "Affilorg"). This yields a list of publications where at least one author has an affiliation that matches the search criteria. Having executed a complete test download in October 2019, between April 23 and 27, 2020, we downloaded all scientific publications that contain the names of our sample firms in the affiliation name field. To reduce false positives from the outset, we impose the criterion that in the case of multiple name components, the name components must follow consecutively (e.g., "General Electric") and without permutations for the publication to be downloaded. However, in cases where we suspect that names might be permuted by the authors (e.g., "Eli Lilly" or "Lilly Eli"), we query both name orders.

Second, we complete the affiliation information in the downloaded publications with an additional specific search from the Scopus Affiliation Search. Publication records, when downloaded via the Scopus API, report the Scopus affiliation identifier and only partial information regarding the affiliated institution. For instance, location information is reported in a separate field and cannot be unambiguously matched at the affiliated institution level. To circumvent this limitation, we reconstruct the complete affiliation information by searching for it, starting from the Scopus affiliation identifiers, for all affiliated institutions named in our full set of downloaded publications. This search yields a list of affiliations with well-structured information on the affiliated institution name, name variants, nearly complete and clean location data (city and country), and a total count of publications attributed to each affiliated institution in Scopus.

Third, we classify all affiliated institutions as either a firm or other type (academic, government, etc.). We do this by applying a combination of methods. First, we search for a list of keywords that rather unambiguously imply that one affiliated institution belongs to a specific category (e.g., "university", "college", "school", "faculty" versus "inc.", "corp.", etc.). Then, we match at a highly precise level of string similarity all our affiliated institutions with institutions in the Global Research Identifier Database (GRID; https://www.grid.ac), which contains classifications for different categories. Finally, for the remaining affiliated institutions not yet classified, we retrieve the Scopus classification, which is available via an affiliation-level API (Scopus Affiliation Retrieval API). The remaining unclear affiliation types are sorted by frequency, screened, and assigned manually up to a frequency of 15 observations. For the remainder, we prioritize recall and classify ambiguous cases as firm affiliations. Next, we reshape the publication data to identify and maintain for each publication only the affiliation with the focal firm that we search for (whether the match is correct or not). We do this by maintaining for each publication only affiliations classified as firm affiliations unless all affiliations are classified as nonfirm. For any remaining multiple affiliations, we retain the one with the highest string similarity to the searched firm. We verify that this process identifies the affiliation of the author of the retrieved publication in virtually all cases.

Step 4: Identifying incorrect matches with a machine learning approach

The above steps yield a long list of affiliations paired to firm–name pairs, which we deduplicate at the level of the firm only. At this stage, this list is still characterized by high recall (close to 100% in practice) but low precision. We proceed further by using a machine learning–(ML-) based approach to identify incorrect or incomplete matches. In particular, we use a supervised ML algorithm based on the Python package dedupe to train a model to estimate the probability of a correct match on the basis of the following variables: the similarity of the affiliated institution name and company names, the classification of the affiliated institution as a firm or otherwise, and the coincidence or similarity of the affiliated institution's and company's country (U.S.) and, separately, city. Moreover, we include a number of additional predictors to increase the accuracy of a match. First, we use the total number of publications attributed to one affiliated institution in Scopus. We compare this number with the number of publications associated with the same institution that our search has yielded for a given focal firm. We compute the ratio of these two numbers. The intuition behind this exercise is that the two numbers should be balanced if the match is correct and may be largely unbalanced if it is not correct. Second, we use the ratio between the total number of publications retrieved for a company and its total R&D expenditure. A disproportionate ratio could signal that the company has been incorrectly matched as an author affiliation. Finally, we include interactions between all variables. We set up the algorithm to train the model on chunks of the sample of firms obtained from the deciles by total company R&D expenditure. This is akin to further considering that the relevance of some indicators may differ for different companies. For instance, the match in terms of location should be a more important indicator for small companies, while for large multinational companies, it may have very little predictive power. Dedupe returns examples recursively to be manually classified as correct or incorrect matches, focusing progressively on more ambiguous cases. We provide approximately 200 examples for each chunk. The algorithm returns a list of pairs of potential matches over the full sample with a probability score for each chunk. We bring back this information to our original list of company–affiliation pairs.

