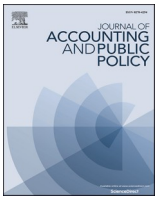
J. Account. Public Policy 45 (2024) 107210

Contents lists available at ScienceDirect

Journal of Accounting and Public Policy

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



Full length article

Narrative R&D disclosure and insider trading profitability: Evidence from China

Wan Huang [a], Qingwen Liang [b,*]

[a] Business School, Sichuan University, Chengdu, China
[b] School of Accounting, Southwestern University of Finance and Economics, Chengdu, China

ARTICLE INFO

JEL Classifications:
G14
G34
M41

Keywords:
Narrative R&D disclosure
Insider trading profitability
Information asymmetry

ABSTRACT

This paper investigates how narrative R&D disclosures affect firms' information asymmetry from the perspective of insider trading profitability. Inconsistent with the traditional information economics theory that information disclosure can alleviate information asymmetry, we find robust evidence that insiders in firms with more narrative R&D disclosures gain significantly greater returns from their stock selling, but this positive relation is attenuated when the readability of R&D-related texts is higher. We examine two potential channels through which more narrative R&D disclosures increase information asymmetry, thus creating opportunities for insiders to trade profitably. One is managers' obfuscation disclosure of narrative R&D information due to proprietary costs, and the other is the high information processing costs associated with the technical nature of R&D information. We also find that good corporate governance mechanisms can weaken the positive association between narrative R&D disclosure and insider trading profitability. In further analysis, we provide evidence that more narrative R&D disclosures indeed exert high information processing costs for investors and provide opportunities for insiders to trade more before bad news arrives. Overall, our research sheds light on the dark side of narrative R&D disclosure on firms' information environment from the perspective of insider trading.

1. Introduction

Technological innovation is key to a country's long-term economic growth and a firm's future performance (e.g., Chan et al., 2001; Eberhart et al., 2004; Kogan et al., 2017). Given the great importance of R&D projects for firms' productivity and future performance, outside stakeholders are especially concerned about firms' R&D-related information. Moreover, estimating the future cashflows and earnings impacts associated with R&D investments can be difficult for outside investors because R&D-intensive firms usually suffer from serious information asymmetry problems due to the risky and experimental nature of R&D activities (Himmelberg and Petersen, 1994; Aboody and Lev, 2000). Owing to its great value in relation to a firms' future performance, R&D-related information disclosures

* Corresponding author at: Postal address: 555, Liutai Avenue, Wenjiang District, Chengdu City, Sichuan Province 611130, China.
E-mail addresses: hw1022@scu.edu.cn (W. Huang), Freya_Liang@outlook.com (Q. Liang).

https://doi.org/10.1016/j.jaccpubpol.2024.107210

0278-4254/© 2024 Elsevier Inc. All rights reserved.

always attract great attention from the capital market. However, the rigorous criteria for R&D investment capitalization set by accounting standards of most countries worldwide further limit the transmission of R&D-related information (da Silva and de Carvalho MoraisCurto, 2013; Huang et al., 2023),[1] leading to additional needs for interpretations of firms' R&D projects. In addition to quantitative R&D disclosures in public firms' financial reports, managers usually voluntarily provide a large number of descriptions and explanations of firms' R&D activities in words (hereafter narrative R&D disclosure), which provide detailed and exhaustive, yet hard to validate, R&D-related information. Although there is an increasing amount of narrative R&D disclosures, until recently, few studies have examined whether more narrative R&D disclosures truly convey incremental decision-relevant information to outside stakeholders and contribute to better information transparency in the capital market. Answering this question is highly important for investors, regulators and firm managers. In this paper, we seek to provide empirical evidence on how narrative R&D disclosures affect firms' information asymmetry from the perspective of insider selling profitability, as previous literature has demonstrated that insider trading profits are related to the degree of information asymmetry between insiders and outside investors (e.g., Kyle, 1985; Baiman and Verrecchia, 1996; Aboody and Lev, 2000; Lakonishok and Lee, 2001; Dai et al., 2015).

Theoretically, there may be two opposite effects of narrative R&D disclosure on insider selling profitability. To some degree, narrative R&D disclosure is useful for helping market participants understand public firms' R&D status by complementing financial information on R&D expenditures and hence providing a more comprehensive picture of firms' R&D endeavors and progress (Jones, 2007; Merkley, 2014). According to traditional information economic theory, as more public information disclosures can improve firms' transparency and reduce the information advantages of privately informed investors (e.g., Diamond and Verrecchia, 1991; Healy and Palepu, 1993; Baiman and Verrecchia, 1996), it can be expected that more narrative R&D disclosures will reduce firms' information asymmetry, prohibiting insiders from selling profitably.

However, more narrative R&D disclosures may facilitate informed insider sales by providing insiders with greater private information advantages when the uniqueness of R&D-related information is taken into account. First, R&D information is related to high proprietary costs, which benefit competitors. Managers have incentives and opportunities to strategically distort or obfuscate textual information disclosures to avoid proprietary costs, leading to obscure and unreliable public narrative disclosures (Merkl-Davies and Brennan, 2007; Kothari et al., 2009; Diouf and Boiral, 2017; Aghamolla and Smith, 2023). In addition, recent empirical literature has documented the information processing costs of public disclosures (e.g., Dedman et al., 2009; Myatt and Wallace, 2012; Blankespoor et al., 2019). These findings suggest that more technical, complicated, and burdensome narrative R&D disclosures can instead prevent outside investors with limited information processing expertise from deriving valuable and useful information from financial reports (Kim and Verrecchia,1994), thereby enhancing the private information advantages of insiders. As a whole, more narrative R&D disclosures do not always contribute to better firm transparency but may lead to a more asymmetric information environment than would exist in their absence. Given the above competing arguments, whether more narrative R&D disclosures increase or decrease insider selling profitability remains an open empirical question.

China provides us with a unique setting in which to test our research question. Although it is quite common for managers to voluntarily conduct narrative R&D disclosures in financial reports,[2] few countries require firms to make comprehensive and extensive narrative R&D disclosures as China does. Since 2001, the China Securities Regulatory Commission (hereafter CSRC) began to require listed companies to disclose information related to R&D activities in their annual reports in accordance with the "No. 2 Standard on the Content and Format of Information Disclosure for Publicly Issued Securities Companies – Content and Format of Annual Reports" (hereafter CFAR). This requirement was amended in subsequent years, and the most significant adjustments were made in 2012. The CSRC requires listed companies to provide thorough information regarding the objectives, progress, and potential impacts on the future performance of existing R&D projects as well as the necessary resources and future R&D investment plans.[3] Huang et al. (2023) showed that there was a significant increase in Chinese firms making narratives of R&D information in their financial reports, indicating that these firms are quite concerned with this regulation policy. Therefore, China's stringent R&D disclosure regime provides us with a larger sample of public firms undertaking ample narrative R&D disclosures to investigate whether more narrative R&D disclosures are beneficial for decreasing information asymmetry. Moreover, China's R&D disclosure regime is unique in that it sets compulsory narrative disclosure requirements for information with high proprietary costs, which allows us to capture managers' incentive to obfuscate narrative disclosures due to proprietary costs. It is difficult to isolate managers' obfuscation behavior under a voluntary disclosure regime where managers do not have to make any disclosures if they perceive proprietary costs to be high.

Our empirical analysis is based on a large sample of Chinese listed firms from 2008 to 2020. Following prior research (Merkley, 2014; Hu et al., 2018), we create a Chinese dictionary of R&D-related keywords from a corpus consisting of Chinese listed firms' annual reports and Chinese government working reports by machine learning methods. Then, we use the textual analysis method to construct two proxies for narrative R&D disclosure: the quantity and the readability of narrative R&D disclosure. While both insider purchases

[1] In accordance with IAS 38, an entity is able to recognize a self-created (at cost) intangible asset if, and only if: 1) It is possible that the future economic benefits attributable to this asset will eventually flow into the enterprise; 2)The cost owning to an asset can be measured reliably; 3)The probability of future economic benefits must be based on reasonable and supportable assumptions about conditions that will exist over the life of the asset; 4) The probability recognition criterion is always considered to be satisfied for intangible assets that are acquired separately or arises from contractual as well as other legal rights.

[2] In an investigation, 18,000 organizations in 40 countries were sampled over a period of five year, i.e., "Reporting of R&D: Disclosure without recognition? (https://www.researchgate.net/publication/371081634)", the researchers found that 53% of the sample did not report any R&D asset or expenditure in financial statements, but meanwhile they did use R&D-related terms in their annual reports.

[3] The detailed requirements of narrative R&D disclosure in China are shown in Appendix A.

and sales can be motivated by private information, the opportunistic motivation of insiders to obtain private benefits from information advantages is mostly demonstrated in stock sales.[4] Therefore, in this paper, we mainly focus on stock sales made by insiders (i.e., senior executives, directors, and supervisors) and their relatives. Referring to prior studies (e.g., Huddart and Ke, 2007; Skaife et al., 2013; Chung et al., 2019), we define the profitability of insider selling as the profits earned after the losses avoided from selling shares, which are the product of three terms: the abnormal return on the stock following the trade, the value of trades and the frequency of trades. We investigate the aggregate profits received by a firm's insiders from all sale trades each year.

Our baseline regression result suggests that insider selling profits are significantly greater for firms with more narrative R&D disclosures, lending support to the hypothesis that more narrative R&D disclosures are not conducive to improving firms' information environment. This result is robust to the inclusion of a variety of firm-level control variables as well as firm and year fixed effects. Moreover, we find that the positive relation between narrative R&D disclosure and insider selling profitability is attenuated when the readability of R&D-related texts is higher, indicating that more readable R&D-related texts can help investors comprehend the information provided by narrative R&D disclosures. We also conduct a series of robustness checks to reinforce our baseline conclusions, such as PSM-DiD analysis, excluding alternative explanations, alternative measures for narrative R&D disclosure and insider selling profitability, and alternative samples. All the empirical results are consistent with our main findings.

In cross-sectional tests, we provide evidence in support of the two hypothesized channels for the positive association between narrative R&D disclosure and insider selling profitability. Our empirical results show that this positive relation is magnified when proprietary costs associated with R&D disclosure are higher, i.e., when firms face a higher degree of product market competition, when firms are high-tech enterprises and when firms belong to high-growth industry. These findings are in accordance with the explanation that managers in firms facing higher proprietary costs are more likely to obfuscate the narrative disclosures of R&D-related information. We also find that the positive relation between narrative R&D disclosure and insider selling profits is weaker when firms receive more attention from sophisticated market participants, such as short sellers, analysts, and media. These results demonstrate that sophisticated market participants can help process and disseminate complex narrative R&D information, thus reducing the information advantage of insiders. Additionally, we find consistent evidence that good corporate governance mechanisms, i.e., more institutional shareholders, better internal control quality and greater board independence, can restrain insiders from trading profitably with the shelter of information asymmetry aggravated by narrative R&D disclosures.

In further analysis, we provide additional evidence that more narrative R&D disclosures indeed decelerate the adjustments of abnormal returns after earnings announcements in the long run, which supports our viewpoint that more narrative R&D disclosures increase investors' information processing costs. Moreover, we document that more narrative R&D disclosures increase the propensity of insiders to trade before negative earnings announcements and result in greater profitability of insider sales.

This paper contributes to the literature in several aspects. First, we expand the burgeoning literature on the substantial effects of narrative R&D disclosures. Prior studies show that more narrative disclosures on R&D activities are useful to market participants through providing incremental information and increasing the transparency of R&D-intensive firms (e.g., Jones, 2007; Merkley, 2014; Bloom et al., 2013). Our study, which takes advantage of China's unique mandatory R&D disclosure regime, complements previous research by documenting the opposite finding. Our results show that more narrative R&D disclosures may be associated with intentional information obfuscation as well as high information processing costs, which can aggravate information asymmetry instead. Moreover, the readability of R&D-related texts can alleviate this negative effect to some extent. These findings advance our understanding of how narrative R&D disclosures affect firms' information environment from the perspective of insider trading, hence providing important implications for other countries where narrative R&D disclosure is not mandated but policymakers are concerned about alleviating the information asymmetry of R&D-intensive companies.

Second, we extend the literature on the determinants of insider trading profits. Prior research has attributed insider trading profitability to their superior information on firms' significant events or long-term prospects (e.g., Aboody and Lev, 2000; Ke et al., 2003; Piotroski and Roulstone, 2005). We add to this line of literature by revealing that narrative R&D disclosures generate an opaque environment for insiders to extract rents from uninformed shareholders. Our research highlights the necessity of introducing more complete and effective rules to regulate firms' narrative R&D disclosures as well as insider trading activities. It is worth noting that our paper differs from that of Aboody and Lev (2000), who document that insider gains in firms conducting R&D activities are substantially greater than insider gains in firms with no R&D activities. In this paper, we focus on narrative R&D disclosures instead of R&D expenditures or the number of patent applications. Firms engaging in R&D activities do not necessarily would like to disclose authentic R&D-related narrative information and vice versa. The theory of our paper contends that the increased information asymmetry comes from deliberate obfuscation as well as the technical or professional traits of narrative R&D disclosures. Therefore, our paper has incremental contributions relative to Aboody and Lev (2000).

Third, our study is related to the growing literature regarding information processing frictions (e.g., Cohen and Lou, 2012; Blankespoor et al., 2019, 2020; Booker et al., 2023). The uniqueness of R&D activities leads to financial reports containing voluminous

[4] In China's capital market, compared with insiders' purchases, the frequency and scale of insiders' sales are extremely larger. For example, during the sample period of this paper, insiders of listed companies conducted a total of 49,165 sale transactions and 23,032 purchase transactions in the secondary market. The market value of sale transactions was up to 19.6 trillion RMB, while the market value of purchase transactions was only 72 billion RMB. In addition, insider purchases in China's capital market are more motivated by the political motivation of stabilizing the market or the market value management motivation of maintaining the firms' stock price, rather than the signaling or market timing motivation in the general sense (Shen et al., 2011; Hu et al., 2020). Additionally, the unreported results suggest that more narrative R&D disclosures cannot increase insiders' purchasing profitability.

jargon and technical words, which increase the difficulty of communicating R&D-related information. Our research extends this line of literature by suggesting that more narrative R&D disclosures might be an important contributor to firms' information processing costs. Due to the sophisticated nature of narrative R&D disclosures, investors are prohibited from acquiring valuable information from public disclosures in an efficient way, further widening the information gap between insiders and investors.

The remainder of this paper is organized as follows. Section 2 relates our study to the existing literature. Section 3 develops the hypotheses. Section 4 presents the data source, sample selection, and variable construction. Section 5 reports the baseline results as well as related robustness and cross-sectional tests. We conduct further analysis in Section 6 and conclude the paper in Section 7.

2. Literature review

2.1. Narrative R&D disclosure

Innovation activities are important for firms' growth, but our understanding of R&D information disclosure is limited, especially for narrative R&D disclosure, which has recently become an emerging research topic. R&D-related information disclosure is quite unique not only because of the risky and opaque nature of innovation activities but also because of its high proprietary costs. In this regard, the narrative disclosure of R&D-related information can have two opposite economic consequences. On the one hand, narrative disclosures on R&D activities may be useful to market participants by providing additional explanations for firms' financial numbers of R&D investments or patent applications, hence increasing the transparency of R&D-intensive firms. With manually collected extensive data on R&D-related disclosures, Jones (2007) finds that managers would like to voluntarily disclose specific information on firms' R&D activities, which is associated with lower analyst forecast errors for both earnings and sales. Merkley (2014) also provides evidence that narrative R&D disclosure significantly affects sell-side analyst behavior, indicating that market participants view narrative R&D disclosure as informative. Nekhili et al. (2016) argue that better-governed firms disclose more R&D narrative information in their annual reports, which leads to low information asymmetry and desirable economic consequences. On the basis of this viewpoint, they find a positive association between narrative R&D disclosure and firms' market value. Saidi and Zaldokas (2021) confirm that increased innovation disclosures help firms switch lenders, resulting in lower costs of debts and facilitating their access to syndicated-loan and public capital markets. Huang et al. (2021) investigate how the enactment of patent protection legislation affects firms' narrative R&D disclosures by adopting a DiD design. They document that the reinforcement of patent protection aggravates information asymmetry by reducing narrative R&D disclosures.

However, proprietary costs may prevent managers from disclosing accurate and precise information, thus reducing the informativeness of R&D-related information disclosure. Entwistle (1999) undertakes a series of interviews with firms' executives, covering questions about effective R&D disclosure management, and finds that firms are reluctant to disclose innovation-related information, which may be improperly used by competitors. Newman and Sansing (1993) find that managers may distort voluntary disclosure to discourage potential new competitors from entering the market, indicating that managers are motivated to provide less useful R&D-related information because of proprietary costs. Consistent with their research, Glaeser and Landsman (2021) also show that firms strategically use disclosures of patents to deter product market competition. Glaeser et al. (2020) demonstrate that managers choose to supply more discretionary R&D disclosures when outside investors perceive managers to have a short horizon. Andreou et al. (2021) propose a managerial rhetoric channel for stock price crash risk, stating that managers exploit narrative R&D disclosures to conceal bad news. Rawson (2022) finds that overconfident CEOs are willing to provide more narrative R&D disclosures because they are likely to perceive proprietary costs to be lower.

Collectively, considering the proprietary cost traits of R&D-related information, the existing literature provides mixed evidence on whether more narrative R&D disclosures can improve firms' information environment. We contribute to the literature on this topic by examining the association between narrative R&D disclosure and insider selling profitability, which is directly related to the degree of information asymmetry between insiders and outside investors.

