# Disclosure and the Pace of Drug Development*

Colleen Cunningham† Florian Ederer‡ Charles Hodgson§ Zhichun Wang¶

**Abstract**

Policies that mandate disclosure of innovative project outcomes aim to increase innovation by limiting wasteful duplicative innovation. Yet, such policies change not only the ex-post information environment but also firms' ex-ante innovation incentives. Firms may slow down their own innovation efforts in anticipation of increased disclosure by others. We examine the innovation-related impacts of the 2017 FDA Final Rule amendment, which mandates disclosure of clinical trial results for pharmaceutical firms. We show that the policy hastened and increased disclosure of results for clinical trials post-completion, but also increased the time to completion of clinical trials, the time between early phases of clinical trials, and delays in development-related investments. We provide evidence consistent with mandated disclosure leading firms to wait to learn from their competitors. Our results suggest that mandating disclosure may slow innovation when there is value to waiting.

**JEL Classification**: O31, O34, L65, D83, I18

**Keywords**: Innovation, Disclosure, Drug Development, Learning, Competition

*We thank Panle Jia Barwick, Victor Bennett, Vivek Bhattacharya, Federico Ciliberto, Rob Clark, Jeff Dyer, Liran Einav, Matthew Gentzkow, Avi Goldfarb, Tim Gubler, Yutao He, Matt Higgins, Philip Haile, Mitsuru Igami, Tomas Larroucau, Anran Li, Alexander MacKay, Sally Sadoff, Andrew Steck, Nicholas Vreugdenhil, Stefan Wagner, and seminar participants at the AEA Meetings, the Alpine IO Symposium, ASU, the BYU Winter Strategy Conference, DISS, Toronto, Queen's, Stanford, UCSD, UVA, and Yale for helpful comments and suggestions. All errors are our own.

†University of Utah, Eccles School of Business, colleen.cunningham@utah.edu

‡Boston University Questrom School of Business, CEPR, ECGI, and NBER, ederer@bu.edu

§Yale Department of Economics and NBER, charles.hodgson@yale.edu

¶Yale Department of Economics, zhichun.wang@yale.edu

# 1 Introduction

Innovation relies on cumulative knowledge and learning from others' experiences (Nelson and Winter, 1982; Cohen and Levinthal, 1990; Griliches, 1992). However, firms often have incentives to withhold critical information about their research and development (R&D) efforts to maintain their competitive advantage (Arrow, 1962; Anton and Yao, 2002, 2004). To counteract such tendencies, policymakers have implemented mandatory disclosure policies designed to enhance knowledge diffusion, reduce redundant efforts, and promote broader innovation, particularly in industries where the development of new products has large welfare consequences (Jaffe and Lerner, 2001).

Mandated disclosure can enable firms to learn what does and does not work, and thereby reallocate their R&D resources more efficiently. However, if R&D investments and outcomes, such as the discovery of new technologies or products, are public, firms may free ride by learning from rivals' discoveries and delaying their own costly innovation efforts (Bolton and Farrell, 1990; Bolton and Harris, 1999). In this paper, we study how policies that mandate disclosure influence firms' disclosure decisions and, subsequently, their innovation behavior.

Our setting is drug development, an industry in which disclosure mandates are particularly relevant (Zarin et al., 2016, 2017) and in which we can track innovation project progress through clinical trial phase progression. Although firms have been required to publicly register clinical trials pre-initiation since the early 2000s and to disclose the results of selected trials since 2007, the latter requirements entailed relatively weak enforcement (DeVito et al., 2020). The 2017 FDA Final Rule amendment (henceforth Final Rule) broadened and clarified results reporting as of 2017 and expanded requirements for timely disclosure of clinical trial results, aiming to improve transparency in drug development.

Our empirical analysis documents that the Final Rule significantly increased the reporting of completed Phase 2 clinical trials, particularly those with unsuccessful outcomes. Before the policy change, only 68% of completed Phase 2 trials were disclosed within 24 months of completion. Following the implementation of the Final Rule, this figure rose to 81%, representing a 18% increase in compliance. The increase is largely driven by increased reporting of failed or inconclusive trials, with the rate of reporting of unsuccessful trial results increasing from 56% to 75% while the reporting rate for successful results increased from 91% to 95%. This empirical pattern is consistent with firms' strategic incentive to withhold R&D failures from competitors (unless they are forced to disclose them) because disclosure enables competitors to avoid redundant efforts and to adjust their innovation strategies accordingly (Krieger, 2021).

With expectations of increased disclosure by others, firms may adjust the timelines of their own innovation investments, and strategically delay when the value of waiting to learn outweighs the costs. Indeed, we find that the consequences of mandatory disclosure extend beyond increased disclosure. Firms also adapt their strategic decisions regarding the timing and structure of their own R&D pipelines. Following the policy's implementation, the completion time for Phase 2 clinical

trials increased significantly, with the median rising from approximately 700 days to around 950 days. At the same time, the probability of completing a trial within one year dropped from 20-25% pre-policy to just 15% post-policy. Moreover, firms exhibited a marked slowdown in transitioning from Phase 1 to Phase 2 trials. Prior to the Final Rule, nearly 70% of Phase 1 trials were followed by a subsequent trial within a year. The transition rate fell to below 50% post-policy. This delay is consistent with firms waiting for disclosed results from other trials before committing resources to new studies.[1] The data suggest that the enhanced disclosure environment altered firms' incentives, leading them to slow both trial progression and follow-up investments, likely as a strategic response to the increased availability of information.

The patterns so far provide consistent but still indirect evidence that firms are adopting a slower investment approach. To provide more direct evidence and to rule out shifting priorities across medical development areas as a potential explanation, we focus on within-trial investment behavior and examine how firms adjust their investment in active trials by adding clinical trial sites. Using detailed data from Phase 2 trials, we find that initiated trials after the policy change exhibited a substantial reduction in the number of active sites compared to those in the pre-policy period. Specifically, for trials that had been active for two years, the number of active sites declined by more than 35%, from an average of approximately 11 sites per trial before the policy to fewer than 7 sites per trial after its implementation. This trend suggests that firms, rather than aggressively expanding trials to generate results quickly, opted for a more measured approach, possibly to limit knowledge spillovers and maintain a competitive edge in an environment of heightened transparency.

Yet, even these more direct patterns remain merely suggestive of firms' purposefully waiting. While, helpfully, the changes appear discrete at the time of the policy rather than as continuous trends, because the above results rely on pre-post comparisons, the findings are open to alternative explanations, including systemic cost increases or trial-related technological change. To provide evidence of learning, we first examine how firms incorporate learning from disclosed trial outcomes into their innovation strategies. Abstracting from the policy effects, we find firms adjust their likelihood of starting new trials based on observed successes or failures of related drugs. Specifically, a one-standard deviation increase in successful Phase 2 trials of drugs with a similar mechanism of action increases the probability of a subsequent Phase 2 trial by 1.99 percentage points. Conversely, an increase in unsuccessful Phase 2 trials reduces the probability of a subsequent Phase 2 trial by 2.18 percentage points. These effects are present within the same mechanism of action rather than across broader disease groups, indicating that firms primarily learn from drugs that share core technological similarities. In contrast, the effects of trial outcomes from drugs in the same

[1] Around the same time as the FDA Final Rule, the International Council for Harmonisation (ICH) introduced its E6(R2) guideline amendments, adopted in late 2016 and rolled out gradually across jurisdictions in 2017–2018. The FDA issued guidance in March 2018, while the EU acted earlier in June 2017. These amendments primarily modernized Good Clinical Practice standards to reflect technological changes and aimed to standardize best practices internationally. The revisions did not substantively alter core trial processes or impose major new requirements likely to affect trial speed or costs for U.S.-based industry sponsors in our sample. As such, they are unlikely to explain the sharp break in investment behavior we observe in 2017.

disease group follow the opposite pattern. A one-standard deviation increase in successful Phase 2 trials by a competing drug in the same disease group reduces the probability of a Phase 2 Start by 4.32 percentage points while the same increase in unsuccessful Phase 2 trials increases it by 5.81 percentage points. We also find that firms tend to wait for new disclosures before proceeding with their own trials, suggesting that mandated transparency policies induce firms to free-ride on competitors' results rather than invest in their own exploratory research.

Finally, we show that firms strategically time their trial investments to capitalize on information from ongoing disclosures. The probability of completing a Phase 2 trial and the number of active trial sites is significantly lower when there are related Phase 2 trials that have completed but not yet reported results, implying that firms anticipate upcoming disclosures and delay investment accordingly. Importantly, this effect is only present in the post-policy period starting in 2017, when the Final Rule strengthened reporting requirements and increased the anticipated flow of information about trial outcomes. Controlling for group-time and disease fixed effects, trials that started in 2017 or after had about 0.9 fewer active sites on average than trials started before 2017 and this effect was amplified where at least one related drug was pending disclosure in the same mechanism-of-action group.

The broader implications of these results highlight the dual effects of disclosure mandates on innovation. On the one hand, increased transparency enhances knowledge diffusion, allowing firms to learn from each other's successes and failures, which can improve overall R&D efficiency and reduce wasteful duplication. However, this benefit comes with a countervailing force: the potential discouragement of risk-taking behavior in R&D. This trade-off underscores the complexity of disclosure mandates. Their net effect on innovation depends not only on regulatory enforcement but also on industry characteristics, such as the degree of competition and the strategic value of proprietary research outcomes.

Overall, our paper makes three contributions. First, it provides empirical evidence on how mandated disclosure policies influence firms' innovation behavior. Specifically, it shows that the 2017 FDA Final Rule amendment significantly increased the reporting of completed clinical trial results, particularly for unsuccessful Phase 2 trials, highlighting how transparency alters firms' disclosure incentives. Second, it documents an unintended consequence of disclosure mandates on innovation timing in the pharmaceutical industry, an R&D-intensive sector crucial to consumer and social welfare. Firms strategically delay clinical trial completions, particularly when there are competing trials with results (and their disclosure) pending, suggesting that mandatory transparency can introduce incentives to slow innovation. Third, the paper contributes to the broader policy debate on the trade-offs of disclosure regulation, demonstrating that while transparency facilitates knowledge diffusion and reduces redundant effort, it can also lead to strategic responses that temper the expected benefits of mandated disclosure.

Our paper contributes to the broader literature on firms' R&D incentives (Arrow, 1971; Dasgupta and Stiglitz, 1980; Spence, 1984), with a particular focus on the relationship between disclosure and

innovation. Specifically, it shows that mandatory disclosure of research outcomes improves the information available to firms when they make innovation decision but can also lead to delay due to free-riding incentives highlighted in previous theoretical work (Hendricks and Kovenock, 1989; Bolton and Farrell, 1990; Bolton and Harris, 1999). This particular trade-off is also the focus of several papers in the literature on oil & gas exploration, including Levitt (2009), Lin (2013), Covert (2015), Steck (2018), Fetter et al. (2018), Agerton (2020), Hodgson (2024), Covert and Sweeney (2022), and Fetter (2022).

In the context of drug development, our paper is related to previous contributions by Magazzini et al. (2012) and Krieger (2021) that investigate how much pharmaceutical companies learn from observing the drug trial failures of other firms. Competitive dynamics and externalities shape firms' innovation and disclosure decisions in pharmaceuticals. Kao (2024) shows that firms strategically withhold clinical trial results following a competitor's drug approval to maintain competitive advantage, while Dix and Lensman (2024) document that firms underinvest in combination therapies that involve other firms' drugs due to positive market expansion externalities, suggesting both selective disclosure and under-innovation stem from incentives to protect or free ride on competitive positions. More broadly, by studying the effect of disclosure mandates on inventive activity, our work is related to papers that study how changes in patent disclosure law affect knowledge diffusion and innovation (Scotchmer and Green, 1990; Scotchmer, 1991; Murray and Stern, 2007; Williams, 2013; Galasso and Schankerman, 2015; Hegde and Luo, 2018). Our findings contribute to this literature by highlighting how increased transparency requirements influence not only firms' decisions to disclose but also the timing and structure of their innovation efforts.

The rest of the paper is organized as follows. Section 2 analyzes a simple model of social learning, competition, and innovation delay that is intended to guide our empirical analysis. Section 3 provides background on the regulatory framework and the role of disclosure mandates in drug development. Section 4 describes our data sources. Section 5 presents our main findings on disclosure patterns and innovation responses. Section 6 shows how the empirical findings are related to firm learning. Section 7 and 8 develop and estimate a dynamic model of trial investment. Section 9 concludes.

## 2 A Simple Theoretical Framework

Innovation in drug development involves both uncertainty and strategic interaction. Firms must decide not only whether to invest in clinical trials, but also when to do so, taking into account both the potential to learn from others and the competitive consequences of their actions. In many settings, the outcome of one firm's experimentation provides valuable information about the prospects of related projects pursued by other firms. At the same time, firms compete for market share, so that simultaneous success may reduce the profitability of innovation.

Disclosure policies directly affect this environment by shaping how much firms can learn from one

another and when such learning occurs. By increasing the availability and timeliness of information about trial outcomes, disclosure can improve decision-making by reducing uncertainty. However, it can also alter firms' incentives over the timing of investment by increasing the value of waiting to observe the results of other firms' experimentation.

To capture these trade-offs, we introduce a simple theoretical framework that combines two key features: informational spillovers across firms and product market competition. The model highlights how disclosure influences firms' incentives to invest early versus delay. In particular, it shows that greater disclosure can increase the attractiveness of waiting, as firms anticipate learning from others' outcomes or from the resolution of competitive uncertainty. While highly stylized, the framework isolates the core mechanisms through which disclosure policies shape the pace of innovation in drug development. All the proofs for the model are in Appendix C.

### 2.1 Setup

There are two periods and two firms. In the first period, firms $f \in \{A, B\}$ each choose simultaneously whether to run a trial, $a_{f1} \in \{0, 1\}$. If $a_{f1} = 1$ then firm $f$ pays a cost, $c$, observes a signal $s_f \in \{0, 1\}$, where $\Pr(s_f = 1) = p$. With probability $\rho$, the signals $s_f$ and $s_{-f}$ are perfectly correlated. With probability $(1-\rho)$ they are independent. The parameter $\lambda \in [0, 1]$ captures the intensity of product market competition. When both firms succeed, each firm's payoff is scaled by $\lambda$, so lower values of $\lambda$ correspond to stronger business-stealing effects and a smaller effective market size. If $a_{f1} = 1$ then firm $f$ obtains a payoff $\pi_{f1}$ given by

$$\pi_{f1} = \begin{cases} s_f \pi \lambda & if\, a_{-f1} = 1 \;\&\; s_{-f} = 1 \\ s_f \pi & otherwise \end{cases}$$

where $p\pi - c > 0$. Therefore, it is optimal for a firm to invest in the trial if there is no competitor, but it may not be optimal if firm $f$ knows $a_{-f1} = 1$.

In the second period, the firm that did not invest in the first period can choose $a_{f2} \in \{0, 1\}$. There are two cases. First, if neither firm invested in the first period, then the game is the same as in the first period, and

$$\pi_{f2} = \begin{cases} s_f \pi \lambda & if\, a_{-f2} = 1 \;\&\; s_{-f} = 1 \\ s_f \pi & otherwise \end{cases}$$

Second, if $a_{-f1} = 1$, so firm $-f$ invested in the first period, then firm $f$ observes firm $-f$'s signal from the first period with the disclosure probability $\alpha$ and updates its beliefs about the distribution of $s_f$ as given by

$$\Pr(s_f = 1 | s_{-f} = 0) = (1 - \rho)p$$

$$\Pr(s_f = 1 | s_{-f} = 1) = \rho + (1 - \rho)p$$

Firm $f$ then chooses $a_{f2} \in \{0, 1\}$. If $a_{f2} = 1$ then firm $f$ pays a cost $c$, observes a signal $s_f \in \{0, 1\}$,

and obtains a payoff $\pi_{f2}$ given by

$$\pi_{f2} = \begin{cases} s_f \pi & if\ s_{-f} = 0 \\ s_f \pi \lambda & if\ s_{-f} = 1 \end{cases}$$

Firm $-f$ which invested in the first period receives a payoff of $\pi_{-f2}$, where:

$$\pi_{-f2} = \begin{cases} s_{-f} \pi \lambda & if\ a_{f2} = 1\ \&\ s_f = 1 \\ s_{-f} \pi & otherwise \end{cases}$$

Both firms discount the second period by factor $\delta \in (0, 1)$.

### 2.2 Delay from Informational Free Riding

We first consider how informational free riding affects development decisions. To isolate this effect, we set $\lambda = 1$ such that there is no product market competition between the two firms.

**Proposition 1.** *If $\lambda = 1$, then there is a threshold, $\rho^*$ such that for $\rho < \rho^*$ there is a unique pure strategy SPNE in which both firms set $a_{f1} = 1$. For $\rho > \rho^*$ there are two pure strategy SPNE, one where ($a_{A1} = 1$, $a_{B1} = 0$) and one where ($a_{B1} = 1$, $a_{A1} = 0$), and a mixed-strategy equilibrium where $\Pr(a_{f1} = 1) = \gamma \in (0, 1)$.*

Proposition 1 formalizes a simple but important intuition: when firms can learn sufficiently from each other's experimentation, they may prefer to wait rather than invest immediately. The parameter $\rho$ governs how informative another firm's trial is about one's own prospects. When $\rho$ is low, signals are largely independent, and there is little to be gained from waiting. In this case, both firms optimally invest in the first period, as each firm's trial generates primarily private information.

In contrast, when $\rho$ is high, the outcome of one firm's trial is highly informative about the other firm's prospects. In this case, a firm can benefit from postponing its own investment decision until it observes the outcome of its rival's trial. This creates a coordination problem: each firm prefers that the other conducts the initial experiment while it waits and learns. As a result, the model admits equilibria in which only one firm invests in the first period, as well as mixed-strategy equilibria in which firms randomize over early investment. The emergence of these equilibria reflects a classic informational free-riding problem.

When learning spillovers are strong, delaying investment becomes privately optimal even though it may slow down the overall pace of innovation. In particular the incentive to delay due to informational free riding also increases with the disclosure probability $\alpha$.

**Proposition 2.** *If $\lambda = 1$, $\frac{\partial \gamma}{\partial \alpha} < 0$.*

An increase in $\alpha$ raises the expected benefit of waiting, as it increases the likelihood that a firm can condition its second-period decision on informative signals generated by others. As a result, the

equilibrium probability of early investment $\gamma$ decreases with $\alpha$. Intuitively, greater transparency strengthens incentives to delay because it makes learning from others more reliable. This comparative static captures the central mechanism we explore empirically: policies that increase disclosure can induce firms to postpone investment in order to free ride on the information generated by competitors.

### 2.3 Delay from Competition

We now turn to analyzing how competition for the market also influences the incentives to delay clinical trials. To isolate this effect, we consider the case where there is no social learning ($\rho = 0$).

**Proposition 3.** *If $\rho = 0$, then there is a threshold, $\lambda^*$ such that for $\lambda > \lambda^*$ there is a unique pure strategy SPNE in which both firms set $a_{f1} = 1$. For $\lambda < \lambda^*$ there are two pure strategy SPNE, one where ($a_{A1} = 1$, $a_{B1} = 0$) and one where ($a_{B1} = 1$, $a_{A1} = 0$), and a mixed-strategy equilibrium where $\Pr(a_{f1} = 1) = \gamma_1 \in (0, 1)$ and $\Pr(a_{f2} = 1) = \gamma_2 \in (0, 1)$.*

The parameter $\lambda$ captures the intensity of competition when both firms succeed. When $\lambda$ is close to one, competition is weak and firms can profitably coexist. In this case, both firms optimally invest early, as there is little downside to simultaneous success. However, when $\lambda$ is low, competition is intense and the market effectively supports only one profitable entrant. In this environment, firms have an incentive to delay investment in order to avoid incurring costs when the rival may already have secured the market.

As in Proposition 1, this leads to equilibria in which firms stagger their investments, with one firm investing early and the other waiting. The underlying logic differs, however. Delay is driven not by learning but by the desire to avoid wasteful duplication under strong business-stealing effects. If the market is "small" such that only one firm can enter profitably ($\lambda < \lambda^*$), then there is an incentive to delay the cost of investment, $c$, until you are sure that the rival firm has failed.

**Proposition 4.** *If $\rho = 0$, $\frac{\partial \gamma_1}{\partial \alpha} < 0$.*

Proposition 4 shows that disclosure also amplifies delay in this competitive setting: the probability of delay increases ($\gamma_1$ decreases) when the disclosure probability increases. Even in the absence of informational spillovers ($\rho = 0$), observing whether a rival has succeeded allows a firm to condition its investment decision on the state of competition. An increase in $\alpha$ raises the likelihood that a firm learns whether the rival has already captured the market, thereby increasing the value of waiting.

Consequently, the probability of early investment declines as disclosure increases. This result highlights that the effect of disclosure on delay is not limited to informational free-riding. Even when firms cannot learn about their own success probabilities from others, disclosure can still induce delay by revealing information about competitive conditions.

### 2.4 Welfare

To evaluate the welfare implications of disclosure, we adopt a utilitarian welfare criterion that sums firms' expected discounted payoffs and consumer surplus from successful innovation. Specifically, each successful project generates consumer surplus $s$ in the period in which success occurs. This captures the idea that innovation creates value not only through firm profits but also through improved health outcomes and earlier access to effective treatments.

Disclosure affects welfare through two opposing forces. On the one hand, greater disclosure improves the *screening* of investment by allowing firms to condition their decisions on more informative signals, thereby reducing inefficient experimentation. On the other hand, disclosure increases the *option value of waiting*, leading firms to delay investment in order to learn from others. Because both profits and consumer surplus are discounted, such delays are socially costly as they postpone the realization of value from successful innovation.

**Proposition 5.** *Suppose $\lambda = 1$ and welfare is defined as the sum of firms' expected discounted payoffs and consumer surplus. There exist parameter values $(\delta, \rho, p, \pi, c, s)$ such that the socially optimal disclosure probability chosen by a social planner satisfies $\alpha^{SP} < 1$.*

Proposition 5 shows that full disclosure need not be socially optimal. While disclosure improves the allocation of investment through better screening, it also induces delay by raising the value of waiting. When the social cost of delay which arises from the postponed realization of both profits and consumer benefits, outweighs the incremental gains from improved screening, the planner prefers a disclosure regime with $\alpha < 1$.

A sufficiently large consumer surplus $s$ is necessary for full disclosure to be suboptimal. Higher values of $s$ increase the social cost of delay, as postponing successful innovation delays not only firm profits but also consumer benefits. The result is also more likely to obtain when discounting is strong (i.e., $\delta$ is low), since the cost of delaying experimentation is then larger. Finally, the strength of learning, governed by $\rho$, plays a central role. When $\rho$ is high, trial outcomes are highly informative, which increases the screening benefits of disclosure but also strengthens the incentive to delay with the latter effect dominating when the value of earlier innovation is sufficiently high.

### 2.5 Mapping Theory to Empirics

Propositions 1 through 4 show that increased disclosure can slow investment through two complementary channels: informational free-riding and strategic delay under competition. In both cases, greater transparency raises the option value of waiting relative to immediate investment.

