DISCLOSURE AND THE PACE OF DRUG DEVELOPMENT

By

Colleen Cunnningham, Florian Ederer,

Charles Hodgson and Zhichun Wang

October 2025

COWLES FOUNDATION DISCUSSION PAPER NO. 2465

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

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 Victor Bennett, Vivek Bhattacharya, Federico Ciliberto, Jeff Dyer, Liran Einav, Tim Gubler, Matt Higgins, Philip Haile, Tomas Larroucau, Alexander MacKay, Nicholas Vreugdenhil, and seminar participants at the Alpine IO Symposium, ASU, the BYU Winter Strategy Conference, and UVA 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 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 78%, representing a 15% 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 60% to 75% while the reporting rate for successful results increased from 85% to 92%. 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 900 days. At the same time, the probability of completing a trial within two years 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 oncology 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 50%, from an average of approximately 8 sites per trial before the policy to fewer than 4 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 of successful Phase 2 trials of drugs with a similar mechanism of action increases the probability of a subsequent Phase 2 trial by 1.65 percentage points. Conversely, an increase in unsuccessful Phase 2 trials reduces the probability of a subsequent Phase 2 trial by 1.17 percentage points. These effects are present within the same mechanism of action rather than across broader therapeutic classes, 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.

therapeutic class follow the opposite pattern. A one-standard deviation increase in successful Phase 2 trials by a competing drug in the same therapeutic class reduces the probability of a Phase 2 Start by 1.08 percentage points while the same increase in unsuccessful Phase 2 trials increases it by 1.86 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 starts to capitalize on information from ongoing disclosures. The probability of starting a Phase 2 trial is significantly lower when many related Phase 2 trials 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 drug fixed effects and the number of unreported trial results, trials that started in 2017 or after had about 1.5 fewer active sites on average than trials started before 2017 and this effect is amplified the more

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 provides background on the regulatory framework and the role of disclosure mandates in drug development. Section 3 describes our data sources. Section 4 presents our main findings on disclosure patterns and innovation responses. Section 5 shows how the empirical findings are related to firm learning. Section 7 concludes.

2 Industry and Institutional Background

2.1 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 with a fourth optional phase. 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 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).

2.2 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

[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.

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 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.

3 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

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

Table 1: Trial Sequence Data: Summary Statistics

		Mean	SD	Median	Min	Max	N
Trial Length	Phase 1	549.19	622.20	308	1	8513	13128
	Phase 2	913.77	697.63	731	1	9953	12720
	Phase 3	944.38	727.10	748	8	7018	9754
Reported	Phase 1	0.21	0.41	0	0	1	9909
	Phase 2	0.71	0.46	1	0	1	8731
	Phase 3	0.80	0.40	1	0	1	7201
Reporting Delay	Phase 1	486.76	473.60	350	1	3683	4935
	Phase 2	432.23	456.58	322	1	3908	6540
	Phase 3	363.72	354.72	289	2	3479	5961
	Time To Approval	1833.76	994.34	1707	19	9681	938
Approved Drugs	Phase 1 Count	2.56	4.12	1	0	29	938
	Phase 2 Count	1.97	3.23	1	0	32	938
	Phase 3 Count	3.08	3.60	2	0	31	938
	Share Reported	0.59	0.33	1	0	1	938
	Share Successful	0.64	0.33	1	0	1	879
Unapproved Drugs	Phase 1 Count	1.33	1.84	1	0	26	7715
	Phase 2 Count	0.94	2.02	0	0	54	7715
	Phase 3 Count	0.45	1.41	0	0	43	7715
	Share Reported	0.23	0.33	0	0	1	7715
	Share Successful	0.33	0.39	0	0	1	4789

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

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.

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

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 started between 2010 and 2019. The first set of rows records summary statistics on trial length. The mean Phase 1 trial runs for close to 550 days or about one and a half years. Phase 2 and 3 trials are longer, about two and half 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 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 between one and two years 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 Phase 1 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 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 oncology trials in our main Trialtrove dataset.

Table 2 summarizes our site count data. There are 1,294 trials in the data, with an average trial length of 16 months. Trials take place over a large number of sites. The average number of

[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 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.