2.2. Information asymmetry and insider trading profitability

Although some insider trades may be attributed to their liquidity and portfolio rebalancing objectives, a large part of insider trades are driven by insiders' desire to profit from their superior information over other market participants about firm prospects (e.g., Kyle, 1985; Lang and Lundholm, 1993; Huddart and Ke, 2007). Prior research has long argued that information asymmetry enables insiders to trade profitably by increasing their information advantages. Aboody and Lev (2000) provide evidence that R&D activities increase the information asymmetry between insiders and investors, thus allowing insiders from firms with high R&D expenditures to reap more profits from their trades than insiders from firms without any R&D investments. Frankel and Li (2004) find that financial statement informativeness is negatively associated with the frequency of insider trading. Gu and Li (2007) indicate that managers are privy to good news shortly before disclosure. Bowen et al. (2018) suggest that corporate insiders may trade on private information acquired around private in-house meetings. Chung et al. (2019) and Jia and Gao (2021) both provide evidence that insiders opportunistically exploit the corporate opacity arising from aggressive tax activities to profit from trading in their own firms' shares. Arif et al. (2022) examine insider trading in a tight window around the audit report date when insiders know the audit findings but outside market participants do not. Their findings provide novel evidence that insiders appear to exploit private information about the audit process for opportunistic gains. In addition, some research has demonstrated that information collection activities by sophisticated capital market participants can reduce information asymmetry, thus limiting insiders' ability to trade profitably. For example, Dai et al. (2015) document that media reports can reduce insiders' information advantages by disseminating news on prior insider trades.

Chen et al. (2022) provide evidence that short sellers significantly reduce insider trading profitability through their effect on the market's information discovery.

Overall, the collective evidence on this strand of literature demonstrates that insiders can earn great trading profits by exploiting their private information priorities when there is greater information asymmetry between firms and outsider investors. However, there is still a wider range of research space about the factors that add to information asymmetry and insiders' private information advantages with the advancement of textual data analysis.

3. Hypothesis development

Numerous studies have shown that information asymmetry can enhance insiders' ability to trade profitably over outside market participants (e.g., Aboody and Lev, 2000; Huddart and Ke, 2007; Chung et al., 2019). Factors resulting in greater information asymmetry between insiders and outsiders increase insiders' private information advantages. Theoretically, there may be two opposite effects of more narrative R&D disclosures on insider trading profitability given their unique nature. On the one hand, traditional information economics theory insists that increasing public information disclosures can improve firms' transparency and reduce the information advantage of informed investors (e.g., Diamond and Verrecchia, 1991; Healy and Palepu, 1993; Baiman and Verrecchia, 1996). Consistent with this view, prior studies show that narrative R&D disclosures convey incremental information and can improve the information environment of listed firms (Jones, 2007; Merkley, 2014; Saidi and Zaldokas, 2021; Huang et al., 2023). Under these circumstances, more narrative R&D disclosures will enable outside investors to obtain access to more R&D-related information of firms, such as the progress of present projects, changes in R&D expenditures or future plans for innovations. Therefore, compared to firms with less narrative R&D disclosures, firms with more detailed narrative R&D disclosures can decrease insiders' private information advantages and thus reduce their selling profitability.

However, the substantial effects of narrative R&D disclosure can be different after considering the particularity of R&D-related information. R&D-related information disclosures usually lead to high proprietary costs, which will benefit firms' rivals. Managers' decisions regarding how to make narrative R&D disclosures are expected to be affected by the conflicting objectives of reducing information asymmetry and avoiding proprietary costs (e.g., Dye,1985; Verrecchia, 2001; Healy and Palepu, 2001). In the scenario of this paper where the requirements and supervision of narrative R&D disclosure are compulsory and stringent, we argue that more narrative R&D disclosures can increase firms' information asymmetry and enhance insiders' selling profitability through two potential channels.

First, managers are compelled to communicate more detailed and thorough information about firms' R&D activities in annual reports even though they may not be willing to do so. Considering that R&D-related information is strongly associated with significant proprietary costs, which can incur fierce competition and undercut incumbents' profits, most public firms prefer to keep R&D-relate information secret. As a result, managers may strategically make voluminous invalid narratives to ensure their compliance with mandatory requirements. They may maintain the same disclosure rate as their counterparts but meanwhile use obscure words or perplexing sentences to make R&D-related information more difficult to understand.[5] A large body of empirical research demonstrates that managers who are concerned about proprietary costs have incentives to obfuscate disclosures of sensitive information or disclose uninformed information despite potential rigorous regulations discouraging competitors from segmenting industry profits (e.g., Newman and Sansing, 1993; Gigler,1994; Ellis et al., 2012; Rawson, 2022). Additionally, compared to quantitative R&D investment numbers, narrative R&D disclosure is more susceptible to managers' manipulations because of its subjectivity and lack of verifiability (Huang et al., 2014). Managers have great discretions over the degree of specific details, textual vagueness, complexity of words, metaphors and description styles of narrative R&D disclosures (Davis and Tama-Sweet, 2012; Loughran and McDonald, 2014; Henry and Leone, 2016). Prior studies have confirmed that unstructured textual narratives exhibit more irregularities and ambiguities and are prone to managerial opportunism (e.g., Merkl-Davies and Brennan, 2007; Li, 2010a). Therefore, managers' obfuscation disclosures of narrative R&D information due to the existence of proprietary costs can hinder outside investors from effectively evaluating the prospects and risks of R&D projects and further damage their judgments on firms' future cash flow and earnings sustainability. The information gap between insiders and outside investors will facilitate insiders to sell their firms' stocks profitably when there is negative news.

In addition, the excessively irrelevant, redundant and technical words or sentences in narrative R&D disclosures can also lead to a more asymmetric information environment than would exist in their absence (Zhang, 2001).[6] Public firms' stakeholders are usually exposed to high processing costs, which prevent them from absorbing all the valuable information from public disclosures effectively (e.g., Hirshleifer and Teoh, 2003; Myatt and Wallace, 2012; Ertugrul et al., 2017; Dyer et al., 2017; Blankespoor et al., 2020). Distinct from regular business activities such as manufacturing, material purchasing or product selling, the exploratory, uncertain and experimental nature of R&D activities can make communicating R&D-related information quite difficult (Chan et al., 2001; Fung, 2006; Zhang, 2015), impairing firms' information environment instead (e.g., Dvir et al., 1993; Dedman et al., 2009; Lawrence, 2013; Lim et al., 2018). Aboody and Lev (2000) substantiate that while R&D projects are likely to proliferate firm performance and future

[5] In the Online appendix, we provide some anecdotal evidence to illustrate that Chinese listed firm makes numerous narrative R&D disclosures with ambiguous and uncertainty sentences which are useless for investors to obtain incremental information.

[6] In the Online appendix, we provide some anecdotal evidence to illustrate that Chinese listed firms make numerous parallel sentences, long noun expressions and technical words in narrative R&D disclosures, which made it difficult for non-professional investors to extract key information at first.

cash flow, R&D activities are often firm-specific and idiosyncratic, and the information reports for these types of activities are unreliable to a greater degree. It is difficult for outside investors to verify the information contained in narrative R&D disclosures due to their limited time and cognitive abilities. Moreover, more complicated narrative R&D information will distract investors from other valuable information, which further increases the information asymmetry between insiders and outside investors. In particular, the information processing costs pertinent to complex narrative R&D disclosures could be particularly serious in China's capital market where unsophisticated individual investors dominate.[7] Therefore, insiders, with their comprehensive and detailed knowledge of firms' R&D activities and other information, can earn excess profits by selling down their stock holdings in advance of future bad news.[8]

According to the above analysis, we expect that narrative R&D disclosures are more likely to increase information asymmetry and create more opportunities for insiders to sell firms' stocks profitably. Our first hypothesis is formulated as follows:

H1: *The quantity of narrative R&D disclosure is positively associated with insider selling profitability.*

We further explore the role of another critical trait of narrative R&D disclosures, textual readability, which has recently attracted increasing attention from scholars and regulators. We contend that the readability of R&D-related texts has a mediating effect on the relation between the quantity of narrative R&D disclosure and the profitability of insider sales.

First, narrative R&D disclosures with higher readability indicate that managers have less incentive to obfuscate the narrative disclosures of R&D-related information. In other words, the high readability of R&D-related texts acts as a signal of managers' proactive attitudes toward communicating authentic information about their R&D activities. Therefore, outside investors are more likely to obtain access to credible R&D-related information from disclosing firms through more readable narrative R&D disclosures. In addition, abundant prior studies document that annual report readability and other textual properties of financial disclosures can strongly impact investors' ease of processing information as well as firms' efficiency in communicating value-relevant information (e.g., Miller, 2010; Loughran and McDonald, 2011, 2014; Lee, 2012). The readability of narrative R&D disclosures may interfere with investors' ability to comprehend the information of firms' R&D activities. In line with the above literature, we argue that R&D-related texts that are difficult to read reduce their usefulness in conveying new information and impair an investor's ability to process and extract useful information related to R&D activities in a more efficient way, further increasing insiders' information advantages. In contrast, narrative R&D disclosures with greater readability help capital market participants incorporate key information, thereby reducing insiders' selling profits with information advantages. Therefore, we propose the second hypothesis as follows:

H2*: The positive relation between narrative R&D disclosure and insider selling profitability is attenuated when the readability of narrative R&D disclosure is greater.*

4. Research design

4.1. Sample selection and data source

We examine the association between narrative R&D disclosure and insider selling profitability by considering a sample of listed firms on the Shanghai and Shenzhen Stock Exchanges in China covering the period from 2008 to 2020. We obtain insider trading information from the China Stock Market & Accounting Research (CSMAR) database and focus only on sales of common shares by top managers (i.e., senior executives, directors, supervisors, and their relatives). We download the firms' annual reports from the Juchao Information Collection Website[9] via a crawler program and China's government working reports from the official website of the State Council.[10] To facilitate subsequent processing, we use the *Solid Converter PDF* to convert the PDF-version annual reports to HTML-version ones and delete all the images in every report. We use machine learning and textual analysis techniques to construct proxies for narrative R&D disclosures. To capture information-driven trading activities, we follow prior literature (e.g., Jagolinzer et al., 2011; Gao et al., 2014) and include only open-market insider selling transactions. For each transaction, we require at least 1,000 shares to be traded at any particular time to reduce the effect of insiders' liquidity motives (Lakonishok and Lee, 2001) and exclude observations with missing information on the transaction date, transaction price and transaction shares. In addition, we use various filters to clean up our sample. We exclude (1) firms in the financial sector because of their unique disclosure requirements and accounting rules; (2) firms with more liabilities than assets; (3) firms with "special treatment" ("ST") status or with "particular transfer" ("PT") status; and (4) firms without sufficient data for calculating control variables. Overall, these filters enable us to construct a final sample consisting of 27,023 firm-year observations for 3,577 nonfinancial firms, including 49,165 sale transactions. All the data used in this paper are mainly from the CSMAR database and the Chinese Research Data Services (CNRDS) database. We winsorize all continuous variables at the 1st and 99th percentile levels to mitigate the possible influence of outliers. The sample selection criteria are shown in Appendix B.

[7] According to data released by the China Securities Depository and Clearing Corporation (CSDC), by the end of 2022, the number of investors in China's capital market exceeded 2 trillion, of which individual investors account for 99%.

[8] We find some cases showing that insiders sell their stocks just prior to the release of negative news about firms' R&D projects. For example, Beijing Konruns Pharmaceutical Co., Ltd. (603590.SH) issued an announcement on April 22, 2020, claiming that they decided to terminate a new drug project which had been developing for many years, indicating the R&D investment on this new drug had failed. Following the announcement, Konruns' share price fell 19% in five consecutive trading days. Meanwhile, we find that in the 30 days before the announcement, Konruns' insiders intensively sell their stocks up to 122.63 million RMB. However, there were no insider sales in this company during the same period last year.

[9] Link: https://www.cninfo.com.cn/new/fulltextSearch?notautosubmit=&keyWord=833637.

[10] Link: https://www.gov.cn/zhengce/zhengcewenjianku/index.htm.

4.2. Measures of narrative R&D disclosure

Referring to Merkley (2014) and Hu et al. (2018), we create a vocabulary of R&D-related keywords in the Chinese context from a corpus through a word embedding approach[11] called word2vec.[12] Then, we use the textual analysis method to construct proxies for narrative R&D disclosures. First, we filtered out six seed words for R&D-related information by manually reading firms' annual reports and following previous studies (e.g., Merkley, 2014; James and Shaver, 2016; Hu et al., 2018; Rawson, 2022).[13] Afterward, we randomly selected 3,000 copies of firms' annual reports and China's government work reports from 2007 to 2020 as the corpus, applying the continuous bag of words (CBOW)[14] algorithm to produce word embeddings. Then, we calculate the word similarity in the word embeddings to expand the seed keyword list. After doing so, we obtained a Chinese R&D-related dictionary containing 395 words or phrases.

Next, we construct two proxies for narrative R&D disclosure, which are the quantity (*RDText*) and the readability (*ReadaV*) of narrative R&D disclosure. We parse the Chinese words or phrases in annual reports using "jieba" in Python[15] and drop the stop words. We use the R&D dictionary to identify and count R&D-related disclosures. We measure *RDText* as the number of R&D-related keywords or phrases divided by the total words of every annual report. To avoid the magnitude of *RDText* from being too small, we multiply it by 1,000. Since we are mainly concerned about the readability of R&D-related texts in this paper rather than the readability of the whole annual report, identifying R&D-related texts is required first. Specifically, R&D-related text is identified via the keyword searching method with Python. We searched for R&D-related keywords in annual reports with Python codes. Then, we extracted 100 words before and after the R&D keywords as R&D-related texts.[16] Finally, we use the total common words in R&D-related texts divided by the total words in the R&D-related texts to measure readability (*ReadaV*). According to the study of Meng et al. (2017), the common word list is from the "General standardized Chinese Characters", which was formulated by the State Language Commission.

4.3. Measures of insider selling profitability

Following prior studies (e.g., Huddart and Ke, 2007; Skaife et al., 2013), we define insider selling profits as the losses avoided from selling company shares. The profit from insider trading over a particular period is determined by (i) the stock returns after each transaction, (ii) the dollar value of each transaction, and (iii) the frequency of these transactions. To construct our measure, we first aggregate all trading transactions by insiders of the same firm on the same day and treat multiple transactions made on the same firm-day as a single transaction. We then refer to existing research (Ke et al., 2003; Chung et al., 2019) and compute the three-month, six-month and nine-month buy-and-hold abnormal returns for the period beginning one day after the transaction date. The gain from sales is computed by multiplying the abnormal return by the value sold and then multiplying by −1, so that losses avoided on stock sales are interpreted the same as gains. Finally, we aggregate individual selling trades at the firm-year level to obtain an aggregate profitability measure of all insider selling transactions during the fiscal year. We construct the following equation:

$$SellProfit_{it} = \sum_{j=1}^{n}(-1) \times SellBHAR_{ijt} \times SellValue_{ijt} \quad (1)$$

where *SellProfit* is the aggregate profitability of all insider sales transactions in firm *i* during fiscal year *t*. *SellBHAR* equals the buy-and-hold abnormal return of specific month periods (including three months, six months, nine months), which is computed for the period starting one day after transaction date *j*. We calculate the buy-and-hold abnormal return by the market adjustment method. *SellValue* equals the total value of shares sold by all insiders on day *j*, and *n* is the total number of firm-days with insider sales transactions during the firm-year. This measure is multiplied by −1 so that losses avoided on sales can be interpreted as gains. We construct three measures, *SellProfit1*, *SellProfit2*, and *SellProfit3*, which are computed with three-month (six-month, nine-month) buy-

[11] A word embedding format generally tries to map a word using a dictionary to a vector. In this approach, words and documents are represented in the form of numeric vectors allowing similar words to have similar vector representations. The extracted features are fed into a machine learning model to work with text data and preserve the semantic and syntactic information.

[12] Word2Vec is a word embedding method developed by Google in 2013.

[13] Merkley (2014) and James and Shaver (2016) construct R&D dictionary in English context. Therefore, we find their Chinese counterparts using Google translate.

[14] Word2vec contains two groups of word embedding methods: continuous bag of words and skip-grams. CBOW includes inputs such as words or phrases that are taken by the neural network model. It predicts the targeted word that closely relates to the context of different words fed as input. Once trained, the CBOW model generates numerical vectors, which capture the semantics of words in a continuous vector space and can be used in various natural language processing (NLP) tasks.

[15] It's an open-source package for Chinese text segmentation.