Our theoretical framework thereby provides a simple mapping between model primitives and the empirical objects we study. The disclosure parameter $\alpha$ corresponds to the change induced by the Final Rule, which increases the likelihood and timeliness with which firms observe the outcomes of other trials. The informativeness parameter $\rho$ captures the extent to which outcomes of related trials provide useful information for a firm's own project; empirically, this is reflected in learning

across drugs that share a mechanism of action. Finally, the competition parameter $\lambda$ governs the degree of business-stealing across firms and is most naturally linked to interactions among drugs which treat the same disease. The model predicts that an increase in $\alpha$ raises the option value of waiting, leading firms to delay investment, particularly in environments where learning opportunities (high $\rho$) or competitive pressures (low $\lambda$) are strong. These predictions guide our empirical analysis, where we examine how changes in disclosure affect the timing of trial starts, trial completion, and within-trial investment, as well as how these effects vary with the scope for learning and competition.

# 3 Industry and Institutional Background

## 3.1 Drug Development Background

New drug development involves several steps. After identifying potentially promising drug compounds through routine discovery processes, firms run preliminary screening in vitro and/or in vivo to explore both efficacy and toxicity before any clinical trials in humans. Following these preclinical evaluations, drugs undergo clinical trial testing in humans. The U.S. Food and Drug Administration (FDA) regulates and evaluates the safety and effectiveness of drugs pre- and post-market. Drug developers must submit an Investigation New Drug (IND) application to the FDA before starting clinical trials, which must include animal study and toxicity data, manufacturing information, clinical protocols (i.e., study plans), data from any prior human research, and information about the investigator. The FDA provides a detailed overview of the process (https://www.fda.gov/patients/drug-development-process/step-3-clinical-research) which we briefly describe below.

The clinical trial process is systematically divided into three phases. Phase 1 trials mark the first introduction of a drug into humans, typically involve 20 to 100 healthy volunteers or individuals with the targeted condition, and last several months. The primary goal is to assess safety, determine appropriate dosage ranges, and monitor how the drug interacts with the human body. About 70% of drugs successfully complete this phase. Next, Phase 2 trials expand the participant pool to several hundred individuals with the disease or condition under investigation. Conducted over a period of up to two years, these studies focus on evaluating efficacy while continuing to monitor safety. Only about a quarter to a third of drugs entering in this phase demonstrate sufficient promise to move forward.[2] Phase 3 trials, which last one to four years, involve 300 to 3,000 participants. These studies aim to confirm a drug's effectiveness, monitor side effects, and compare it to existing treatments. Phase 3 trials typically form the basis of FDA approval decisions. In the range of a half to two-thirds of drugs that enter Phase 3 trials are launched.[3] Overall, about 10% of drugs that enter clinical trials eventually enter the market. A significant proportion of drugs do not advance beyond each phase due to various challenges encountered during clinical testing. The continuation

[2]These numbers differ slightly from Mullard (2016) who report lower success rates.

[3]Further, according to the FDA, approximately 25-30% of drugs also run Phase 4 trials, which focus on post-approval safety monitoring.

rates across phases can vary depending on numerous factors, including the drug's therapeutic area and observed safety and efficacy outcomes.

Drug development is costly. Each stage entails substantial financial and time investments. Estimates for the total costs of bringing a new drug to market range from just over $100 million to more than $1 billion DiMasi et al. (2003), Adams and Brantner (2006), Morgan et al. (2011), and Dubois et al. (2015).

### 3.2 Policy Changes

For several decades FDA policies have focused on encouraging the public disclosure of clinical trial registrations and trial results. The FDA first required the registration of selected clinical trials in 1997, with the FDA Modernization Act. This requirement applied to all trials conducted under an IND application for a drug to treat a serious or life-threatening disease or condition and for trials to test drug effectiveness. The policy led the National Institutes for Health to create the online trial repository `ClinicalTrials.gov`, which was made publicly available on February 29, 2000.

Two subsequent policies significantly increased clinical trial registrations. First, in 2005 the International Committee of Medical Journal Editors (ICMJE) began requiring trial registration at the time of trial initiation as a condition of later journal publication. Immediately following the implementation of the ICMJE policy, trial registrations increased by 73% (Zarin et al., 2005). After the initial jump, new trial registrations per week normalized to 220 in 2006 and 2007, up from 30 in the pre-ICMJE policy period (Zarin et al., 2007). Second, in 2007 the FDA Amendments Act (FDAAA) required trial registration for Phase 2 and later trials for products subject to FDA regulation. The FDAAA also required submission of summary results and outlined penalties for noncompliance, including fines of $10,000 per day of noncompliance and the withholding of NIH grant money. However, while registrations increased as a result, the regulations around timely posting of results were vague and not enforced (Zarin et al., 2017; DeVito et al., 2020). As a result, the reporting of results was spotty (DeVito and Goldacre, 2021). After the 2007 FDAAA came into effect, only 39.5% of trials reported results within a year and 66.4% of registrations were late within a year for applicable clinical trials.

Further, and importantly, trials that involved drugs that were neither approved at the time of the trial nor subsequently approved by the FDA were not required by the 2007 FDAAA to report results. Subsequently, the FDA added the "Final Rule" (FR) amendment to the FDAAA, which was announced in 2014 and enacted in 2017.[4] The rule was designed to improve transparency in medical research and ensure greater public access to clinical trial data (Zarin et al., 2016, 2017). The FR has several important features. First, it requires all applicable clinical trials (ACTs) to post their results in a timely manner *regardless* of FDA approval status. This rule applies to trials with a primary completion date on or after January 18, 2017. Second, the FR also added requirements

[4]For details see `https://www.federalregister.gov/documents/2016/09/21/2016-22129/clinical-trials-registration-and-results-information-submission`.

to registrations which had implications for the details of posted results. Third, the FR specifies that for each outcome measure, registrations must include details on the metric used (e.g., percent change in total cholesterol), time points of assessment (e.g., 4 and 6 weeks), and increasing clarity on what is being measured. Fourth, the FR mandates the inclusion of baseline characteristics such as race, ethnicity, and other measures relevant to primary outcome analysis, ensuring that trial results are properly interpreted. Additionally, it strengthens adverse event reporting by making previously optional fields mandatory and requiring detailed documentation of deaths. Finally, to enhance study transparency, the regulation mandates the submission of full protocols and statistical analysis plans (SAPs), making study designs and amendments readily accessible for evaluation.

In what follows we evaluate how the FR affected the timely disclosure of trial results and how expectations of increased disclosure influenced subsequent innovation-related decisions.

# 4 Data

Our analysis uses two data sources, Pharmaprojects and Trialtrove, both purchased from Informa. Pharmaprojects tracks pharmaceutical development projects at the drug level using data collected directly from pharmaceutical companies and researchers (Blume-Kohout and Sood, 2013) and from public sources (press releases, patent filings, conference proceedings, regulatory agencies' reports, and the medical literature) and has been used in earlier research studying drug development, including among many others Kyle (2007), Blume-Kohout and Sood (2013), Adams and Brantner (2006), Branstetter et al. (2014), and Cunningham et al. (2021). Pharmaprojects tracks all candidate drugs developed or under development for eventual sale in the US market.

Trialtrove includes detailed data on individual clinical trials, including phase, the date the trial was started, the date the trial was completed, the date the results were made public, and details about trial outcomes. We merge the two datasets using a crosswalk file provided by Informa. Clinical trials often test combinations of drugs. We use the crosswalk to associate each trial with the "primary" drug being tested, as recorded in Trialtrove. We classify each trial result as either successful or unsuccessful. The former category includes only positive trial results whereas the latter contains negative and inconclusive results and early terminations.[5] The resulting dataset allows us to track the history of clinical trial starts, completions, and outcome reporting for each drug in the sample.

Table 1 summarizes the data on trial length and reporting for the set of all drugs for which some industry-sponsored pre-market clinical trials are completed between 2010 and 2019. The first set of rows records summary statistics on trial length. The mean Phase 1 trial runs for close to 440 days or about 15 months. Phase 2 and 3 trials are longer, about two and a quarter years on average. There is significant variance in trial length for all phases. The duration of the trial is largely a function of the target outcome variable. For example, trials intended to test for same-day reactions

[5]For the details of our classification procedure see Table A1 in the appendix.

**Table 1:** Trial Sequence Data: Summary Statistics

| | | Mean | SD | Median | Min | Max | N |
|---|---|---|---|---|---|---|---|
| Trial Length | Phase 1 | 440.22 | 527.72 | 218 | 1 | 5,660 | 9,787 |
| | Phase 2 | 831.41 | 643.94 | 669 | 1 | 9,953 | 8,524 |
| | Phase 3 | 848.92 | 658.24 | 669 | 8 | 6,489 | 7,066 |
| Reported | Phase 1 | 0.21 | 0.41 | 0 | 0 | 1 | 9,802 |
| | Phase 2 | 0.72 | 0.45 | 1 | 0 | 1 | 8,528 |
| | Phase 3 | 0.81 | 0.39 | 1 | 0 | 1 | 7,067 |
| Reporting Delay | Phase 1 | 486.90 | 473.73 | 350 | 1 | 3,683 | 4,931 |
| | Phase 2 | 432.32 | 456.73 | 322 | 1 | 3,908 | 6,528 |
| | Phase 3 | 364.20 | 355.00 | 290 | 2 | 3,479 | 5,946 |
| | Time To Approval | 1,805.37 | 998.12 | 1,670 | 38 | 9,681 | 915 |
| Approved Drugs | Phase 1 Count | 2.24 | 3.75 | 1 | 0 | 26 | 915 |
| | Phase 2 Count | 1.62 | 2.64 | 1 | 0 | 25 | 915 |
| | Phase 3 Count | 2.67 | 3.30 | 2 | 0 | 30 | 915 |
| | Share Reported | 0.65 | 0.33 | 0.68 | 0 | 1 | 915 |
| | Share Successful | 0.66 | 0.33 | 0.71 | 0 | 1 | 859 |
| Unapproved Drugs | Phase 1 Count | 1.26 | 1.77 | 1 | 0 | 23 | 4,823 |
| | Phase 2 Count | 0.85 | 1.59 | 0 | 0 | 36 | 4,823 |
| | Phase 3 Count | 0.38 | 1.32 | 0 | 0 | 37 | 4,823 |
| | Share Reported | 0.41 | 0.41 | 0.33 | 0 | 1 | 4,823 |
| | Share Successful | 0.37 | 0.41 | 0.20 | 0 | 1 | 3,956 |

Notes: Sample includes all industry sponsored pharmaceutical trials completed between 2010 and 2019. Trial length, reporting delay, and time to approval are reported in days.

will naturally be shorter than trials intended to measure 20-year mortality.

The second and third sets of rows record the share of trials for which results are reported, and the average delay between trial completion and reporting among reported trials. Only about 20% of Phase 1 trials have reported results. Results are much more frequently reported for phases 2 and 3, with reporting rates around 70% and 80% respectively. This finding is due to the exclusion of Phase 1 trials from FDA reporting requirements. The mean delay between trial completion and reporting is about one year or slightly more for all trial phases with significant dispersion. To benchmark this delay, we report the distribution of the time between the start of a drug's first observed clinical trial and FDA approval. The mean approved drug took around 5 years, with some drugs taking up to 26 years to obtain approval.

The final two sets of rows of Table 1 record the average number of trials per drug for drugs that were ultimately approved and drugs that have not been approved. For approved drugs, these statistics only count trials that took place before approval.[6] Approved drugs have more trials in all

[6]The data includes all trials including those on previously approved drugs. These trials are typically for new indications (diseases) or new drug combinations. In the analysis below we include these post-approval trials. Note

**Table 2:** Site Count Data: Summary Statistics

| | Mean | SD | Median | Min | Max | N |
|---|---|---|---|---|---|---|
| Trial Length | 27.80 | 19.06 | 23 | 1 | 145 | 5,280 |
| Any Active Sites | 0.61 | 0.49 | 1 | 0 | 1 | 5,280 |
| Active Site Count | 21.66 | 35.15 | 7 | 0 | 748 | 39,030 |
| Number of Changes | 2.03 | 2.42 | 1 | 1 | 41 | 3,201 |
| Change Magnitude | 14.24 | 30.01 | 3 | -169 | 518 | 6,446 |
| Maximum Sites | 30.05 | 40.66 | 16 | 1 | 748 | 3,201 |

Notes: Sample includes all US industry sponsored Phase 2 trials that started between 2010 and 2019. Active site count is measured at the trial-month level. Changes are recorded by comparing the number of active sites for a trial in consecutive months. Statistics on change magnitude include all non-zero changes in the data. Statistics on number of changes and maximum sites are measured at the trial level. Trial length is measured in months.

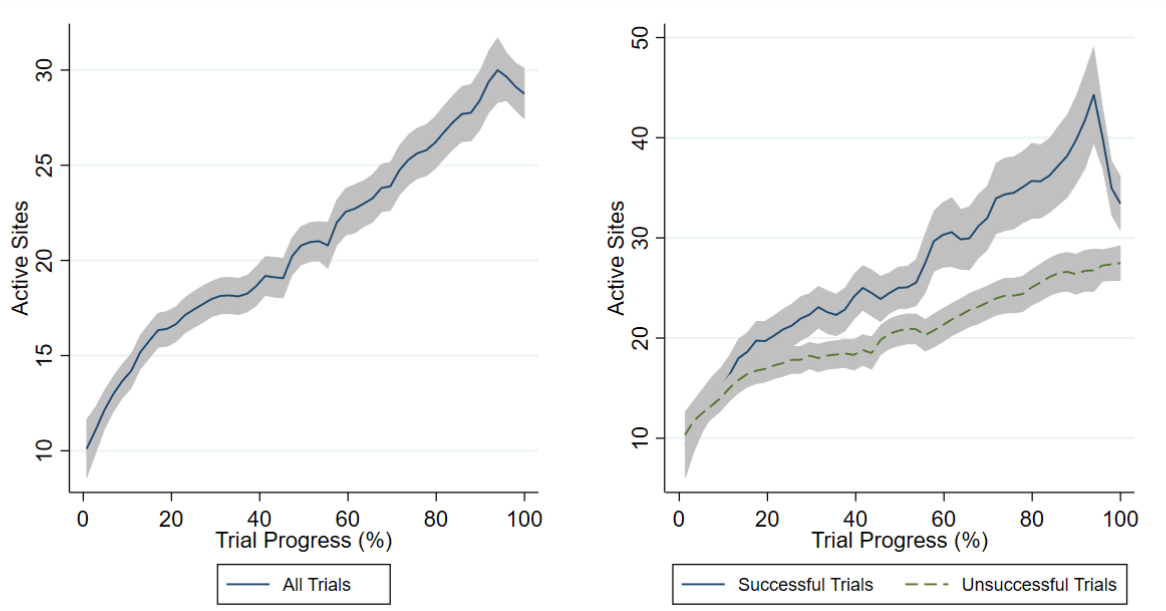
phases on average than unapproved drugs. Moreover, most trials on drugs that are not approved are in Phase 1 or Phase 2, whereas the modal trial on an approved drug is in Phase 3. This suggests that firms often abandon drug projects before Phase 3, and that the results of Phase 1 and 2 trials, which concern both safety and efficacy, contain important information for this decision.

We supplement the trial-level data with within-trial data collected from ClinicalTrials.gov. In addition to deciding when to start clinical trials, an important magnitude of investment for firms is the size of a trial (i.e, how many patients are enrolled in the trial). We measure trial size using the number of *active sites* recorded in each trial's entry on ClinicalTrials.gov. An active site is a location (e.g., a hospital) where patients are recruited, treated, and outcomes are measured. We construct a time series of the number of active sites over the duration of each trial using the change log of each trial's ClinicalTrials.gov entry. We therefore observe, at any point in time after the start of a trial, the number of active sites and consequently the timing of site additions and removals. We collected these data for all Phase 2 trials in our main Trialtrove dataset that we were able to match to ClinicalTrials.gov.

Table 2 summarizes our site count data. There are 5,280 Phase 2 trials in the data, with an average trial length of about 28 months. A notable feature of the data is that 39% of trials have no recorded active sites at any point during their duration (i.e., the "Any Active Sites" indicator is zero for those trials). This arises for two reasons. First, many trials are terminated early, before any sites become active. Second, the reporting of site information to ClinicalTrials.gov has been incomplete, particularly in earlier years of our sample. Among trials that do have some site information, the average number of active sites at the trial-month level is about 22, with some trials operating across hundreds of sites. The number of active sites also varies over time within a trial, changing an average of about 2 times over a trial's duration. On average, site counts increase over time, with the average addition being about 14 sites per change, although sites are also sometimes

that even for drugs that were ultimately approved, many trials have unreported results. Per the Code of Federal Regulations Code of Federal Regulations (2025), all relevant clinical trial data must be submitted to the FDA. However, there is no requirement for *public* disclosure of trial results at the approval stage.

**Figure 1:** Within-trial Addition of Active Sites



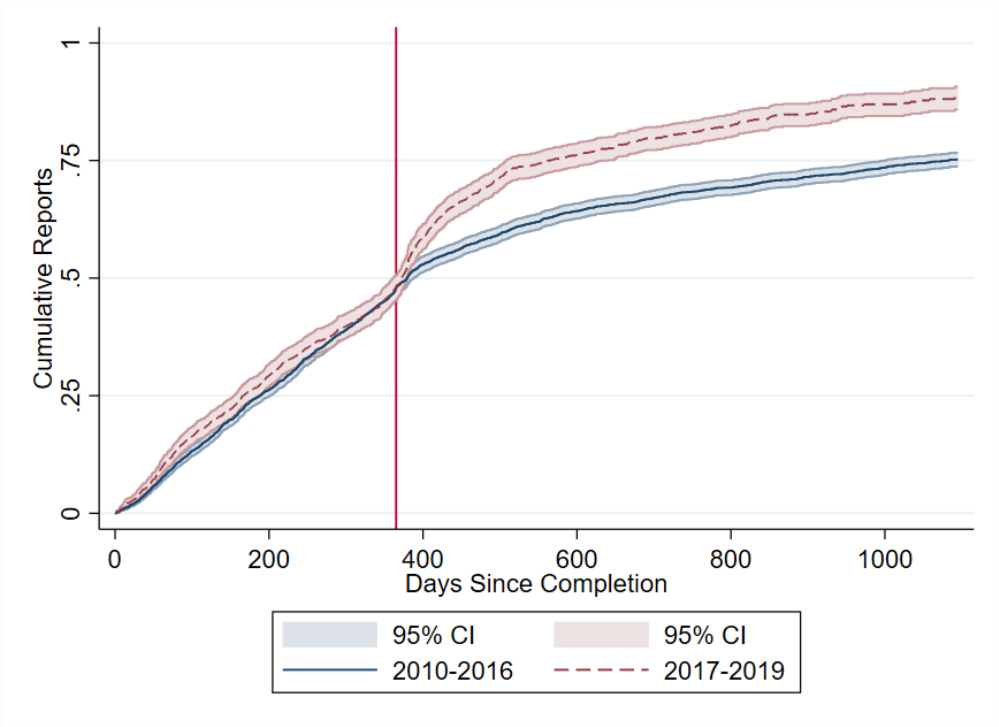

Notes: Local polynomial regression of active site count on trial progress. Sample includes all US industry sponsored Phase 2 trials that started between 2010 and 2019. Trial progress defined in the text.

removed. These descriptive statistics are consistent with trials often starting at a small scale and ramping up in size over time.

The prevalence of trials with no recorded site information is a potential concern for any analysis that relies on site counts as a measure of investment, since the missing data could in principle bias the comparison between pre- and post-policy periods. Appendix Table A2 reports the share of trials with no recorded active sites for different subsamples, allowing us to assess whether this is a concern. While about 29% of successful trials have no site records, the key finding is in the last column: the share of successful trials with no recorded sites does not change significantly across the two policy periods (the coefficient on Start Year $\geq 2017$ is small and not statistically significant). This indicates that the missing data does not differentially affect our pre- and post-policy comparisons for the trials that are most central to our analysis.

To further illustrate the evolution of the number of active sites over the duration of a trial, Figure 1 plots a local polynomial regression of the number of active sites on a measure of *trial progress*, $p_{in}$. If a trial $i$ lasts $N_i$ months in total, then at month $n$ after the start of the trial, trial progress is $p_{in} = 100\frac{n}{N_i}$. The results show that the number of active sites tends to increase over the duration of a trial. The average number of active sites increases monotonically from around 10 at the 0% progress to around 30 at 100% progress. Finally, there is a small decrease in the number of active sites as trials are close to completion.

**Figure 2:** Cumulative Results Disclosure




Notes: Kaplan-Meier failure curves recording the cumulative probability of results reporting for Phase 2 industry trials with at least one site in the United States. The two time periods refer to the year of trial completion which is relevant for reporting.

# 5 Empirical Analysis

## 5.1 Disclosure of Trial Results

We first investigate whether the 2017 policy change led to a meaningful shift in the information environment for drug development projects. We compare reporting patterns before and after the policy change. Figure 2 presents estimated cumulative density functions of the time between trial completion and results reporting for completed industry-sponsored Phase 2 trials.

We separately report the distributions for trials completed before the policy change (2010-2016) and after the policy's implementation (2017-2019). The CDFs are Kaplan-Meier failure curves where the censoring date is the date on which the database entry for each drug was last updated.

The results suggest that the policy change substantially increased the rate of reporting. The share of Phase 2 trials with reported results two years after completion is 81% for trials that were completed in 2017-2019, up from 68% for 2010-2016. In Appendix Figure A1 we repeat this exercise for non-industry trials. Among these trials, the two-year disclosure probability increases from around 21% to around 65%.

The vertical red line in Figure 2 marks the point at which one year has passed after trial completion. This point marks the post-2017 reporting deadline after which penalties for non-reporting

of trials apply. For the first year the probability of reporting is initially very similar for the pre- and post-policy periods. Only after the one-year deadline do these reporting probabilities significantly diverge from each other. In addition, the abrupt increase in the red line shows that there is a significant mass of trials in the post-2017 period that were reported at or immediately following the deadline which suggests that the policy change and associated penalties for late reporting of trial results indeed caused the shift in reporting patterns.

Figure 3 plots histograms of the number of days until the outcome of a completed Phase 2 industry trial is reported for the post-policy period. These histograms are plotted separately for successful and unsuccessful trials. First, positive trials are reported sooner than negative trials as is evident from the red bars for successful trials exceeding the green bars for unsuccessful trials until about 200 days after completion of a trial. In the post-policy period, the mass of all trials reported following the one-year deadline that leads to the divergence visible in Figure 2, is driven by a large spike in the reporting of unsuccessful trials.

Among successful trials, the share with results reported two years after completion is 91% for trials completed between 2010 and 2016, and 95% for trials completed after 2017. For unsuccessful trials, the corresponding shares are 56% for 2010-2016 and 75% for 2017-2019.