Table 2: Site Count Data: Summary Statistics

	Mean	SD	Median	Min	Max	N
Trial Length	16.10	15.33	0	1	112	1294
Active Site Count	17.38	33.92	5	0	748	20836
Number of Changes	2.14	2.86	1	1	41	1294
Change Magnitude	11.06	28.77	2	-121	518	2769
Maximum Sites	25.01	42.85	10	1	748	1294

Notes: Sample includes all industry sponsored Phase 2 oncology trials that were active between 2010 and 2019. Active site count is measured at the trial-month level. Changes are recorded be 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 at the trial level.

active sites at the trial-month level is 17.38 with some trials taking place over hundreds of sites. The number of active sites also varies over time within a trial. The third and fourth row record statistics on monthly changes in the number of active sites. The number of active sites changes an average of 2.14 times over a trial's duration (including the initial site openings that occur when the trial starts). On average, the number of sites increases over time, with the average change being an addition of 11.06 sites, although sites are also sometimes removed. These descriptive statistics show that trials often start at a small scale with just a few active sites and then ramp up in magnitude as trials progress.

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 8 at the 0% progress to around 25 at 100% progress. Finally, there is a small decrease in the number of active sites as trials are close to completion.

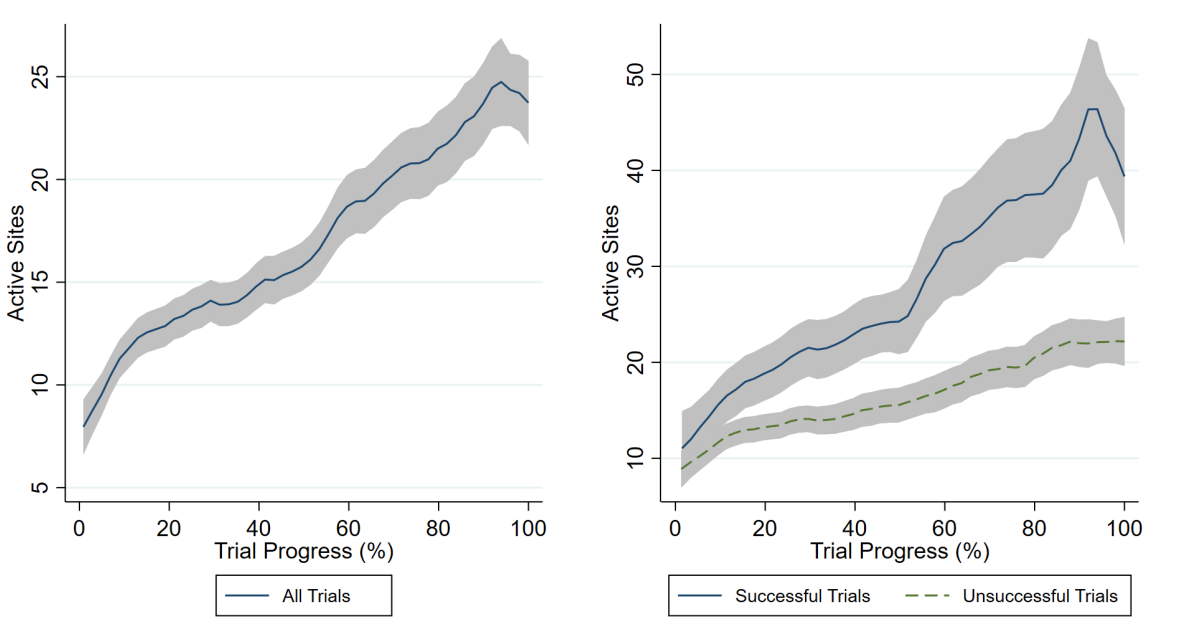
4 Empirical Analysis

4.1 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.