[16] We choose a total of 200 words around R&D-related keywords based on our manual reading of 500 copies of annual reports, suggestions from the financial staff of listed firms and experience judgment. The R&D-related texts vary in their lengths, and the number 200 is just an average estimation. Additionally, when we choose 300 words and 400 words around R&D keywords as the R&D-related texts to calculate the readability, the empirical results still support our Hypothesis 2, which are tabulated in Table A1 of Online appendix.

and-hold abnormal returns respectively, and we take the natural logarithm of insider selling profit.[17] Following Huddart and Ke (2007) and Skaife et al. (2013), we set *SellProfit* equal to 0 in firm-years with no reported insider sales transactions in our sample.[18]

4.4. Regression models

To test Hypothesis 1 that more narrative R&D disclosures enhance insiders' ability to profit from their shares selling, we specify the regression model as follows:

$$SellProfit_{i,t} = \alpha_0 + \alpha_1 \times RDText_{i,t-1} + \alpha \times Controls_{i,t} + \sum Firm + \sum Year + \varepsilon_{i,t} \quad (2)$$

where $SellProfit_{i,t}$ is the measure of insider selling profit for firm i in year t (see Subsection 4.3) and $RDText_{i,t-1}$ is the proxy for the narrative R&D disclosure of firm i in year t−1 (see Subsection 4.2). Since the annual report of a firm in fiscal year t is generally published in the next year, we use the *RDText* of the firm in year t−1 in the regression model. We select an extensive set of control variables that are associated with insider selling profitability based on prior literature (e.g., Huddart and Ke, 2007; Chung et al., 2019; Jia and Gao, 2021). We control for firm size using the market value of equity (*SizeMV*), leverage (*Lev*), return on total assets (*ROA*), sales growth (*Growth*), property rights (*SOE*), firm age (*Age*), Big 4 auditing firm (*Big4*), the independent director ratio (*Indr*), board size (*Board*), return volatility (*RetVol*), accrual earnings management (*Disacc*), the median absolute abnormal return over past earnings announcements (*MAGAR*),[19] and financial statement informativeness (*FSINFORM*).[20] In addition, we control for R&D expenditures (*RD*) and the number of patents (*Patent*), which are documented to be associated with the firm's information environment (Aboody and Lev, 2000) and are also directly associated with the firm's narrative R&D disclosure, to capture the incremental effect of narrative R&D disclosure on insider trading profitability. We also include firm fixed effects (*Firm*) and year fixed effects (*Year*). Finally, we use firm clustered standard errors to alleviate cross-sectional correlations in the data. All variable definitions are available in Appendix C.

To test Hypothesis 2, we define *Reada* as a dummy variable that equals to 1 if the readability of R&D-related texts (*ReadaV)* is higher than the sample median and 0 otherwise. We modify Eq. (2) to include *Reada* and the interaction term between *RDText* and *Reada*.

4.5. Descriptive statistics and correlations

Panel A of Table 1 displays the distribution of narrative R&D disclosure and insider sales transactions across sample years. We present the mean and median of narrative R&D disclosures (*RDText*) from 2008 to 2020. It is obvious that the amount of textual R&D information disclosed by listed firms is increasing annually, which is in line with China's regulatory policy. With respect to insider sales transactions, the number of sales transactions, the number of insiders who have made sale trades and the number of firms involved in insider selling all show an increasing trend, indicating that insider trading is still an important problem for listed companies.[21]

Panel B of Table 1 presents the results of the univariate analysis of insider sales in firms with more narrative R&D disclosure (*High_RDText*) versus firms with less narrative R&D disclosure (*Low_RDText)*. We compare the mean difference in insider sales frequency, the share number of insider sales, the share value of insider sales and insider sales profit between these two groups. The trading profits are calculated based on various windows, i.e., three months, six months, and nine months. As observed from this table, the mean value of insider sale transactions and insider sale profitability are generally greater for *High_RDText* firms, and all differences between the two groups are almost statistically significant at the 1 % level. Taking *Sales Profit3* as an example, the mean sales profit in *Low_RDText* firms is ¥263,310, while the mean sales profit in *High_RDText* firms is ¥483,032; this difference has economic significance, implying that insiders in *High_RDText* firms can make more profits from their shares selling.

Panel C of Table 1 reports the descriptive statistics of the main variables used in this study. The mean value of *RDText* is 3.973, indicating that on average, R&D-related words account for 3.973 ‰ of the total number of words in the annual report. The *RDText*

[17] We utilize the ratio of insider selling profit to the market value of equity at the end of year t−1 in robustness test.

[18] Since insiders are unlikely to trade on their private information if they expect the sales transactions to be unprofitable, and therefore, they refrain from trading if they do not possess superior information (Frankel and Li, 2004). In the robustness test, we re-examine an alternative sample excluding firm-years with no reported insider sales transactions.

[19] Huddart and Ke (2007) document evidence that market reactions to earnings announcements are strongly related to trading profits, thus we control for cumulative abnormal returns around quarterly earnings announcements over the past 5 years.

[20] Prior literature suggests that financial statement informativeness affects the insider trades' predictive ability for future stock returns (Frankel and Li, 2004; Skaife et al., 2013).

[21] Notably, there are large jumps for insider sales in 2013 and 2019 which may be caused by increases in the number of firms as well as the lock-up expiration of stocks. In 2013, insiders of ChiNext companies (firms listed on the Growth Enterprise Market) were the main force of insider sales. The market value of insider sales of ChiNext listed companies increased by 274% over the previous year. This phenomenon is closely related to the construction of China's capital market. In October 2009, China launched the Growth Enterprise Market, which aims to provide a direct financing platform for start-ups. However, the ChiNext market faces various challenges due to its relatively immature issuance, auditing, and regulation systems, such as the "three-highs" phenomenon observed in new issues. All these problems are closely linked to the overvaluation of ChiNext IPO firms, which eventually leads to performance reversal in subsequent years. As a result, insiders of ChiNext companies overvalued during the IPO period are more inclined to reduce their shareholdings after the IPO lock-up period expires in two to three years. Similarly, in 2019, numerous shares restricted by IPO or private placements were also unlocked collectively. After excluding the sample observations in the two years, the unreported results still show that our main conclusion remains unchanged.

Table 1

Descriptive statistics. This table presents the sample distribution and descriptive statistics for the main variables used in this paper. Panel A reports the distribution of narrative R&D disclosure and insider sales transactions across sample years. Panel B reports the summary statistics for insider sales transactions partitioned by narrative R&D disclosure. We divided narrative R&D disclosure into two groups based on the mean value of the whole sample, i.e., the *High_RDText* group and *Low_RDText* group. Panel C presents descriptive statistics for the variables used in the main tests. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively. The variables are defined in Appendix C. All the continuous variables are winsorized at the 1% and 99% levels.

Panel A: Distribution of narrative R&D disclosure and insider sales transactions across years

Year	Narrative R&D disclosure		Sales transactions		
	Mean	Median	No. of sales	No. of insiders	No. of firms
2008	2.503	1.560	969	615	322
2009	2.685	1.722	2,201	1,030	432
2010	2.896	1.887	2,073	1,005	394
2011	3.305	2.187	2,144	1,034	422
2012	3.668	2.328	2,157	1,081	473
2013	3.917	2.706	4,067	2,040	687
2014	3.882	2.691	4,197	2,565	854
2015	3.682	2.627	3,934	2,659	969
2016	4.039	3.024	2,975	2,100	883
2017	4.331	3.268	2,689	1,624	759
2018	4.478	3.461	3,936	1,596	766
2019	4.648	3.611	8,665	3,015	1,143
2020	4.713	3.606	9,158	3,083	1,111

Panel B: Descriptive statistics partitioned by narrative R&D disclosure

Variable	*High_RDText*		*Low_RDText*		Mean-Diff	P-value
	Obs.	Mean	Obs.	Mean		
RDText	9,442	7.231	17,581	2.224	5.007	0.000***
ReadaV	8,479	7.346	15,992	8.567	−1.221	0.000***
Reada	8,479	0.438	15,992	0.533	−0.095	0.000***
Sales Frequency	9,442	2.415	17,581	1.317	1.098	0.000***
Sales Share	9,442	504,758	17,581	261,112	243,646	0.000***
Sales Value¥	9,442	7,952,702	17,581	4,011,436	3,941,266	0.000***
Sales Profit1¥	9,442	79,015	17,581	30,691	48,324	0.042**
Sales Profit2¥	9,442	328,568	17,581	175,222	152,246	0.000***
Sales Profit3¥	9,442	483,032	17,581	263,310	219,722	0.000***

Panel C: Descriptive statistics for the full sample

Variable	Obs.	Mean	Min	P25	P50	P75	Max	Std
RDText	27,023	3.973	0.342	1.557	2.856	5.239	15.580	3.354
ReadaV	24,471	8.144	0.000	5.764	7.155	9.245	40.840	3.985
Reada	24,471	0.050	0.000	0.000	0.000	1.000	1.000	0.500
SellProfit1	27,023	0.688	−16.290	0.000	0.000	0.000	16.260	6.772
SellProfit2	27,023	0.908	−16.450	0.000	0.000	0.000	16.790	6.941
SellProfit3	27,023	1.043	−16.670	0.000	0.000	0.000	17.050	7.043
SizeMV	27,023	15.590	13.750	14.890	15.460	16.150	18.500	0.964
Lev	27,023	0.430	0.050	0.270	0.426	0.582	0.893	0.202
ROA	27,023	0.035	−0.285	0.013	0.035	0.065	0.194	0.065
Growth	27,023	0.162	−0.570	−0.029	0.102	0.259	2.573	0.400
SOE	27,023	0.399	0.000	0.000	0.000	1.000	1.000	0.490
Age	27,023	2.155	0.693	1.609	2.303	2.773	3.296	0.749
Big4	27,023	0.055	0.000	0.000	0.000	0.000	1.000	0.228
Indr	27,023	0.380	0.250	0.333	0.364	0.429	0.600	0.071
Board	27,023	2.384	1.792	2.303	2.398	2.485	2.944	0.225
Retools	27,023	0.066	0.027	0.048	0.060	0.077	0.223	0.026
Disacc	27,023	0.010	−0.318	−0.036	0.011	0.058	0.302	0.094
MAGAR	27,023	0.030	0.008	0.021	0.028	0.037	0.079	0.013
FSINFORM	27,023	0.060	−0.442	−0.051	0.004	0.129	0.913	0.172
RD	27,023	13.810	0.000	14.710	17.320	18.390	21.180	7.515
Patent	27,023	1.551	0.000	0.000	1.386	2.773	6.014	1.627

values are volatile, ranging from 0.342 to 15.580, with a standard deviation of 3.354 in our sample. These statistics indicate that there are great differences in narrative R&D disclosures among listed companies in China. The measures of insider sales profit (i.e., *SellProfit1*, *SellProfit2* and *SellProfit3*) have means of 0.688, 0.908, and 1.043, respectively, indicating that insiders can generally gain abnormal returns from selling company shares, which leaves it an important topic for restraining insiders' opportunistic behaviors. In terms of the other control variables, the sample firms are relatively large (mean of *SizeMV* = 15.590), have heavy debt (mean of *Lev* = 0.430), make a decent profit (mean of *ROA* = 0.035), and post modest growth (mean of *Growth* = 0.162). On average, 39.9 % of the sample firms are state-owned, only 5.5 % of the sample firms are audited by the Big Four auditing firms, and the percentage of

independent directors is 38.0 %. In addition, there are, on average, approximately 10.85 (=$e^{2.384}$) directors on a firm's board, ¥994,504 (=$e^{13.810}$) for each firm spent on R&D activities per year, and 4.72 (=$e^{1.551}$) for patent applications owned per firm.

Table 2 presents pairwise correlations between the narrative R&D disclosure measure, insider trading profit variables, and control variables used in the main regression model. Both the Pearson and Spearman correlations between *RDText* and all three measures of insider sales profit are significantly positive, which provides preliminary evidence for our Hypothesis H1, indicating that more narrative R&D disclosures may enhance insiders' ability to profit from their shares selling. In addition, the three insider trading profit proxies are positively correlated with each other, and the correlations are all above 0.50, suggesting that these measures capture common dimensions of insider sales profit. Notably, some of the correlations between control variables are slightly high. Thus, we calculate the values of the variance expansion factor (VIF) of each variable used in the main regression analysis and find that they are all less than 5.0, indicating that there is no serious multicollinearity concern in this paper.

5. Empirical results

5.1. Baseline regressions

5.1.1. Test of Hypothesis 1

Table 3 reports the baseline regression results. In Columns (1) to (3), where only *RDText* is included and *SellProfit1*, *SellProfit2* and *SellProfit3* are the dependent variables, the coefficients on *RDText* are all positive and statistically significant at the 1 % level. In Columns (4) to (6), where all control variables and firm and year fixed effects are included, the coefficients on *RDText* are 0.066 (t = 1.78), 0.087 (t = 2.33), and 0.116 (t = 2.90), indicating that insiders in firms with more narrative R&D disclosures earn higher selling profits than insiders in firms with less narrative R&D disclosures. These results are also economically significant. For instance, the coefficient of 0.066 (0.087, 0.116) on *RDText* implies that a one-standard deviation increase in narrative R&D disclosure is associated with a 32.18 % (32.14 %, 37.30 %) increase in insider sales profitability relative to the mean, respectively,[22] suggesting that the economic impact of narrative R&D disclosure on insider trading gains is remarkable. These findings support Hypothesis 1 of this paper.

With regard to control variables, our results generally conform to those of previous research. Specifically, we find that the coefficient on firm size (*SizeMV*) is significantly negative, consistent with the finding that insiders from small firms profit more than insiders from large firms (Lakonishok and Lee, 2001). We also find that insider selling profitability is negatively correlated with return on assets (*ROA*) and past sales growth (*Growth*), indicating that insider selling profitability is weakened in firms with better performance and greater growth. The coefficients of *MAGAR* are significantly positive, consistent with the interpretation of this proxy as a measure of information asymmetry (Huddart and Ke, 2007). In addition, we find that insider selling profit is positively correlated with discretionary accruals, which is consistent with the findings of prior research (Park and Park, 2004) that earnings management is regarded as a useful tool for insiders to gain abnormal returns from selling company shares.

5.1.2. Test of Hypothesis 2

Table 4 presents the empirical results for Hypothesis 2. We find that the coefficients on *RDText* × *Reada* are −0.057, −0.050, and −0.025, respectively, and that almost all the coefficients are significant. This result is consistent with our prediction that the higher readability of narrative R&D disclosure can to some extent reduce the information processing costs for investors and alleviate the information asymmetry caused by professional and complex R&D activities, hence weakening the positive association between narrative R&D disclosure and insider selling profitability. These findings support Hypothesis 2 of this paper and indicate the great importance of readability for narrative R&D disclosure.

5.2. Robustness checks

5.2.1. Changes in R&D information disclosure requirements

In September 2012, the CSRC made significant revisions to the CFAR (hereafter CFAR2012), requiring listed companies not only to fully disclose the objectives, progress, potential impacts on future performance and other relevant information on existing R&D projects but also the resources and their future plans regarding R&D activities. The release of the CFAR2012 would increase firms' narrative R&D disclosures, but this disclosure regime change was not affected by firm-level insider trading profits. Therefore, we predict that the positive association between narrative R&D disclosure and insider selling profitability will be accentuated for treatment firms after the enactment of the CFAR2012, which can strengthen the identification of a causal relationship of our main results. Since the disclosure regime change was made at the country level and applied to all the firms, we do not have any natural treatment or control groups for our analysis. However, this disclosure change does not affect all firms in the same manner; firms with more innovation activities before the reform are more likely to be affected by the disclosure change than firms with less innovation activities. Thus, we can use this difference to construct our treatment and control groups. Notably, this identification strategy is consistent with the prior literature (Vig, 2013; Campelo and Larrain, 2016). Therefore, we exploit the pretreatment cross-sectional variation in the amount of R&D expenditure to identify those companies that are subject to greater disclosure regime change as

[22] The economic significance is computed as 0.066 (coefficient on *RDText* when *SellProfit1*is the dependent variable) ×3.354 (the sample standard deviation of *RDText*) ÷0.688 (the sample mean of *SellProfit1*)=32.18 percent. The comparative statics for *SellProfit2* and *SellProfit3* are computed analogously.

Table 2
Correlation matrix. This table reports the Pearson (Spearman) correlation between the variables used in the regression analysis on the upper (lower) diagonal. Correlation coefficients that are significant at the 5% or better level are in bold. The detailed definitions of the variables are provided in Appendix C.