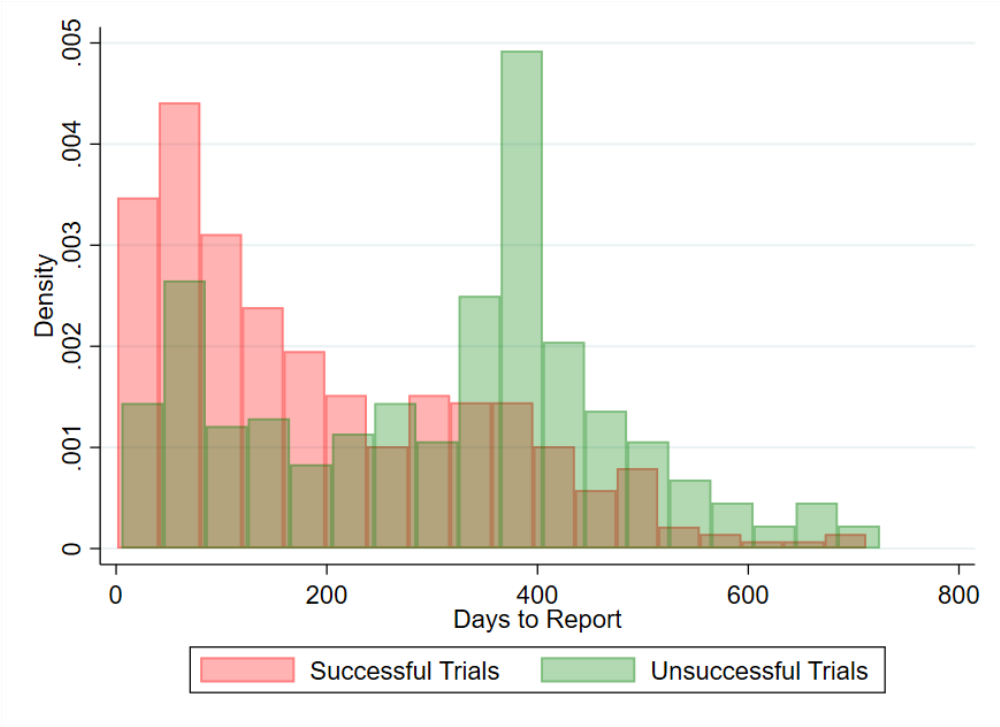
Together with Figure 2, these results suggest that the enforcement of the one-year reporting deadline brought results reporting forward in time and that this aggregate increase in reporting was primarily due to increased reporting of trials with negative outcomes.

### 5.2 Time to Start and Time to Complete Trials

Our results so far show that the final rule increased the speed of disclosure of clinical trial results. We now investigate whether this reporting requirement and the resulting change in the information environment also led to changes in firms' innovation and development choices. Firms may have an incentive to delay investment in clinical trials if they expect to learn from the results of other trials which is more likely under the new reporting rules. Such delay may come from several sources including slower completion of trials or slower transitions between different development phases.

We first show that under the new reporting regime trials take substantially longer to complete and are less likely to complete within a short time frame. The left panel of Figure 4 records the cumulative completion rate and the associated confidence intervals of Phase 2 trials over time for the two policy periods. For trials started in the pre-policy period, the median trial takes around 700 days to complete. The completion rate is substantially lower in the post-policy period. For trials started in the post-policy period, the median time to completion is around 950 days, roughly 250 days longer. This increase in the time to completion of Phase 2 trials is also apparent in the right panel of Figure 4 which reports point estimates and confidence intervals for the one-year completion probability by trial start year with the red vertical line denoting the time when the new reporting policy went into effect. Whereas in the pre-policy period between 20% and 25% of Phase 2 trials were completed within one year of their start, the same completion probability dropped to only 15%

**Figure 3:** Results Disclosure by Trial Outcome 2017-2019



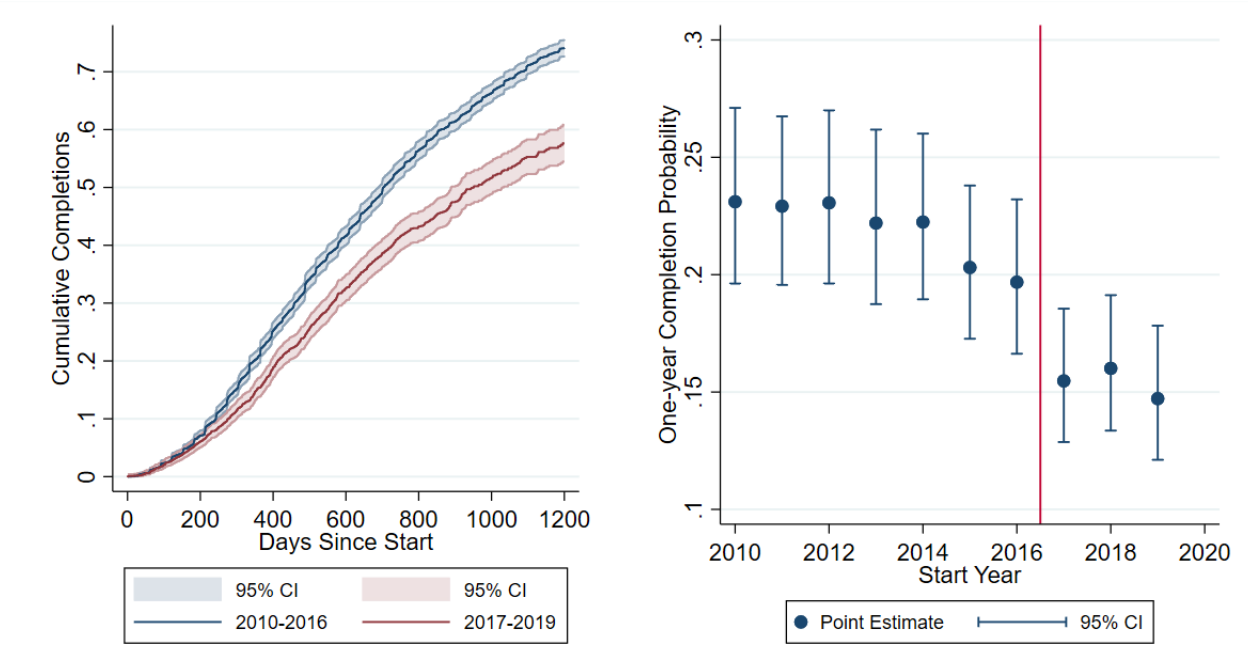

Notes: Histograms of days to report for US Phase 2 industry trials completed between 2017 and 2019 with reported outcomes. Histogram plotted separately for successful and unsuccessful trials. Trials with time to report greater than two years are dropped.

for trials begun in 2017 and remained stable for trials started in later years. Although the decrease in the one-year completion probability is particularly pronounced right after the final rule went into effect in 2017, there is also a smaller decrease for trials started in 2015 and 2016 after the final rule was announced but had not yet taken effect.

Appendix Figure A3 provides complementary evidence on the behavior of non-industry trials in response to the Final Rule. Panel (i) and Panel (ii) show that, unlike industry-sponsored trials, there is markedly less or very little evidence of a post-2017 slowdown in the rate of trial completions among non-industry trials, especially when compared to industry-sponsored trials. In contrast, Appendix Figure A1 demonstrates a clear increase in timely result disclosure for this group following the implementation of the Final Rule. These patterns suggest that non-industry sponsors complied with the new reporting requirements and were thus responsive to the regulatory change. However, their investment behavior remained stable, likely reflecting a lesser degree of strategic responsiveness to the incentives created by disclosure. This divergence supports the interpretation that the slowdown in trial investment we observe for industry trials stems from forward-looking, strategic behavior, rather than from mechanical compliance with the policy.

Given the changed reporting rules and increased availability of information, firms may also learn more from the successes and failures of other firms' trials about the likely success of trials which

**Figure 4:** Time to Complete Phase 2




**(i)** Time to Complete by Policy Period **(ii)** One-year Completion Probability by Start Year

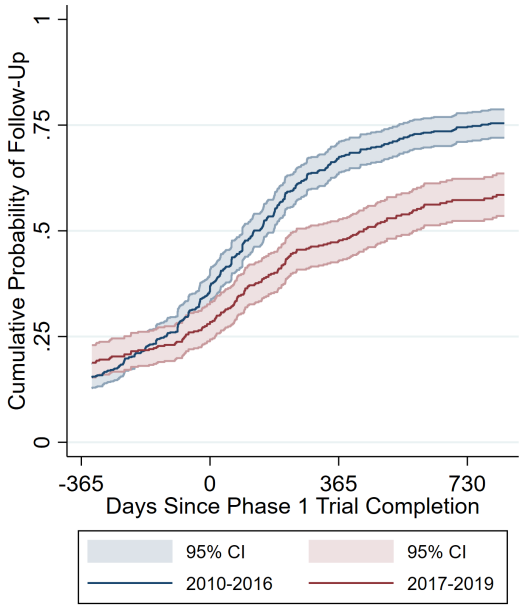
Notes: Panel (i) presents Kaplan-Meier failure curves recording the cumulative probability of completion for Phase 2 industry trials with at least one site in the United States. The two time periods refer to the start year. Panel (ii) records point estimates and 95% confidence intervals of one-year completion probability by trial start year.

could influence how quickly they begin follow-up trials. We therefore investigate how much time passes before firms start follow-up trials after completing a Phase 1 trial. The left panel of Figure 5 reports information about the probability of a follow-up trial start. Specifically, for every completed Phase 1 trial of a drug that is not already approved at the time of that trial, the figure plots the cumulative probability that a follow-up trial for the same drug is started as a function of time since the completion of the Phase 1 trial. That is, the figure shows the probability distribution of the time between the completion of a Phase 1 trial and the start of another trial for the same drug. This amount of time can be negative because firms can start additional trials even before the completion of a Phase 1 trial. As is apparent in the left panel of Figure 5, firms take substantially longer to start new trials in the post-policy period than in the pre-policy period with the two curves diverging around 0 days since the Phase 1 trial completion.

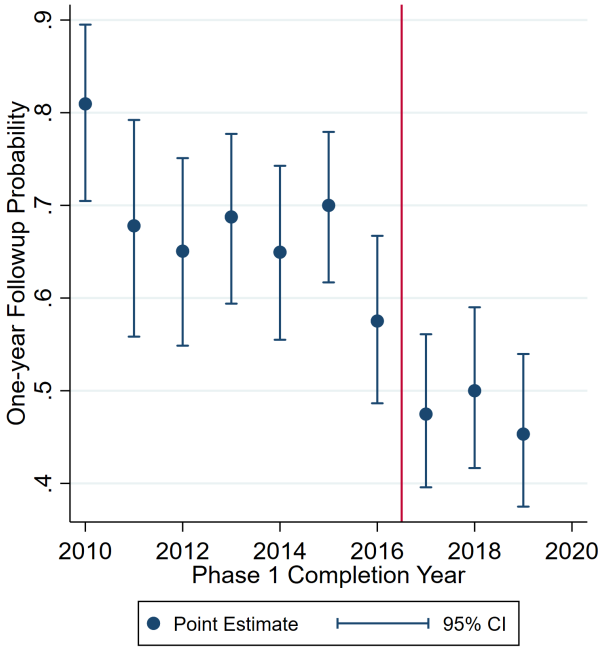
This slowdown in starting follow-up trials is driven by a sharp drop in the trial follow-up probability around the implementation of the final rule. The right panel of Figure 5 plots point estimates and confidence intervals of the one-year follow-up probability by year of trial start. Whereas in the pre-policy period almost 70% of all Phase 1 trials are followed by another trial within a year, less than 50% of trials have a follow-up within a year in the post-policy period.[7]

[7] As before with trial completions, the slowdown in trial starts already begins a little earlier, during the period

**Figure 5:** Time to Start Next Trial after Phase 1




**(i)** Days to Start After Phase 1




**(ii)** One-year Follow-up Probability by Year

Notes: Panel (i) presents Kaplan-Meier failure curves recording the cumulative probability of a follow-up trial for the same drug after the completion of a Phase 1 trial that completed before drug approval. The two time periods refer to the completion year of the Phase 1 trial. Panel (ii) records point estimates and 95% confidence intervals of one-year follow-up probabilities by Phase 1 trial completion year.

A potential concern is that the observed increase in Phase 2 duration may be driven by strategic manipulation of reported trial completion dates, rather than real delays. If firms were backdating or otherwise altering reported end dates to exaggerate trial length (perhaps to signal more extensive testing or better align with disclosure deadlines) then the time between the end of Phase 2 and the start of Phase 3 should systematically shorten. In contrast, we find that the Phase 2–to–Phase 3 transition interval does not decrease following the policy change. This suggests that firms are not merely shifting dates to make trials appear slower. Rather, the observed delays reflect actual reductions in the pace of development. Moreover, trial start dates and site-level investment measures such as the number and growth of active sites are unlikely to be manipulable in the same way and they also exhibit consistent slowdown patterns, further corroborating our interpretation.

Relatedly, in the pre-policy environment firms may have engaged in selective reporting or $p$-hacking across sequential trials by conducting multiple trials and selectively disclosing favorable results to secure approval. This form of strategic behavior has been discussed extensively in the the-

when the final rule was announced but not yet implemented. In addition, the point estimate for 2010 needs to be interpreted with some care as it is due to particularly selected sample which only contains follow-up projects from Phase 1 trials completed and started in 2010. For comparison, the same-year completion rate is around 45% of the sample for later years.

oretical literature (Henry, 2009; Henry and Ottaviani, 2019) and is supported by empirical patterns. For example, the prevalence of null results in large clinical trials increased significantly after the early 2000s, following the introduction of trial preregistration and reporting requirements, such as those implemented via `clinicaltrials.gov` (Kaplan and Irvin, 2015). These reforms limited firms' ability to selectively disclose only favorable outcomes, thereby curbing opportunities for $p$-hacking. Because our sample period begins several years after these reforms were introduced, such practices are likely much more limited in our data. Nevertheless, some room for strategic non-disclosure may persist. Our empirical strategy helps mitigate these concerns by focusing on site-level investment and timing patterns—such as the number of active sites and trial duration—that are not easily manipulated ex post. Moreover, the implementation of mandatory result disclosure further reduces the strategic value of selective reporting after the policy change.

Together, the patterns in Figure 4 and Figure 5 indicate that after the implementation of the final rule, Phase 2 trials took substantially longer to complete and the time between Phase 1 trial completions and follow-up trials increased suggesting that firms were less willing to push the pace of drug discovery and approval. Cox proportional hazard models (reported in Appendix Table A3) indicate that the difference between time periods is statistically significant and not driven by changes in the composition of trials by disease.

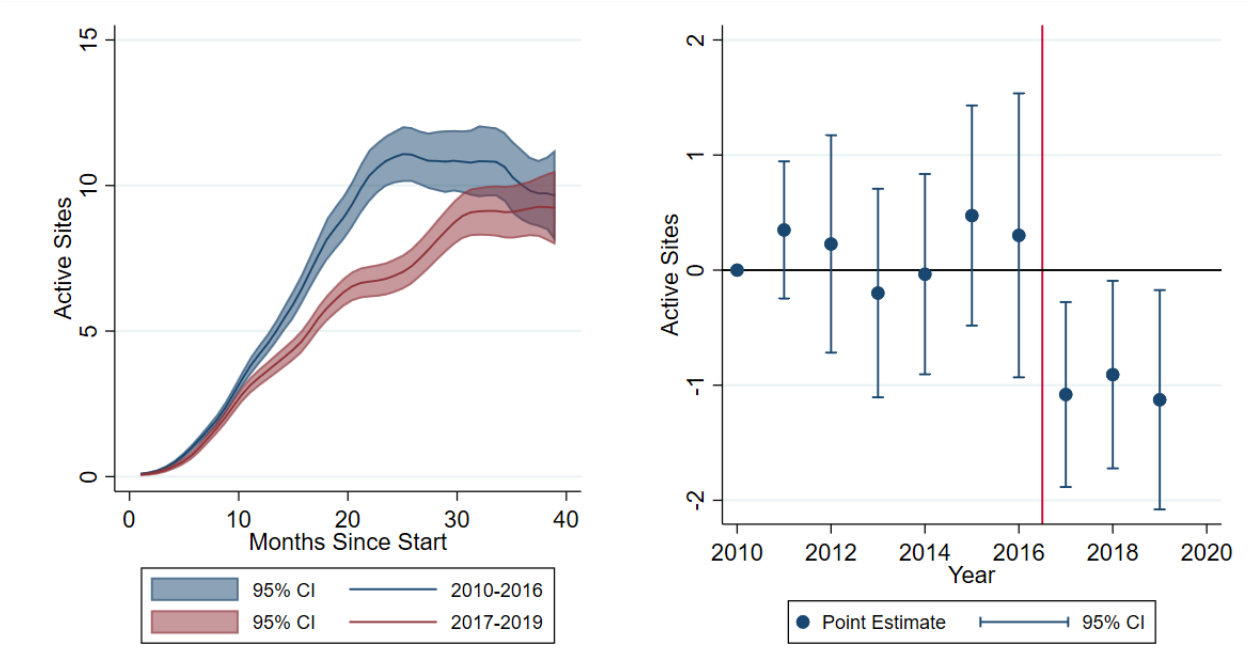
### 5.3 Within-trial Investment

Our analysis so far focuses on broad data patterns on trial reporting, starts, and completion and shows a slowdown in pharmaceutical development following the implementation of new reporting requirements. To further investigate the mechanisms driving the slowdown in trial completions, we now explore more granular data from Industry-sponsored Phase 2 trials and compare the pace at which sites are added to active trials in the pre- and post-policy periods. In order to access a large number of patients, Phase 2 trials typically take place at multiple sites. Opening additional sites is a costly investment that yields additional data on the safety and efficacy of the drug. One mechanism by which trials could be delayed is by opening fewer sites or opening sites more slowly.

The left panel of Figure 6 records the number of and confidence intervals for active sites by number of months since the start of a Phase 2 trial for the two policy periods using data from US industry-sponsored trials, the sample for which we have information on the number of active sites. The count of active sites is consistently higher for trials operating in the pre-policy period than for trials operating after the policy went into effect, particularly for the first three years since the Phase 2 trial start. Moreover, the magnitude of this difference is quite large. For Phase 2 trials active for two years, those operating in the pre-policy period had around 11 active sites on average whereas those operating in the post-policy period had fewer than 7 active sites, a reduction in site count of more than a third. This evidence is consistent with the slowdown in trial completions leading to a slowdown in the pace of investment for active trials.

The right panel of Figure 6 provides additional empirical evidence in support of this link. Specif-

**Figure 6:** Trial Site Dynamics




**(i)** Active Sites by Months Since Trial Start

**(ii)** Year Fixed Effects from Sites Regression

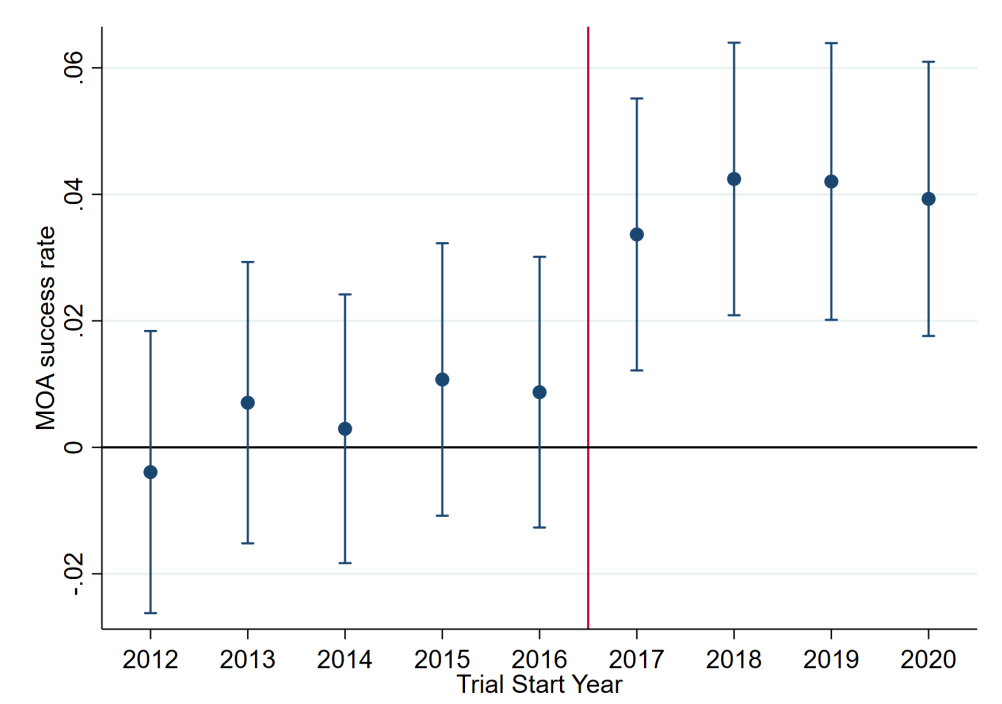
Notes: Panel (i) plots the average number of active sites by months since the start of a Phase 2 trial, separately for ongoing trials before and after the implementation of the Final Rule. Shaded areas represent 95% confidence intervals. Panel (ii) shows the estimated coefficients and 95% confidence intervals from regressions of active sites on year-fixed effects and months since trial start. Standard errors two-way clustered at the trial-month level. Estimates are conditional on months since start and reflect differences in active site counts across years.

ically, it reports the coefficients of year-fixed effects from a regression of active sites on year and months since trial start, effectively comparing the number of active sites since the trial start across years.[8] Following the implementation of the new reporting regime in 2017, there is a significant drop in the number of active sites compared to the pre-policy period with each trial having approximately one fewer site on average in the post-policy period.

Figure 7 plots estimated year effects on the MOA-level Phase 2 trial success rate, with the red vertical line marking the 2017 policy change. In the pre-policy period, the estimated year effects are close to zero with wide confidence intervals, indicating no systematic upward or downward trend in success rates relative to the omitted baseline year. Starting in 2017, the MOA-level success rate rises by approximately 3–4 percentage points and remains elevated through 2020. This post-policy increase in average trial quality is consistent with firms becoming more selective about which MOAs to pursue: by waiting to observe the results of related trials before committing their own resources, firms can reallocate resources towards more successful MOAs, raising the average success rate of the trials that do proceed. The data generally support the interpretation that this pattern reflects

[8] Loosely speaking, the right panel shows the average difference between the two lines in the left panel when drawn year by year.

**Figure 7:** Change in MOA-level Success Rate




Notes: The figure plots year-specific coefficient estimates and 95% confidence intervals from a regression of MOA-group-level Phase 2 trial success rates on year fixed effects. The omitted base year is 2011. The vertical red line marks the implementation of the Final Rule in 2017.

a response to the policy change rather than the continuation of a pre-existing trend.

# 6 Learning from Trial Results

As we showed in the previous sections, the final rule induced a change in the information environment faced by drug development projects, bringing the disclosure of trial results forward in time, and led to firms starting and completing trials more slowly as well as investing in fewer active trial sites. In terms of a simple model of learning and investment (see Section 2), one can think of the policy as increasing the observability of the results of other firms' trials. Such a change in the information environment will affect firms' trial investment decisions if there is learning across drugs—that is, if the outcome of a trial on one drug is informative about the potential success of related drugs.

## 6.1 Trial Responses to Results Disclosure

To investigate whether firms' behavior is consistent with cross-drug learning, we define two sets of related drugs for every drug $j$ in the data. $M_j$ is the set of drugs with the same mechanism of action as drug $j$. These are drugs that are technologically related to drug $j$ because they target the same mechanism within the patient's body. The success or failure of a trial on a drug $k \in M_j$ is likely

to be informative about the probability of success of a trial on drug $j$. $D_j$ is the set of drugs that treat the same disease as drug $j$. These are drugs that are likely to compete with drug $j$ in the same market if they are both approved. The outcomes from trials on drugs $k \in D_j$ are therefore potentially informative about future competition.