Figure 1: Within-trial Addition of Active Sites



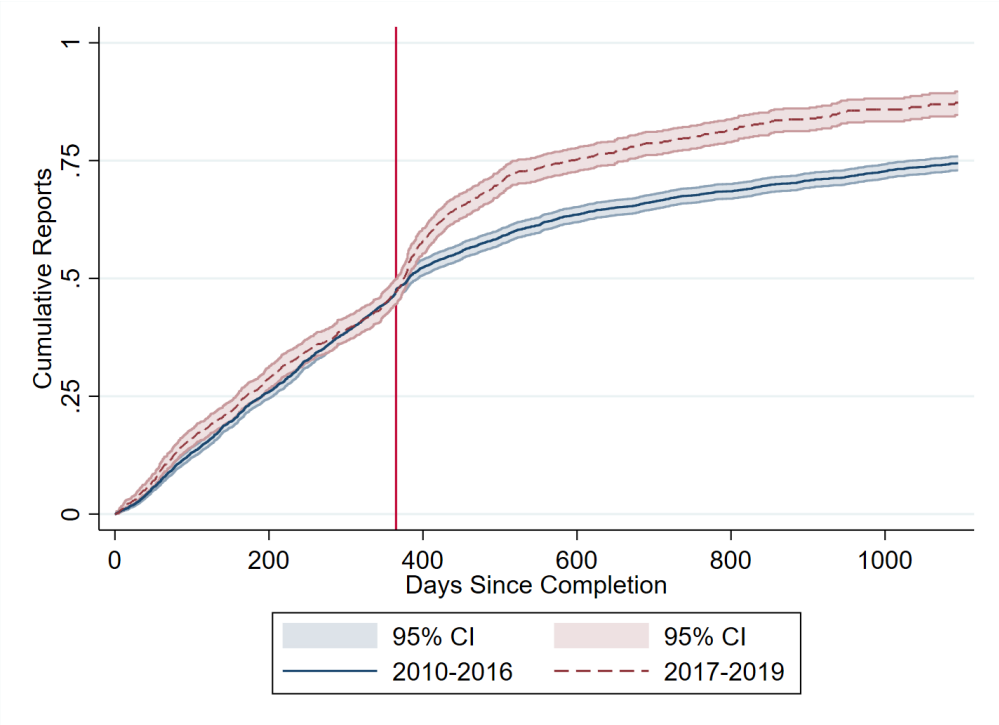
Notes: Local polynomial regression of active site count on trial progress. Sample includes all industry sponsored oncology trials that were active between 2010 and 2019. Trial progress defined in the text.

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 78% for trials that were completed in 2017-2019, up from 68% for 2010-2016. In Appendix Figure A2 we repeat this exercise for non-industry trials. Among these trials, the two-year disclosure probability increases from around 20% to around 60%.

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 adds support to the hypothesis that the policy change and associated penalties for late reporting of trial results 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.

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.

Among successful trials, the share with results reported two years after completion is 85% for trials started between 2010 and 2016, and 92% for trials started after 2017. For unsuccessful trials, the corresponding shares are 60% for 2010-2016 and 75% for 2017-2019.

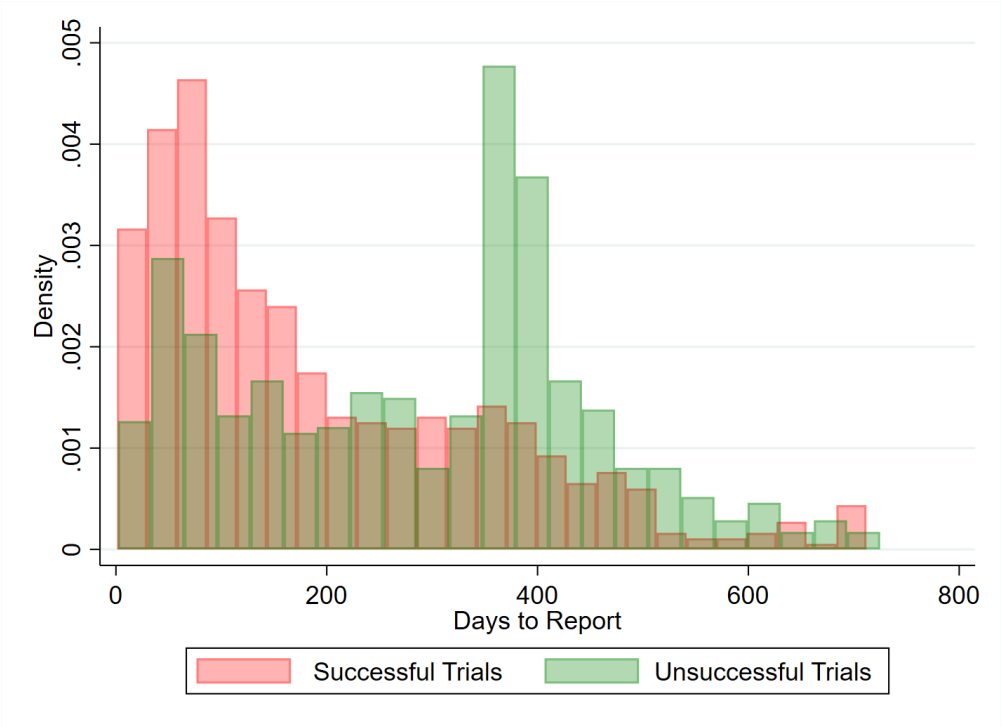
Together with Figure 2, these results suggest that the enforcement of the one-year reporting deadline brought results reporting forward in time and that this aggregate increase in reporting was primarily due to increased reporting of trials with negative outcomes.