		1	2	3	4	5	6	7	8	9	10
1	*RDText*		**−0.156**	**0.046**	**0.054**	**0.056**	**0.043**	**−0.225**	**0.088**	**0.093**	**−0.290**
2	*ReadaV*	**−0.155**	1	0.002	0.003	−0.006	**−0.161**	**−0.050**	**0.025**	0.007	−0.060
3	*SellProfit1*	**0.035**	−0.009		**0.612**	**0.510**	**−0.041**	**−0.029**	**−0.054**	**−0.066**	**−0.049**
4	*SellProfit2*	**0.039**	−0.004	**0.630**		**0.706**	**−0.031**	**−0.027**	**−0.055**	**−0.062**	**−0.062**
5	*SellProfit3*	**0.037**	−0.011	**0.525**	**0.725**		−0.011	**−0.023**	**−0.051**	**−0.065**	**−0.068**
6	*SizeMV*	**0.028**	**−0.093**	**−0.053**	**−0.044**	**−0.025**		**0.066**	**0.274**	**0.146**	**0.098**
7	*Lev*	**−0.194**	**0.014**	**−0.025**	**−0.022**	**−0.017**	**0.066**		**−0.421**	−0.002	**0.275**
8	*ROA*	**0.018**	**0.013**	**−0.068**	**−0.071**	**−0.069**	**0.270**	**−0.365**		**0.334**	**−0.142**
9	*Growth*	**0.045**	−0.005	**−0.050**	**−0.048**	**−0.049**	**0.132**	**0.018**	**0.239**		**−0.081**
10	*SOE*	**−0.234**	0.008	**−0.040**	**−0.054**	**−0.059**	**0.110**	**0.276**	**−0.072**	**−0.063**	
11	*Age*	**−0.223**	**−0.048**	**−0.032**	**−0.043**	**−0.037**	**0.171**	**0.322**	**−0.182**	**−0.077**	**0.425**
12	*Big4*	**−0.109**	**0.080**	**−0.025**	**−0.030**	**−0.032**	**0.292**	**0.097**	**0.052**	−0.005	**0.124**
13	*Indr*	**0.073**	**−0.031**	0.000	0.010	**0.017**	**0.040**	**−0.062**	**0.034**	0.001	**−0.145**
14	*Board*	**−0.086**	**0.014**	0.005	−0.000	0.005	**0.152**	**0.147**	**−0.066**	−0.002	**0.238**
15	*RetVol*	**0.059**	**0.027**	0.005	**0.015**	**0.035**	**−0.026**	0.001	**−0.120**	**0.035**	**−0.083**
16	*Disacc*	**−0.036**	0.005	**−0.019**	**−0.014**	**−0.014**	**0.081**	**−0.105**	**0.395**	**0.107**	−0.004
17	*MAGAR*	**0.104**	0.007	0.012	**0.026**	**0.029**	**−0.102**	**−0.043**	**−0.021**	**0.083**	**−0.146**
18	*FSINFORM*	**−0.033**	0.002	−0.006	**−0.017**	**−0.015**	**−0.020**	**0.081**	**−0.107**	−0.009	**0.038**
19	*RD*	**0.372**	**−0.112**	**0.017**	**0.033**	**0.030**	**0.220**	**−0.134**	**0.065**	**0.015**	**−0.234**
20	*Patent*	**0.256**	**0.021**	**0.015**	**0.018**	**0.016**	**0.139**	**−0.054**	**0.117**	0.010	**−0.111**
		11	12	13	14	15	16	17	18	19	20
1	*RDText*	**−0.276**	**−0.126**	**0.091**	**−0.088**	**0.078**	**−0.026**	**0.105**	**−0.032**	**0.478**	**0.351**
2	*ReadaV*	**−0.123**	**0.033**	**−0.020**	**−0.018**	**0.054**	**0.017**	**0.028**	**−0.023**	**−0.132**	−0.006
3	*SellProfit1*	**−0.046**	**−0.028**	0.004	0.001	0.006	−0.009	**0.013**	−0.004	0.007	**0.023**
4	*SellProfit2*	**−0.059**	**−0.033**	**0.014**	−0.005	**0.024**	−0.003	**0.028**	−0.012	**0.018**	**0.026**
5	*SellProfit3*	**−0.055**	**−0.034**	**0.020**	0.002	**0.044**	−0.001	**0.031**	−0.009	**0.021**	**0.025**
6	*SizeMV*	**0.175**	**0.240**	**0.021**	**0.148**	**−0.047**	**0.083**	**−0.098**	0.004	**0.405**	**0.089**
7	*Lev*	**0.307**	**0.098**	**−0.065**	**0.139**	−0.003	**−0.115**	**−0.036**	**0.076**	−0.004	**−0.080**
8	*ROA*	**−0.226**	**0.042**	**0.037**	**−0.076**	**−0.104**	**0.285**	0.012	**−0.138**	**0.126**	**0.120**
9	*Growth*	**−0.165**	0.003	**0.020**	**−0.045**	**−0.013**	**0.110**	**0.074**	**−0.049**	**0.079**	**0.071**
10	*SOE*	**0.435**	**0.124**	**−0.149**	**0.229**	**−0.089**	−0.012	**−0.139**	**0.053**	**−0.135**	**−0.136**
11	*Age*		**0.058**	**−0.111**	**0.165**	**−0.118**	**−0.066**	**−0.117**	**0.134**	**−0.055**	**−0.261**
12	*Big4*	**0.056**		−0.002	**0.061**	**−0.100**	**−0.020**	**−0.117**	−0.008	**0.097**	**0.033**
13	*Indr*	**−0.106**	0.004		**−0.154**	**0.024**	−0.003	**0.016**	0.001	**0.065**	**0.059**
14	*Board*	**0.166**	**0.064**	**−0.160**		**−0.067**	**−0.016**	**−0.091**	**0.030**	**0.037**	**−0.033**
15	*RetVol*	**−0.084**	**−0.087**	**0.023**	**−0.048**		**−0.066**	**0.307**	**−0.013**	**−0.119**	**−0.039**
16	*Disacc*	**−0.066**	**−0.014**	0.002	**−0.017**	**−0.066**		0.002	**−0.051**	**−0.050**	−0.004
17	*MAGAR*	**−0.143**	**−0.108**	**0.022**	**−0.092**	**0.280**	−0.009		−0.007	**−0.081**	**−0.018**
18	*FSINFORM*	**0.095**	**−0.017**	−0.002	**0.026**	**0.030**	**−0.048**	**0.023**		−0.002	**−0.035**
19	*RD*	**−0.186**	**0.016**	**0.085**	−0.010	**−0.065**	**−0.055**	**−0.033**	**−0.036**		**0.435**
20	*Patent*	**−0.210**	**0.058**	**0.059**	**−0.023**	**−0.053**	**−0.018**	**−0.022**	**−0.037**	**0.381**	

treatment firms. Specifically, we construct the following difference-in-differences (DiD) regression model:

$$SellProfit_{i,t} = \alpha_0 + \alpha_1 \times Treat_i \times Post_t + \alpha \times Controls_{i,t} + \sum Firm + \sum Year + \varepsilon_{i,t} \quad (3)$$

where *Treat* is a dummy variable that equals 1 for treatment firms and 0 for control firms. We define the treatment firms as those whose average R&D expenditures to the total assets over the three years before the disclosure requirements change was in the top third of the sample, as we believe these companies will be more affected by the new R&D information disclosure policy. We define the control firms as those whose average R&D expenditures to the total assets over the three years before the disclosure requirements change was in the bottom third of the sample. *Post* is a dummy variable that equals 1 if the firm-year observation is after fiscal year 2013 and 0 otherwise since CFAR2012 is effective for fiscal years beginning after January 1st, 2013. We expect the interaction term *Treat* × *Post* to be positive.

Additionally, to avoid other factors leading to pre-event differences between the treated firms and the control firms, we conduct the DiD test based on the PSM paired sample. Specifically, we match each treated firm with a control firm with similar firm characteristics, including firm size (*SizeMV*), leverage (*Lev*), return on total assets (*ROA*), sales growth (*Growth*), firm age (*Age*), industry growth (*IndGrowth*), product market competition (*HHI*), and high-tech firm (*HighTec*). This matching was performed using a nearest-neighbor propensity score with a 0.005 caliper. After matching, the standardized differences in covariates between the treatment group and the control group are less than 10 %, and the differences are generally insignificant, suggesting that our matched sample produces a good

Table 3
Narrative R&D disclosure and insider selling profitability: Test of Hypothesis 1.This table presents the results of Eq. (2) accounting for the relationship between narrative R&D disclosure and insider selling profitability. The dependent variables are *SellProfit1*, *SellProfit2* and *SellProfit3*, which refer to the losses avoided after 3 months, 6 months and 9 months of sales transactions, respectively. *RDText* proxies for the quantity of narrative R&D disclosure based on machine learning methods. In Columns (1) through (3), only *RDText* is included. In Columns (4) through (6), we include all the control variables and firm and year fixed effects. The detailed definitions of the variables are provided in Appendix C. The t statistics, based on standard errors clustered at the firm level and corrected for heteroscedasticity, are reported in parentheses. *, ** and *** denote statistical significance at the 10 %, 5 % and 1 % levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	SellProfit1	*SellProfit2*	*SellProfit3*	*SellProfit1*	*SellProfit2*	*SellProfit3*
RDText	0.069***	0.077***	0.078***	0.066*	0.087**	0.116***
	(4.82)	(5.13)	(4.93)	(1.78)	(2.33)	(2.90)
SizeMV				−0.981***	−0.671***	−0.046
				(−8.44)	(−5.95)	(−0.39)
Lev				−1.017**	−0.191	−0.428
				(−2.28)	(−0.41)	(−0.88)
ROA				−6.359***	−7.077***	−7.371***
				(−5.53)	(−6.50)	(−6.80)
Growth				−0.352***	−0.453***	−0.594***
				(−3.28)	(−4.16)	(−5.46)
SOE				−0.076	−0.012	−0.054
				(−0.25)	(−0.04)	(−0.16)
Age				0.300	0.797***	1.155***
				(1.33)	(3.52)	(4.89)
Big4				−0.598	−0.655	−0.876**
				(−1.56)	(−1.41)	(−2.00)
Indr				−0.735	−0.998	0.111
				(−0.96)	(−10.31)	(0.14)
Board				0.217	0.248	0.340
				(0.81)	(0.91)	(1.24)
RetVol				−18.923***	−19.040***	−11.058***
				(−6.62)	(−6.54)	(−3.91)
Disacc				0.983*	1.418**	1.132**
				(1.81)	(2.51)	(1.98)
MAGAR				7.603	12.237**	14.269***
				(1.54)	(2.44)	(2.90)
FSInform				−0.072	−0.507*	−0.385
				(−0.26)	(−1.87)	(−1.35)
RD				−0.002	0.016*	0.007
				(−0.24)	(1.83)	(0.80)
Patent				−0.007	−0.006	0.030
				(−0.14)	(−0.11)	(0.53)
Constant	0.413***	0.604***	0.733***	16.137***	10.667***	−0.251
	(6.86)	(9.62)	(11.05)	(9.17)	(6.02)	(−0.14)
Firm FE	No	No	No	Yes	Yes	Yes
Year FE	No	No	No	Yes	Yes	Yes
N	27,023	27,023	27,023	27,023	27,023	27,023
Adj. R^2	0.001	0.001	0.001	0.020	0.026	0.024

balance in all covariates.[23]

Table 5 reports the regression results of Eq. (3) based on the PSM paired sample. In Column (1), the dependent variable is the quantity of narrative R&D disclosure (*RDText*). In Columns (2) through (4), the dependent variables are *SellProfit1*, *SellProfit2* and *SellProfit3*, respectively. As shown in this table, we find that the quantity of narrative R&D disclosure increases significantly after the release of the CFAR2012. More importantly, we find that the coefficients on *Treat* × *Post* are significantly positive, suggesting that more narrative R&D disclosure indeed leads to greater insider selling profitability.[24]

5.2.2. Mitigating omitted variable concerns

In our main analysis, we include an extensive set of control variables that have been documented in prior literature to be correlated with insider trading profitability, and we also include firm fixed effects to control for time-invariant firm characteristics. To further mitigate concerns that omitted variables could drive our results, we conduct the following tests.

[23] The results of covariate balance are available in Online appendix Table A2.

[24] We also conduct the dynamic DiD test based on this disclosure requirements change. The results reported in Table A3 of Online appendix suggest that in three years before the release of CFAR2012, there is no significant difference between the treatment and control group in terms of insider selling profitability, but after the release of CFAR2012, the insider selling profitability of the treatment firms is significantly higher. This results confirm the parallel trend for our DiD design, further supporting the main finding.

Table 4
Narrative R&D disclosure and insider selling profitability: Test of Hypothesis 2. This table presents the empirical results of the role of readability in the relation between narrative R&D disclosure and insider selling profitability. We use the total common words in R&D-related texts divided by the total words in the R&D-related texts to measure readability. *Reada* is a dummy variable that equals 1 if the readability of R&D text is higher than the median of the sample and 0 otherwise. The detailed definitions of all the variables are provided in Appendix C. The t statistics, based on standard errors clustered at the firm level and corrected for heteroscedasticity, are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

	SellProfit1	*SellProfit2*	*SellProfit3*
	(1)	(2)	(3)
RDText	0.071	0.078***	0.110***
	(1.53)	(4.10)	(5.50)
RDText × *Reada*	−0.057***	−0.050***	−0.025
	(−3.67)	(−5.95)	(−0.88)
Reada	0.347***	0.292**	0.117
	(4.29)	(2.46)	(0.74)
SizeMV	−0.903***	−0.630***	−0.022
	(−5.75)	(−4.36)	(−0.13)
Lev	−1.071**	−0.448**	−0.645***
	(−2.69)	(−2.37)	(−4.05)
ROA	−6.364***	−6.502***	−7.047***
	(−8.91)	(−9.67)	(−6.89)
Growth	−0.316**	−0.415***	−0.539***
	(−2.84)	(−5.10)	(−6.36)
SOE	−0.478*	−0.129	−0.452**
	(−1.88)	(−0.45)	(−2.85)
Age	0.304*	0.617***	1.051***
	(2.03)	(3.78)	(5.08)
Big4	−0.431	−0.512	−0.707
	(−0.86)	(−0.66)	(−1.31)
Indr	−1.077**	−1.179	0.244
	(−2.21)	(−1.28)	(0.48)
Board	0.266	0.278	0.458*
	(0.86)	(1.01)	(1.85)
RetVol	−17.679***	−18.809***	−10.306***
	(−7.21)	(−13.57)	(−3.50)
Disacc	0.826	0.968	0.932
	(1.41)	(1.36)	(1.75)
MAGAR	9.564*	14.036**	16.722**
	(2.00)	(2.54)	(2.69)
FSINFORM	−0.268	−0.633***	−0.551***
	(−0.91)	(−6.61)	(−3.72)
RD	−0.004	0.013	0.003
	(−0.55)	(1.32)	(0.37)
Patent	−0.027	−0.026	0.032
	(−1.46)	(−0.65)	(1.04)
Constant	15.425***	10.046***	−1.621
	(5.38)	(3.37)	(−0.50)
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes
N	24,471	24,471	24,471
Adj. R^2	0.058	0.069	0.074

First, to truly capture what the manager "says" about R&D activities in firms' annual reports rather than firm fundamentals related to R&D activities, we further control for another three innovation characteristic variables. The first variable is innovation dispersion (*InnovDisp*), which characterizes the degree of dispersion of corporate patents in different fields. According to the study of Makri et al. (2010), this indicator is calculated as follows:

$$InnovDisp = 1 - \sum_{j}^{n} B_{ij}^{2} \tag{4}$$

where B_{ij} is the proportion of patents of firm i under category j. Specifically, China's patent classification adopts the IPC classification established by the International Patent Classification. Referring to the practice of Huang and Chen (2010), we adopt the first three digits of the IPC main classification code to define the categories of technical fields. A larger value of *InnovDisp* indicates that the firm's patents are distributed in more diverse fields. The patent classification number data of listed companies are manually collected from the CNRDS database. The second variable is exploratory innovation (*ExpInnov*), which characterizes the degree to which new technologies are explored in corporate innovation activities. Inspired by Huang and Chen (2010), we measure *ExpInnov* with the number of categories of patent technologies newly entered by firms' innovation. According to the IPC classification, if a firm's patent enters a technology category that it has not been involved in the past, it means that the firm has explored new technological fields and

Table 5
Changes in R&D information disclosure requirements: PSM-DiD analysis. This table shows the identification of a causal relationship between narrative R&D disclosure and insider selling profitability by using difference-in-differences analysis. We conduct the DiD test based on the PSM paired sample. The matching was performed using a nearest-neighbor propensity score with a 0.005 caliper. *Treat* is a dummy variable that equals 1 for firms whose average R&D expenditures to the total assets over the three years before the disclosure change was in the top third of the sample and 0 for firms in the bottom third of the sample. *Post* is a dummy variable for the years after the changes in R&D information disclosure requirements occurred (i.e., after 2013). In Column (1), the dependent variable is the quantity of narrative R&D disclosure (*RDText)*. In Columns (2) through (4), the dependent variables are *SellProfit1*, *SellProfit2* and *SellProfit3*. The detailed definitions of all the variables are provided in Appendix C. For brevity, we do not report the regression results of the control variables, which are denoted as *Controls*. The t statistics, based on standard errors clustered at the firm level and corrected for heteroscedasticity, are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	RDText	*SellProfit1*	*SellProfit2*	*SellProfit3*
Treat × *Post*	0.159**	1.001***	0.805*	0.838*
	(1.99)	(2.59)	(1.95)	(1.93)
Constant	−5.981***	9.036*	−1.742	−6.819
	(−6.54)	(1.83)	(−0.38)	(−1.36)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	4,930	4,930	4,930	4,930
Adj. R^2	0.028	0.019	0.025	0.019

created new knowledge. The third variable is innovation efficiency (*InnovEffi*), which characterizes the input–output efficiency of corporate R&D activities. Referring to the study of Lee and Park (2011), we use the DEA (data envelopment analysis) method to calculate innovation efficiency by taking innovation expenditure and the R&D workforce ratio as the input variables and yearly patent applications as the output variables. The empirical results are shown in Panel A of Table 6. Notably, after we control for these three variables, our main results still hold.

Second, we include industry fixed effects and Industry-Year interaction fixed effects, as opposed to firm fixed effects in our earlier specifications, to control for the effect of unobservable time-invariant and time-variant industry characteristics. Panel C of Table 6 shows the estimated results. We find that the coefficients on *RDText* are all positive and significant at the 1 % level, suggesting that our baseline regression results are unlikely to be driven by omitted variables.

5.2.3. Alternative measures and alternative samples

First, we construct the other two measures of narrative R&D disclosures. *NumRDText* refers to the natural logarithm of the number of R&D-related keywords or phrases in annual reports. *MDA_RDText* refers to the natural logarithm of the number of R&D-related keywords or phrases in the MD&A section of the annual reports. The empirical results shown in Panel A of Table 7 still support our main findings.