Step 5: Cleaning matches: Firm–publication combinations

The probability score estimated according to the procedure above is used to trim matches that can be considered correct or incorrect with high confidence. For the ambiguous matches, we refrain from setting an arbitrary threshold of match acceptance and complete the matching exercise with extensive semimanual cleaning. This additional cleaning is still supported by the information constructed, as detailed in the previous sections and existing benchmark data on companies' publications. Specifically, we use as benchmarks scientific publication information provided by Arora et al. (2021), which stems from the alternative raw data source of Thomson's Web of Science, and scientific publication numbers based on an earlier manual matching exercise of Compustat names with the Scopus database (Simeth and Cincera, 2016).

We proceed as follows to define whether the publication–firm matches are correct. Using information on the affiliation type, we impose the criterion that the name component that forms the basis of the match actually relates to an affiliated institution classified as a company. This crucial automatic cleaning step ensures that false matches are excluded where academic institutions have a name component similar to the names of our Compustat firms of interest. While this intervention proves powerful in identifying false positive matches, the presence of publications matched on the basis of other firms with similar name components poses a concern. Such cases are more difficult to detect. To identify false positives based on similar firm names, we first take advantage of various heuristics and plausibility checks using aggregate firm-level indicators. We create flags for differences in firm-level publication counts with the two publication benchmark indicators, unusual publication-R&D productivity ratios, and a low percentage of firm publications that have (i) a U.S. affiliation and (ii) a city overlap with the firm's headquarters location. A low geographic overlap can indicate foreign R&D locations and point to unrelated foreign companies with similar name components. Subsequently, based on the created flags, we manually screen potentially problematic cases and mark incorrect affiliation names to be removed when we compute final publication counts. A prominent example for a firm with both many correct and many incorrect matches is Johnson & Johnson. The firm name "Johnson" matches records such as "Johnson Matthey Biomedical Research", "Johnson Controls", "Robert Wood Johnson", "Johnson Matthey", or "R.W. Johnson", among others. In such cases, we leverage further information from corporate websites to disambiguate correct and incorrect matches.

Tables

Table AT1
Summary statistics for the OLS sample.
This table presents summary statistics for the ordinary least squares (OLS) sample including U.S. public firms with at least one year of positive R&D expenditures and at least one patent over the period 1997-2014.

Variable	Mean	1%	10%	25%	50%	75%	90%	99%	SD	*N*
LN_PUB	1.465	0.000	0.000	0.000	1.099	2.398	2.398	2.398	1.658	20,826
LN_COV	1.574	0.000	0.000	0.780	1.626	2.329	2.900	3.479	0.986	20,826
LN_AT	5.905	1.914	3.274	4.332	5.731	7.345	8.793	11.258	2.099	20,826
RDTA	0.101	0.000	0.004	0.018	0.057	0.129	0.246	0.701	0.131	20,826
LN_AGE	2.646	1.099	1.609	2.079	2.708	3.296	3.664	3.912	0.770	20,826
PPETA	0.181	0.006	0.033	0.070	0.142	0.253	0.387	0.660	0.147	20,826
CAPEXTA	0.040	0.001	0.008	0.016	0.029	0.051	0.086	0.203	0.038	20,826
ROA	0.035	-0.935	-0.245	-0.000	0.100	0.159	0.220	0.369	0.230	20,826
LEV	0.151	0.000	0.000	0.001	0.105	0.255	0.383	0.633	0.161	20,826
Q	2.323	0.645	0.982	1.239	1.724	2.698	4.375	10.389	1.771	20,826
DELAYCON	-0.011	-0.189	-0.120	-0.068	0.000	0.029	0.102	0.230	0.085	20,826
HINDEX	0.243	0.048	0.063	0.112	0.185	0.306	0.508	0.960	0.191	20,826
$HINDEX^2$	0.095	0.002	0.004	0.013	0.034	0.094	0.258	0.923	0.162	20,826
INSTOWN	0.533	0.002	0.094	0.290	0.583	0.778	0.886	0.990	0.287	20,826

Table AT2
OLS regressions of corporate publications on analyst coverage.