Following the logic of Krieger (2021) we estimate regression equations of the form

$$Start_{jt} = \beta_1^{Same} Suc_{jt} + \beta_2^{Same} UnSuc_{jt} + \beta_1^{Dis} \sum_{k \in D_j} Suc_{kt} + \beta_2^{Dis} \sum_{k \in D_j} UnSuc_{kt} \quad (1)$$
$$+ \beta_1^{MOA} \sum_{k \in M_j} Suc_{kt} + \beta_2^{MOA} \sum_{k \in M_j} UnSuc_{kt} + \alpha_j + \gamma_t + \epsilon_{jt}$$

where $Start_{jt}$ is an indicator for whether a trial on drug $j$ started in month $t$. $Suc_{jt}$ is the cumulative count of reported positive outcomes from trials on drug $j$ up to date $t$ measured in terms of standard deviations. $UnSuc_{jt}$ is the equivalent count for unsuccessful trials. We run various specifications in which we include different trial phases on the right- and left-hand sides of the regression equation. Table 3 reports the estimated coefficients from specifications in which we include only Phase 2 trials in the counts in $Suc_{jt}$ and $UnSuc_{jt}$.[9]

The results reported in Table 3 are consistent with learning within and across drugs. Column (1) reports the estimated coefficients when the dependent variable is the start of a Phase 2 trial. A one-standard deviation of successful Phase 2 trial counts increases the probability of a subsequent Phase 2 trial on the same drug by 2.28 percentage points. Conversely, an additional unsuccessful Phase 2 trial reduces the probability of a subsequent Phase 2 trial by 3.89 percentage points. Columns (2) and (3) report similar patterns for the effect of Phase 1 and 3 trials, though with slightly smaller magnitudes. These estimated coefficients are consistent with learning within a drug. Successful research on drug $j$ encourages more follow-on research for drug $j$ and vice versa for unsuccessful research. The magnitudes of the coefficients also suggest a large effect because the mean of $Start_{jt}$ for Phase 2 trials is 0.019.

The coefficients in rows 5 and 6 of column (1) report the estimated coefficients on trial outcomes for drugs $k \in M_j$ (i.e., drugs that share a mechanism of action with $j$). An additional successful trial of a drug $k \in M_j$ increases the probability of starting a Phase 2 trial on drug $j$ by 1.99 percentage points. There is a similar negative effect of unsuccessful trials which reduces the probability of starting a Phase 2 trial by 2.18 percentage points. A similar pattern holds for the effect on starting Phase 1 trials, but with a smaller magnitude. In contrast, the effects of trial outcomes from drugs $k \in D_j$ (i.e., drugs treating the same disease) reported in rows 3 and 4 follow the opposite pattern. A one-standard deviation increase in successful Phase 2 trials by a competing drug in the same disease group reduces the probability of a Phase 2 start by 4.32 percentage points while the same increase in unsuccessful Phase 2 trials increases it by 5.81 percentage points. The same pattern also

[9]For this exercise, we divide all drugs in the sample into disjoint mechanism and disease groups. That is, there is no overlap between $M_j$ and $M_k$ and $D_j$ and $D_k$ for $j \neq k$. The method we use to construct these group is described in Appendix A. Both $M_j$ and $D_j$ exclude drugs sponsored by the same firm as $j$.

**Table 3:** Response of Trial Start to Results Disclosure

| | | Phase 2 Start | Phase 1 Start | Phase 3 Start |
|---|---|---|---|---|
| Same Drug | Successful Phase 2 | 0.0228 | 0.0104 | 0.0071** |
| | | (0.0183) | (0.0096) | (0.0030) |
| | Unsuccessful Phase 2 | -0.0389*** | -0.0177** | -0.0126*** |
| | | (0.0121) | (0.0069) | (0.0027) |
| Same Disease | Successful Phase 2 | -0.0432** | -0.0528*** | -0.0190*** |
| | | (0.0204) | (0.0125) | (0.0058) |
| | Unsuccessful Phase 2 | 0.0581** | 0.0562*** | 0.0222*** |
| | | (0.0242) | (0.0144) | (0.0063) |
| Same Mechanism of Action | Successful Phase 2 | 0.0199** | 0.0157*** | -0.0035 |
| | | (0.0096) | (0.0056) | (0.0033) |
| | Unsuccessful Phase 2 | -0.0218** | -0.0164*** | 0.0016 |
| | | (0.0094) | (0.0056) | (0.0031) |
| Time FE | | YES | YES | YES |
| Drug FE | | YES | YES | YES |
| $N$ | | 809,108 | 809,108 | 809,108 |
| $R^2$ | | 0.1852 | 0.0905 | 0.0682 |
| Mean of Dep. Var. | | 0.0187 | 0.0249 | 0.0118 |

Notes: The table reports coefficient estimates of the regressions described in equation (1). The dependent variables for the respective columns are the number of Phase 2, Phase 1, and Phase 3 trial starts in a drug-month. Regressors are measured in standard deviations. Standard errors are clustered at the drug level. Stars indicate statistical significance: * 90%, ** 95%, *** 99%.

holds for Phase 1 and Phase 3 starts, though again with lower magnitudes.

These results are consistent with the firm developing drug $j$ learning from trials on drugs that are technologically related, in the sense of using the same mechanism of action, and also with learning from trials on drugs that are competitive substitutes because they treat the same disease. The marked difference in the effect of same-mechanism and same-disease drugs is likely the result of a competitive preemption effect for drugs in the same disease group. A successful trial of a competing drug $k$ in the same disease group $D_j$ raises the probability that there will be a viable competitor for drug $j$ and thus lowers the potential profits of drug $j$ making further costly development efforts such as clinical trials less attractive.

These effects are most pronounced when the right-hand side variables are Phase 2 outcomes. In Appendix Tables A5 and A6 we present regressions using the outcomes of Phase 1 and Phase 3 trials as explanatory variables and find less significant effects. The coefficients for Phase 3 outcomes tend to have the opposite sign. This suggests that most of the learning in drug development occurs during Phase 2 and Phase 3 being the final hurdle before approval. That is, success of another drug in Phase 3 potentially has a significant competitive preemption effect that offsets any remaining learning effects.

### 6.2 Response of Trial Completions and Research Site Developments to Unreported Trials

Having documented the increase in reporting, the slowdown in drug development and investments, and the presence of significant social learning effects in drug development we now investigate if these phenomena are related. In particular, if the slowdown in investment is driven by incentives to delay the disclosure of development results to other firms (as illustrated by the model in Section 2), then differences in investment between the pre- and post-policy periods should be heterogeneous depending on the potential to learn about the outcomes of related drugs.

For each drug $j$, consider the set $M_j$ of technologically related drugs.[10] Let $Unrep_{tj}$ be an indicator equal to one if there is at least one drug in $M_j$ for which a Phase 2 trial has been completed but the results are not yet reported at date $t$. $Unrep_{tj}$ therefore captures whether there are any trials related to drug $j$ that are "at risk" of being reported at date $t$. If firms delay investment in response to expected future information flow from trial reports, we should expect to see fewer sites being opened and lower completion rates when $Unrep_{tj} = 1$. Furthermore, the response of firms to $Unrep_{tj}$ should be greater in the post-policy period because, as illustrated by Figure 2, the Final Rule substantially increased the rate of reporting for completed trials. The presence of any unreported related trial therefore should have a larger effect on the present discounted value of future information flow after the policy change.

To test these hypotheses we estimate the following regression specifications using OLS

$$Completion_{tj} = \beta_0 + \beta_1 Post_{tj} + \beta_2 Unrep_{tj} + \beta_3 Post_{tj} \times Unrep_{tj} + \alpha_{d(j)} + \gamma_{tm(j)} + \epsilon_{tj} \quad (2)$$

$$Active_{tj} = \beta_0 + \beta_1 Post_{tj} + \beta_2 Unrep_{tj} + \beta_3 Post_{tj} \times Unrep_{tj} + \alpha_{d(j)} + \gamma_{tm(j)} + \epsilon_{tj} \quad (3)$$

where $Completion_{tj}$ indicates the completion of trial $j$ and $Active_{tj}$ is the count of active sites for trial $j$, $t$ months since the trial started. We run two versions of this second specification: one in which $Post_{tj}$ is an indicator for the start year of trial $j$ being after 2016, and one in which $Post_{tj}$ is an indicator for post-2016 at the $(t, j)$ level. This second definition means that $Post_{tj}$ can vary within trial $j$ if that trial is ongoing when the policy change takes place. We include drug and technology-group-time fixed effects to control for any compositional changes between the pre- and post-policy periods.

We report the estimated coefficients from the regressions described in equations (2) and (3) in Table 4. The table has two stacked panels: the upper panel shows trial completion outcomes and the lower panel shows active sites outcomes across the same six columns. The sample excludes the top 5% of trials by maximum site count to prevent outlier trials with very large site counts from dominating the regressions.[11]

Column (1) establishes our baseline result: trials operating in 2017 or later have both a lower

[10] We use the same disjoint mechanism groups as Table 3.

[11] We return to these trials in Appendix Table A7.

**Table 4:** Heterogeneous Responses to Policy Period

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Trial Completion | | | | | |
| Year $\geq$ 2017 | -0.0039*** | -0.0034** | 0.0094*** | 0.0051 | -0.0030** | 0.0102*** |
| | (0.0014) | (0.0014) | (0.0033) | (0.0064) | (0.0015) | (0.0033) |
| MOA Unreported $> 0$ | | -0.0026* | 0.0001 | 0.0064** | | 0.0001 |
| | | (0.0016) | (0.0017) | (0.0032) | | (0.0017) |
| MOA Unreported $> 0\times$ | | | -0.0148*** | -0.0111** | | -0.0147*** |
| Year $\geq$ 2017 | | | (0.0038) | (0.0048) | | (0.0038) |
| Small Market $\times$ | | | | | -0.0049** | -0.0047** |
| Year $\geq$ 2017 | | | | | (0.0023) | (0.0023) |
| Mean of Dep. Variable | 0.029 | 0.029 | 0.029 | 0.029 | 0.029 | 0.029 |
| | Active Sites | | | | | |
| Year $\geq$ 2017 | -0.8768*** | -0.9120*** | 0.7635 | 1.5420* | -0.6790** | 0.9369 |
| | (0.2864) | (0.2972) | (0.6241) | (0.7861) | (0.3182) | (0.6407) |
| MOA Unreported $> 0$ | | 0.1741 | 0.5200 | 1.2808*** | | 0.5195 |
| | | (0.2930) | (0.3302) | (0.4600) | | (0.3300) |
| MOA Unreported $> 0\times$ | | | -1.9282*** | -2.1806*** | | -1.9068*** |
| Year $\geq$ 2017 | | | (0.6568) | (0.7837) | | (0.6544) |
| Small Market $\times$ | | | | | -1.0359* | -1.0087* |
| Year $\geq$ 2017 | | | | | (0.5939) | (0.5935) |
| Mean of Dep. Variable | 4.058 | 4.058 | 4.058 | 4.059 | 4.058 | 4.058 |
| Time FE | YES | YES | YES | YES | YES | YES |
| MOA Group FE | YES | YES | YES | NO | YES | YES |
| Disease FE | YES | YES | YES | NO | YES | YES |
| Trial FE | NO | NO | NO | YES | NO | NO |
| $N$ | 121,890 | 121,890 | 121,890 | 121,845 | 121,890 | 121,890 |

Notes: The table reports coefficient estimates of the regressions described in equations (2) and (3). The sample includes all trial-months before 2020 for all US phase 2 industry trials started between 2010 and 2019, excluding the 5% largest trials by maximum site count. Stars indicate statistical significance: * 90%, ** 95%, *** 99%. Standard errors two way clustered at the trial-date level.

probability of completion and fewer active sites, consistent with the patterns documented in Figures 4 and 6. Column (2) adds the indicator variable MOA Unreported $> 0$ for whether at least one related trial has been completed but not yet reported. Its coefficient is negative for both outcomes, suggesting that investment is lower whenever any related trial is awaiting disclosure. Column (3) adds the interaction between MOA Unreported $> 0$ and the post-policy indicator. This interaction is negative and statistically significant for both trial completion and active sites. Firms reduce investment more strongly when related trials are pending disclosure in the post-policy period during which the expected rate of disclosure is substantially higher. Once this interaction is included, the direct effect of the post-policy indicator shrinks considerably, indicating that much of the post-policy

investment slowdown is explained by heightened sensitivity to pending disclosures rather than by other secular changes. Column (4) replaces MOA group and disease fixed effects with trial fixed effects, yielding similar qualitative patterns.

Columns (5) and (6) test an additional implication of the model. Proposition 3 predicts that competition-based delay is stronger in smaller markets where business-stealing effects are more pronounced. We capture this by interacting the post-policy indicator with an indicator for whether the drug's disease area is small (a "small market"). The coefficient on Small Market $\times$ Year $\geq$ 2017 is negative and statistically significant for both trial completions and active sites, consistent with delay being amplified in more competitive markets with fewer potential entrants. Column (6) includes all interactions together and confirms that both the unreported trials channel and the small market competition channel operate independently.

Figure A4 in the appendix addresses the concern that our main results may be driven by disruptions associated with the COVID-19 pandemic. We re-estimate the key outcomes after truncating all trial starts and completions at March 2020, effectively treating the data as if no information were observed thereafter. The patterns remain qualitatively unchanged: industry-sponsored trials exhibit a post-2017 slowdown in both trial starts and completions, while disclosure increases following the Final Rule. We conclude that our findings are not driven by pandemic-related disruptions to clinical research activity, but instead reflect development responses to policy changes that predate COVID-19.

Appendix Table A7 presents the same specifications estimated on the full sample including the top 5% of trials by maximum site count. The trial completion results are robust to the inclusion of these large trials. In contrast, the active sites results lose statistical significance. This asymmetry likely reflects the fact that very large trials have very different investment dynamics from the typical trial in the sample: their site counts are so large that they exert disproportionate influence on the active sites regression, masking the pattern that is clearly present in the 95% of trials in our main sample. Because these outlier trials do not appear to follow the same strategic logic as the rest of the data, we exclude them from the main analysis.

Appendix Table A11 augments the baseline specifications with controls for ongoing (not yet completed) trials in the same MOA group: an indicator for whether there is at least one ongoing trial in $M_j$ and its interaction with the post-policy indicator. The coefficient on Ongoing $\times$ Post is negative and statistically significant, with a magnitude similar to the effect of the unreported (completed but not yet disclosed) indicator. This makes intuitive sense: ongoing trials represent a future source of information in the same way that completed-but-unreported trials do, and firms anticipate and respond to both. The similarity in magnitudes suggests that what drives delay is the broader expectation of forthcoming information, whether from trials that have completed but not yet reported or from trials that are still actively running.

The findings in Section 6.2 demonstrate that firms strategically adjust their investment in clinical trials based on the expected future availability of information from other trials. Active trial sites as

well as trial completions decline significantly when there are more related but unreported trials. This suggests that firms anticipate forthcoming disclosures and delay their own investments, especially in the post-policy period where increased disclosure requirements increase the expected flow of information. These findings are in line with theoretical predictions that increased transparency can lead to strategic delay, as firms seek to minimize redundant investments while maximizing the benefits of learning from the successes and failures of competitors.

# 7 Dynamic Model of Trial Investment

## 7.1 Setup

We first consider the basic tradeoff between the cost of running trials and learning from other trials. For the moment we will ignore the impact of competition which causes drug development to reduce the payoff of other drugs in the same disease group.

Let $j$ be a drug in MOA group $m$. The ex-ante unknown effectiveness of a drug $j$ is given by $\lambda_j = \mu_m + \lambda_j^o$ where $\mu_m$ is the component common to all drugs in $m$ and $\lambda_j^o$ is the drug-specific component. Trials in the same group are technologically related and allow for learning. Observing the outcome of other trials in the same group leads firms to update their posterior about the group-level productivity $\mu_m$. We assume that both the individual drug-specific component and the group-level productivity are normally distributed with the following distributions.

$$\begin{aligned} \mu_m &\sim \mathcal{N}(\mu_M, \sigma_M) && \text{group-level productivity} \\ \lambda_j^o &\sim \mathcal{N}(0, \sigma_\lambda) && \text{drug-specific productivity} \end{aligned}$$

Thus, observing $\lambda_k$ where $k \neq j$ is informative about group-level productivity $\mu_m$ which in turn is informative about the drug effectiveness $\lambda_j$. Firms have the same identities as drugs, denoted by $j$, and invest in trial sites to obtain signals about $\lambda_j^o$.

Denote the state of firm $j$ by $S_j$ containing the following elements

$$S_j = \left\{ \tilde{\mu}_m, \tilde{\sigma}_m, \tilde{\lambda}_j^o, \tilde{\sigma}_j; A, O, C \right\}$$

where $\tilde{\mu}_m$ and $\tilde{\sigma}_m$ describe the firm's current beliefs about group-level productivity $\mu_m$. $\tilde{\lambda}_j^o$ and $\tilde{\sigma}_j$ denote the beliefs about the mean and variance of drug-specific effectiveness $\lambda_j^o$. Given this setup, we have

$$E(\lambda_j) = \tilde{\mu}_m + \tilde{\lambda}_j^o$$

$A$ denotes the number of active trial sites of the drug, $O$ is the number of ongoing other trials for drugs $k \in m$ (i.e., drugs in the same MOA group), and $C$ is the number of completed but not yet reported trials for drugs $k \in m$.

In each period $t$, the firm $j$ observes a signal from its own drug trial sites:

$$s_{jt} \sim \mathcal{N}\left(\lambda_j, \frac{\sigma_o}{A}\right)$$

Adding trial sites $A$ reduces the variance of the signal $s_{jt}$ that a firm receive about the effectiveness of its drug project. In particular, with our specification, the signal $s_{jt}$ is completely uninformative when the number of active sites $A$ is equal to zero and perfectly informative with an infinite number of active sites.

In each period there is also a chance that completed but unreported trials $k \in C$ disclose their results. If a trial $k$ is disclosed, then firm $j$ observes

$$S_{kt} = \frac{1}{\sum_{\tau=0}^{t} A_{k\tau}} \sum_{\tau=0}^{t} A_{k\tau} s_{k\tau}.$$

This is the average of all own-signals observed by $k$ before it completed where $A_{k\tau}$ is the number of active sites for trial $k$ in period $\tau$. It means that when trial results for $k$ are disclosed, all active trials observe the signals that were previously private information to $k$. The distribution of this information is

$$S_k \sim \mathcal{N}\left(\mu_m + \lambda_k^o, \frac{\sigma_o}{\sum_{\tau=0}^{t} A_{k\tau}}\right).$$

where $\sum_{\tau=0}^{t} A_{k\tau}$ is the total number of site-days completed by trial $k$. Thus, $S_{kt}$ is informative about the MOA-level efficacy component, $\mu_m$. To simplify the firm's problem, we assume that firms do not keep track of the total site days of all undisclosed trials, but instead behave as if all disclosed trial signals were generated by $\bar{A}$ site-days, which can be thought of as the average number of site-days over all trials. Thus, the perceived distribution of $S_k$ is

$$S_k \sim \mathcal{N}\left(\mu_m + \lambda_k^o, \frac{\sigma_o}{\bar{A}}\right). \tag{4}$$

Firm $j$ has perceptions over the rates at which other trials are started, completed, and eventually reported. We define these rates as follows. First, trials of other firms are started (i.e., they enter $O$) with probability $\alpha_S$ which denotes the trial start rate. Second, $\alpha_C$ is the completion rate at which ongoing trials $O$ convert to completed but not yet reported trials $C$. Third, completed trial results in $C$ are disclosed with probability $\alpha_D$. To summarize, these transition rates are

$$\begin{aligned}
\alpha_S &= \text{start rate of trials} \\
\alpha_C &= \text{completion rate (how } O \text{ converts to } C) \\
\alpha_D &= \text{disclosure rate of trials } (S_k \text{ becomes observable})
\end{aligned}$$

Given the firm's own signal $s_{jt}$ from its drug trials and the observation of $\lambda_k$ for any disclosed

trials of other technologically related drugs $k$, firm $j$ updates beliefs using rational priors and Bayesian updating. We assume that firms only observe the outcomes of trials that have been completed and reported, but they do not observe the outcomes of trials that have been completed but not reported. In addition, we assume that firms do not infer anything about the outcomes of trials that have been started but not yet reported.

The value function of firm $j$ undertaking trials for drug $j$ is given by:

$$V(S) = \max\left\{\beta E\left[V(S')\right] - cA, \max_{A^* \in [0,\bar{A}]}\left\{\beta E\left[V(S')|A^*\right] - cA - \delta\right\}, \pi(S)\right\} \tag{5}$$

where $\beta$ is the discount factor, $S'$ is the state in the next period, $c$ is the per-period cost of running an active trial site, and $\delta$ is the adjustment cost of changing the number of active trial sites $A$. The profit function $\pi(S)$ is the maximum of the expected effectiveness of drug $j$, $E(\lambda_j)$, or its shutdown value (normalized to 0) and thus given by

$$\pi = \max\{\tilde{\mu}_m + \tilde{\lambda}_j^o, 0\}$$

The value function described in equation (5) thus captures the following three broad choices for the firm. First, it can wait and keep the current number of active trial sites $A$ operational which comes at a cost of $cA$. However, it also gives the firm the opportunity to learn from other trial outcomes represented by the state $S'$ which yields the new value function $V(S')$ discounted by $\beta$. Second, the firm can adjust the number of active sites $A$ for its trials between 0 and an upper bound $\bar{A}$ at an optimal value $A^*$. However, changing the number of active sites $A$ requires an adjustment cost of $\delta$. This choice also allows the firm to learn from the outcomes of other trials.[12] Third, the firm can end the trial on drug $j$ this period and obtain the terminal payoff $\pi(S)$ which depends on the current state of nature $S$.

This value function captures the essential trade-off of our analysis. On the one hand, a firm can invest and obtain information from running its own trials. On the other hand, it can wait and learn from the successes and failures of other firms. Investment in a firm's own trials involves adding trial sites. We model this channel by assuming that adding sites reduces the variance of the signal that a firm receive about the productivity of their drug project.

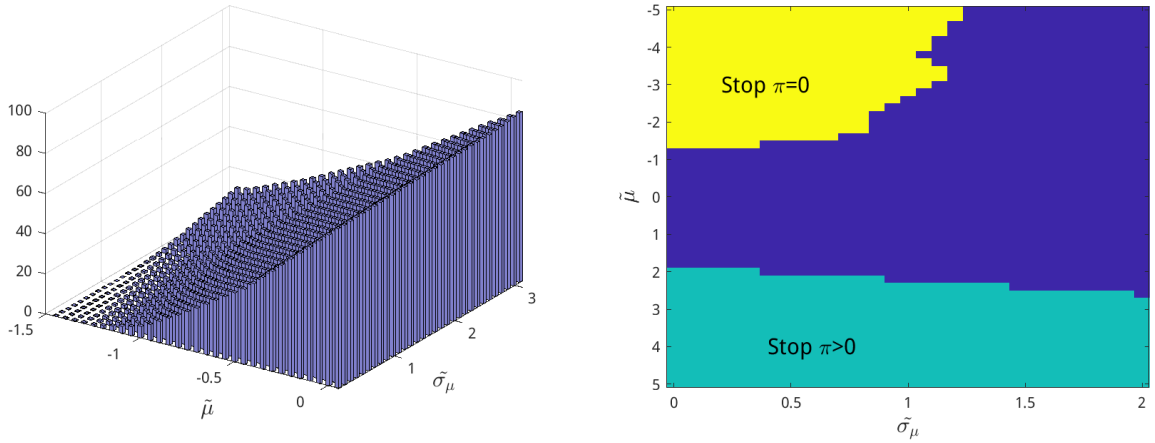
### 7.2 Solving the Model

We solve the model using value function iteration, where the value function is approximated by a neural network on each iteration of the Bellman equation.[13] This approach allows us to handle the high-dimensional state space while maintaining sufficient flexibility to capture complex patterns in the decision-making environment.