Second, we construct two other measures of insider selling profitability. Following Chung et al. (2019), we rebuild the measure as the ratio of the insider selling profit to the market value of equity at the end of the previous year (*New_SellProfit*). In addition, to mitigate the effect of the nonopportunistic motivation of insider sales transactions, such as liquidity purposes or portfolio rebalancing objectives, we focus only on the profitability of opportunistic insider sales transactions (*Opp_SellProfit*). Inspired by prior research (Huddart and Ke, 2007; Cohen et al., 2012), we exploit abnormal returns one month after insider sales to distinguish opportunistic sales from routine sales. A negative abnormal return indicates that insiders probably use private information to reduce shareholdings in advance. Therefore, sales transactions followed by negative abnormal returns are classified as opportunistic sales; otherwise, they are classified as routine sales. Panel B of Table 7 reports the empirical results. We still observe significant positive relationships between narrative R&D disclosures and insider selling profitability, suggesting that our main baseline findings are robust.

Third, in our baseline regression, we include firm-years with no reported insider sales transactions; in this section, we exclude firm-years with no reported insider sales transactions as a robustness check. The empirical results are reported in Columns (1) through (3) of Panel C of Table 7, and we find that our main conclusions remain unchanged.[25] In addition, we re-examine the baseline regression model based on a sample of individual insiders' trades instead of aggregating insider trading at the firm level. We add a series of control variables at the insider individual level in addition to the control variables in the baseline regression model. *Male* is a dummy variable that equals to 1 if the insider is a male and 0 otherwise. *Comp* is a dummy variable that equals to 1 if the insider is paid by the listed company and 0 otherwise. *InsiderAge* refers to the natural logarithm of insider's age. *lnJob* refers to the natural logarithm of the number of companies where insiders work part-time. *Excu* is a dummy variable that equals to 1 if the insider is an executive and 0 otherwise. *Dire* is a dummy variable that equals to 1 if the insider is a member of the board and 0 otherwise. *Superv* is a dummy variable that equals to 1 if the insider is a supervisor and 0 otherwise. We control for the month of the insiders' trading date instead of the year fixed effect

[25] Since some insider sales transactions may occur before the previous year annual report release, we re-examine the baseline regression model after excluding the insider transactions in current year that occur before the previous year annual report release. The empirical results which are shown in Table A4 of Online appendix suggest that our main findings still hold.

Table 6

Mitigating omitted variable concerns. This table presents the baseline regression results after mitigating omitted variable concerns. Panel A shows the robustness of our baseline regression results after controlling for the other three innovation characteristics. *InnovDisp* refers to the degree to which corporate patents are dispersed among different fields. *ExpInnov* refers to the degree to which new technologies are explored in corporate innovation activities. *InnovEffi* refers to the input–output efficiency of corporate R&D activities. Panel B reports the baseline regression results after we add the industry fixed effect and Industry-Year interaction fixed effect. The detailed definitions of all the variables are provided in Appendix C. For brevity, we do not report the regression results of the control variables, which are denoted as *Controls*. The t statistics, based on standard errors clustered at the firm level and corrected for heteroscedasticity, are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Panel A: Controlling for the other three innovation characteristics

	SellProfit1	*SellProfit2*	*SellProfit3*
	(1)	(2)	(3)
RDText	0.091**	0.110**	0.133***
	(2.04)	(2.45)	(2.81)
InnovDisp	−0.138	−0.384	−0.564
	(−0.31)	(−0.85)	(−1.29)
ExpInnov	0.201	0.326*	0.244
	(1.14)	(1.83)	(1.37)
InnovEffi	−0.134	0.463	0.594
	(−0.06)	(0.19)	(0.25)
Constant	20.857***	13.306***	0.606
	(7.23)	(4.40)	(0.20)
Controls	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes
N	21,171	21,171	21,171
Adj. R^2	0.025	0.030	0.028

Panel B: Industry-Year interaction fixed effect

	SellProfit1	*SellProfit2*	*SellProfit3*	*SellProfit1*	*SellProfit2*	*SellProfit3*
	(1)	(2)	(3)	(4)	(5)	(6)
RDText	0.057***	0.041**	0.043**	0.052	0.078**	0.103**
	(3.35)	(2.40)	(2.38)	(1.38)	(2.07)	(2.54)
Constant	9.289***	9.030***	6.027***	18.113***	12.221***	0.942
	(9.78)	(8.95)	(5.91)	(9.21)	(6.12)	(0.46)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	No	No	No
Industry# Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	27,023	27,023	27,023	27,023	27,023	27,023
Adj. R^2	0.031	0.041	0.042	0.062	0.077	0.084

to alleviate the potential impact of the interim reports. Columns (4) through (6) of Panel C in Table 7 report the regression results with this alternative sample, which are qualitatively consistent with our main results.

5.2.4. Ruling out alternative explanations

Given that managers have great discretions over the specific details of narrative R&D disclosures, it is possible that insiders opportunistically increase or exaggerate narrative R&D disclosure to distract investors while trading on other bad news about firms' current performance. To rule out this alternative explanation, we first test the impact of current earnings performance on the quantity and tone of narrative R&D disclosure. We exploit three variables to measure a firm's earnings performance, namely return on total assets (*ROA*), return on equity (*ROE*), and unexpected earnings (*RSUE*). The *RSUE* is measured as the decile rank of standardized unexpected earnings (*SUE*). *SUE* refers to the difference between actual earnings per share and the median of earnings forecast of analysts divided by the year-end stock price. In addition, we construct the measure of the narrative R&D disclosure tone (*RDTone*) by referring to the sentiment dictionary constructed by Yao et al. (2021). We count the number of positive and negative words appearing in the R&D-related texts separately. The calculation method for *RDTone* is shown in the following formula.

$$RDTone = \frac{Positivewords - Negativewords}{Positivewords + Negativewords} \tag{4}$$

Referring to Merkley (2014), we test the relation between current earnings performance and narrative R&D disclosure by estimating the following multivariate regression:

$$RDText_{i,t}/RDTone_{i,t} = \beta_0 + \beta_1 Earn_{i,t} + \beta Controls_{i,t} + \sum Firm + \sum Year + \varepsilon_{i,t} \tag{5}$$

The independent variable (*Earn*) is current earnings performance and includes the above three measures, *ROA*, *ROE* and *RSUE*. *Controls* refers to various factors that may relate to firms' R&D information disclosure policy. We add all the control variables in Eq. (2).

Table 7

Alternative measures and alternative samples. Panel A reports the baseline regression results using alternative measures of narrative R&D disclosures. *NumRDText* refers to the natural logarithm of the number of R&D-related keywords or phrases in annual reports. *MDA_RDText* refers to the natural logarithm of the number of R&D-related keywords or phrases in the MD&A section. Panel B reports the results after using alternative measures of insider selling profitability. *New_SellProfit* is measured as the ratio of the insider selling profit to the market value of equity at the end of the previous year. *Opp_SellProfit* refers to the profitability of opportunistic insider sales transactions. Panel C reports the results after using alternative samples. Columns (1) through (3) show the baseline regression results with the sample excluding firm-years with no reported insider sales transactions. Columns (4) through (6) report the baseline regression results based on a sample of individual insiders' trades. *Insider_Ctrls* refers to a series of control variables at the insider individual level. The detailed definitions of all the variables are provided in Appendix C. For brevity, we do not report the regression results of the control variables, which are denoted as *Controls*. The t statistics, based on standard errors clustered at the firm level and corrected for heteroscedasticity, are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Panel A: Alternative measures for narrative R&D disclosure

	SellProfit1	*SellProfit2*	*SellProfit3*	*SellProfit1*	*SellProfit2*	*SellProfit3*
	(1)	(2)	(3)	(4)	(5)	(6)
NumRDText	0.331***	0.457***	0.547***			
	(3.09)	(4.15)	(4.76)			
MDA_RDText				0.006	0.101*	0.174***
				(0.13)	(1.86)	(3.85)
Constant	15.293***	9.510***	−1.639	14.709***	8.096***	−3.519
	(8.64)	(5.32)	(−0.89)	(6.04)	(3.18)	(−1.38)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	27,023	27,023	27,023	22,754	22,754	22,754
Adj. R^2	0.021	0.026	0.025	0.062	0.073	0.079

Panel B: Alternative measures for insider selling profitability

	New_ SellProfit1	*New_ SellProfit2*	*New_ SellProfit3*	*Opp_ SellProfit1*	*Opp_ SellProfit2*	*Opp_ SellProfit3*
	(1)	(2)	(3)	(4)	(5)	(6)
RDText	0.039	0.113*	0.167**	0.064**	0.075**	0.050
	(0.92)	(1.87)	(2.17)	(1.99)	(2.26)	(1.49)
Constant	17.008***	14.717***	0.671	7.214***	5.173***	−2.897*
	(8.11)	(4.93)	(0.18)	(4.68)	(3.28)	(−1.80)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	26,810	26,810	26,810	27,023	27,023	27,023
Adj. R^2	0.073	0.086	0.092	0.011	0.017	0.017

Panel C: Alternative samples

	Excluding firm-years with no insider sales			Insider-level analyses		
	SellProfit1	*SellProfit2*	*SellProfit3*	*SellProfit1*	*SellProfit2*	*SellProfit3*
	(1)	(2)	(3)	(4)	(5)	(6)
RDText	0.200	0.274**	0.281**	0.033	0.247**	0.320**
	(1.58)	(2.19)	(2.14)	(0.30)	(1.98)	(2.46)
Constant	68.591***	49.708***	12.913**	63.679***	41.757***	15.865*
	(1.73)	(7.74)	(1.96)	(9.38)	(4.89)	(1.86)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Insider_Ctrls	–	–	–	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	No	No	No
Year-Month	No	No	No	Yes	Yes	Yes
N	7,840	7,840	7,840	48,206	48,206	48,206
Adj. R^2	0.069	0.082	0.068	0.083	0.097	0.089

Moreover, we control for industry growth (*IndGrowth*), product market competition (*HHI*) and high-tech firms (*HighTec*), which are related to proprietary costs of narrative R&D disclosures. If firms opportunistically increase or exaggerate narrative R&D disclosure to obfuscate bad earnings news, then β_1 is expected to be significantly negative.

Panel A of Table 8 reports the empirical results on the relation between earnings performance and narrative R&D disclosure. In Columns (1) through (3), the dependent variable is total narrative R&D disclosure quantity. In Columns (4) through (6), the dependent variable is tone of narrative R&D disclosure. The empirical results show that earnings performance is not significantly or negatively associated with total narrative R&D disclosure quantity. Conversely, all three measures of earnings performance are positively and significantly related to R&D disclosure tone. Overall, these results suggest that current earnings performance does not correlate negatively with narrative R&D disclosure quantity or tone in our sample period, indicating that firms may not obfuscate other bad news by manipulating narrative R&D disclosures.

Table 8

Ruling out alternative explanations. Panel A presents the regression results of the relation between earnings performance and narrative R&D disclosure. *RDText* proxies for the quantity of narrative R&D disclosure. *RDTone* proxies for the tone of narrative R&D disclosure. The independent variables are *ROA*, *ROE* and *RSUE*, referring to return on total assets, return on equity, and unexpected earnings, respectively. Panel B presents the regression results of the relation between earnings performance and narrative R&D disclosure when insiders have strong incentives to trade on bad earnings news. *LockShareR* proxies for insiders' selling motivations, which is calculated as the total amount of restricted shares that expire in the next year divided by the total share capital. The detailed definitions of all the variables are provided in Appendix C. For brevity, we do not report the regression results of the control variables, which are denoted as *Controls*. The t statistics, based on standard errors clustered at the firm level and corrected for heteroscedasticity, are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Panel A: Earnings performance and narrative R&D disclosure

	RDText	*RDText*	*RDText*	*RDTone*	*RDTone*	*RDTone*
	(1)	(2)	(3)	(4)	(5)	(6)
ROA	−0.486			0.226***		
	(−1.52)			(13.63)		
ROE		−0.065			0.094***	
		(−.54)			(12.66)	
RSUE			−0.004			0.002***
			(−0.83)			(7.92)
Constant	−1.912**	−1.818**	−1.843**	0.628***	0.623***	0.598***
	(−2.31)	(−2.19)	(−2.21)	(16.68)	(16.57)	(15.70)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	23,258	23,223	23,258	23,258	23,223	23,258
Adj. R^2	0.100	0.100	0.100	0.122	0.119	0.112

Panel B: Insider sales motivations, earnings performance and narrative R&D disclosure

	RDText	*RDText*	*RDText*	*RDTone*	*RDTone*	*RDTone*
	(1)	(2)	(3)	(4)	(5)	(6)
$ROA \times LockShareR_{i,t+1}$	−1.680			0.130		
	(−1.19)			(1.63)		
$ROE \times LockShareR_{i,t+1}$		−0.119			0.116**	
		(−0.15)			(2.07)	
$RSUE \times LockShareR_{i,t+1}$			0.022			−0.000
			(1.04)			(−0.39)
LockShareR	0.149	0.072	−0.052	−0.002	−0.005	0.008
	(1.58)	(0.81)	(−0.41)	(−0.35)	(−0.91)	(1.10)
ROA	−0.521			0.204***		
	(−1.27)			(10.47)		
ROE		−0.145			0.081***	
		(−0.95)			(8.56)	
RSUE			−0.009*			0.002***
			(−1.80)			(6.41)
Constant	−1.079	−1.019	−0.991	0.626***	0.620***	0.594***
	(−1.21)	(−1.15)	(−1.11)	(15.48)	(15.40)	(14.54)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	18,686	18,684	18,686	18,686	18,684	18,686
Adj. R^2	0.088	0.087	0.087	0.105	0.104	0.098

Next, we examine whether insiders increase narrative R&D disclosures or exaggerate firms' future prospects to distract investors when they have incentives to trade on bad earnings news. To test this issue, we construct the following Eq. (6):

$$RDText_{i,t}/RDTone_{i,t} = \beta_0 + \beta_1 Earn_{i,t} + \beta_2 Earn_{i,t} \times LockShareR_{i,t+1} + \beta_3 LockShareR_{i,t+1} + \beta \times Controls_{i,t} + \sum Firm + \sum Year + \varepsilon_{i,t} \quad (6)$$

where *LockShareR* is the proxy for insiders' selling motivations, which is calculated as the total amount of restricted shares that expire in the next year divided by the total share capital. Prior research shows that the number of shares that lockup expiration is positively associated with insiders' tendency to sell company shares (Ofek and Yermack, 2000). In addition, the stock lock-up period is stipulated by relevant laws. In other words, the expiration time of restricted shares is exogenous to a firm's narrative R&D disclosure. Therefore, we can use the total number of shares of lockup expiration as a proxy for insider selling motivations. If insiders intentionally manipulate narrative R&D disclosure to distract investors, then β_2 is expected to be significantly negative. The results reported in Panel B of Table 8 show that there are no significantly negative coefficients on the interaction term $Earn \times LockShareR$, implying that insiders may not distract investors by increasing or exaggerating narrative R&D disclosure when a firm has poor earnings performance.

Taken together, the above results indicate that our conclusions are less subject to this alternative explanation.

5.3. Cross-sectional analysis

As discussed above, we find robust evidence that insiders in firms with more narrative R&D disclosures make greater abnormal profits from their sale trades. To further support the potential channels we hypothesized, we explore conditions where the positive relationship between narrative R&D disclosure and insider selling profitability could be attenuated or accentuated. Specifically, we conduct a series of cross-sectional tests and expect the positive relation between narrative R&D disclosure and insider selling profitability to be more pronounced for firms with greater information proprietary costs and less pronounced for firms with more sophisticated market participants. In addition, since the greater profitability of insider sales in firms with more narrative R&D disclosures is related to managerial opportunism, we expect that good corporate governance mechanisms can weaken this positive relation.

5.3.1. Proprietary costs

We identify several proprietary cost proxies for R&D information. First, prior studies document that the proprietary cost is greater for firms in more competitive environments because their rivals may opportunistically exploit proprietary information, which damages the firm's competitive position (Harris, 1998; Li, 2010b; Cao et al., 2018). Therefore, we use the degree of product market competition (measured by the Herfindahl–Hirschman industry concentration index; the higher the index is, the lower the degree of market

Table 9

Cross-sectional analysis: Information proprietary costs. The table compares the baseline regression results between firms with higher information proprietary costs and those with lower information proprietary costs. We examine the proprietary cost channel from three aspects: product market competition (*PMC*), high-tech firms (*Hightech*), and the industry growth cycle (*IndCycle*). We report the p value of the F test (two-sided) for the difference in the coefficient on *RDText* between the two groups in the last row. The detailed definitions of all the variables are provided in Appendix C. For brevity, we do not report the regression results of the control variables, which are denoted as *Controls*. The t statistics, based on standard errors clustered at the firm level and corrected for heteroscedasticity, are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Panel A: Product market competition						
	SellProfit1		*SellProfit2*		*SellProfit3*	
	PMC = 1	*PMC* = 0	*PMC* = 1	*PMC* = 0	*PMC* = 1	*PMC* = 0
	(1)	(2)	(3)	(4)	(5)	(6)
RDText	0.135*	0.050	0.161**	0.075	0.221***	0.127
	(1.81)	(0.64)	(2.17)	(0.93)	(2.89)	(1.49)
Constant	30.400***	14.554***	22.044***	7.490**	4.880	0.544
	(7.44)	(4.15)	(5.26)	(2.31)	(1.14)	(0.16)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	9,126	8,938	9,126	8,938	9,126	8,938
Adj. R^2	0.030	0.021	0.043	0.023	0.039	0.020
Diff. in coef. on *RDText*	p-value = 0.095		p-value = 0.082		p-value = 0.078	
Panel B: High-tech firms						
	SellProfit1		*SellProfit2*		*SellProfit3*	
	Hightech = 1	*Hightech* = 0	*Hightech* = 1	*Hightech* = 0	*Hightech* = 1	*Hightech* = 0
	(1)	(2)	(3)	(4)	(5)	(6)
RDText	0.112**	−0.029	0.126***	0.019	0.134***	0.118
	(2.38)	(−0.46)	(2.70)	(0.27)	(2.73)	(1.58)
Constant	21.361***	11.398***	14.977***	6.557**	0.119	−1.334
	(7.85)	(3.74)	(5.46)	(2.16)	(0.04)	(−0.42)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	16,170	10,853	16,170	10,853	16,170	10,853
Adj. R^2	0.027	0.016	0.035	0.017	0.030	0.017
Diff. in coef. on *RDText*	p-value = 0.005		p-value = 0.040		p-value = 0.410	
Panel C: Industry growth cycle						
	SellProfit1		*SellProfit2*		*SellProfit3*	
	IndCycle = 1	*IndCycle* = 0	*IndCycle* = 1	*IndCycle* = 0	*IndCycle* = 1	*IndCycle* = 0
	(1)	(2)	(3)	(4)	(5)	(6)
RDText	0.172**	0.041	0.140	0.028	0.230**	0.099
	(2.04)	(0.50)	(1.47)	(0.31)	(2.35)	(1.08)
Constant	23.065***	10.642***	8.995*	7.222*	−5.486	0.756
	(4.91)	(2.91)	(1.81)	(1.88)	(−1.11)	(0.19)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	6,650	7,987	6,650	7,987	6,650	7,987
Adj. R^2	0.044	0.021	0.036	0.028	0.041	0.032
Diff. in coef. on *RDText*	p-value = 0.043		p-value = 0.067		p-value = 0.050	

competition) to measure the proprietary costs associated with the market competition environment. Specifically, we define *PMC* as a dummy variable that equals to 1 if a firm's product market competition is in the upper third of the sample and 0 if it is in the lower third of the sample. Panel A of Table 9 reports the grouping results based on *PMC*. The results show that the positive relation between narrative R&D disclosure and insider selling profitability is more pronounced for firms with greater product market competition. The difference in the *RDText* coefficients between these two groups is statistically significant.