This table presents the ordinary least squares (OLS) results of the regression of corporate scientific publications on analyst coverage over the sample period from 1997 to 2014. Robust standard errors are clustered by firm (in parentheses). The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Dep. Var.: LN_PUB	(1)	(2)	(3)	(4)	(5)	(6)
LN_COV	0.728***	0.154***	0.118***	0.022*	0.026**	0.030**
	(0.038)	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)
LN_AT				0.261***	0.255***	0.261***
				(0.022)	(0.021)	(0.022)
RDTA				0.625***	0.655***	0.652***
				(0.117)	(0.118)	(0.118)
LN_AGE				-0.101**	-0.102**	-0.099**
				(0.043)	(0.043)	(0.043)
PPETA				0.441***	0.418***	0.413***
				(0.119)	(0.118)	(0.118)
CAPEXTA				-0.385*	-0.323	-0.307
				(0.221)	(0.217)	(0.219)
ROA				-0.145**	-0.132**	-0.132**
				(0.061)	(0.061)	(0.062)
LEV				0.048	0.030	0.025
				(0.074)	(0.074)	(0.074)
Q					-0.014***	-0.013**
					(0.005)	(0.005)
DELAYCON					0.032	0.034
					(0.102)	(0.102)
HINDEX					-1.049***	-1.042***
					(0.308)	(0.307)
HINDEX^2					1.003***	0.999***
					(0.290)	(0.288)
INSTOWN						-0.071
						(0.069)
Firm fixed effects	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	Yes
Number of obs.	20,826	20,826	20,826	20,826	20,826	20,826
Adjusted R^2	0.187	0.906	0.910	0.914	0.915	0.915

Table AT3

Regressions of corporate publications on analyst coverage shocks: Conditioning on merger/closure characteristics.

This table replicates the baseline difference-in-difference (DiD) regression results from Table 2, column 7, but splits the sample of brokerage house closures/mergers based on the following characteristics: whether the broker disappeared during a crisis period (calendar years 2000, 2001, 2002, 2008 and 2009) or noncrisis period (columns 1 and 2); whether the broker is among the top 10 brokers ranked by the number of firms that lose analyst or is outside the top 10 (columns 3 and 4); whether the broker disappearance is due to a merger or a closure (columns 5 and 6); and whether the broker is included in the Kelly and Ljungqvist (2012) list or is not included (columns 7 and 8). Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms assembled through the sample construction procedure described in Section 4.1. See Table 2 for control variables and the Online Appendix, Section A.2, for variable definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Crisis?		Top 10?		Merger?		KL?	
Sample	Yes	No	Yes	No	Yes	No	Yes	No
Dep. Var.: LN_PUB	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TREATED × POST	0.102***	0.170***	0.085**	0.158***	0.130***	0.109***	0.127***	0.100***
	(0.035)	(0.053)	(0.037)	(0.038)	(0.039)	(0.037)	(0.037)	(0.038)
POST	0.023	0.007	0.020	0.009	-0.006	0.020*	0.010	-0.006
	(0.019)	(0.011)	(0.014)	(0.012)	(0.011)	(0.012)	(0.011)	(0.019)
TREATED	-0.030	-0.099	-0.042	-0.071	0.020	-0.089**	-0.086*	0.076
	(0.056)	(0.086)	(0.050)	(0.050)	(0.068)	(0.044)	(0.045)	(0.075)
Controls included	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Number of obs.	69,792	52,637	42,870	79,559	66,993	55,436	86,977	35,452
Adjusted R^2	0.905	0.909	0.908	0.901	0.905	0.903	0.902	0.913

Table AT4

Regressions of corporate publications on analyst coverage shocks: Miscellaneous robustness tests.

This table replicates the baseline difference-in-difference (DiD) regression results from Table 2, column 7, but requires firms to be present in the sample over the entire pre- and posttreatment period (column 1), augments the specification by including industry × year fixed effects (column 2), or uses scientific publication data from Web of Science instead of Scopus or alternative clustering schemes (columns 4 though 6). The Web of Science data are from Arora et al. (2021). Industries are defined at the two-digit SIC code level. The sample includes the treated and control firms assembled through the sample construction procedure described in Section 4.1. See Table 2 for control variables and the Online Appendix, Section A.2, for variable definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Test	Balanced Sample	Industry × Year FEs	WoS Data	Alternative clustering schemes		
Dep. Var.: LN_PUB	(1)	(2)	(3)	(4)	(5)	(6)
TREATED × POST	0.114***	0.105***	0.088***	0.116***	0.116***	0.116***
	(0.038)	(0.031)	(0.031)	(0.014)	(0.024)	(0.022)
POST	-0.000	0.010	-0.005	0.010	0.010**	0.010
	(0.011)	(0.009)	(0.008)	(0.008)	(0.005)	(0.008)
TREATED	-0.056	-0.033	-0.054	-0.060***	-0.060*	-0.060**
	(0.051)	(0.039)	(0.041)	(0.016)	(0.035)	(0.029)
Controls included	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry × year fixed effects	No	Yes	No	No	No	No
Clustered by firm	Yes	Yes	Yes	No	No	No
Clustered by event	No	No	No	No	Yes	No
Clustered by event × firm	No	No	No	No	No	Yes
Number of obs.	87,054	122,429	122,429	122,429	122,429	122,429
Adjusted R^2	0.908	0.907	0.880	0.903	0.903	0.903