[12] In our data we observe that firms both increase and decrease the number of active sites $A$.

[13] See Bertsekas and Tsitsiklis (1996). Hodgson and Lewis (2025) apply this approach to a similar dynamic learning problem.

**Figure 8:** Value Function and Optimal Stopping Rule



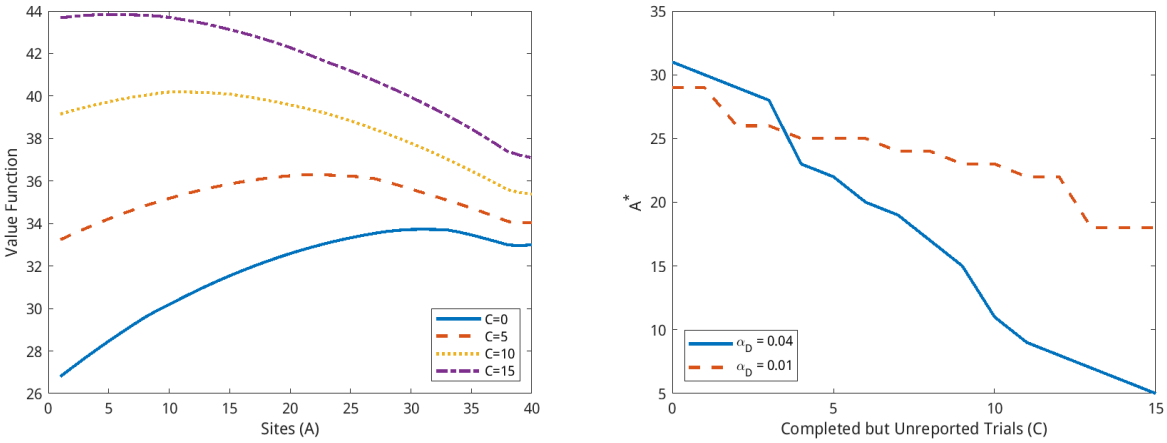

**(i)** Value Function **(ii)** Optimal Stopping Rule

Notes: Panel (i) presents a 3D bar plot of the value function over the space of belief means ($\tilde{\mu}_M$) and standard deviations ($\tilde{\sigma}_M$), representing the continuation value of a trial given current beliefs. The value function increases with both $\tilde{\mu}_M$ and $\tilde{\sigma}_M$, reflecting the combined effect of higher expected returns and greater option value from uncertainty. Panel (ii) shows the optimal stopping rule as a heatmap over the same state space. Yellow areas indicate states where the trial is terminated without seeking approval, teal areas represent termination with approval, and dark blue areas denote continuation. The figure illustrates how firms dynamically balance expected returns and the value of learning when making trial continuation decisions.

We calibrate the model parameters to match several key empirical moments observed in the data. Specifically, we target the average number of active sites, the average duration of clinical trials, the frequency with which new sites are added during trials, and the sensitivity of site behavior to the outcomes of other trials. Matching these moments ensures that the model replicates critical features of the real-world setting we seek to study. The calibration procedure also informs the future estimation and identification strategy. In particular, the moments selected for calibration highlight the dimensions along which variation in the data will be most informative for identifying structural parameters.

It is useful to conceptualize two important components of the model. We treat the disclosure rate $\alpha_D$ as a policy parameter, reflecting choices made by the principal or regulator, while $C$ is an endogenous equilibrium object that emerges from the strategic interactions of agents within the model. While one might consider modeling firms' disclosure decisions as endogenous, we treat the disclosure regime as an exogenous policy parameter for both tractability and institutional realism. The 2017 FDA Final Rule imposed legally binding requirements on trial result reporting, limiting firms' ability to strategically withhold information post-policy. Modeling disclosure behavior endogenously would require additional assumptions about firms' beliefs and expectations regarding enforcement, public scrutiny, and market reactions. Instead, by treating the disclosure environment as a shift in policy, we can focus on how firms adjust real investment behavior in response, which is the key object of interest in our analysis.

**Figure 9:** Value Function and the Role of Disclosure




**(i)** Value Function over $A$ and $C$ **(ii)** Optimal $A^*$ for different $\alpha_D$ and $C$

Notes: Panel (i) plots the value function as a function of the number of active sites $A$ for different values of the number of completed but unreported trials $C$, holding the probability of disclosure fixed at $\alpha_D$. As $C$ increases, the value function peaks for a lower number of active sites and thus the optimal number of active sites $A^*$ is lower. This pattern is illustrated for two different values of $\alpha_D$ in panel (ii).

We first plot the value function and the optimal stopping rule in $(\tilde{\mu}_M; \tilde{\sigma}_M)$ space which are shown in Figure 8. It consists of two complementary panels. In the left panel, a 3D bar plot shows the value function over the grid of $(\tilde{\mu}_M, \tilde{\sigma}_M)$ values. The vertical axis represents the continuation value of the trial as a function of both the current belief about the drug's effectiveness (mean) and the uncertainty surrounding that belief (standard deviation). The value function increases monotonically with both $\tilde{\mu}_M$ and $\tilde{\sigma}_M$, reflecting higher expected returns when the treatment looks more promising (higher mean) and/or when uncertainty creates greater option value from continued experimentation. The right panel shows the optimal stopping rule. The heatmap segments the $(\tilde{\mu}_M, \tilde{\sigma}_M)$ space into regions based on the optimal stopping decision. The yellow region labeled "Stop $\pi = 0$" corresponds to states in which the expected payoff from continuation is sufficiently low that it is optimal to terminate the trial with no approval. In contrast, the teal region labeled "Stop $\pi > 0$" indicates states where the firm stops and seeks approval (because expected payoff is high enough). The intermediate dark blue region represents the continuation region, where the trial proceeds and further information is gathered. Together, these two panels illustrate the trade-offs underlying the firm's dynamic experimentation decision. In particular, the stopping policy reflects how both the mean belief and uncertainty jointly determine whether further investment is warranted. The shapes of the stopping boundaries highlight the model's dynamic structure and the value of information in marginal cases.

Figure 9 illustrates how the value function and the optimal number of active sites $A^*$ vary with the extent of external information availability, captured by the number of completed but unreported trials $C$. Panel (i) shows that, for a given value of the disclosure probability $\alpha_D$, the value function

peaks at lower levels of $A$ as $C$ increases. This reflects the idea that when more external information is available, firms have weaker incentives to generate costly private information through their own trial sites. Panel (ii) formalizes this relationship by plotting the optimal number of active sites $A^*$ as a function of $C$, for two different values of $\alpha_D$. In both cases, $A^*$ declines monotonically with $C$, and the drop is markedly steeper when the disclosure probability $\alpha_D$ is higher.

The relationships depicted in Figure 9 demonstrate that disclosure policies affect optimal firm behavior by crowding out private investment in information acquisition and are consistent with our empirical observations showing a slowdown in site investment following the implementation of disclosure policies that mandate greater information sharing. The dynamic model thus captures a key mechanism by which transparency policies can affect firms' strategic investment behavior.

# 8 Estimation and Identification

We estimate several parameters directly from the data without using the solution to the structural model. In particular, we estimate $\alpha_S$, $\alpha_C$ and $\alpha_D$ from the empirical rates of trial starts, completions, and results disclosure. We estimate a separate disclosure rate for the pre- and post-policy periods which we call $\alpha_{D1}$ and $\alpha_{D2}$. We also estimate the average site-days per completed trial, $\bar{A}$, which we use to define firm's beliefs about the distribution of other firms' signals (see equation 4).

The remaining parameters are those that control the distribution of group-level efficacies, $(\mu_M, \sigma_M)$, the variance of idiosyncratic efficacy, $\sigma_\lambda$, the cost parameters $(c, \delta)$, and the own-signal noise parameter, $\sigma_0$. Call the vector of parameters $\theta$.

The model is estimated using simulated method of moments and indirect inference. For each candidate parameter vector, $\theta$, the model-implied moments come from forward simulation of the trial investment problems of all trials in a MOA group, $m$. The simulation is repeated for many MOA groups and the moments are computed for each group and then averaged over groups. The unit of observation used to construct the moments is therefore the MOA group $m$, and an observation is the history of trials for that group. The simulation proceeds as follows:

1. Draw $\mu_m$ from the distribution of group level productivities, $\mu_m \sim \mathcal{N}(\mu_M, \sigma_M)$

2. Start at $t = 1$, and forward simulate for $T$ periods. We divide time into two eras, a first era of length $T_1$ and a second era of length $T_2$ where $T_1 + T_2 = T$. $\alpha_D$ is allowed to be different in the two periods. Denote the era $e \in \{1, 2\}$ and the corresponding disclosure parameter $\alpha_{De}$.

3. Each period, a new trial $j$ enters with probability $\alpha_S$. When a new trial enters, draw its idiosyncratic efficacy, $\lambda_j^o \sim \mathcal{N}(0, \sigma_\lambda)$.

4. Trials that are active choose to adjust their number of sites, $A_{jt}$, or end the trial according to their policy function.

5. Trials that are active draw a signal $s_{jt} \sim \mathcal{N}\left(\lambda_j, \frac{\sigma_o}{A}\right)$ and update their beliefs.

**Table 5:** Simulated Moments

| Moment | Description | Definition | Related Parameters |
|---|---|---|---|
| 1-2 | Average number of active sites in era $e$ | $\frac{1}{M}\sum_m \frac{1}{J_{me}}\sum_j \sum_{t=T_{0j}}^{T_j} A_{jt}1(t \in e)$ | $c$ |
| 3-4 | Average trial length in era $e$ | $\frac{1}{M}\sum_m \frac{1}{J_{me}}\sum_j (T_j - T_{0j})$ | $\sigma_M + \sigma_\lambda$ |
| 5 | Average max sites | $\frac{1}{M}\sum_m \frac{1}{J_m}\sum_j \max_t\{A_{jt}\}$ | $\mu_M$ |
| 6 | Average min sites | $\frac{1}{M}\sum_m \frac{1}{J_m}\sum_j \min_t\{A_{jt}\}$ | $\mu_M$ |
| 7 | Average number of site changes | $\frac{1}{M}\sum_m \frac{1}{J_m}\sum_j \sum_{t=T_{0j}}^{T_j} 1(A_{jt} \neq A_{jt-1})$ | $\delta$ |
| 8 | Interaction coefficient from regression 3 | $\hat{\beta}_{3m}$ | $\sigma_M, \sigma_\lambda$ |
| 9 | Interaction coefficient from regression 2 | $\hat{\beta}_{3m}$ | $\sigma_M, \sigma_\lambda$ |
| 10 | Slope coefficient from regression 7 | $\hat{\beta}_{1m}$ | $\sigma_o$ |

Notes: The table lists the 10 moments used in the simulated method of moments estimation. Moments 1–2 and 3–4 are computed separately for era $e \in \{1, 2\}$ corresponding to the pre- and post-policy periods. $M$ is the number of MOA groups, $J_{me}$ is the number of trials in group $m$ in era $e$, $J_m$ is the total number of trials in group $m$, and $T_{0j}$ and $T_j$ are the start and end periods of trial $j$. $\hat{\beta}_{3m}$ denotes the MOA-group-level interaction coefficient and $\hat{\beta}_{1m}$ the slope coefficient from the indicated regression, run on the simulated data. The "Related Parameters" column indicates which structural parameters each moment is primarily informative about.

6. Trials $k$ that have ended disclose results with probability $\alpha_{De}$. Active trials observe $S_k = \sum_t s_{kt}$ and update their beliefs accordingly. Trials disclose results only once.

We repeat the simulation for $R$ MOA groups. For each simulated MOA history, $m$, we record a vector of statistics, $g_m(\theta)$. These include the average number of active sites, average trial length, and analogs of the coefficients in the regressions described by equations 2 and 3 run on the simulated data. The exact moments used are described in Table 5. Let $g(\theta) = \frac{1}{R}\sum_{m=1}^{R} g_m(\theta)$ and $\hat{g} = \frac{1}{M}\sum_{m=1}^{M} \hat{g}_m$, where $M$ is the number of MOA groups in the data and $\hat{g}_m$ is the moment vector computed from the data for MOA group $m$. The estimator is then defined as

$$\theta^* = \arg\min_\theta \left(g(\theta) - \hat{g}\right)' W \left(g(\theta) - \hat{g}\right), \tag{6}$$

where $W$ is the optimal weighting matrix given by $W = Cov(\hat{g}_m)^{-1}$.

## 8.1 Identification

There are 6 model parameters identified by 10 moments. Although all parameters affect all moments, we provide a heuristic discussion of which parameters are likely to be more informed by changes in one moment, holding the others fixed. Most obviously, the average number of active sites (moments 1 and 2) should decrease in the per-site cost, $c$, and the average trial length (moments 3 and 4) should

**Table 6:** Monte Carlo Results

| Parameter | True | $M = 100$ | | | $M = 1000$ | | |
|---|---|---|---|---|---|---|---|
| | | Mean | SD | $\Pr(\theta \in CI_{95})$ | Mean | SD | $\Pr(\theta \in CI_{95})$ |
| $\mu_M$ | -2 | -2.301 | 0.711 | 0.95 | -2.172 | 0.567 | 0.98 |
| $\sigma_M$ | 7 | 7.134 | 0.421 | 0.95 | 7.067 | 0.259 | 0.98 |
| $\sigma_\lambda$ | 5 | 5.030 | 0.178 | 0.96 | 5.017 | 0.084 | 0.95 |
| $c$ | 0.05 | 0.048 | 0.004 | 0.93 | 0.049 | 0.009 | 0.98 |
| $\delta$ | 0.3 | 0.295 | 0.013 | 0.93 | 0.298 | 0.009 | 0.91 |
| $\sigma_o$ | 60 | 59.884 | 1.933 | 0.95 | 59.936 | 1.164 | 0.98 |

Notes: Entries report, for each parameter, the Monte Carlo mean of the estimator, the Monte Carlo standard deviation across replications, and empirical coverage of the nominal 95% confidence interval. $M$ denotes the number of observations in each simulated dataset. For each exercise, we perform 200 Monte Carlo replications. Estimation uses $R = 200$ simulations.

increase in the total uncertainty $\sigma_M + \sigma_\lambda$. Moments 5-7 are the average minimum and maximum number of sites in a trial, and the number of site changes. The number of site changes decreases with the adjustment cost, $\delta$. Fixing the average number of sites, the minimum and maximum are informative about $\mu_M$. If priors are pessimistic (low $\mu_M$), then there is an incentive to start with a low level of investment and only scale up after receiving positive information.

To separate $\sigma_M$ and $\sigma_\lambda$, we rely on the regression coefficients (moments 8 and 9) which come from analogs of regression specifications 2 and 3. These coefficients measure the differential effect of the policy change on trials for which there are more or fewer outstanding unreported results in the same MOA. If the correlation of efficacy across trials within an MOA is low, then there is limited scope for cross-trial learning, and we should not expect a differential change in the completion rate or investment level across these groups of trials. Conversely, when trial outcomes are highly correlated, then the incentive to delay is increasing in the number of unreported trials. The covariance of efficacy for a pair of trials in the same group is $\sigma_M$. Finally, moment 10 comes from a regression of trial completion on the number of sites,

$$Completion_{tj} = \beta_0 + \beta_1 Active_{tj} + \alpha_j + \gamma_t + \epsilon_{tj} \tag{7}$$

The extent to which more sites increase stopping probability is informative about the per-site noise, $\sigma_o$.

To illustrate identification, we conduct a Monte Carlo exercise in which we estimate the model on simulated datasets for a fixed parameter vector. We set $\alpha_{D1} = 0.01$ and $\alpha_{D2} = 0.02$. Each simulation runs for $T = 500$ periods with $T_1 = 200$. Table 6 records the results for two sample sizes, $M = 100$ and $M = 1000$. For each set of results we simulate 200 datasets and present the mean and standard deviation of the point estimates across datasets as well as the probability that the true value of the parameter falls in the 95% coverage interval.

# 9 Conclusion

Using data on pharmaceutical clinical trials initiated from 2010-2019, we provided evidence of an increase in the timely disclosure of trial results associated with the 2017 FDA final rule and of a subsequent slowdown in the initiation and completion of trials. We further showed that firms learn from trial results and respond to the information about trial successes and failures by adjusting the scale, speed, and intensity of their development efforts. One explanation for the collection of our results is that mandatory disclosure allows firms to free ride on the information provided by competitors and therefore cut back on and slow their own innovation efforts.

# References


**Adams, Christopher P. and Van V. Brantner**, "Estimating the Cost of New Drug Development: Is It Really $802 Million?," *Health Affairs*, 2006, *25* (2), 420–428.

**Agerton, Mark**, "Learning Where to Drill: Drilling Decisions and Geological Quality in the Haynesville Shale," *USAEE Working Paper*, 2020.

**Anton, James J. and Dennis A. Yao**, "The Sale of Ideas: Strategic Disclosure, Property Rights, and Contracting," *Review of Economic Studies*, 2002, *69* (3), 513–531.

**Anton, James J. and Dennis A. Yao**, "Little Patents and Big Secrets: Managing Intellectual Property," *RAND Journal of Economics*, 2004, *35* (1), 1–22.

**Arrow, Kenneth**, "Economic Welfare and the Allocation of Resources for Invention," in "The Rate and Direction of Inventive Activity: Economic and Social Factors," Princeton University Press: Princeton, NJ, 1962, pp. 609–626.

**Arrow, Kenneth J.**, *Essays in the Theory of Risk-Bearing*, Amsterdam: North-Holland, 1971.

**Bertsekas, Dimitri P and John N Tsitsiklis**, *Neuro-Dynamic Programming*, Athena Scientific, 1996.

**Blume-Kohout, Margaret E. and Neeraj Sood**, "Market Size and Innovation: Effects of Medicare Part D on Pharmaceutical Research and Development," *Journal of Public Economics*, January 2013, *97*, 327–336.

**Bolton, Patrick and Christopher Harris**, "Strategic Experimentation," *Econometrica*, 1999, *67* (2), 349–374.

**Bolton, Patrick and Joseph Farrell**, "Decentralization, Duplication, and Delay," *Journal of Political Economy*, 1990, *98* (4), 803–826.

**Branstetter, Lee, Chirantan Chatterjee, and Matthew J. Higgins**, "Starving (or Fattening) the Golden Goose?: Generic Entry and the Incentives for Early-Stage Pharmaceutical Innovation," Working Paper 20532, National Bureau of Economic Research September 2014.

**Code of Federal Regulations**, "Content and format of an NDA," 2025. Title 21, 314.5.

**Cohen, Wesley M. and Daniel A. Levinthal**, "Absorptive capacity: a new perspective on learning and innovation," *Administrative science quarterly*, 1990, pp. 128–152.

**Covert, Thomas**, "Experiential and Social Learning in Firms: The Case of Hydraulic Fracturing in the Bakken Shale," *SSRN Working Paper*, 2015.

**Covert, Thomas R and Richard L Sweeney**, "Secrecy Rules and Exploratory Investment: Theory and Evidence from the Shale Boom," *NBER Working Paper*, 2022.

**Cunningham, Colleen, Florian Ederer, and Song Ma**, "Killer Acquisitions," *Journal of Political Economy*, 2021, *129* (3), 649–702.

**Dasgupta, Partha and Joseph Stiglitz**, "Uncertainty, Industrial Structure, and the Speed of R&D," *Bell Journal of Economics*, 1980, *11* (1), 1–28.

**DeVito, Nicholas J and Ben Goldacre**, "Evaluation of compliance with legal requirements under the FDA amendments act of 2007 for timely registration of clinical trials, data verification, delayed reporting, and trial document submission," *JAMA Internal Medicine*, 2021, *181* (8), 1128–1130.

**DeVito, Nicholas J, Seb Bacon, and Ben Goldacre**, "Compliance with legal requirement to report clinical trial results on clinicaltrials.gov: a cohort study," *The Lancet*, 2020, *395* (10221), 361–369.

**DiMasi, Joseph A., Ronald W. Hansen, and Henry G. Grabowski**, "The Price of Innovation: New Estimates of Drug Development Costs," *Journal of Health Economics*, 2003, *22* (2), 151–185.

**Dix, Rebekah and Todd Lensman**, "Combining Complements: Theory and Evidence from Cancer Treatment Innovation," *Working Paper*, 2024.

**Dubois, Pierre, Olivier de Mouzon, Fiona Scott Morton, and Paul Seabright**, "Market Size and Pharmaceutical Innovation," *The RAND Journal of Economics*, 2015, *46* (4), 844–871.

**Fetter, T Robert**, "Fracking, Toxics, and Disclosure," *SSRN Working Paper*, 2022.

**Fetter, T Robert, Andrew L Steck, Christopher Timmins, and Douglas Wrenn**, "Learning by Viewing? Social Learning, Regulatory Disclosure, and Firm Productivity in Shale Gas," *NBER Working Paper*, 2018.

**Galasso, Alberto and Mark Schankerman**, "Patent Rights and Innovation by Small and Large Firms," *American Economic Journal: Economic Policy*, 2015, *7* (4), 90–134.

**Griliches, Zvi**, "The Search for R&D Spillovers," *Scandinavian Journal of Economics*, 1992, *94* (S1), S29–S47.

**Hegde, Deepak and Hong Luo**, "Patent publication and the market for ideas," *Management Science*, 2018, *64* (2), 652–672.

**Hendricks, Kenneth and Dan Kovenock**, "Asymmetric Information, Information Externalities, and Efficiency: The Case of Oil Exploration," *The RAND Journal of Economics*, 1989, *20* (2), 164–182.

**Henry, Emeric**, "Strategic disclosure of research results: The cost of proving your honesty," *The Economic Journal*, 2009, *119* (539), 1036–1064.

**Henry, Emeric and Marco Ottaviani**, "Research and the approval process: The organization of persuasion," *American Economic Review*, 2019, *109* (3), 911–955.

**Hodgson, Charles**, "Information Externalities, Free Riding, and Optimal Exploration in the UK Oil Industry," *NBER Working Paper*, 2024.

**Hodgson, Charles and Gregory Lewis**, "You Can Lead a Horse to Water: Spatial Learning and Path Dependence in Consumer Search," *Working Paper*, 2025.

**Jaffe, Adam B. and Josh Lerner**, "Reinventing Public R&D: Patent Policy and the Commercialization of National Laboratory Technologies," *RAND Journal of Economics*, 2001, *32* (1), 167–198.

**Kao, Jennifer**, "Information Disclosure and Competitive Dynamics: Evidence from the Pharmaceutical Industry," *Management Science*, 2024.

**Kaplan, Robert M and Veronica L Irvin**, "Likelihood of null effects of large NHLBI clinical trials has increased over time," *PloS one*, 2015, *10* (8), e0132382.

**Krieger, Joshua L**, "Trials and terminations: Learning from competitors' R&D failures," *Management Science*, 2021, *67* (9), 5525–5548.

**Kyle, Margaret K.**, "Pharmaceutical Price Controls and Entry Strategies," *The Review of Economics and Statistics*, 2007, *89* (1), 88–99.

**Levitt, Clinton J**, "Learning through oil and gas exploration," *Working Paper*, 2009.

**Lin, C-Y Cynthia**, "Strategic decision-making with information and extraction externalities: A structural model of the multistage investment timing game in offshore petroleum production," *Review of Economics and Statistics*, 2013, *95* (5), 1601–1621.

**Magazzini, Laura, Fabio Pammolli, and Massimo Riccaboni**, "Learning from failures or failing to learn? Lessons from pharmaceutical R&D," *European Management Review*, 2012, *9* (1), 45–58.