Second, innovation is particularly important for high-tech firms and is an important driving force for them to maintain their core competitiveness. Therefore, compared with non-high-tech firms, high-tech firms face greater proprietary costs when disclosing R&D information. We divide the full sample into high-tech firms and non-high-tech firms. *Hightech* is a dummy variable that equals to 1 when the firm belongs to a high-tech industry and 0 otherwise. Panel B of Table 9 reports the grouping results based on *Hightech*. The results show that the positive relation between narrative R&D disclosure and insider selling profitability is muted for non-high-tech firms but remains significant for high-tech firms. The difference in the *RDText* coefficients between these two groups is also statistically significant.

Table 10

Cross-sectional analysis: Sophisticated market participants following. The table compares the baseline regression results between firms with more sophisticated market participants following and those with less sophisticated market participants following. We consider the information dissemination role of three types of sophisticated market participants: short sellers (*Short*), analysts following (*Analyst*), and media coverage (*Media*). We report the p value of the F test (two-sided) for the difference in the coefficient on *RDText* between the two groups in the last row. The detailed definitions of all the variables are provided in Appendix C. For brevity, we do not report the regression results of the control variables, which are denoted as *Controls*. The t statistics, based on standard errors clustered at the firm level and corrected for heteroscedasticity, are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Panel A: Deregulation of short selling

	SellProfit1		*SellProfit2*		*SellProfit3*	
	Short = 1	*Short* = 0	*Short* = 1	*Short* = 0	*Short* = 1	*Short* = 0
	(1)	(2)	(3)	(4)	(5)	(6)
RDText	−0.019	0.170***	0.037	0.146***	0.074	0.146**
	(−0.34)	(3.11)	(0.66)	(2.59)	(1.28)	(2.48)
Constant	20.040***	19.523***	15.210***	10.134***	2.010	−0.011
	(6.87)	(6.57)	(5.14)	(3.35)	(0.66)	(−0.00)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	11,933	15,090	11,933	15,090	11,933	15,090
Adj. R^2	0.025	0.022	0.031	0.025	0.028	0.025
Diff. in coef. on *RDText*	p-value = 0.000		p-value = 0.005		p-value = 0.025	

Panel B: Analyst following

	SellProfit1		*SellProfit2*		*SellProfit3*	
	Analyst = 1	*Analyst* = 0	*Analyst* = 1	*Analyst* = 0	*Analyst* = 1	*Analyst* = 0
	(1)	(2)	(3)	(4)	(5)	(6)
RDText	0.078	0.130**	0.023	0.159***	0.085	0.174***
	(0.80)	(2.27)	(0.26)	(2.72)	(0.94)	(2.72)
Constant	25.701***	18.548***	24.834***	9.836***	4.262	2.145
	(5.54)	(5.23)	(5.46)	(2.75)	(0.94)	(0.59)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	8,742	10,513	8,742	10,513	8,742	10,513
Adj. R^2	0.031	0.013	0.031	0.013	0.031	0.013
Diff. in coef. on *RDText*	p-value = 0.143		p-value = 0.007		p-value = 0.050	

Panel C: Median coverage

	SellProfit1		*SellProfit2*		*SellProfit3*	
	Media = 1	*Media* = 0	*Media* = 1	*Media* = 0	*Media* = 1	*Media* = 0
	(1)	(2)	(3)	(4)	(5)	(6)
RDText	−0.129	0.067	−0.097	0.137**	0.054	0.169**
	(−1.63)	(1.04)	(−1.24)	(2.04)	(0.69)	(2.30)
Constant	18.313***	19.647***	13.970***	14.400***	−3.568	5.110
	(5.05)	(5.05)	(3.85)	(3.72)	(−1.00)	(1.33)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	8,965	9,177	8,965	9,177	8,965	9,177
Adj. R^2	0.022	0.020	0.022	0.020	0.022	0.020
Diff. in coef. on *RDText*	p-value = 0.000		p-value = 0.000		p-value = 0.020	

Third, the industry life cycle may also influence the proprietary costs of R&D-related disclosure. Companies in an industry that is at the growth stage face a greater threat of potential competitors' entry, which may turn into higher proprietary costs. We define *IndCycle* as a dummy variable that equals to 1 if the industry sales growth rate is in the upper third of the whole sample and 0 if it is in the lower third of the sample. Panel C of Table 9 reports the grouping results based on *IndCycle*. The results show that the positive relation between narrative R&D disclosure and insider selling profitability is accentuated for firms in industries with high sales growth rates, and the difference in the coefficients of *RDText* between these two groups is also significant.

Overall, these results are consistent with our theoretical inferences that managers may obfuscate R&D disclosure to avoid proprietary costs, which can deter competitors and increase information asymmetry, thus providing insiders with the opportunity to gain from private information.

5.3.2. Sophisticated market participants following

We consider the information dissemination role of three types of sophisticated capital market participants, namely short sellers,

Table 11

Cross-sectional analysis: Corporate governance. The table compares the baseline regression results between firms with better corporate governance mechanisms and those with poor corporate governance mechanisms. We examine the governance role of institutional investor shareholding (*InstiH*), internal control quality (*ICQ*) and board independence (*Board*). We report the p value of the F test (two-sided) for the difference in the coefficient on *RDText* between the two groups in the last row. The detailed definitions of all the variables are provided in Appendix C. For brevity, we do not report the regression results of the control variables, which are denoted as *Controls*. The t statistics, based on standard errors clustered at the firm level and corrected for heteroscedasticity, are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Panel A: Institutional investor shareholdings

	SellProfit1		*SellProfit2*		*SellProfit3*	
	InstiH = 0	*InstiH* = 1	*InstiH* = 0	*InstiH* = 1	*InstiH* = 0	*InstiH* = 1
	(1)	(2)	(3)	(4)	(5)	(6)
RDText	0.221***	−0.047	0.173**	−0.017	0.154**	0.095*
	(2.94)	(−1.02)	(2.27)	(−0.36)	(2.00)	(1.95)
Constant	29.711***	15.142***	16.512***	13.077***	−1.634	−1.535
	(6.21)	(5.31)	(3.41)	(4.41)	(−0.33)	(−0.51)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
N	8,999	8,985	8,999	8,985	8,999	8,985
Adj. R^2	0.035	0.086	0.052	0.072	0.053	0.081
Diff. in coef. on *RDText*	p-value = 0.000		p-value = 0.003		p-value = 0.240	

Panel B: Internal control quality

	SellProfit1		*SellProfit2*		*SellProfit3*	
	ICQ = 0	*ICQ* = 1	*ICQ* = 0	*ICQ* = 1	*ICQ* = 0	*ICQ* = 1
	(1)	(2)	(3)	(4)	(5)	(6)
RDText	0.107*	0.049	0.106*	0.062	0.170***	0.047
	(1.79)	(0.84)	(1.75)	(1.04)	(2.80)	(0.78)
Constant	11.052***	20.380***	4.994	18.138***	−3.399	4.050
	(3.11)	(5.79)	(1.39)	(5.05)	(−0.95)	(1.11)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	9,014	8,999	9,014	8,999	9,014	8,999
Adj. R^2	0.102	0.091	0.119	0.106	0.137	0.104
Diff. in coef. on *RDText*	p-value = 0.177		p-value = 0.290		p-value = 0.030	

Panel C: Board independence

	SellProfit1		*SellProfit2*		*SellProfit3*	
	Board = 0	*Board* = 1	*Board* = 0	*Board* = 1	*Board* = 0	*Board* = 1
	(1)	(2)	(3)	(4)	(5)	(6)
RDText	0.101**	0.035	0.044	0.074	0.113**	0.045
	(1.99)	(0.56)	(0.84)	(1.16)	(2.14)	(0.69)
Constant	11.760***	19.741***	4.016	17.630***	−4.350	1.430
	(3.50)	(4.94)	(1.16)	(4.33)	(−1.24)	(0.35)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	12,003	8,389	12,003	8,389	12,003	8,389
Adj. R^2	0.078	0.078	0.070	0.093	0.080	0.103
Diff. in coef. on *RDText*	p-value = 0.053		p-value = 0.230		p-value = 0.070	

media coverage and analysts. First, short sellers have expertise in collecting and processing complicated information, and their trading behaviors convey valuable news to the capital market, thus helping improve the information environment. Therefore, it can be expected that the presence of short sellers, who act as informed and sophisticated market participants, will weaken the information advantage of insiders arising from narrative R&D disclosure and thus reduce insider selling profitability. In the Chinese capital market, where short selling is partially restricted, not all public firms can be short-selling targets. Therefore, we define *Short* as a dummy variable that equals to 1 if the firm's stocks are allowed to be sold short in the current year and 0 otherwise. The full sample is divided into groups that are short-selling targets and those that are not. Panel A of Table 10 reports the grouping results based on *Short*. The results show that the positive relation between narrative R&D disclosure and insider selling profitability is weakened for firms that are allowed to be short sold, which is consistent with our inference. The difference in the *RDText* coefficients between these two groups is statistically significant.

Second, analysts are regarded as important capital market information intermediaries. They can process and disseminate the public information of firms by virtue of their professional analysis ability and information advantage, and can also dig out additional undisclosed information about firms through site visits (Bushman et al., 2004). Therefore, it can be expected that insiders in firms with more analysts revealing and disseminating their information to the public have fewer opportunities to trade profitably on private information. We define *Analyst* as a dummy variable that equals to 1 if the number of analysts following the firm is in the upper third of the sample and equals 0 for the lower third of the sample. Panel B of Table 10 reports the grouping results based on *Analyst*. As anticipated, we find that the positive relation between narrative R&D disclosure and insider selling profitability is attenuated in firms with more analyst followings.

Third, prior research documents the role of media in disseminating news, which can improve firms' information environment (Bushee et al., 2010; Tetlock, 2010), and reduce the profitability of insiders' trades (Dai et al., 2015). Media coverage is measured by the number of news items mentioning the firm in the content. We define *Media* as a dummy variable that equals to 1 if the amount of news about the firm is in the upper third of the sample and 0 if it is in the lower third of the sample. Panel C of Table 10 reports the grouping results on *Media*. The results show that the positive relation between narrative R&D disclosure and insider selling profitability is weaker in firms with more media coverage.

Overall, these cross-sectional analyses further strengthen our main inference that more narrative R&D disclosures may not alleviate firms' information asymmetry, but it inversely increases information processing costs and prevents outside investors from accurately assessing firms' value. With more sophisticated market participants processing firms' public disclosures, insiders' information advantages can be mitigated.

5.3.3. Corporate governance

Prior studies document that better corporate governance can weaken insiders' information advantages, hence curbing their opportunistic trading behaviors (Jagolinzer et al., 2011; Lee et al., 2014). Therefore, if insiders use the information advantages derived from narrative R&D disclosures to carry out opportunistic stock sales, it can be expected that a good corporate governance mechanism can weaken the positive correlation between narrative R&D disclosure and insider selling profitability. In this paper, we examine the moderating effects of three corporate governance mechanisms, i.e., institutional investor shareholding, internal control quality and board independence, which have been widely demonstrated in prior literature (Chen et al., 2007; Chalmers et al., 2019; Peng, 2004).

Table 11 shows the moderating effects of corporate governance. Panel A reports the moderating effect of institutional investor shareholding. *InstiH* is a dummy variable that equals to 1 when a firm's institutional investor shareholding is in the upper third of the sample and 0 for the lower third of the sample. Panel B reports the moderating effect of internal control quality. We use the internal control index from the DIB database to measure internal control quality and define *ICQ* as a dummy variable that equals to 1 if the firm's internal control quality is in the upper third of the sample and 0 if it is in the lower third of the sample.[26] Panel C reports the moderating effect of board independence, which is measured by the proportion of independent directors. Similarly, we construct the dummy variable *Board* based on the upper and lower thirds of the sample. The results suggest that the positive relation between narrative R&D disclosure and insider selling profitability is more pronounced in firms with fewer institutional investor shareholdings, poorer internal control quality and lower board independence. In addition, the difference in the *RDText* coefficients between these two groups was statistically significant. Overall, this consistent evidence indicates that good corporate governance mechanisms can restrain insiders from opportunistically trading on firms' private information arising from more narrative R&D disclosures.

6. Further analysis

Thus far, we have demonstrated that managers' obfuscation disclosures of narrative R&D information and the high information processing costs associated with the technical nature of R&D information would increase information asymmetry, and create opportunities for insiders to trade profitably. In this section, we perform two additional tests to directly examine the information processing costs associated with more narrative R&D disclosures and test the timing of insider trading/profits.

[26] The DIB Business Risk Management Inc. (DIB) in China adopts the publicly available information to assess internal control disclosure among listed firms. Concretely, it follows both the Committee of Sponsoring Organizations (COSO) framework and Basic Standards of Enterprise Internal Control to develop a point scoring system, which gives (or deducts) scores to a firm based on five components of internal control (control environment, risk assessment, control activities, information and communication, and monitoring).

6.1. Narrative R&D disclosure and post-earnings announcement return pattern

Theoretically, if investors incur additional information processing costs related to more narrative R&D disclosures, there should be a predictable post-announcement return pattern (Huang et al., 2018; Blankespoor et al., 2020). To examine the relation between narrative R&D disclosure and post-announcement abnormal returns, we estimate the following multivariate regression:

$$CAR = \beta_0 + \beta_1 RSUE_{it} + +\beta_2 RSUE_{it} \times RDText_{it} + \beta_3 RDText_{it} + \beta_4 Controls + \sum Firm + \sum Year + \varepsilon_{it} \quad (7)$$

where *CAR* is the cumulative abnormal return over the long window ([+2, +70]; [+2, +80]) starting two days after the earnings announcement date.[27] We calculate CAR by the market model method. *RSUE* is the decile rank of the standardized unexpected earnings *SUE*, which is measured as the difference between actual earnings per share and the median of earnings forecasts of analysts divided by the year-end stock price. Referring to the study of Huang et al. (2018), we add a series of control variables, including firm size (*SizeMV)*, profitability (*ROA*), the book-to-market ratio (*BM*), stock returns (*RET)*, volatility of stock return (*RetVol*), discretionary accruals (*Disacc)*, and analyst following (*LnAna*).

Table 12 reports the empirical results. In Columns (1) and (2), we find that there is a significantly negative relationship between *RSUE* and long-term *CAR*, indicating that investors in China's capital market tend to overreact to earnings news at the time of the earnings announcement, followed by a reversal in abnormal returns. Therefore, if more narrative R&D disclosures result in higher processing costs for investors, decelerating the adjustments of abnormal returns in the long run, then we can expect a stronger return reversal after earnings announcements. The results reported in Columns (3) and (4) show that the coefficients on *RSUE* × *RDText* are significantly negative. Overall, these findings support our theory assumption that more narrative R&D disclosures increase investors' information processing costs.

6.2. Narrative R&D disclosure and pre-earnings announcement trades

If more narrative R&D disclosures enable insiders to earn excess profits by selling down their stock holdings in advance of future bad news, we expect to see more insider sales prior to negative earnings announcements in firms with greater narrative R&D disclosures. We then examine the timing of insider trades before the disclosure of quarterly earnings announcements. Following Fu et al. (2020), we first identify insider sales occurring 20, 30, 40, or 50 days prior to earnings announcement dates (hereafter referred to as pre-earnings announcement trades). Next, we use the average percentage of pre-earnings announcement trades in a firm-year to examine insider trading timing (*PreSale*). We use the average profitability of insider trades taking place 30 days before earnings announcement dates in a firm-year to examine insider trading profits (*PSellProfit*). In addition, we construct *N_Anno* to measure the number of negative earnings announcements in a firm-year. A negative earnings announcement refers to quarterly earnings that are significantly lower than those in the previous period or when there is a loss. We conduct this examination based on the subsample with insider sales.