Table AT5

Regressions of corporate publications on analyst coverage shocks: Collapsing the panel at the event–firm level.

This table replicates the baseline difference-in-difference (DiD) regression results from Table 2, column 7, but collapses the panel at the event–firm level into two effective periods: before and after the coverage shock. The dependent variable, LN_PUB^{*} is the natural logarithm of (one plus) the average number of scientific publications per year in the three-year window either preceding or following the coverage shock. The control variables are measured in the pre-event year (year -1) or the postevent year (year $+1$). The sample includes the treated and control firms assembled through the sample construction procedure described in Section 4.1. See Table 2 for control variables and the Online Appendix, Section A.2, for variable definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Dep. Var.: LN_PUB*	(1)	(2)	(3)	(4)	(5)	(6)	(7)
TREATED × POST	0.122***	0.145***	0.138***	0.136***	0.130***	0.130***	0.130***
	(0.035)	(0.034)	(0.033)	(0.033)	(0.032)	(0.023)	(0.022)
POST	0.087***	0.084***	0.044***	0.045**	0.052***	0.052***	0.052***
	(0.012)	(0.012)	(0.009)	(0.019)	(0.018)	(0.014)	(0.017)
TREATED	1.502***	-0.098**	-0.044	-0.042	-0.069*	-0.069*	-0.069**
	(0.122)	(0.046)	(0.043)	(0.043)	(0.040)	(0.040)	(0.029)
Controls included	No	No	No	No	Yes	Yes	Yes
Firm fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Event fixed effects	No	No	No	Yes	Yes	Yes	Yes
Clustered by firm	Yes	Yes	Yes	Yes	Yes	No	No
Clustered by event	No	No	No	No	No	Yes	No
Clustered by event × firm	No	No	No	No	No	No	Yes
Number of obs.	43,642	43,642	43,642	43,642	43,642	43,642	43,642
Adjusted R^2	0.042	0.939	0.941	0.941	0.944	0.944	0.944

Table AT6

Regressions of corporate publications on analyst coverage shocks: Alternative sample frames.

This table replicates the baseline difference-in-difference (DiD) regression results from Table 2, column 7, using alternative pre- and posttreatment contrasts. Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms assembled through the sample construction procedure described in Section 4.1. See Table 2 for control variables and the Online Appendix, Section A.2, for variable definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Pre-period	$[-4,-2]$	$[-5,-3]$	$[-3,-1]$	$[-3,-1]$	$[-2,-1]$	$[-4,-1]$	$[-5,-1]$
Post-period	$[+1,+3]$	$[+1,+3]$	$[+2,+4]$	$[+3,+5]$	$[+1,+2]$	$[+1,+4]$	$[+1,+5]$
Dep. Var.: LN_PUB	(1)	(2)	(3)	(4)	(5)	(6)	(7)
TREATED × POST	0.107***	0.129***	0.121***	0.144***	0.080***	0.113***	0.132***
	(0.031)	(0.034)	(0.031)	(0.035)	(0.024)	(0.029)	(0.031)
POST	0.005	0.004	0.002	-0.007	0.011	0.005	0.004
	(0.010)	(0.012)	(0.010)	(0.011)	(0.009)	(0.007)	(0.007)
TREATED	-0.050	-0.066**	-0.055*	-0.066**	-0.027	-0.049	-0.064**
	(0.033)	(0.032)	(0.032)	(0.032)	(0.035)	(0.031)	(0.028)
Controls included	Yes	Yes	Yes	Yes	Yes	Yes	
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes
Number of obs.	146,029	135,814	148,593	142,288	105,663	195,598	230,183
Adjusted R^2	0.901	0.900	0.902	0.901	0.909	0.901	0.899

Table AT7

Regressions of corporate publications on analyst coverage shocks: Alternative R&D firm definitions.