**Morgan, Steve, Paul Grootendorst, Joel Lexchin, Colleen Cunningham, and Devon Greyson**, "The cost of drug development: a systematic review," *Health policy*, 2011, *100* (1), 4–17.

**Mullard, Asher**, "Parsing clinical success rates," *Nature Reviews Drug Discovery*, 2016, *15* (7), 447–448.

**Murray, Fiona and Scott Stern**, "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? Evidence from Patent-Paper Pairs," *Journal of Economic Behavior & Organization*, 2007, *63* (4), 648–687.

**Nelson, Richard R. and Sidney G. Winter**, *An Evolutionary Theory of Economic Change*, Harvard University Press, 1982.

**Scotchmer, Suzanne**, "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives*, 1991, *5* (1), 29–41.

**Scotchmer, Suzanne and Jerry Green**, "Novelty and Disclosure in Patent Law," *RAND Journal of Economics*, 1990, pp. 131–146.

**Spence, Michael**, "Cost Reduction, Competition, and Industry Performance," *Econometrica*, 1984, *52* (1), 101–121.

**Steck, Andrew L**, "Industry Dynamics with Social Learning: Evidence from Hydraulic Fracturing," *Duke Economics Working Paper*, 2018.

**Traag, Vincent A., Ludo Waltman, and Nees Jan van Eck**, "From Louvain to Leiden: Guaranteeing Well-Connected Communities," *Scientific Reports*, 2019, *9* (1), 5233.

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

**Zarin, Deborah A., Nicholas C. Ide, Tony Tse, William R. Harlan, Joyce C. West, and Donald A. B. Lindberg**, "Issues in the Registration of Clinical Trials," *JAMA*, 2007, *297* (19), 2112–2120.

**Zarin, Deborah A., Tony Tse, and Nicholas C. Ide**, "Trial Registration at ClinicalTrials.gov between May and October 2005," *New England Journal of Medicine*, 2005, *353* (26), 2779–2787.

**Zarin, Deborah A, Tony Tse, Rebecca J Williams, and Sarah Carr**, "Trial Reporting in ClinicalTrials.gov—The Final Rule," *New England Journal of Medicine*, 2016, *375* (20), 1998–2004.

**Zarin, Deborah A, Tony Tse, Rebecca J Williams, and Thiyagu Rajakannan**, "Update on trial registration 11 years after the ICMJE policy was established," *New England Journal of Medicine*, 2017, *376* (4), 383–391.

# A Constructing Mechanism-of-Action and Disease Groups

This section describes how we construct the Mechanism-of-Action (MOA) and disease groups used to define related drugs. Pharmaprojects lists, for each drug, one or more MOA labels. Similarly, for each trial, Trialtrove lists one or more disease labels.[14] MOA and disease labels often partially overlap between drugs, so these fields do not cleanly partition drugs into groups. We therefore cluster drugs by the similarity of their MOA and disease label sets, so that every drug belongs to a disjoint group. Table A10 summarizes the resulting MOA and disease groups used in our main specification.

We partition the network of distinct label sets with the Leiden algorithm (Traag et al., 2019) under the Constant Potts Model (CPM), which chooses the groups $c$ that maximize

$$\sum_c \left[ E_c - \gamma \binom{n_c}{2} \right],$$

Where $E_c$ is the total edge weight within group $c$, $n_c$ its number of label sets, and $\gamma$ is the resolution. Higher resolution parameters increase the threshold for group membership, and create more, smaller groups.

The steps to calculate the edge weight, $e(A, B)$, for any two label sets are as follows:

1. Measure the rarity of each distinct label within label sets.

   Each label $\ell$ receives an inverse-document-frequency (IDF) weight

   $$w_\ell = \log \frac{N}{n_\ell},$$

   where $N$ is the total number of label sets and $n_\ell$ the number that contain $\ell$. Without IDF weights, $w_\ell = 1$.

2. Measure the Jaccard similarity for each edge.

   For each edge, the Jaccard similarity of two label sets is

   $$J(A, B) = \frac{\sum_{\ell \in A \cap B} w_\ell}{\sum_{\ell \in A \cup B} w_\ell}.$$

3. Apply the exponential parameter, $j_{exp}$, to the Jaccard similarity measured in step 2.

   The final edge weight between two label sets is then

   $$e(A, B) = J(A, B)^{j_{exp}} \cdot \frac{\bar{J}}{\bar{J}^{j_{exp}}},$$

[14] To move disease labels from the trial level to the drug level, we take all trials in which the drug was the primary drug studied and assign the drug its most common disease label across these trials. We also assign any other label that appears in at least 60% of the drug's trials. Drugs that never appear as the primary drug in a trial therefore have no disease label and are not assigned to a disease group.

where the bars denote means over all edges. The exponent $j_{exp}$ sharpens the contrast between close and loose matches, and the rescaling keeps the mean edge weight equal to the mean of $J$.

Our main specification uses IDF weights, $j_{exp} = 3$ and $\gamma = 0.30$ for both MOA and disease groups. Tables A8 and A9 re-estimate Table 4 with MOA groups from 30 alternative definitions: with or without IDF weights, $j_{exp}$ of 1, 2 or 3, and $\gamma$ between 0.05 and 0.40.

**Example.** The MOA label sets $A = \{$Tumour necrosis factor ligand 4 antagonist, Immune checkpoint inhibitor$\}$ and $B = \{$Tumour necrosis factor ligand 4 antagonist, Immuno-oncology therapy$\}$, which share only "Tumour necrosis factor ligand 4 antagonist". Of the $N = 5{,}503$ label sets, only 3 contain "Tumour necrosis factor ligand 4 antagonist", while 167 contain "Immune checkpoint inhibitor" and 689 contain "Immuno-oncology therapy", so the IDF weights are

$$w_{\text{TNFL4}} = \log \tfrac{5503}{3} = 7.51, \qquad w_{\text{ICI}} = \log \tfrac{5503}{167} = 3.50, \qquad w_{\text{IO}} = \log \tfrac{5503}{689} = 2.08.$$

With $j_{exp} = 3$, the edge between $A$ and $B$ works out as follows:

| | IDF | No IDF |
|---|---|---|
| $J(A,B)$ | $\frac{7.51}{7.51+3.50+2.08} = 0.574$ | $\frac{1}{3} = 0.333$ |
| $J(A,B)^3$ | 0.189 | 0.0370 |
| Rescaling $\bar{J}/\overline{J^3}$ (over all 331,008 edges) | 17.0 | 11.9 |
| Edge weight $e(A,B)$ | 3.22 | 0.440 |

# B Additional Figures and Tables

**Table A1:** Trial Outcomes Classification

| Trial Outcome | Positive | Negative |
|---|---|---|
| Completed, Early positive outcome | 1 | 0 |
| Completed, Early positive outcome; Completed, Positive outcome/primary endpoint(s) met | 1 | 0 |
| Completed, Early positive outcome; Terminated, Business decision - Other | 1 | 0 |
| Completed, Negative outcome/primary endpoint(s) not met | 0 | 1 |
| Completed, Negative outcome/primary endpoint(s) not met; Completed, Positive outcome/primary endpoint(s) met | 1 | 0 |
| Completed, Outcome indeterminate | 0 | 0 |
| Completed, Outcome indeterminate; Completed, Outcome unknown | 0 | 0 |
| Completed, Outcome indeterminate; Completed, Positive outcome/primary endpoint(s) met | 1 | 0 |
| Completed, Outcome unknown | 0 | 0 |
| Completed, Outcome unknown; Completed, Positive outcome/primary endpoint(s) met | 1 | 0 |
| Completed, Positive outcome/primary endpoint(s) met | 1 | 0 |
| Terminated, Business decision - Drug strategy shift | 0 | 0 |
| Terminated, Business decision - Drug strategy shift; Terminated, Business decision - Other | 0 | 0 |
| Terminated, Business decision - Drug strategy shift; Terminated, Lack of efficacy | 0 | 1 |
| Terminated, Business decision - Drug strategy shift; Terminated, Safety/adverse effects | 0 | 1 |
| Terminated, Business decision - Other | 0 | 0 |
| Terminated, Business decision - Other; Terminated, Lack of efficacy | 0 | 1 |
| Terminated, Business decision - Other; Terminated, Safety/adverse effects | 0 | 1 |
| Terminated, Business decision - Pipeline reprioritization | 0 | 0 |
| Terminated, Business decision - Pipeline reprioritization; Terminated, Lack of efficacy | 0 | 1 |
| Terminated, Business decision - Pipeline reprioritization; Terminated, Safety/adverse effects | 0 | 1 |
| Terminated, Lack of efficacy | 0 | 1 |
| Terminated, Lack of efficacy; Terminated, Safety/adverse effects | 0 | 1 |
| Terminated, Lack of funding | 0 | 0 |
| Terminated, Lack of funding; Terminated, Safety/adverse effects | 0 | 1 |
| Terminated, Poor enrollment | 0 | 0 |
| Terminated, Poor enrollment; Terminated, Safety/adverse effects | 0 | 1 |
| Terminated, Safety/adverse effects | 0 | 1 |

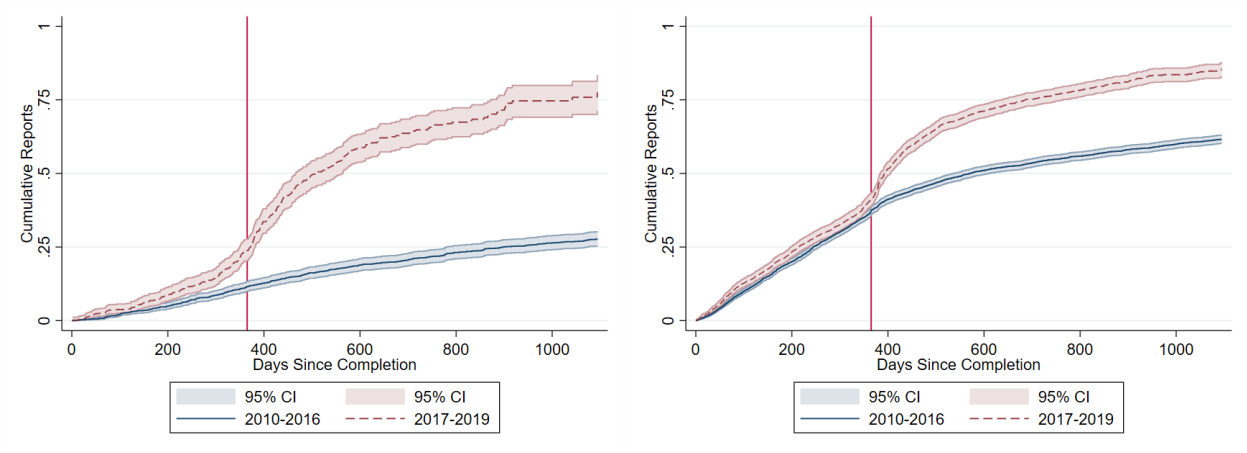
Notes: The table summarizes trial outcomes and their categorization as either positive or negative. The successful category used in the paper contains only positive trial outcomes.

**Table A2:** Share of Zero-Site Trials by Period

| | No Recorded Active Sites | | | | |
|---|---|---|---|---|---|
| Start Year $\geq$ 2017 | 0.144*** | 0.025 | 0.041* | 0.084*** | 0.020 |
| | (0.014) | (0.016) | (0.024) | (0.025) | (0.032) |
| Constant | 0.346*** | 0.338*** | 0.312*** | 0.374*** | 0.292*** |
| | (0.008) | (0.008) | (0.009) | (0.011) | (0.013) |
| $N$ | 5,280 | 4,477 | 3,012 | 2,328 | 1,549 |
| Sample | All | Completed | Reported | Unsuccessful | Successful |

Notes: The table reports OLS estimates from regressing an indicator for no recorded active sites on a constant and Start Year $\geq$ 2017. Each column restricts the sample to the indicated subset of Phase 2 US industry trials started between 2010 and 2019. The constant captures the share of trials with no recorded active sites in the pre-2017 period. Standard errors in parentheses. Stars indicate statistical significance: * 90%, ** 95%, *** 99%.

**Figure A1:** Cumulative Results Disclosure: Non-Industry and All Trials




**(i)** Non-Industry Trials **(ii)** All Trials

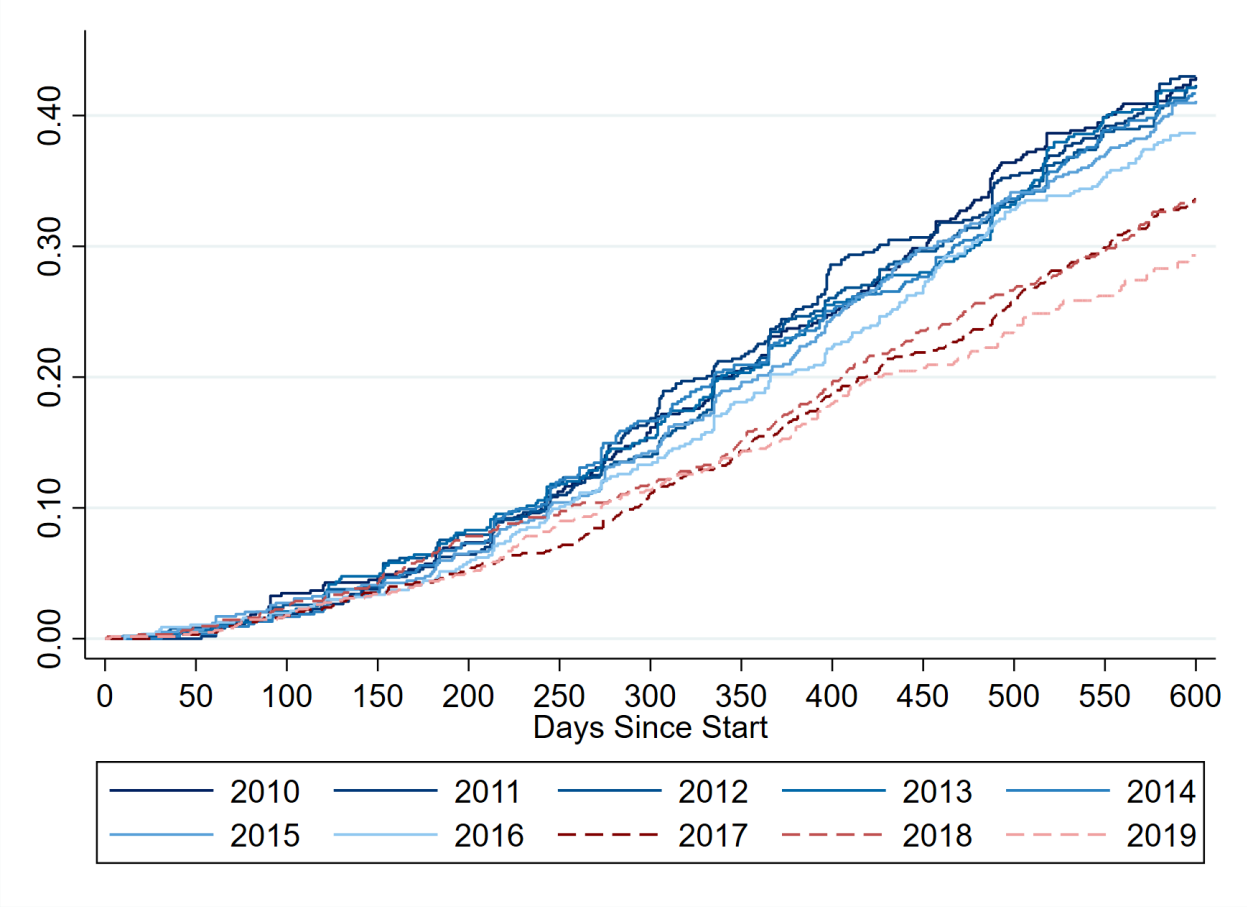
Notes: Kaplan-Meier failure curves recording the cumulative probability of results reporting for Phase 2 trials with at least one site in the United States. The two time periods refer to the year of trial completion which is relevant for reporting.

**Table A3:** Cox Model Estimates

| Start Year | Complete | Complete | Start | Start |
|---|---|---|---|---|
| 2011 | 1.022 | 1.035 | | |
| | (0.065) | (0.066) | | |
| 2012 | 0.969 | 0.931 | 0.988 | 1.010 |
| | (0.062) | (0.060) | (0.177) | (0.182) |
| 2013 | 0.948 | 0.989 | 0.954 | 0.979 |
| | (0.062) | (0.065) | (0.168) | (0.174) |
| 2014 | 0.927 | 0.898* | 0.832 | 0.839 |
| | (0.060) | (0.058) | (0.147) | (0.150) |
| 2015 | 0.893* | 0.893* | 0.822 | 0.824 |
| | (0.057) | (0.057) | (0.141) | (0.142) |
| 2016 | 0.733*** | 0.765*** | 0.715* | 0.707** |
| | (0.049) | (0.052) | (0.125) | (0.125) |
| 2017 | 0.606*** | 0.661*** | 0.588*** | 0.589*** |
| | (0.042) | (0.046) | (0.102) | (0.103) |
| 2018 | 0.643*** | 0.667*** | 0.578*** | 0.595*** |
| | (0.048) | (0.050) | (0.104) | (0.108) |
| 2019 | 0.590*** | 0.615*** | 0.557*** | 0.556*** |
| | (0.056) | (0.058) | (0.100) | (0.101) |
| Disease FE | NO | YES | NO | YES |

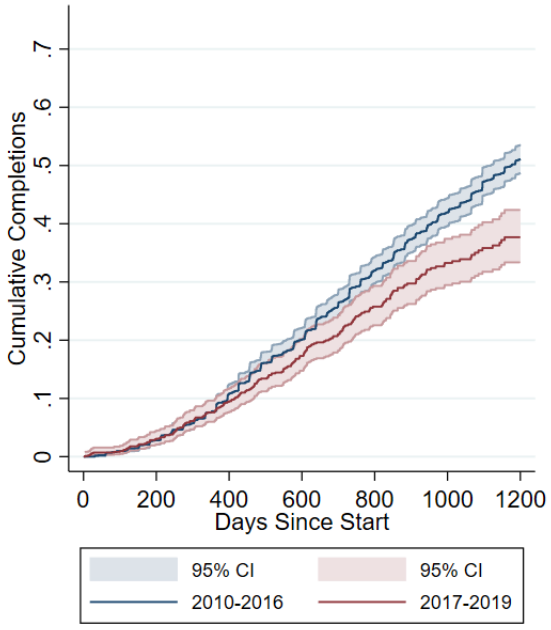
Notes: Estimates from a cox proportional hazards model. The outcome in the first two columns is Phase 2 trial completion, and year refers to the start year of the Phase 2 trial. The outcome of the second two columns is a trial start after Phase 1 completion, and year refers to the potential start year of the Phase 2 trial. Stars indicate statistical significance: * 90%, ** 95%, *** 99%.

**Figure A2:** Cumulative Completion Rates: Phase 2 Industry Trials

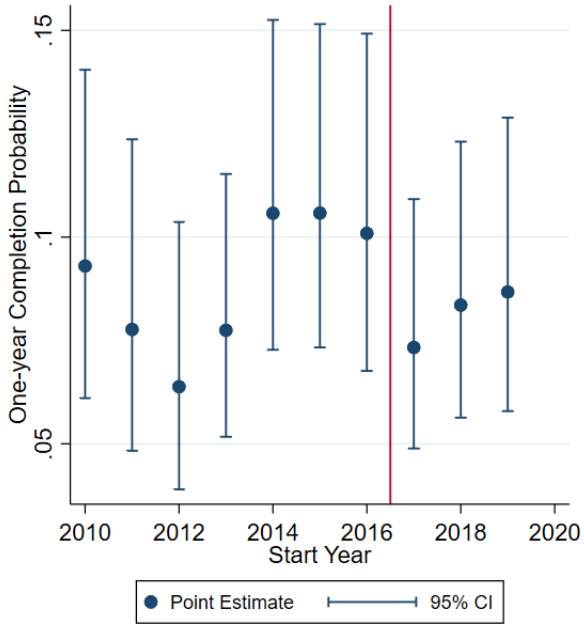



Notes: Kaplan-Meier failure curves recording the cumulative probability of Phase 2 industry trial completions with at least one site in the United States. Lines shaded in blue are for the pre-policy period (2010-2016) and dashed lines shaded in red are for the post-policy period (2017-2019).

**Figure A3:** Time to Complete Phase 2: Non-Industry



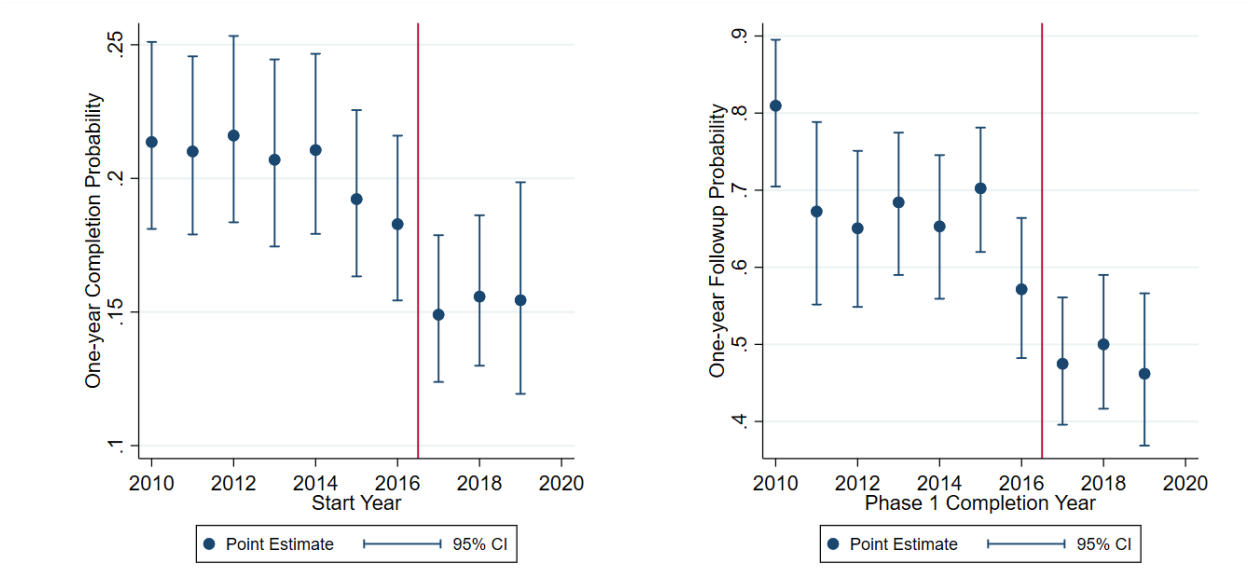

**(i)** Time to Complete by Policy Period




**(ii)** One-year Completion Probability by Start Year

Notes: Panel (i) presents Kaplan-Meier failure curves recording the cumulative probability of completion for Phase 2 non-industry trials with at least one site in the United States. The two time periods refer to the start year. Panel (ii) records point estimates and 95% confidence intervals of one-year completion probability by trial start year.