Panel A of Table 13 reports the regression results for narrative R&D disclosure and pre-earnings announcement trading propensity. The coefficients on *RDText* × *N_Anno* are significantly positive, suggesting that more narrative R&D disclosures increase the propensity of insiders to trade within two months immediately before negative earnings announcements. The regression results in Panel B of Table 13 suggest that more narrative R&D disclosures indeed increase the profitability of insider sales before negative earnings announcements, which provides further support for the timing of insider trades. Overall, these results indicate that more narrative R&D disclosures strengthen insiders' information advantages and result in more opportunistic pre-earnings announcement insider trades.

7. Conclusion

This paper investigates how narrative R&D disclosures affect firms' information asymmetry from the perspective of insider trading profitability. Exploiting machine learning and textual analysis techniques to construct proxies for narrative R&D disclosures, we find strong evidence that insider selling profits are significantly greater in firms with more narrative R&D disclosures, and this positive relation is attenuated when the readability of R&D-related texts is higher. Moreover, the positive relation between narrative R&D disclosures and insider selling profitability is magnified when proprietary costs associated with narrative R&D disclosure are higher, which is in accordance with our proposition that managers in firms facing higher proprietary costs are more likely to obfuscate the narrative disclosures of R&D information. This positive relation is also weakened when the firms are followed by more sophisticated market participants, indicating that the information dissemination role of sophisticated capital market participants can decrease the information processing costs associated with narrative R&D disclosures for investors, thus mitigating insiders' information advantages. We also find consistent evidence that good corporate governance mechanisms, i.e., more institutional investor shareholding, better internal control quality and greater board independence, can restrain insiders from trading profitably with the shelter of information asymmetry aggravated by narrative R&D disclosures. Finally, we provide evidence that more narrative R&D disclosures indeed decelerate the adjustments of abnormal returns after earnings announcements in the long run, which suggests that more narrative R&D

[27] The empirical results remain unchanged if we calculate the cumulative abnormal return (CAR) over the long windows ([+2, +50]; [+2, +60]) starting two days after the earnings announcement date.

Table 12
Narrative R&D disclosure and post-announcement period CAR. This table reports the empirical results of narrative R&D disclosure on post-announcement abnormal returns. *CAR* is the cumulative abnormal returns over the long window ([+2, +70]; [+2, +80]) starting two days after the earnings announcement date. *RSUE* is the decile rank of standardized unexpected earnings. *RDText* proxies for the quantity of narrative R&D disclosure for the current year. The detailed definitions of all the variables are provided in Appendix C. The t statistics, based on standard errors clustered at the firm level and corrected for heteroscedasticity, are reported in parentheses. *, ** and *** denote statistical significance at the 10 %, 5 % and 1 % levels, respectively.

	CAR [+2, +70]	*CAR* [+2, +80]	*CAR* [+2, +70]	*CAR* [+2, +80]
	(1)	(2)	(3)	(4)
RSUE	−1.189***	−1.396***	−0.856***	−0.986***
	(−8.01)	(−8.77)	(−3.89)	(−4.17)
RSUE × *RDText*			−0.086**	−0.106**
			(−2.04)	(−2.33)
RDText			0.285	0.281
			(0.76)	(0.69)
SizeMV	−3.149***	−2.777**	−3.124***	−2.722**
	(−2.66)	(−2.20)	(−2.72)	(−2.20)
ROA	13.797	13.012	14.319	13.631
	(1.24)	(1.09)	(1.31)	(1.16)
BM	−4.610	−3.508	−4.687	−3.584
	(−1.40)	(−0.99)	(−1.43)	(−1.02)
RET	2.622***	3.256***	2.659***	3.298***
	(2.60)	(3.03)	(2.78)	(3.21)
RetVol	−0.683	−6.218	−0.632	−6.092
	(−0.03)	(−0.23)	(−0.03)	(−0.23)
Disacc	−5.934	−7.072	−5.934	−7.072
	(−1.35)	(−1.49)	(−1.32)	(−1.46)
LnAna	−0.425	−0.414	−0.431	−0.419
	(−0.79)	(−0.72)	(−0.80)	(−0.73)
Constant	62.182**	52.695*	60.421**	50.254*
	(2.34)	(1.86)	(2.36)	(1.83)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	20,312	20,312	20,312	20,312
Adj. R^2	0.016	0.019	0.025	0.020

disclosures increase investors' processing costs. We also find that more narrative R&D disclosures increase the propensity of insiders to trade before negative earnings announcements and result in greater profitability of insider sales.

Our paper provides important implications for policymakers, regulators, and other market participants by shedding light on the dark side of narrative R&D disclosures on firms' information environment from the perspective of insider trading. With respect to regulators, our findings underscore the need for enhanced supervision of narrative information disclosures, as imprecise managerial narratives in corporate reports featuring technology and innovation activities may contribute to information asymmetry. Specifically, policymakers should develop more concrete and detailed standards on narrative R&D information disclosure to provide listed firms with more detailed guidelines. Regulators should not only focus on the quantity of narrative information disclosure but should also pay attention to improve the quality and readability of narrative disclosures, leveling the playing ground for outside investors. In addition, the incentives of managers to obfuscate narrative R&D disclosures in this paper may also apply to other qualitative disclosures of firms' core information in a similar sense, such as firms' risk information or ESG information disclosures. From this perspective, our research can offer some insights to policymakers who are considering more regulations on those types of information. Additionally, our research reminds investors of the need for more sophisticated capacity to process increasingly complicated public disclosures and the awareness about manager's obfuscation disclosure of R&D information.

Our conclusions are subject to several limitations. First, in this paper, we attempt to examine the incremental impact of narrative R&D disclosures beyond real R&D activities on firms' information environment. We make great efforts to capture what the manager "says" about R&D activities in firms' annual reports by exploiting machine learning and textual analysis techniques. Despite our desperate efforts, we cannot completely rule out the confounding effect of a firm's fundamentals related to R&D activities. Second, due to the data availability, we are not able to directly examine what types of innovation information that insiders in firms with more narrative R&D disclosures trade on. Future research is warranted to determine in greater detail what type of private innovation information insiders profitably trade on when more granular data is available. Third, in addition to narrative information regarding firms' R&D activities, other nonfinancial information, such as firms' risk information or ESG information, also deserve more attention when examining the strategic information disclosure associated with insider trading.

CRediT authorship contribution statement

Wan Huang: Conceptualization, Methodology, Investigation, Data curation, Project administration. **Qingwen Liang:**

Table 13

Narrative R&D disclosure and pre-earnings announcement trades. This table presents the results of narrative R&D disclosure and pre-earnings announcement trades based on the subsample with insider sales. The dependent variables in Panel A are the average percentage of pre-earnings announcement trades in a firm-year. Pre-earnings announcement trades are defined as insider sales that occurred 20, 30, 40, or 50 days before earnings announcement dates. The dependent variables in Panel B are the average profitability of insider sales occurring 30 days before earnings announcement dates in a firm-year. *N_Anno* refers to the number of negative earnings announcements in a firm-year. The detailed definitions of all the variables are provided in Appendix C. For brevity, we do not report the regression results of the control variables, which are denoted as *Controls*. The t statistics, based on standard errors clustered at the firm level and corrected for heteroscedasticity, are reported in parentheses. *, ** and *** denote statistical significance at the 10 %, 5 % and 1 % levels, respectively.

Panel A: Narrative R&D disclosure and pre-earnings announcement trading propensity

	PreSale1	*PreSale2*	*PreSale3*	*PreSale4*
	(1)	(2)	(3)	(4)
RDText	−0.042	−0.051	−0.058	−0.045
	(−0.49)	(−0.42)	(−0.40)	(−0.28)
RDText × *N_Anno*	0.065***	0.116***	0.110**	0.118**
	(2.59)	(3.00)	(2.50)	(2.50)
N_Anno	−0.157	−0.378	−0.457	−0.516
	(−0.93)	(−1.57)	(−1.57)	(−1.60)
Constant	5.688	22.318***	27.127***	26.438***
	(1.20)	(3.32)	(3.25)	(2.79)
Controls	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
N	6,666	6,666	6,666	6,666
Adj. R^2	0.012	0.022	0.017	0.014

Panel B: Narrative R&D disclosure and pre-earnings announcement trading profitability

	PSellProfit1	*PSellProfit2*	*PSellProfit3*
	(1)	(2)	(3)
RDText	0.048	−0.010	−0.010
	(0.68)	(−0.14)	(−0.12)
RDText × *N_Anno*	0.038	0.063***	0.060***
	(1.62)	(2.85)	(2.61)
N_Anno	0.120	−0.222	−0.326**
	(0.85)	(−1.58)	(−2.27)
Constant	28.416***	26.796***	2.954***
	(6.91)	(6.53)	(5.02)
Controls	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes
N	6,666	6,666	6,666
Adj. R^2	0.036	0.051	0.037

Conceptualization, Methodology, Investigation, Data curation, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

We sincerely appreciate the helpful comments from Marco Trombetta (the editor), Kathleen Hanley (the editor), and two anonymous reviewers. The author Wan Huang acknowledges financial support from the National Natural Science Foundation of China (Project No. 72302167), China Postdoctoral Science Foundation (Project No. 2022M722240) and the Fundamental Research Funds for the Central Universities (Project No. skbsh2023-56).

Appendix A. Requirements of narrative R&D disclosure in China

This table describes the requirements of the CFAR for the narrative R&D disclosure of listed companies, which were revised in 2005, 2012 and 2021.

Version of CFAR	Requirements
CFAR2001	Public firms' disclosures are supposed be decision-making related and focused on major investments, R&D projects, construction in progress and such information which is helpful for investors to know more about firms' future development.
CFAR2005	Public firms are supposed to notify their investors about future opportunities or challenges that concern the management, and disclose the development strategies, as well as the new products to be developed, and new R&D projects to be invested.
CFAR2012	Public firms are supposed to clearly state the purposes, progress, and targets of current R&D projects and their expected impact on the future development, as well as their future plans for innovation during the reporting period. Specifically, public firms are supposed to illustrate the total amount of R&D expenditures in the current year and its proportion on latest audited net assets and operating revenues. If the proportion changes are above 30 %, they should clarify the reasons for those changes. Public firms are also supposed to disclose important changes in their core competitiveness (such as key technical dominance, proprietary equipment, patents, non-patented technologies, etc.) and their impact on the company during the reporting period. In addition, public firms are supposed to disclose their staff or teams assigned to their R&D projects. If there is a great change in their R&D staff team, they are supposed to explain why those changes happen and clarify what impact it has on firms' future progress.
CFAR2021	The same as CFAR2012, but new requirements for public firms in ChiNext and STAR Market to targeted disclose information that can reflect the industry competitiveness, such as technology, R&D outputs, R&D investments, R&D staff team, etc.

Appendix B. Sample selection

Sample selection criteria	# of firms	# of insider sales transactions	# of firm-year observations
A-share listed firms, 2008–2020	3,992	74,583	37,609
Delete: insider sales transactions beyond the secondary market	1	11,551	899
Delete: insider sales transactions less than 1000 shares in each transaction	0	3,710	1,170
Delete: insider sales transactions with missing information on transaction date, transaction price and the number of transaction shares	1	391	73
Delete: firms in financial sectors	122	479	1,107
Delete: firms with more liabilities than assets	1	370	423
Delete: firms with "special treatment" ("ST") status or with "particular transfer" ("PT") status	2	2,352	435
Delete: firms with missing data to calculate variables	288	6,565	6,479
Final sample for baseline regression test	**3,577**	**49,165**	**27,023**

Appendix C. Variable definitions

	Definition
Variables used in the baseline regression	
RDText	The quantity of narrative R&D disclosure (see Subsection 4.2).
ReadaV	The readability of narrative R&D disclosure (see Subsection 4.2).
Reada	A dummy variable that equals to 1 if the readability of R&D text is higher than the median of the sample, and 0 otherwise.
SellProfit1	Aggregate profitability of all insider sales transactions during the fiscal year, computed as follows: $SellProfit_{it} = \sum_{j=1}^{n} SellBHAR_{itj} \times SellValue_{itj} \times (-1)$where *SellProfit* is the aggregate profitability of all insider sales transactions in firm *i* during the fiscal year *t*. *SellBHAR* equals the buy-and-hold abnormal return of three months, which is computed for the period starting one day after transaction date *j*. *SellValue* equals the total value of shares sold by all insiders on day *j*, *n* is the total number of firm-days with insider sales transactions during firm-year *it*. We take the natural logarithm of insider selling profit.
SellProfit2	Same definition as *SellProfit1* except *SellBHAR* equals six months buy-and-hold abnormal return computed for the period starting one day after transaction date *j*.
SellProfit3	Same definition as *SellProfit1* except *SellBHAR* equals nine months buy-and-hold abnormal return computed for the period starting one day after transaction date *j*.
SizeMV	Natural logarithm of market value of equity at the end of year.
Lev	The ratio of total liabilities to total assets at the end of year.
ROA	Return on assets, measured as net profit divided by total assets at the end of year.
Growth	The growth rate in operating income from the beginning of the year to the end of year.
SOE	A dummy variable that equals to 1 if the firm is a state-owned enterprise at the end of year, and 0 otherwise.
Age	Natural logarithm of firms' listed years.
Big4	A dummy variable that equals to 1 if the firm's incumbent auditor is one of the Big 4 (E&Y, KPMG, PricewaterhouseCoopers, and D&T), and 0 otherwise.
Indr	The ratio of the number of independent directors to the total number of directors on the board at the end of year.
Board	Natural logarithm of the number of board members at the end of year.
RetVol	Stock return volatility, measured as the standard deviation of a stock's weekly return.
Disacc	Discretionary accruals at fiscal year-end, based on the cross-sectional modified Jones' model.
MAGAR	The median of absolute market reaction to prior quarterly earnings announcements, where market reaction is measured as the cumulative size-adjusted return from two days before to the day of the earnings announcement (Huddart and Ke, 2007); the median is measured over the 20-quarter period ending with the fourth quarter of the current fiscal year.

(*continued on next page*)

(*continued*)

	Definition
FSINFORM	Financial statement informativeness, computed as the adjusted R^2 from a firm-specific time-series regression of price per share on book value per share and earnings per share using quarterly data from CSMAR for the 20-quarter period ending with the fourth quarter of the current fiscal year.
RD	Natural logarithm of research and development expenses in the current year.
Patent	Natural logarithm of the number of patents in the current year.
Other variables used in the endogeneity concerns and robustness checks	
Treat	A dummy variable that equals to 1 for treatment firms, and 0 for control firms.
Post	A dummy variable that equals to 1 if the firm-year observation is after fiscal year 2013 and 0 otherwise.
IndGrowth	The sales growth rate of industry in which the firm operates.
HHI	Herfindahl–Hirschman index based on total sales.
InnovDisp	The degree of dispersion of corporate patents in different fields. Referring to the study of Makri et al. (2010), this indicator is calculated as follows: $InnovDisp = 1 - \sum_j^n B_{ij}^2$ where B_{ij} is the proportion of patents of firm i under category j. Specifically, China's patent classification adopts the IPC classification established by the International Patent Classification. Referring to the practice of Huang and Chen (2010), we adopt the first three digits of the IPC main classification code to define the category of technical fields.
ExpInnov	The degree of exploration of new technologies in corporate innovation activities. Following Huang and Chen (2010), we measure *ExpInnov* with the number of categories of patent technologies newly entered by firms' innovation.
InnovEffi	The input–output efficiency of corporate R&D activities. Referring to the study of Lee and Park (2011), we use the DEA (Data Envelopment Analysis) method to calculate the innovation efficiency by taking innovation expenditure and R&D workforce ratio as the input variables and yearly patent applications as the output variables.
NumRDText	The natural logarithm of the number of R&D-related keywords or phrases in the annual report.
MDA_RDText	The natural logarithm of the number of R&D-related keywords or phrases in the MD&A section of the annual report.
New_ SellProfit	Aggregate profitability of all insider sales transactions during the fiscal year, computed as follows: $New\_SellProfit_{it} = \frac{(-1) \times \sum_{j=1}^{n} SellBHAR_{itj} \times SellValue_{itj}}{MV_{t-1}}$ where *SellBHAR* equals three (six, nine) months buy-and-hold abnormal return computed for the period starting one day after transaction date *j*, respectively; *SellValue* equals the total value of shares sold by all insiders on day *j*, *n* is the total number of firm-days with insider sales activity during firm-year *it*, MV_{t-1} is the market value of equity at the end of year *t–1*.
Opp_SellProfit	Same definition as *SellProfit*, while the insider sales transactions in firm i during the fiscal year t only include the opportunistic insider sales. Inspired from prior research (Huddart and Ke, 2007; Cohen et al., 2012), we exploit the abnormal returns one month after insider sales to distinguish the opportunistic sales and routine sales.
Male	A dummy variable that equals to 1 if the insider is a male, and 0 otherwise.
Comp	A dummy variable that equals to 1 if the insider is paid by the listed company, and 0 otherwise.
InsiderAge	The natural logarithm of insider's age.
lnJob	The natural logarithm of the number of companies where insiders working part-time in.
Excu	A dummy variable that equals to 1 if the insider is an executive, and 0 otherwise.
Dire	A dummy variable that equals to 1 if the insider is a member of board, and 0 otherwise.
Superv	A dummy variable that equals to 1 if the insider is a supervisor, and 0 otherwise.
ROE	Return on equity, measured as net profit divided by total equity at the end of year.
RSUE	The decile rank of the standardized unexpected earnings.
RDTone	The measure of the narrative R&D disclosure tone by referring to the sentiment dictionary constructed by Yao et al. (2021).
LockShareR	The total amount of restricted shares that expire in the next year divided by the total share capital.
Other variables used in the cross-sectional analysis	
PMC	A dummy variable that equals to 1 if firm's product market competition is in the upper third of the sample and equals to 0 for the lower third of the sample.
Hightech	A dummy variable that equals to 1 if the firm belongs to high-tech industry, and 0 otherwise.
IndCycle	A dummy variable that equals to 1 if the industry sales growth rate is in the upper third of the sample and equals to 0 for the lower third of the sample.
Short	A dummy variable that equals to 1 if the firm is allowed to be sold short in the current year, and 0 otherwise.
Analyst	A dummy variable that equals to 1 if the number of analysts following the firm is in the upper third of the sample and equals to 0 for the lower third of the sample.
Media	A dummy variable that equals to 1 if the number of news about the firm is in the upper third of the sample and equals to 0 for the lower third of the sample.
InstiH	A dummy variable that equals to 1 if firm's institutional investors shareholding is in the upper third of the sample and equals to 0 for the lower third of the sample.
ICQ	A dummy variable that equals to 1 if the firm's internal control quality is in the upper third of the sample and equals to 0 for the lower third of the sample.
Board	A dummy variable that equals to 1 if the firm's proportion of independent directors in a board is in the upper third of the sample and equals to 0 for the lower third of the sample.
Other variables used in the further analysis	
CAR	The cumulative abnormal return over the long windows ([+2, +70]; [+2, +80]) starting two days after the earnings announcement date.
BM	The book-to-market ratio.
LnAna	The natural logarithm of the number of analysts following the firm.
PreSale	The number of pre-earnings announcement trades to the total number of sale transactions in a firm-year. Pre-earnings announcement trades are defined as insider sales that occurred in 20/30/40/50 days before earnings announcement dates, respectively.
PSellProfit	The average profitability of insider sales occurring in 30 days before earnings announcement dates in a firm-year.
N_Anno	The number of negative earnings announcements in a firm-year.