This table replicates the baseline difference-in-difference (DiD) regression results from Table 2, column 7, using alternative definitions for R&D-active firms. Our basic approach follows Arora et al. (2021) and defines R&D firms as those with at least one patent and at least one year of positive R&D expense during our sample period (1997-2014). Column 1 requires R&D firms to have at least one patent. Column 2 requires R&D firms to have at least one year of positive R&D spending. Column 3 requires R&D firms to have at least one scientific publication. Column 4 requires R&D firms to have at least one year of positive R&D spending and at least one scientific publication. Column 5 requires R&D firms to have at least one patent and at least one scientific publication. Column 6 requires R&D firms to have at least one year of positive R&D spending, at least one patent, and at least one scientific publication. Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms assembled through the sample construction procedure described in Section 4.1. See Table 2 for control variables and the Online Appendix, Section A.2, for variable definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

R&D firm?	Pat.	R&D exp.	Pub.	R&D exp. & Pub.	Pat. & Pub.	R&D exp. & Pat. & Pub.
Dep. Var.: LN_PUB	(1)	(2)	(3)	(4)	(5)	(6)
TREATED × POST	0.098***	0.104***	0.083***	0.100***	0.100***	0.115***
	(0.027)	(0.029)	(0.029)	(0.031)	(0.030)	(0.032)
POST	0.009	0.008	0.007	0.009	0.011	0.012
	(0.008)	(0.008)	(0.009)	(0.010)	(0.009)	(0.010)
TREATED	-0.037	-0.056	-0.030	-0.045	-0.030	-0.049
	(0.033)	(0.037)	(0.033)	(0.039)	(0.036)	(0.041)
Controls included	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	153,582	139,904	141,551	116,913	124,406	106,146
Adjusted R^2	0.904	0.904	0.889	0.891	0.890	0.891

Table AT8
Regressions of corporate publications on analyst coverage shocks: Additional publication measures.

This table replicates the baseline difference-in-difference (DiD) regression results from Table 2, column 7, using additional publication measures. Column 1 uses LN_PUB_CIT5, the natural logarithm of (one plus) the total number of publications weighted by the number of citations that each publication receives in a five-year window. Column 2 uses LN_PUB_CIT7, the natural logarithm of (one plus) the total number of publications weighted by the number of citations that each publication receives in a seven-year window. Column 3 uses LN_PUB_JIF, the natural logarithm of (one plus) the total number of publications weighted by the journal impact factor (JIF). Column 4 uses $LN\_PUB\_JIF \geqslant 50\%$, the natural logarithm of (one plus) the total number of publications in top 50% journals ranked by the JIF each year. Column 5 uses $LN\_PUB\_JIF < 50\%$, the natural logarithm of (one plus) the total number of publications in bottom 50% journals ranked by the JIF each year. Column 6 uses LN_PUB_COL, the natural logarithm of (one plus) the total number of publications coauthored with university-based scientists. Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms assembled through the sample construction procedure described in Section 4.1. See Table 2 for control variables and the Online Appendix, Section A.2, for variable definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable	LN PUB CIT5 (1)	LN PUB CIT7 (2)	LN PUB JIF (3)	LN PUB JIF⩾50% (4)	LN PUB JIF<50% (5)	LN PUB COL (6)
TREATED × POST	0.228***	0.235***	0.253***	0.210***	0.105***	0.170***
	(0.052)	(0.054)	(0.033)	(0.030)	(0.031)	(0.029)
POST	0.005	0.009	-0.001	-0.004	0.011	0.006
	(0.018)	(0.019)	(0.008)	(0.007)	(0.010)	(0.008)
TREATED	-0.070	-0.066	-0.161***	-0.136***	-0.053	-0.094**
	(0.073)	(0.076)	(0.049)	(0.042)	(0.038)	(0.038)
Controls included	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	122,429	122,429	122,429	122,429	122,429	122,429
Adjusted R^2	0.824	0.818	0.864	0.850	0.888	0.857