**Figure A4:** Time to Complete and Start Next Trial Truncated at COVID Onset

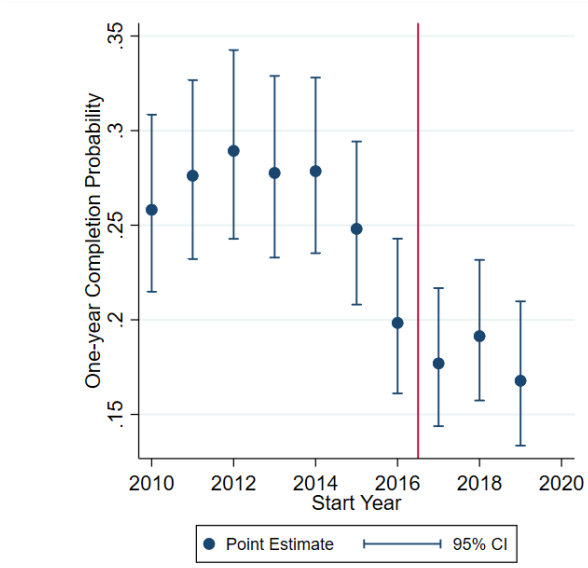



**(i)** One-year Completion Probability by Year

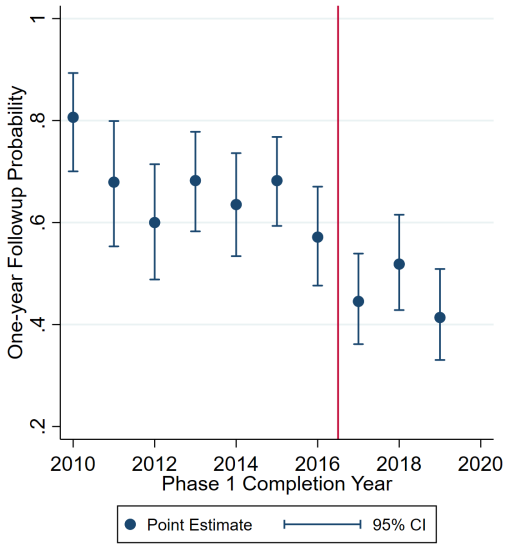
**(ii)** One-year Follow-up Probability by Year

Notes: Data truncated in March 2020. Panel (i) records point estimates and 95% confidence intervals of one-year completion probabilities by Phase 1 trial start year. Panel (ii) records point estimates and 95% confidence intervals of one-year follow-up probability by trial completion year.

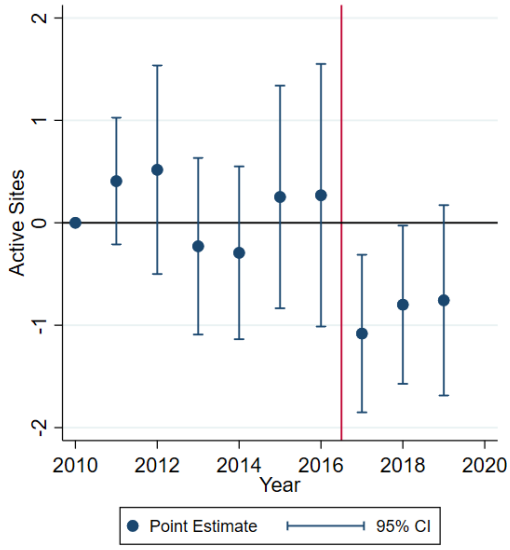
**Figure A5:** Time to Complete and Start Excluding Trials with EU Sites




**(i)** One-year Completion Probability by Year




**(ii)** One-year Follow-up Probability by Year




**(iii)** Year Fixed Effects from Sites Regressions

Notes: Sample includes all Phase 2 industry sponsored trials with some US sites and no EU sites. Panel (i) records point estimates and 95% confidence intervals of one-year completion probabilities by Phase 1 trial start year. Panel (ii) records point estimates and 95% confidence intervals of one-year follow-up probability by trial start year. Panel (iii) shows the estimated coefficients and 95% confidence intervals from regressions of active sites on year-fixed effects and months since trial start. Standard errors are two-way clustered at the trial-month level. Estimates are conditional on months since start and reflect differences in active site counts across years.

**Table A4:** Response of Trial Start to Results Disclosure: Robustness

| | | Phase 2 Start | | | |
|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) |
| Same Drug | Successful Phase 2 | 0.0229 | 0.0147 | 0.0229 | 0.0147 |
| | | (0.0183) | (0.0125) | (0.0183) | (0.0125) |
| | Unsuccessful Phase 2 | -0.0390*** | -0.0128 | -0.0388*** | -0.0128 |
| | | (0.0121) | (0.0122) | (0.0121) | (0.0122) |
| Same Disease | Successful Phase 2 | -0.0430** | 0.0125 | -0.0441** | 0.0130 |
| | | (0.0204) | (0.0120) | (0.0207) | (0.0120) |
| | Unsuccessful Phase 2 | 0.0580** | 0.0003 | 0.0594** | 0.0001 |
| | | (0.0242) | (0.0140) | (0.0246) | (0.0142) |
| Same Mechanism of Action | Successful Phase 2 | 0.0183 | 0.0307** | -0.0048 | 0.0188** |
| | | (0.0119) | (0.0120) | (0.0099) | (0.0088) |
| | Unsuccessful Phase 2 | -0.0331*** | -0.0133** | -0.0159** | -0.0192*** |
| | | (0.0110) | (0.0060) | (0.0078) | (0.0066) |
| | Completed Suc. Phase 2 | -0.0080 | 0.0015 | | |
| | | (0.0074) | (0.0098) | | |
| | Completed Unsuc. Phase 2 | 0.0207*** | -0.0176 | | |
| | | (0.0051) | (0.0110) | | |
| | Started Phase 3 | | | 0.0240** | 0.0048 |
| | | | | (0.0120) | (0.0094) |
| Time FE | | YES | YES | YES | YES |
| Drug FE | | YES | NO | YES | NO |
| Drug-Year FE | | NO | YES | NO | YES |
| $N$ | | 809,108 | 808,625 | 809,108 | 808,625 |
| $R^2$ | | 0.1853 | 0.3348 | 0.1853 | 0.3348 |

Notes: The table reports coefficient estimates of the regressions described in equation (1). The dependent variable is the number of Phase 2 trial starts in a drug-month. Regressors are measured in standard deviations. Standard errors are clustered at the drug level. Stars indicate statistical significance: * 90%, ** 95%, *** 99%.

**Table A5:** Response of Trial Start to Results Disclosure: Phase 1

| | | Phase 2 Start | Phase 1 Start | Phase 3 Start |
|---|---|---|---|---|
| Same Drug | Successful Phase 1 | 0.0063 | -0.0015 | 0.0011 |
| | | (0.0089) | (0.0049) | (0.0021) |
| | Unsuccessful Phase 1 | -0.0101*** | -0.0279*** | -0.0071*** |
| | | (0.0033) | (0.0021) | (0.0017) |
| Same Disease | Successful Phase 1 | 0.0110*** | 0.0117*** | 0.0064*** |
| | | (0.0029) | (0.0023) | (0.0016) |
| | Unsuccessful Phase 1 | -0.0010 | -0.0066*** | -0.0034** |
| | | (0.0027) | (0.0022) | (0.0017) |
| Same Mechanism of Action | Successful Phase 1 | 0.0031** | 0.0006 | 0.0041*** |
| | | (0.0013) | (0.0011) | (0.0014) |
| | Unsuccessful Phase 1 | -0.0077*** | -0.0056*** | -0.0080*** |
| | | (0.0021) | (0.0015) | (0.0018) |
| Time FE | | YES | YES | YES |
| Drug FE | | YES | YES | YES |
| $N$ | | 809,108 | 809,108 | 809,108 |
| $R^2$ | | 0.1799 | 0.0959 | 0.0683 |
| Mean of Dep. Var. | | 0.0187 | 0.0249 | 0.0118 |

Notes: The table reports coefficient estimates of the regressions described in equation (1). The dependent variables for the respective columns are the number of Phase 2, Phase 1, and Phase 3 trial starts in a drug-month. Regressors are measured in standard deviations. Standard errors are clustered at the drug level. Stars indicate statistical significance: * 90%, ** 95%, *** 99%.

**Table A6:** Response of Trial Start to Results Disclosure: Phase 3

| | | Phase 2 Start | Phase 1 Start | Phase 3 Start |
|---|---|---|---|---|
| Same Drug | Successful Phase 3 | 0.0033 | -0.0067*** | -0.0166*** |
| | | (0.0037) | (0.0023) | (0.0015) |
| | Unsuccessful Phase 3 | -0.0086** | -0.0017 | -0.0108*** |
| | | (0.0041) | (0.0020) | (0.0015) |
| Same Disease | Successful Phase 3 | -0.0047** | -0.0077*** | -0.0004 |
| | | (0.0023) | (0.0019) | (0.0012) |
| | Unsuccessful Phase 3 | 0.0131** | 0.0069* | 0.0037** |
| | | (0.0061) | (0.0037) | (0.0017) |
| Same Mechanism of Action | Successful Phase 3 | -0.0104** | -0.0058 | -0.0072** |
| | | (0.0048) | (0.0046) | (0.0030) |
| | Unsuccessful Phase 3 | 0.0073* | 0.0048 | 0.0048* |
| | | (0.0043) | (0.0042) | (0.0028) |
| Time FE | | YES | YES | YES |
| Drug FE | | YES | YES | YES |
| $N$ | | 809,108 | 809,108 | 809,108 |
| $R^2$ | | 0.1791 | 0.0896 | 0.0765 |
| Mean of Dep. Var. | | 0.0187 | 0.0249 | 0.0118 |

Notes: The table reports coefficient estimates of the regressions described in equation (1). The dependent variables for the respective columns are the number of Phase 2, Phase 1, and Phase 3 trial starts in a drug-month. Regressors are measured in standard deviations. Standard errors are clustered at the drug level. Stars indicate statistical significance: * 90%, ** 95%, *** 99%.

**Table A7:** Heterogeneous Responses to Policy Period: Full Sample

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Trial Completion | | | | | |
| Year $\geq$ 2017 | -0.0044*** | -0.0039*** | 0.0078** | 0.0039 | -0.0037** | 0.0085*** |
| | (0.0014) | (0.0014) | (0.0033) | (0.0064) | (0.0014) | (0.0032) |
| MOA Unreported $>$ 0 | | -0.0025 | -0.0001 | 0.0063** | | -0.0001 |
| | | (0.0016) | (0.0017) | (0.0032) | | (0.0017) |
| MOA Unreported $>$ 0$\times$ | | | -0.0135*** | -0.0102** | | -0.0135*** |
| Year $\geq$ 2017 | | | (0.0038) | (0.0048) | | (0.0038) |
| Small Market $\times$ | | | | | -0.0042* | -0.0040* |
| Year $\geq$ 2017 | | | | | (0.0022) | (0.0022) |
| Mean of Dep. Variable | 0.029 | 0.029 | 0.029 | 0.029 | 0.029 | 0.029 |
| | Active Sites | | | | | |
| Year $\geq$ 2017 | -1.3315*** | -1.2572*** | -0.1275 | -0.0303 | -0.8296* | 0.3260 |
| | (0.4516) | (0.4560) | (0.9004) | (1.0335) | (0.4789) | (0.9155) |
| MOA Unreported $>$ 0 | | -0.3639 | -0.1294 | 0.3000 | | -0.1291 |
| | | (0.4345) | (0.5038) | (0.5699) | | (0.5037) |
| MOA Unreported $>$ 0$\times$ | | | -1.3026 | -1.2357 | | -1.2535 |
| Year $\geq$ 2017 | | | (0.9312) | (0.9714) | | (0.9290) |
| Small Market $\times$ | | | | | -2.6772*** | -2.6562*** |
| Year $\geq$ 2017 | | | | | (1.0078) | (1.0072) |
| Mean of Dep. Variable | 5.263 | 5.263 | 5.263 | 5.265 | 5.263 | 5.263 |
| Time FE | YES | YES | YES | YES | YES | YES |
| MOA Group FE | YES | YES | YES | NO | YES | YES |
| Disease FE | YES | YES | YES | NO | YES | YES |
| Trial FE | NO | NO | NO | YES | NO | NO |
| $N$ | 126,297 | 126,297 | 126,297 | 126,250 | 126,297 | 126,297 |

Notes: The table reports coefficient estimates of the regressions described in equations (2) and (3). The sample includes all trial-months before 2020 for all phase 2 industry trials started between 2010 and 2019, including the top 5% of trials by maximum site count excluded from Table 4. Stars indicate statistical significance: * 90%, ** 95%, *** 99%. Standard errors two way clustered at the trial-date level.

**Table A8:** Heterogeneous Responses to Policy Period by MOA Group Definition: IDF Weights

| | | Trial Completion | | | Active Sites | | |
|---|---|---|---|---|---|---|---|
| $j_{exp}$ | Resolution | (3) | (4) Trial FE | (6) | (3) | (4) Trial FE | (6) |
| 1 | 0.05 | -0.0183*** | -0.0193*** | -0.0181*** | -1.8811** | -2.0646* | -1.8252** |
| | | (0.0053) | (0.0070) | (0.0054) | (0.9164) | (1.1344) | (0.9133) |
| | 0.10 | -0.0170*** | -0.0178*** | -0.0168*** | -1.9475** | -2.4957** | -1.8925** |
| | | (0.0051) | (0.0060) | (0.0051) | (0.8472) | (1.0667) | (0.8433) |
| | 0.20 | -0.0158*** | -0.0145*** | -0.0157*** | -1.5509** | -1.8195** | -1.5266** |
| | | (0.0037) | (0.0047) | (0.0037) | (0.6672) | (0.7907) | (0.6645) |
| | 0.30 | -0.0156*** | -0.0117** | -0.0154*** | -1.5819** | -1.7243** | -1.5565** |
| | | (0.0037) | (0.0045) | (0.0037) | (0.6423) | (0.7657) | (0.6394) |
| | 0.40 | -0.0124*** | -0.0091** | -0.0122*** | -1.0449* | -1.3265* | -1.0145* |
| | | (0.0033) | (0.0041) | (0.0033) | (0.5975) | (0.6994) | (0.5944) |
| 2 | 0.05 | -0.0169*** | -0.0147** | -0.0167*** | -0.4708 | -0.4468 | -0.4135 |
| | | (0.0055) | (0.0073) | (0.0055) | (0.8433) | (0.9655) | (0.8435) |
| | 0.10 | -0.0135*** | -0.0110* | -0.0133*** | -1.6153** | -1.7549* | -1.5758* |
| | | (0.0050) | (0.0062) | (0.0051) | (0.8035) | (0.9606) | (0.8002) |
| | 0.20 | -0.0158*** | -0.0121** | -0.0157*** | -2.0961*** | -2.3213*** | -2.0704*** |
| | | (0.0044) | (0.0053) | (0.0044) | (0.7326) | (0.8790) | (0.7291) |
| | 0.30 | -0.0135*** | -0.0098** | -0.0134*** | -1.1401* | -1.3598* | -1.1188* |
| | | (0.0037) | (0.0048) | (0.0037) | (0.6280) | (0.7536) | (0.6252) |
| | 0.40 | -0.0130*** | -0.0097** | -0.0129*** | -1.1076* | -1.3799* | -1.0832* |
| | | (0.0037) | (0.0046) | (0.0037) | (0.6168) | (0.7367) | (0.6140) |
| 3 | 0.05 | -0.0143** | -0.0105 | -0.0141** | -1.6487* | -2.0421* | -1.6038* |
| | | (0.0056) | (0.0070) | (0.0056) | (0.9538) | (1.1865) | (0.9495) |
| | 0.10 | -0.0166*** | -0.0126** | -0.0164*** | -1.7835** | -1.8723** | -1.7444** |
| | | (0.0050) | (0.0062) | (0.0050) | (0.7341) | (0.8635) | (0.7327) |
| | 0.20 | -0.0156*** | -0.0106** | -0.0155*** | -1.9736*** | -1.9569** | -1.9530*** |
| | | (0.0042) | (0.0048) | (0.0042) | (0.6828) | (0.8242) | (0.6803) |
| | 0.30 | -0.0148*** | -0.0111** | -0.0147*** | -1.9282*** | -2.1806*** | -1.9068*** |
| | | (0.0038) | (0.0048) | (0.0038) | (0.6568) | (0.7837) | (0.6544) |
| | 0.40 | -0.0126*** | -0.0093** | -0.0125*** | -1.3554** | -1.5281** | -1.3282** |
| | | (0.0037) | (0.0045) | (0.0037) | (0.6129) | (0.7346) | (0.6102) |

Notes: Each row re-estimates Table 4 with MOA groups built by Leiden clustering of drugs' MOA label sets, linked by IDF-weighted Jaccard similarity raised to the exponent $j_{exp}$, at the given resolution. Each cell is the MOA Unreported $> 0\times$ Year $\geq 2017$ coefficient from the indicated column of Table 4; column (4) uses trial fixed effects. The main specification is $j_{exp} = 3$, resolution 0.30. Stars indicate statistical significance: * 90%, ** 95%, *** 99%. Standard errors two way clustered at the trial-date level.

**Table A9:** Heterogeneous Responses to Policy Period by MOA Group Definition: No IDF Weights

| | | Trial Completion | | | Active Sites | | |
|---|---|---|---|---|---|---|---|
| $j_{exp}$ | Resolution | (3) | (4) Trial FE | (6) | (3) | (4) Trial FE | (6) |
| 1 | 0.05 | -0.0181*** | -0.0172** | -0.0179*** | -1.7818** | -1.7411* | -1.7469** |
| | | (0.0056) | (0.0070) | (0.0056) | (0.8581) | (1.0439) | (0.8548) |
| | 0.10 | -0.0193*** | -0.0194*** | -0.0191*** | -1.8581** | -2.1392** | -1.8091** |
| | | (0.0048) | (0.0057) | (0.0048) | (0.8165) | (0.9984) | (0.8129) |
| | 0.20 | -0.0148*** | -0.0137** | -0.0147*** | -0.8499 | -1.0601 | -0.8228 |
| | | (0.0050) | (0.0062) | (0.0050) | (0.7074) | (0.8409) | (0.7052) |
| | 0.30 | -0.0133*** | -0.0117** | -0.0132*** | -0.9317 | -0.8160 | -0.9103 |
| | | (0.0040) | (0.0050) | (0.0040) | (0.6321) | (0.7632) | (0.6295) |
| | 0.40 | -0.0149*** | -0.0122*** | -0.0149*** | -1.0379* | -1.1554 | -1.0180* |
| | | (0.0035) | (0.0043) | (0.0035) | (0.5926) | (0.7063) | (0.5896) |
| 2 | 0.05 | -0.0219*** | -0.0212*** | -0.0218*** | -1.9098** | -1.8898 | -1.8761* |
| | | (0.0059) | (0.0076) | (0.0059) | (0.9538) | (1.1664) | (0.9511) |
| | 0.10 | -0.0197*** | -0.0177** | -0.0195*** | -1.8803** | -2.0909** | -1.8321** |
| | | (0.0059) | (0.0071) | (0.0059) | (0.8502) | (1.0192) | (0.8471) |
| | 0.20 | -0.0176*** | -0.0161*** | -0.0175*** | -1.3348* | -1.3777 | -1.3111* |
| | | (0.0047) | (0.0057) | (0.0047) | (0.7464) | (0.8802) | (0.7431) |
| | 0.30 | -0.0156*** | -0.0128*** | -0.0155*** | -0.8491 | -0.7000 | -0.8208 |
| | | (0.0041) | (0.0048) | (0.0041) | (0.6422) | (0.7605) | (0.6394) |
| | 0.40 | -0.0158*** | -0.0141*** | -0.0158*** | -0.8763 | -0.8637 | -0.8627 |
| | | (0.0040) | (0.0048) | (0.0040) | (0.6220) | (0.7463) | (0.6200) |
| 3 | 0.05 | -0.0207*** | -0.0172** | -0.0204*** | -1.6713 | -1.8716 | -1.6093 |
| | | (0.0061) | (0.0081) | (0.0061) | (1.0082) | (1.2386) | (1.0054) |
| | 0.10 | -0.0206*** | -0.0179** | -0.0204*** | -1.6965** | -1.4023 | -1.6417* |
| | | (0.0060) | (0.0072) | (0.0060) | (0.8481) | (1.0222) | (0.8461) |
| | 0.20 | -0.0183*** | -0.0162*** | -0.0181*** | -1.4642* | -1.5442* | -1.4066* |
| | | (0.0050) | (0.0057) | (0.0051) | (0.7526) | (0.8989) | (0.7490) |
| | 0.30 | -0.0177*** | -0.0136*** | -0.0176*** | -0.9453 | -0.8106 | -0.9186 |
| | | (0.0043) | (0.0051) | (0.0043) | (0.6812) | (0.8144) | (0.6780) |
| | 0.40 | -0.0165*** | -0.0146*** | -0.0164*** | -1.3045** | -1.2858 | -1.2840** |
| | | (0.0040) | (0.0050) | (0.0040) | (0.6481) | (0.7927) | (0.6457) |

Notes: As Table A8, with MOA groups built from Jaccard similarity. IDF weights are not applied. Stars indicate statistical significance: * 90%, ** 95%, *** 99%. Standard errors two way clustered at the trial-date level.

**Table A10:** MOA and Disease Groups: Summary Statistics

| | Groups | Drugs per Group | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *N* | Mean | SD | p25 | p50 | p75 | p90 | p99 | Max |
| Disease | 217 | 49.6 | 80.3 | 4 | 19 | 69 | 122 | 303 | 795 |
| MOA | 2,093 | 11.5 | 72.4 | 1 | 3 | 8 | 22 | 106 | 2,766 |
| | Drugs | Peers per Drug | | | | | | | |
| | *N* | Mean | SD | p25 | p50 | p75 | p90 | p99 | Max |
| Disease | 10,714 | 178.4 | 196.0 | 64 | 109 | 205 | 347 | 794 | 794 |
| MOA | 23,997 | 465.5 | 888.9 | 14 | 44 | 170 | 2,765 | 2,765 | 2,765 |

Notes: Groups are the main-specification Leiden groups ($j_{exp} = 3$, resolution 0.30, IDF weights) used in Tables 3 and 4. "Drugs per group" is the number of drugs in a group; "Peers per Drug" is the number of other drugs in the drug's group.

**Table A11:** Response of Investment to MOA State by Policy Period

| | Trial Completion | | Active Sites | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Year $\geq$ 2017 | 0.0130*** | 0.0177** | 1.9218** | 2.6101** |
| | (0.0047) | (0.0087) | (0.7589) | (1.0011) |
| MOA Unreported $>$ 0 | -0.0002 | 0.0051 | 0.4161 | 1.2263*** |
| | (0.0018) | (0.0032) | (0.3306) | (0.4671) |
| MOA Unreported $>$ 0$\times$ | -0.0144*** | -0.0095** | -1.6626** | -1.9982** |
| Year $\geq$ 2017 | (0.0039) | (0.0048) | (0.6564) | (0.7811) |
| MOA Ongoing $>$ 0 | -0.0012 | -0.0055 | 0.6738 | 0.2567 |
| | (0.0029) | (0.0045) | (0.4115) | (0.5425) |
| MOA Ongoing $>$ 0$\times$ | -0.0043 | -0.0162** | -1.5861*** | -1.4040* |
| Year $\geq$ 2017 | (0.0042) | (0.0065) | (0.5650) | (0.7344) |
| Mean of Dep. Variable | 0.029 | 0.029 | 4.058 | 4.059 |
| Time FE | YES | YES | YES | YES |
| MOA Group FE | YES | NO | YES | NO |
| Disease FE | YES | NO | YES | NO |
| Trial FE | NO | YES | NO | YES |
| *N* | 121,890 | 121,845 | 121,890 | 121,845 |

Notes: The table reports coefficient estimates of the regressions described in equations (2) and (3) with additional controls indicating the existence of ongoing trials in $M_j$. The sample includes all trial-months before 2020 for all Industry US phase 2 trials started between 2010 and 2019, excluding the largest 5% of trials by maximum site count. Stars indicate statistical significance: * 90%, ** 95%, *** 99%. Standard errors two way clustered at the trial-date level.