4.2 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 now 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

Figure 3: Results Disclosure by Trial Outcome 2017-2019

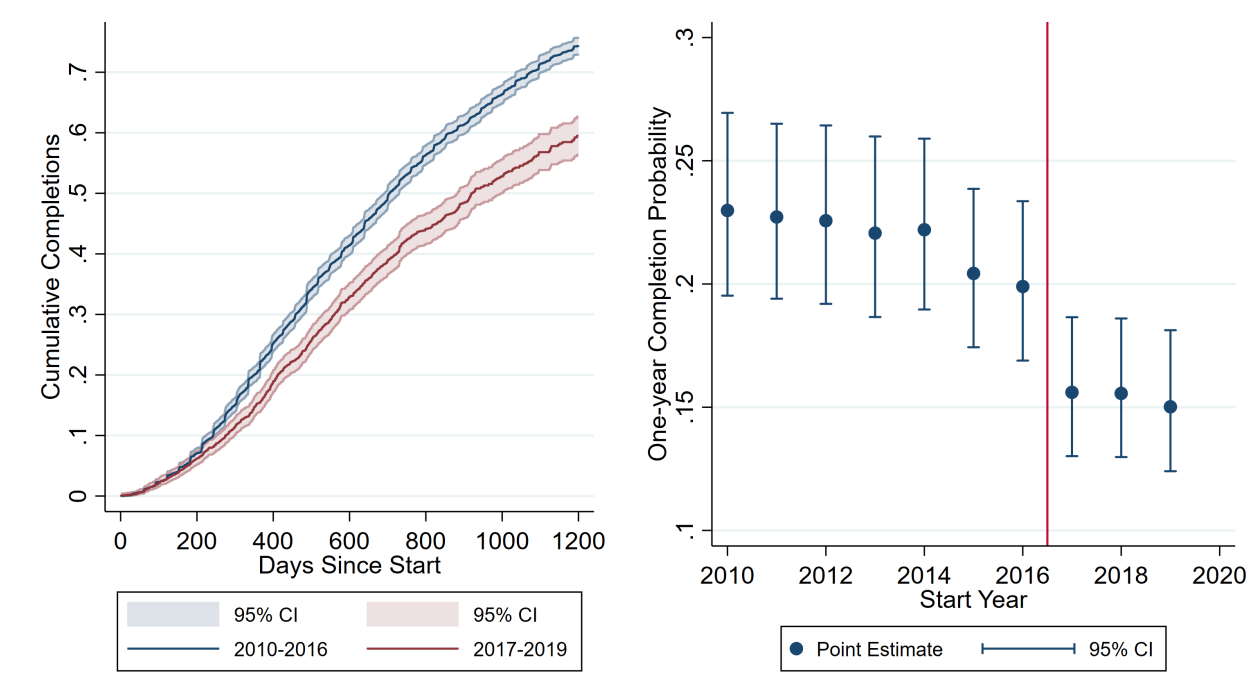


Notes: Histograms of days to report for Phase 2 industry trials completed after 2017 with reported outcomes. Histogram plotted separately for successful and unsuccessful trials. Trials with time to report greater than two years are dropped.