Appendix D. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jaccpubpol.2024.107210.

References

Aboody, D., Lev, B., 2000. Information asymmetry, R&D, and insider gains. J. Financ. 55 (6), 2747–2766.
Aghamolla, C., Smith, K., 2023. Strategic complexity in disclosure. J. Account. Econ. 76 (2–3), 101635.
Andreou, P.C., Lambertides, N., Magidou, M., 2021. Stock price crash risk and the managerial rhetoric channel. Evidence from Narrative r&d Disclosure. Available at SSRN 3891736.
Arif, S., Kepler, J.D., Schroeder, J., Taylor, D., 2022. Audit process, private information, and insider trading. Rev. Acc. Stud. 27 (3), 1125–1156.
Baiman, S., Verrecchia, R.E., 1996. The relation among capital markets, financial disclosure, production efficiency, and insider trading. J. Account. Res. 34 (1), 1–22.
Blankespoor, E., 2019. The impact of information processing costs on firm disclosure choice: evidence from the XBRL mandate. J. Account. Res. 57 (4), 919–967.
Blankespoor, E., Dehaan, E., Marinovic, I., 2020. Disclosure processing costs, investors' information choice, and equity market outcomes: a review. J. Account. Econ. 70 (2–3), 101344.
Bloom, N., Schankerman, M., Van Reenen, J., 2013. Identifying technology spillovers and product market rivalry. Econometrica 81 (4), 1347–1393.
Booker, A., Curtis, A., Richardson, V.J., 2023. Investor disagreement, disclosure processing costs, and trading volume evidence from social media. Account. Rev. 98 (1), 109–137.
Bowen, R.M., Dutta, S., Tang, S.L., Zhu, P.C., 2018. Inside the "black box" of private in-house meetings. Rev. Acc. Stud. 23 (2), 487–527.
Bushee, B.J., Core, J.E., Guay, W.R., Hamm, S., 2010. The role of the business press as an information intermediary. J. Account. Res. 48, 1–19.
Bushman, R.M., Piotroski, J.D., Smith, A.J., 2004. What determines corporate transparency? J. Account. Res. 42 (2), 207–252.
Campelo, M., Larrain, M., 2016. Enlarging the contracting space: collateral menus, access to credit, and economic activity. Rev. Financ. Stud. 29 (2), 349–383.
Cao, S., Ma, G., Tucker, J.W., Wan, C., 2018. Technological peer pressure and product disclosure. Account. Rev. 93 (6), 95–126.
Chalmers, K., Hay, D., Khlif, H., 2019. Internal control in accounting research: a review. J. Account. Lit. 42, 80–103.
Chan, L.K., Lakonishok, J., Sougiannis, T., 2001. The stock market valuation of research and development expenditures. J. Financ. 56 (6), 2431–2456.
Chen, X., Harford, J., Li, K., 2007. Monitoring: which institutions matter? J. Financ. Econ. 86 (2), 279–305.
Chen, X., Cheng, Q., Luo, T., Yue, H., 2022. Short sellers and insider trading profitability: a natural experiment. J. Account. Public Policy 41, 106936.
Chung, S.G., Goh, B.W., Lee, J., Shevlin, T., 2019. Corporate tax aggressiveness and insider trading. Contemp. Account. Res. 36 (1), 230–258.
Cohen, L., Lou, D., 2012. Complicated firms. J. Financ. Econ. 104 (2), 383–400.
Cohen, L., Malloy, C., Pomorski, L., 2012. Decoding inside information. J. Financ. 67, 1009–1043.
da Silva, S.M.T., de Carvalho Morais, A.I.A.P., Curto, J.D., 2013. Disclosure of R&D activities. Global Business Perspectives 1, 391–417.
Dai, L., Parwada, J.T., Zhang, B., 2015. The governance effect of the media's news dissemination role: evidence from insider trading. J. Account. Res. 53 (2), 331–366.
Davis, A.K., Tama-Sweet, I., 2012. Managers' use of language across alternative disclosure outlets: earnings press releases versus MD&A. Contemp. Account. Res. 29 (3), 804–837.
Dedman, E., Lennox, C., 2009. Perceived competition, profitability and the withholding of information about sales and the cost of sales. J. Account. Econ. 48 (2–3), 210–230.
Diamond, D.W., Verrecchia, R.E., 1991. Disclosure, liquidity, and the cost of capital. J. Financ. 46 (4), 1325–1359.
Diouf, D., Boiral, O., 2017. The quality of sustainability reports and impression management: a stakeholder perspective. Account. Audit. Account. J. 30 (3), 643–667.
Dvir, D., Segev, E., Shenhar, A., 1993. Technology's varying impact on the success of strategic business units within the miles and snow typology. Strateg. Manag. J. 14 (2), 155–161.
Dye, R.A., 1985. Disclosure of nonproprietary information. J. Account. Res. 123–145.
Dyer, T., Lang, M., Stice-Lawrence, L., 2017. The evolution of 10-K textual disclosure: evidence from latent dirichlet allocation. J. Account. Econ. 64 (2–3), 221–245.
Eberhart, A.C., Maxwell, W.F., Siddique, A.R., 2004. An examination of long-term abnormal stock returns and operating performance following R&D increases. J. Financ. 59 (2), 623–650.
Ellis, J.A., Fee, C.E., Thomas, S.E., 2012. Proprietary costs and the disclosure of information about customers. J. Account. Res. 50 (3), 685–727.
Entwistle, G.M., 1999. Exploring the R&D disclosure environment. Account. Horiz. 13 (4), 323–342.
Ertugrul, M., Lei, J., Qiu, J., Wan, C., 2017. Annual report readability, tone ambiguity, and the cost of borrowing. J. Financ. Quant. Anal. 52 (2), 811–836.
Frankel, R., Li, X., 2004. Characteristics of a firm's information environment and the information asymmetry between insiders and outsiders. J. Account. Econ. 37 (2), 229–259.
Fu, X., Kong, L., Tang, T., Yan, X., 2020. Insider trading and shareholder investment horizons. Finance 62, 101508.
Fung, M.K., 2006. R&D, knowledge spillovers and stock volatility. Account. Finance 46 (1), 107–124.
Gao, F., Lisic, L.L., Zhang, I.X., 2014. Commitment to social good and insider trading. J. Account. Econ. 57 (2–3), 149–175.
Gigler, F., 1994. Self-enforcing voluntary disclosures. J. Account. Res. 32 (2), 224–240.
Glaeser, S.A., Landsman, W.R., 2021. Deterrent disclosure. Account. Rev. 96 (5), 291–315.
Glaeser, S., Michels, J., Verrecchia, R.E., 2020. Discretionary disclosure and manager horizon: evidence from patenting. Rev. Acc. Stud. 25, 597–635.
Gu, F., Li, J.Q., 2007. The credibility of voluntary disclosure and insider stock transactions. J. Account. Res. 45 (4), 771–810.
Harris, M.S., 1998. The association between competition and managers' business segment reporting decisions. J. Account. Res. 36, 111–128.
Healy, P.M., Palepu, K.G., 1993. The effect of firms' financial disclosure strategies on stock prices. Account. Horiz. 7 (1), 1–11.
Healy, P.M., Palepu, K.G., 2001. Information asymmetry, corporate disclosure, and the capital markets: a review of the empirical disclosure literature. J. Account. Econ. 31 (1–3), 405–440.
Henry, E., Leone, A.J., 2016. Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. Account. Rev. 91 (1), 153–178.
Himmelberg, C.P., Petersen, B.C., 1994. R&D and internal finance: a panel study of small firms in high-tech industries. Rev. Econ. Stat. 76 (1), 38–51.
Hirshleifer, D., Teoh, S.H., 2003. Limited attention, information disclosure, and financial reporting. J. Account. Econ. 36 (1–3), 337–386.
Hu, N., Zhang, T., & Li, X. Narrative innovation, real innovative activities and earnings persistence: evidence from word2vec (September 30, 2018). Available at SSRN: https://ssrn.com/abstract=3289221.
Hu, C., Zhu, F., Qiu, H., 2020. Share pledges, risk management, and stake raising by large shareholders. J. Financ. Res. 9, 190–206. In Chinese.
Huang, Y.F., Chen, C.J., 2010. The impact of technological diversity and organizational slack on innovation. Technovation 30 (7), 420–428.
Huang, X., Nekrasov, A., Teoh, S.H., 2018. Headline salience, managerial opportunism, and over- and underreactions to earnings. Account. Rev. 93 (6), 231–255.
Huang, R., Li, L., Lu, L.Y., Wu, H., 2021. The impact of the leahy-smith America invents act on firms' R&D disclosure. Europ. Account. Rev. 30 (5), 1067–1104.
Huang, D., Liu, B., Chan, K.C., Chen, Y., 2023. Intended and unintended effects of mandatory R&D disclosure on innovation outcomes. Econ. Model. 119, 106144.
Huang, A.H., Zang, A.Y., Zheng, R., 2014. Evidence on the information content of text in analyst reports. Account. Rev. 89 (6), 2151–2180.
Huddart, S.J., Ke, B., 2007. Information asymmetry and cross-sectional variation in insider trading. Contemp. Account. Res. 24 (1), 195–232.
Jagolinzer, A.D., Larcker, D.F., Taylor, D.J., 2011. Corporate governance and the information content of insider trades. J. Account. Res. 49 (5), 1249–1274.
James, S.D., Shaver, J.M., 2016. Motivations for voluntary public R&D disclosures. Acad. Manage. Discoveries 2 (3), 290–312.
Jia, Y., Gao, X., 2021. Is managerial rent extraction associated with tax aggressiveness? evidence from informed insider trading. Rev. Quant. Finan. Acc. 56, 423–452.
Jones, D.A., 2007. Voluntary disclosure in R&D-intensive industries. Contemp. Account. Res. 24 (2), 489–522.
Ke, B., Huddart, S., Petroni, K., 2003. What insiders know about future earnings and how they use it: evidence from insider trades. J. Account. Econ. 35 (3), 315–346.
Kim, O., Verrecchia, R.E., 1994. Market liquidity and volume around earnings announcements. J. Account. Econ. 17 (1–2), 41–67.
Kogan, L., Papanikolaou, D., Seru, A., Stoffman, N., 2017. Technological innovation, resource allocation, and growth. Q. J. Econ. 132 (2), 665–712.
Kothari, S.P., Shu, S., Wysocki, P.D., 2009. Do managers withhold bad news? J. Account. Res. 47 (1), 241–276.
Kyle, A. S. (1985). Continuous auctions and insider trading. Econometr. J. Econometr. Soc. 1315-1335.
Lakonishok, J., Lee, I., 2001. Are insider trades informative? Rev. Financ. Stud. 14 (1), 79–111.

Lang, M., Lundholm, R., 1993. Cross-sectional determinants of analyst ratings of corporate disclosures. J. Account. Res. 31 (2), 246–271.
Lawrence, A., 2013. Individual investors and financial disclosure. J. Account. Econ. 56 (1), 130–147.
Lee, Y.J., 2012. The effect of quarterly report readability on information efficiency of stock prices. Contemp. Account. Res. 29 (4), 1137–1170.
Lee, I., Lemmon, M., Li, Y., Sequeira, J.M., 2014. Do voluntary corporate restrictions on insider trading eliminate informed insider trading? Finance 29, 158–178.
Lee, H., Park, Y., 2011. An international comparison of R&D efficiency: DEA approach. Asian J. Technol. Innov. 13, 207–222.
Li, F., 2010a. The information content of forward-looking statements in corporate filings: a naive bayesian machine learning approach. J. Account. Res. 48 (5), 1049–1102.
Li, X., 2010b. The impacts of product market competition on the quantity and quality of voluntary disclosure. Rev. Acc. Stud. 15, 663–711.
Lim, E.K., Chalmers, K., Hanlon, D., 2018. The influence of business strategy on annual report readability. J. Account. Public Policy 37 (1), 65–81.
Loughran, T., McDonald, B., 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. J. Financ. 66 (1), 35–65.
Loughran, T., McDonald, B., 2014. Measuring readability in financial disclosures. J. Financ. 69 (4), 1643–1671.
Makri, M., Hitt, M.A., Lane, P.J., 2010. Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions. Strateg. Manag. J. 31 (6), 602–662.
Meng, Q.B., Yang, J.H., Lu, B., 2017. Information content of management discussion and analysis disclosure and stock price crash risk - a study based on text-oriented quantitative method. China Ind. Economy 12, 132–150. In Chinese.
Merkl-Davies, D.M., Brennan, N.M., 2007. Discretionary disclosure strategies in corporate narratives: incremental information or impression management? J. Account. Lit. 27, 116–196.
Merkley, K.J., 2014. Narrative disclosure and earnings performance: evidence from R&D disclosures. Account. Rev. 89 (2), 725–757.
Miller, B.P., 2010. The effects of reporting complexity on small and large investor trading. Account. Rev. 85 (6), 2107–2143.
Myatt, D.P., Wallace, C., 2012. Endogenous information acquisition in coordination games. Rev. Econ. Stud. 79 (1), 340–374.
Nekhili, M., Hussainey, K., Cheffi, W., Chtioui, T., Tchakoute-Tchuigoua, H., 2016. R&D narrative disclosure, corporate governance and market value: Evidence from France. J. Appl. Bus. Res. 32 (1), 111–128.
Newman, P., Sansing, R., 1993. Disclosure policies with multiple users. J. Account. Res. 31 (1), 92–112.
Ofek, E., Yermack, D., 2000. Taking stock: equity-based compensation and the evolution of managerial ownership. J. Financ. 55 (3), 1367–1384.
Peng, M.W., 2004. Outside directors and firm performance during institutional transitions. Strateg. Manag. J. 25, 453–471.
Piotroski, J.D., Roulstone, D.T., 2005. Do insider trades reflect both contrarian beliefs and superior knowledge about future cash flow realizations? J. Account. Econ. 39 (1), 55–81.
Rawson, C., 2022. Manager perception and proprietary investment disclosure. Rev. Acc. Stud. 27 (4), 1493–1525.
Saidi, F., Zaldokas, A., 2021. How does firms' innovation disclosure affect their banking relationships? Manag. Sci. 67 (2), 742–768.
Shen, Y. F., Cu, W. H., & Li, P. G. (2011). Stake-raising: financial motivation vs. political motivation. Accounti. Res. 279(1):52-59. (In Chinese).
Skaife, H.A., Veenman, D., Wangerin, D., 2013. Internal control over financial reporting and managerial rent extraction: evidence from the profitability of insider trading. J. Account. Econ. 55 (1), 91–110.
Tetlock, P.C., 2010. Does public financial news resolve asymmetric information? Rev. Financ. Stud. 23, 3520–3557.
Verrecchia, R.E., 2001. Essays on disclosure. J. Account. Econ. 32 (1–3), 97–180.
Vig, V., 2013. Access to collateral and corporate debt structure: Evidence from a natural experiment. J. Financ. 68 (3), 881–928.
Yao, J.Q., Feng, X., Wang, Z.J., Ji, R.R., Zhang, W., 2021. Tone, sentiment and market impact: based on the financial sentiment dictionary. J. Manage. Sci. 24 (5), 26–46. In Chinese.
Zhang, G., 2001. Private information production, public disclosure, and the cost of capital: Theory and implications. Contemp. Account. Res. 18 (2), 363–384.
Zhang, W., 2015. R&D investment and distress risk. J. Empir. Financ. 32, 94–114.