# C Proofs for Simple Theoretical Framework

In period 2, only the firm that waited in period 1 can invest. Let $U^I$ denote a firm's expected payoff from investing in period 1, and let $U^W$ denote its expected payoff from waiting in period 1.

Let us first consider the best response of the follower in the second period. Suppose firm $-f$ invested in period 1, so $a_{-f1} = 1$, and firm $f$ is the follower in period 2. With disclosure probability $\alpha$, $f$ observes $s_{-f}$ and updates

$$q_1(\rho) \equiv \Pr(s_f = 1 \mid s_{-f} = 1) = \rho + (1-\rho)p$$
$$q_0(\rho) \equiv \Pr(s_f = 1 \mid s_{-f} = 0) = (1-\rho)p.$$

With probability $(1-\alpha)$ it observes nothing and keeps prior $p$.

Conditional on the information available in period 2, the follower invests if and only if the corresponding expected flow payoff $\mathbb{E}[\pi_{f2} \mid \mathcal{I}] - c \geq 0$ is non-negative.

Under $\lambda = 1$ (Propositions 1–2), the follower's expected gross payoff from investing given belief $q$ is $q\pi$. Under $\rho = 0$ (Hypotheses 3–4), the follower's expected gross payoff from investing given that it learns $s_{-f}$ is $p\pi$ if $s_{-f} = 0$ and $p\pi\lambda$ if $s_{-f} = 1$.

## C.1 Proposition 1

When $\lambda = 1$ there is no business-stealing, so any strategic delay must come from the informational option value created by (i) correlation $\rho$ and (ii) disclosure $\alpha$. For sufficiently low $\rho$, the informational value of waiting is small and both firms invest immediately. For sufficiently high $\rho$, waiting becomes attractive if the rival invests, generating two asymmetric pure-strategy equilibria (one firm invests, the other waits) and a symmetric mixed-strategy equilibrium.

Assume $\lambda = 1$. If a firm invests in period 1, it pays $c$ and obtains expected period-1 payoff $p\pi - c$. In period 2 it cannot invest, but it still earns payoff from its own project based on its period-1 signal. Since $\Pr(s_f = 1) = p$, its expected period-2 payoff is $\delta p\pi$. Thus

$$U^I = p\pi - c + \delta p\pi.$$

Now compute the value of waiting in period 1 when the rival invests in period 1 with probability $x$. If the rival invests with probability $x$, then in period 2 the follower invests only when its posterior success probability makes $q\pi - c \geq 0$. Using the posteriors above, the follower's period-2 continuation value (gross of discounting) is

$$V^F(\rho, \alpha) = \alpha\Big( \Pr(s_{-f} = 1) \max\{0, q_1(\rho)\pi - c\} + \Pr(s_{-f} = 0) \max\{0, q_0(\rho)\pi - c\} \Big) + (1-\alpha) \max\{0, p\pi - c\}.$$

Since $\Pr(s_{-f} = 1) = p$ and $\Pr(s_{-f} = 0) = 1 - p$ and $p\pi - c > 0$, this becomes

$$V^F(\rho, \alpha) = \alpha\Big(p \max\{0, q_1(\rho)\pi - c\} + (1-p)\max\{0, q_0(\rho)\pi - c\}\Big) + (1-\alpha)(p\pi - c).$$

If the rival firm does not invest in period 1 (probability $1 - x$), then period 2 is essentially a repeat of period 1, and investing in period 2 yields net payoff $(p\pi - c)$ discounted by $\delta$. The strategic interaction in that subgame is the same as period 1. For the threshold argument below it suffices that the follower can secure at least the discounted value $(p\pi - c)$ by investing in period 2 when it is optimal to do so.

Hence, the payoff from waiting satisfies

$$U^W(x) = \delta\Big(xV^F(\rho, \alpha) + (1-x)(p\pi - c)\Big).$$

Define the gain from investing now rather than waiting when the rival invests for sure ($x = 1$) as follows:

$$\Delta(\rho) \equiv U^I - U^W(1) = p\pi - c + \delta p\pi - \delta V^F(\rho, \alpha).$$

$V^F(\rho, \alpha)$ is weakly increasing in $\rho$ because $q_1(\rho)$ increases in $\rho$ and $q_0(\rho)$ decreases in $\rho$, which raises the informational value of conditioning the period-2 investment decision on $s_{-f}$ because the follower invests more selectively and avoids low-value states. Thus, $\Delta(\rho)$ is weakly decreasing in $\rho$.

At $\rho = 0$, the rival's signal is uninformative about $f$'s success probability, so conditioning is worthless and $V^F(0, \alpha) = p\pi - c$ which yields

$$\Delta(0) = p\pi - c + \delta p\pi - \delta(p\pi - c) = (p\pi - c)(1 - \delta) + \delta p\pi > 0.$$

For $\rho$ close to 1, $s_{-f}$ becomes highly informative, and the follower can avoid paying $c$ in low-posterior states while still investing when the posterior is high. Hence $V^F(\rho, \alpha)$ becomes large enough that $\Delta(\rho)$ can become negative for sufficiently patient firms (large $\delta$) and sufficiently high disclosure (large $\alpha$). By monotonicity of $\Delta(\rho)$, there exists a threshold $\rho^* \in (0, 1)$ such that $\Delta(\rho) > 0$ for $\rho < \rho^*$ and $\Delta(\rho) < 0$ for $\rho > \rho^*$.

Equilibrium implications follow the standard best-response logic:

- If $\rho < \rho^*$, investing is a best response to the rival investing. Since $p\pi - c > 0$, investing is also a best response to the rival waiting. Thus, $(a_{A1}, a_{B1}) = (1, 1)$ is the unique pure-strategy SPNE.

- If $\rho > \rho^*$, waiting is a best response to the rival investing (because the option value dominates), but investing is a best response to the rival waiting (since $p\pi - c > 0$). This generates two asymmetric pure-strategy SPNE: $(1, 0)$ and $(0, 1)$.

- In the same region there is also a symmetric mixed-strategy equilibrium where each firm invests with probability $\gamma \in (0, 1)$, characterized by indifference between investing and waiting.

## C.2 Proposition 2

In the symmetric mixed-strategy equilibrium from Proposition 1, higher disclosure makes waiting more valuable because the follower is more likely to observe $s_{-f}$ and condition its decision on it. Thus, to restore indifference the equilibrium must put less weight on investing in period 1.

In the symmetric mixed-strategy equilibrium, $\gamma$ solves the indifference condition

$$U^I = U^W.$$

Using the expressions above,

$$p\pi - c + \delta p\pi = \delta\Big(\gamma V^F(\rho,\alpha) + (1-\gamma)(p\pi - c)\Big),$$

which can be rearranged as

$$\gamma = \frac{p\pi - (1-\delta)c}{\delta\big(V^F(\rho,\alpha) - (p\pi - c)\big)}.$$

The numerator is always positive. For $\rho > \rho^*$, the denominator is also positive because $V^F(\rho,\alpha) > p\pi - c$. Moreover, $V^F(\rho,\alpha)$ is increasing in $\alpha$ because disclosure makes it more likely that the follower conditions on $s_{-f}$ rather than relying on the prior. Therefore, the denominator increases in $\alpha$ while the numerator does not depend on $\alpha$, implying

$$\frac{\partial\gamma}{\partial\alpha} < 0.$$

## C.3 Proposition 3

### C.3.1 Period 2

Recall that period-1 actions $(a_{A1}, a_{B1})$ are always observable. Disclosure only affects whether signals $(s_A, s_B)$ are observable.

We only have to consider the period-2 decision for the firm if it did not invest in period 1. If it succeeded, then there is no decision because it can invest only once. If it did not succeed in period 1, then it does not invest in period 2 because project success is persistent.

**Other Firm Did Not Invest in Period 1** Assume the other firm did not invest in period 1. Disclosure plays no role because no signals were generated. In period 2, firms $f \in \{A, B\}$ simultaneously choose $a_{f2} \in \{0, 1\}$. If firm $f$ invests ($a_{f2} = 1$), it pays cost $c$, draws $s_f \in \{0, 1\}$ with $\Pr(s_f = 1) = p$, and earns

$$\pi_{f2} = \begin{cases} s_f\pi\lambda, & \text{if } a_{-f2} = 1 \ \&\ s_{-f} = 1, \\ s_f\pi, & \text{otherwise.} \end{cases}$$

Note that under $\rho = 0$, signals are independent across firms.

Let $\gamma_2$ denote the probability that the rival invests in period 2 in a symmetric mixed-strategy equilibrium. If firm $f$ invests in period 2, its expected gross payoff is

$$\begin{aligned}\mathbb{E}[\pi_{f2} \mid a_{f2} = 1] &= (1-\gamma_2)\, p\pi + \gamma_2\, p\pi\big((1-p) + p\lambda\big) \\ &= p\pi - \gamma_2 p^2 \pi(1-\lambda).\end{aligned}$$

Therefore, the expected net payoff from investing is

$$\mathbb{E}[\pi_{f2} \mid a_{f2} = 1] - c = p\pi - \gamma_2 p^2 \pi(1-\lambda) - c.$$

Not investing yields payoff 0. In a symmetric mixed-strategy equilibrium with $\gamma_2 \in (0,1)$, the firm must be indifferent between investing and not investing, so

$$p\pi - \gamma_2 p^2 \pi(1-\lambda) - c = 0,$$

which implies

$$\gamma_2 = \frac{p\pi - c}{p^2\pi(1-\lambda)}.$$

This mixed-strategy equilibrium exists (i.e., $\gamma_2 \in (0,1)$) provided

$$p\pi((1-p) + p\lambda) < c < p\pi,$$

since $p\pi((1-p)+p\lambda) - c$ is the net payoff from investing when the rival invests for sure and $p\pi - c$ is the net payoff from investing when the rival never invests. Equivalently, the condition for $\lambda$ is

$$\lambda < \frac{c - p\pi(1-p)}{p^2\pi} \equiv \lambda_2^*.$$

If $c \leq p\pi((1-p)+p\lambda)$, then investing is a best response even when the rival invests for sure and the unique equilibrium is $(a_{A2}, a_{B2}) = (1,1)$. If $c \geq p\pi$, investing is never profitable and the unique equilibrium is $(a_{A2}, a_{B2}) = (0,0)$.

For $\lambda < \lambda_2^*$ there are also two asymmetric pure-strategy equilibria in the period-2 subgame for $(a_{A1}, a_{B1})$: $(1,0)$ and $(0,1)$.

**Other Firm Invested in Period 1** Suppose firm $-f$ invested in period 1 (so $a_{-f1} = 1$), then it is only a deterministic decision for firm $f$ in period 2. If there was no disclosure, firm $f$ invests if

$$p\pi\big((1-p) + p\lambda\big) \geq c.$$

If disclosure occurs (probability $\alpha$), firm $f$ observes $s_{-f} \in \{0,1\}$. In period 2, firm $f$ chooses

whether to invest, $a_{f2} \in \{0, 1\}$. Conditional on $s_{-f}$, the expected net payoff from investing is

$$\mathbb{E}[\pi_{f2} \mid s_{-f} = 0, a_{f2} = 1] - c = p\pi - c, \qquad \mathbb{E}[\pi_{f2} \mid s_{-f} = 1, a_{f2} = 1] - c = p\pi\lambda - c.$$

Firm $f$ observes $s_{-f}$ with disclosure probability $\alpha$. If disclosure occurs, the period-2 best response is

$$a_{f2} = \begin{cases} 1, & \text{if } s_{-f} = 0 \\ 0, & \text{if } s_{-f} = 1 \end{cases}$$

because $p\pi\lambda < p\pi\big((1-p) + p\lambda\big) < c$ or equivalently $\lambda < \lambda_2^*$

If disclosure does not occur (probability $1 - \alpha$), firm $f$ compares the expected net payoff from investing,

$$\mathbb{E}[\pi_{f2} \mid a_{f2} = 1] - c = p\pi\big((1-p) + p\lambda\big) - c,$$

to zero. Hence without disclosure,

$$a_{f2} = 1 \quad \Longleftrightarrow \quad p\pi\big((1-p) + p\lambda\big) \geq c.$$

If $\lambda < \lambda_2^*$ this condition is not satisfied and the firm does not invest.

**Continuation values in period 2:** If neither firm invested in period 1, then period 2 is the simultaneous-move subgame. In the symmetric mixed equilibrium of that subgame, each firm is indifferent between investing and not investing, and the payoff from not investing is 0. Hence each firm's equilibrium continuation value in this branch is 0.

If instead one firm invested in period 1, the other firm is a follower in period 2. Under $p\pi\big((1-p) + p\lambda\big) < c < p\pi$, without disclosure the follower does not invest. With disclosure, the follower observes $s_{-f}$ and invests if and only if $s_{-f} = 0$ (since $p\pi - c > 0$ and $p\pi\lambda - c < 0$). Therefore, the follower invests with probability $\alpha(1-p)$ and, conditional on investing, obtains expected net payoff $p\pi - c$. The follower's expected period-2 payoff (gross of discounting) is

$$V^F = \alpha(1-p)(p\pi - c).$$

### C.3.2 Period 1

Assume $\rho = 0$ and consider the "competition" parameter region

$$p\pi\big((1-p) + p\lambda\big) < c < p\pi,$$

so investing is profitable if the rival certainly does not invest, but unprofitable if the rival certainly invests.

Let $\gamma_1$ denote the probability the rival invests in period 1 in a symmetric mixed equilibrium.

**Expected payoff from investing in period 1:** If firm $f$ invests in period 1, its expected period-1 gross payoff given rival investment probability $\gamma_1$ is

$$\mathbb{E}[\pi_{f1} \mid a_{f1} = 1] = (1-\gamma_1)p\pi + \gamma_1 p\pi\big((1-p) + p\lambda\big) = p\pi - \gamma_1 p^2\pi(1-\lambda).$$

For the period-2 payoff, if the rival also invested in period 1 (probability $\gamma_1$), firm $f$'s expected period-2 payoff is

$$p\pi\big((1-p) + p\lambda\big).$$

If instead the rival did not invest in period 1 (probability $1-\gamma_1$), the rival is the potential follower in period 2. Whenever $s_f = 1$ the follower does not enter (either because it observes $s_f = 1$ and stays out, or because without disclosure it does not enter at all), so firm $f$ receives the monopoly payoff $\pi$ whenever it succeeds. Thus the expected period-2 payoff in this branch is $p\pi$. Hence the expected discounted period-2 payoff is

$$\delta\Big[\gamma_1\, p\pi\big((1-p) + p\lambda\big) + (1-\gamma_1)\, p\pi\Big] = \delta\, p\pi\Big(1 - \gamma_1 p(1-\lambda)\Big).$$

Combining,

$$U_1^I(\gamma_1) = \Big(p\pi - \gamma_1 p^2\pi(1-\lambda) - c\Big) + \delta\, p\pi\Big(1 - \gamma_1 p(1-\lambda)\Big). \tag{8}$$

**Expected payoff from waiting in period 1.** If firm $f$ waits in period 1, then with probability $1-\gamma_1$ the rival also waits and the period-2 subgame payoff is 0 in the symmetric mixed equilibrium. With probability $\gamma_1$ the rival invested in period 1 and firm $f$ is the follower in period 2, yielding (gross) continuation value $V^F = \alpha(1-p)(p\pi - c)$. Therefore

$$U_1^W(\gamma_1) = \delta\, \gamma_1\, \alpha(1-p)(p\pi - c). \tag{9}$$

**Symmetric mixed equilibrium in period 1.** A symmetric mixed equilibrium with $\gamma_1 \in (0,1)$ satisfies indifference:

$$U_1^I(\gamma_1) = U_1^W(\gamma_1).$$

Using (8)–(9) and collecting terms yields

$$p\pi(1+\delta) - c = \gamma_1\Big(p^2\pi(1-\lambda)(1+\delta) + \delta\,\alpha(1-p)(p\pi - c)\Big),$$

so

$$\gamma_1 = \frac{p\pi(1+\delta) - c}{p^2\pi(1-\lambda)(1+\delta) + \delta\,\alpha(1-p)(p\pi - c)}.$$

In the region where this expression lies in $(0,1)$, the period-1 stage game also admits the two asymmetric pure equilibria $(1,0)$ and $(0,1)$.

## C.4 Proposition 4

From the equation for $\gamma_1$ above, the numerator does not depend on $\alpha$ and the denominator is increasing in $\alpha$ because $p\pi - c > 0$. Thus, we have

$$\frac{\partial \gamma_1}{\partial \alpha} < 0.$$

## C.5 Proof of Proposition 5

In this proof we set $\lambda = 1$, so that there is no product-market competition and delay arises only through informational free-riding. We also adopt the welfare criterion described in the main text: social welfare is the sum of firms' expected discounted payoffs and consumer surplus from successful innovation. Let $s > 0$ denote the consumer surplus generated by a successful project in the period in which that success is realized.

We impose the following conditions:

$$p\pi - c > 0, \tag{10}$$

$$(\rho + (1-\rho)p)\pi - c > 0, \tag{11}$$

$$(1-\rho)p\pi - c < 0, \tag{12}$$

$$\delta p\left[(\rho + (1-\rho)p)\pi - c\right] > p\pi - c. \tag{13}$$

Condition (10) ensures that investment is ex ante profitable. Conditions (11) and (12) imply that, under full disclosure, a firm that waits until period 2 finds it optimal to invest after observing a successful rival signal but not after observing an unsuccessful rival signal. Condition (13) ensures that, under full disclosure, waiting is privately optimal when the rival invests in period 1.

We characterize equilibrium behavior under $\alpha = 0$ and $\alpha = 1$.

**Case 1:** $\alpha = 0$**.** If a firm invests in period 1, its expected payoff is

$$p\pi - c.$$

If instead it waits until period 2, then because there is no disclosure it cannot condition its decision on the rival's signal, so its expected payoff from waiting is

$$\delta(p\pi - c).$$

By (10) and $\delta < 1$, we have

$$p\pi - c > \delta(p\pi - c).$$

Thus investing in period 1 strictly dominates waiting, regardless of the rival's action. Hence under $\alpha = 0$ both firms invest in period 1.

Social welfare under $\alpha = 0$ is therefore

$$W(0) = 2(p\pi - c) + 2ps. \tag{14}$$

The first term is total firm surplus and the second term is total consumer surplus, since both firms invest immediately and each succeeds with probability $p$.

**Case 2:** $\alpha = 1$**.** Now suppose disclosure is perfect. Consider a firm whose rival invests in period 1. If the firm waits, then after observing a successful rival signal its posterior success probability is

$$\Pr(s_f = 1 \mid s_{-f} = 1) = \rho + (1 - \rho)p,$$

whereas after observing an unsuccessful rival signal its posterior success probability is

$$\Pr(s_f = 1 \mid s_{-f} = 0) = (1 - \rho)p.$$

By (11) and (12), the firm invests in period 2 if and only if it observes a successful rival signal. Its expected payoff from waiting is therefore

$$\delta p \left[(\rho + (1 - \rho)p)\pi - c\right].$$

By (13), this exceeds the payoff from investing immediately, $p\pi - c$, when the rival invests in period 1.

If instead the rival waits, then investing in period 1 yields payoff $p\pi - c$, whereas waiting until period 2 yields only $\delta(p\pi - c)$, which is strictly smaller by (10) and $\delta < 1$. Hence if the rival waits, it is optimal to invest immediately. It follows that under $\alpha = 1$ there is a pure-strategy equilibrium in which one firm invests in period 1 and the other waits until period 2 and invests only if the first firm's signal is successful.

Consider such an equilibrium. Welfare under $\alpha = 1$ is

$$W(1) = (p\pi - c) + \delta p \left[(\rho + (1 - \rho)p)\pi - c\right] + ps + \delta p(\rho + (1 - \rho)p)s. \tag{15}$$

The first term is the investing firm's expected payoff in period 1. The second term is the waiting firm's expected discounted payoff in period 2: with probability $p$ the first firm's signal is successful, in which case the second firm invests and has posterior success probability $\rho + (1 - \rho)p$. The third term is consumer surplus from the first firm's success, and the fourth term is discounted consumer surplus from the second firm's success.

Subtracting (15) from (14) yields

$$\begin{aligned} W(0) - W(1) &= [2(p\pi - c) + 2ps] - [(p\pi - c) + \delta p\left((\rho + (1-\rho)p)\pi - c\right) + ps + \delta p(\rho + (1-\rho)p)s] \\ &= (p\pi - c) - \delta p\left[(\rho + (1-\rho)p)\pi - c\right] + p\left[1 - \delta(\rho + (1-\rho)p)\right]s. \end{aligned} \tag{16}$$

Since $\delta < 1$ and $\rho + (1-\rho)p \leq 1$, the coefficient on $s$ in (16) is strictly positive. Therefore $W(0) > W(1)$ whenever

$$s > \frac{\delta p\left[(\rho + (1-\rho)p)\pi - c\right] - (p\pi - c)}{p\left[1 - \delta(\rho + (1-\rho)p)\right]}. \tag{17}$$

Under (13), the numerator in (17) is positive, so this is a non-empty restriction.

Hence there exists a set of parameter values satisfying (10)–(13) and (17) for which

$$W(0) > W(1).$$

In particular, full disclosure yields lower welfare than no disclosure. It follows immediately that $\alpha = 1$ cannot be socially optimal. Therefore the socially optimal disclosure probability satisfies

$$\alpha^{SP} < 1.$$

**Mixed-strategy equilibrium.** The welfare result also extends to the mixed-strategy equilibrium under $\alpha = 1$. Suppose firms mix in period 1 and invest with probability $\gamma \in (0, 1)$. There are three possible outcomes.

With probability $\gamma^2$, both firms invest in period 1, yielding welfare $W(0)$ as defined in (14). With probability $2\gamma(1-\gamma)$, exactly one firm invests in period 1, yielding welfare $W(1)$ as defined in (15). With probability $(1-\gamma)^2$, both firms delay and invest in period 2. In that case, welfare is simply the discounted version of the immediate-investment benchmark, $\delta W(0)$.

Hence welfare in the mixed-strategy equilibrium is

$$W_M(1) = \gamma^2 W(0) + 2\gamma(1-\gamma)W(1) + (1-\gamma)^2\delta W(0). \tag{18}$$

Subtracting (18) from $W(0)$ yields

$$\begin{aligned} W(0) - W_M(1) &= \gamma^2\big[W(0) - W(0)\big] + 2\gamma(1-\gamma)\big[W(0) - W(1)\big] + (1-\gamma)^2\big[W(0) - \delta W(0)\big] \\ &= 2\gamma(1-\gamma)\big[W(0) - W(1)\big] + (1-\gamma)^2(1-\delta)W(0). \end{aligned} \tag{19}$$

Since $\gamma \in (0, 1)$ and $\delta < 1$, the second term in (19) is strictly positive. Moreover, under the condition derived above that $W(0) > W(1)$, the first term is also strictly positive. It follows that

$$W(0) > W_M(1).$$

Thus, even when firms mix in equilibrium under full disclosure, welfare is strictly lower than under no disclosure. Therefore $\alpha = 1$ cannot be socially optimal, implying $\alpha^{SP} < 1$.