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 900 days, more than 200 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 two-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 with two years of their start, the same completion probability dropped to only 15% for trials begun in 2017 and remained stable for trials started in later years. Although the decrease in the two-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 A2 demonstrates a clear increase in timely result disclosure for this group following the

Figure 4: Time to Complete Phase 2



(i) Time to Complete by Policy Period

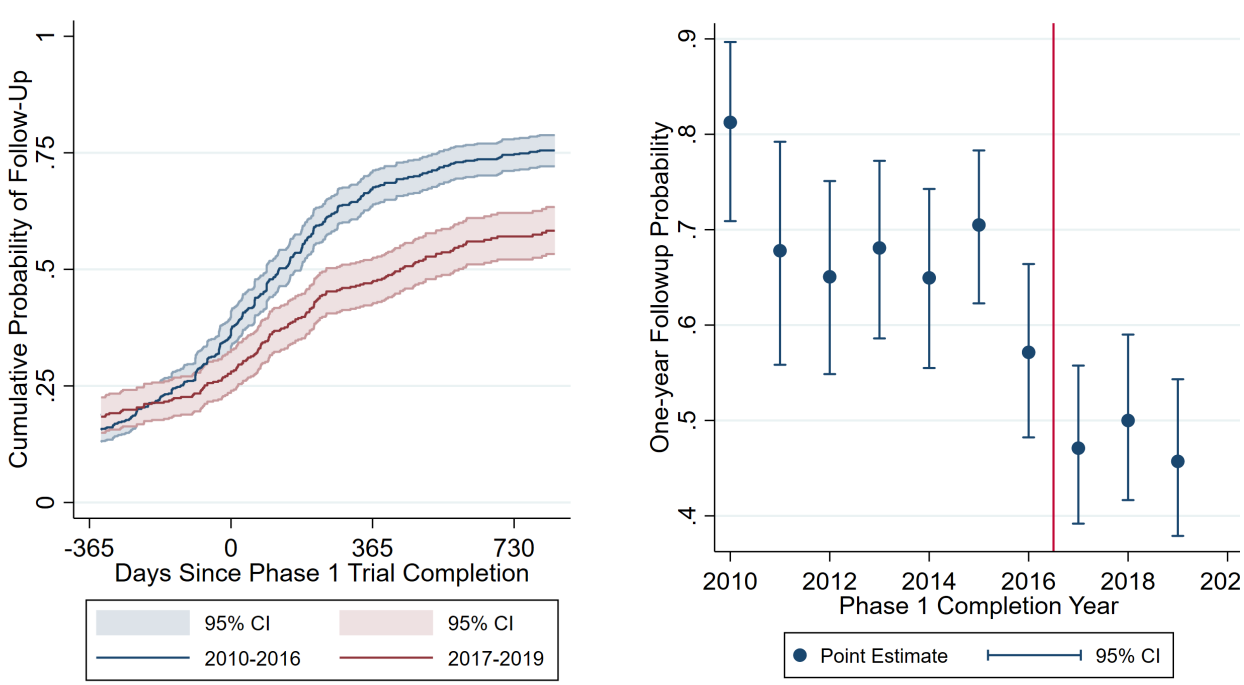
(ii) One-year Completion Probability by Start Year

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.

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 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 (either Phase 1 or Phase 2) 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

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 (Phase 1 or Phase 2) for the same drug after the completion of a Phase 1 trial that started before drug approval. The two time periods refer to the start 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 start year.

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]

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

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

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 theoretical 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 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.

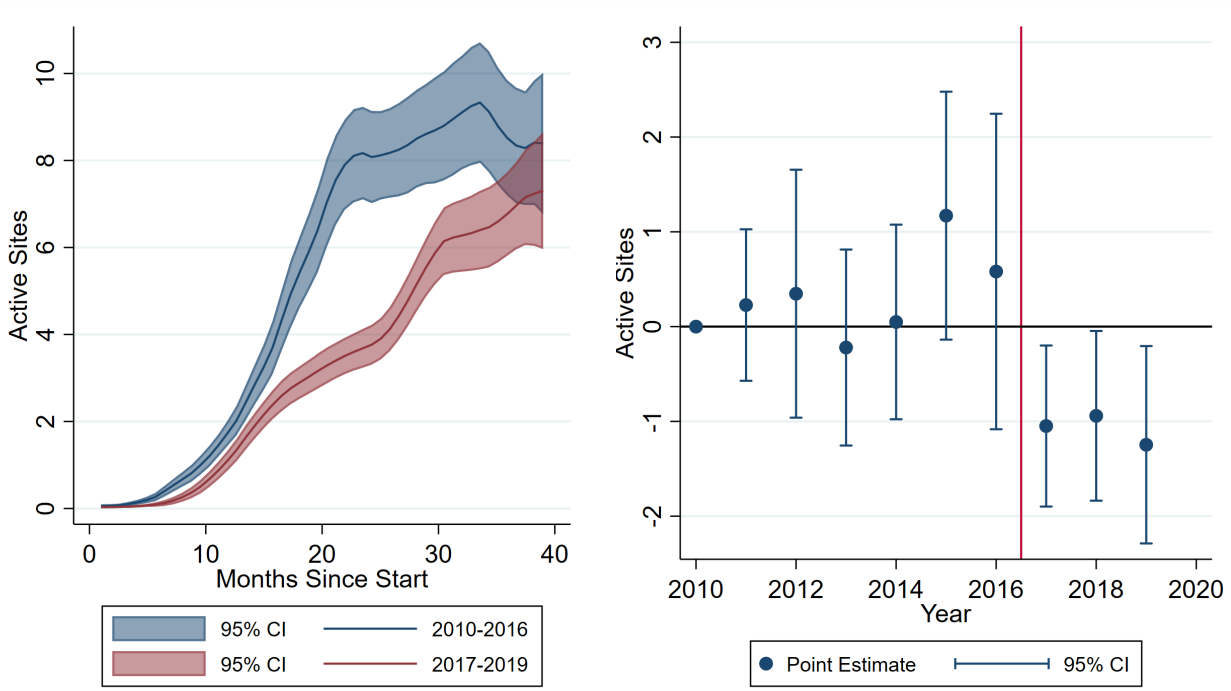
Together, the patterns in Figure 4 and Figure 5 indicate that after the implementation of the final rule, Phase 2 trials took substantially longer to complete and the time between Phase 1 trial completions and follow-up trials increased suggesting that firms were less willing to push the pace of drug discovery and approval. Cox proportional hazard models (reported in Appendix Table A2) indicate that the difference between time periods is statistically significant and not driven by changes in the composition of trials by disease or mechanism of action.

4.3 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 oncology 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, as documented in Section X above. 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

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 for chemical oncology drugs, separately for trials started 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. Estimates are conditional on months since start and reflect differences in active site counts across trial start years.

number of months since the start of a Phase 2 trial for the two policy periods using data on chemical oncology drugs, the sample for which we have information on the number of active sites. The count of active sites is consistently higher for trials started in the pre-policy period than for trials started 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 started in the pre-policy period had around 8 active sites on average whereas those started in the post-policy period had less than 4 active sites, a reduction in site count of over 50%. This evidence is consistent with the hypothesis that the slowdown in trial completions coincided with a slowdown in the pace of investment for active trials.

The right panel of Figure 6 provides additional evidence in support of this hypothesis. Specifically, it reports the coefficients of year-fixed effects from a regression of active sites on year trial start 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

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

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.

5 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 6), 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.

5.1 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. T_j is the set of drugs in the same therapeutic class 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 $k \in T_j$ are therefore potentially informative about future competition.

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

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

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

[9]Each drug may have multiple mechanisms of action and target diseases. In this exercise, we define M_j as the set of all drugs that share at least one mechanism with drug j, and we define T_j as the set of all drugs that shares at least one therapeutic class with drug j. Both M_j and T_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	0.0104	0.0044	0.0029**
		(0.0080)	(0.0042)	(0.0013)
	Unsuccessful Phase 2	-0.0167***	-0.0075**	-0.0053***
		(0.0052)	(0.0030)	(0.0012)
Same Therapeutic Class	Successful Phase 2	-0.0108*	-0.0064	0.0019
		(0.0064)	(0.0042)	(0.0026)
	Unsuccessful Phase 2	0.0186**	0.0084*	0.0021
		(0.0072)	(0.0045)	(0.0025)
Same Mechanism of Action	Successful Phase 2	0.0165***	0.0081***	0.0030*
		(0.0055)	(0.0026)	(0.0016)
	Unsuccessful Phase 2	-0.0117***	-0.0064***	-0.0022*
		(0.0041)	(0.0023)	(0.0013)
Time FE		YES	YES	YES
Drug FE		YES	YES	YES
N		1,617,366	1,617,366	1,617,366
R^2		0.1892	0.0999	0.0719
Mean of Dep. Var.		0.0098	0.0128	0.0061

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%.

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 1.04 percentage points. Conversely, an additional unsuccessful Phase 2 trial reduces the probability of a subsequent Phase 2 trial by 1.67 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 0.014.

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 1.65 percentage points. There is a similar though slightly smaller negative effect of unsuccessful trials which reduces the probability of starting a Phase 2 trial by 1.17 percentage points. Similar patterns hold for the effect on starting Phase 1 and Phase 3 trials, but with smaller magnitudes especially for Phase 3 starts. In contrast, the effects of trial outcomes from drugs $k \in T_j$ (i.e., drugs in the same therapeutic class) 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 therapeutic class reduces the probability of a Phase 2 start by 1.08 percentage points while the same increase in unsuccessful Phase 2 trials increases it by 1.86 percentage points. The same pattern also holds for Phase 1 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 are in the same therapeutic class and thus treat the same disease. The marked difference in the effect of same-mechanism and same-therapeutic class drugs is likely the result of a competitive preemption effect for drugs in the same therapeutic class. A successful trial of a competing drug k in the same therapeutic class T_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 A4 and A5 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 is consistent with the hypothesis 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.

5.2 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 6), 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 $Unrep_{tj}$ denote the logarithm of the count of drugs 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 measures the (natural logarithm of the) number of 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 when $Unrep_{tj}$ is higher. 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. A one-unit increase in $Unrep_{tj}$ therefore should

[10]For this exercise, we divide all drugs in the sample into disjoint mechanism groups. That is, there is no overlap between M_j and M_k for $j \neq k$. The method we use to construct these group is described in Appendix A.

Table 4: Responses to Unreported Trials by Policy Period

	Trial Completion		Active or Recruiting Sites		Active Sites	
Start Year $\geq$ 2017	-0.004***	0.003	-1.375***	0.543	-1.544***	-0.956***
	(0.002)	(0.002)	(0.473)	(0.890)	(0.210)	(0.244)
Log Unreported	-0.002**	0.001	-0.890**	-0.088	-0.289*	-0.043
	(0.001)	(0.001)	(0.354)	(0.350)	(0.156)	(0.205)
Log Unreported $\times$ Start $\geq$ 2017		-0.006***		-1.638***		-0.502**
		(0.002)		(0.494)		(0.208)
Time FE	YES	YES	YES	YES	YES	YES
MOA Group FE	YES	YES	YES	YES	YES	YES
N	64,529	64,529	64,529	64,529	64,529	64,529
Mean of Dep. Variable	0.018	0.018	13.408	13.408	3.813	3.813

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 oncology trials started after 2010. Stars indicate statistical significance: * 90%, ** 95%, *** 99%.

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 $Completion_{jt}$ indicates the completion of trial j and $Active_{jt}$ 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 first column shows that trials starting in 2017 or later ($Post_{tj}$ in the regression equation) had a lower probability of completion relative to earlier trials, with the effect more pronounced when more related trials remain unreported. The second column adds the interaction between $Unrep_{tj}$ and the post-policy indicator $Post_{tj}$ and finds a negative and statistically significant coefficient, suggesting that firms are especially likely to delay trial completion when disclosure of related trials is imminent.

Columns (3) and (4) examine the number of active or recruiting sites and yield similar patterns: site counts fall when more related trials remain unreported, particularly after the policy change. Finally, columns (5) and (6) use the number of active sites as the dependent variable. While the overall effect is somewhat smaller than in previous versions of the table, the interaction between

$Unrep_{tj}$ and $Post_{tj}$ remains negative and statistically significant. This pattern reinforces the interpretation that firms slow investment in anticipation of forthcoming information. Notably, once the interaction is included, the coefficient on $Post_{tj}$ alone shrinks markedly, suggesting that a large proportion of the post-policy investment slowdown is attributable to increased sensitivity to the disclosure environment rather than other secular changes.

Table A6 in the appendix presents additional specifications that include a richer set of controls, specifically the number of completed and started trials. Across all columns, the interaction between $\log(\text{Unreported}_{tj})$ and the post-2017 indicator remains negative and statistically significant for both trial completions and active sites. This robustness check confirms that our core finding that firms reduce investment when they anticipate a greater flow of information from other trials holds even after accounting for potentially confounding factors. The stability of this interaction term across more saturated specifications reinforces our interpretation that the observed slowdown in investment is driven by strategic responses to the enhanced disclosure environment rather than by compositional shifts.

The findings in Section 5.2 demonstrate that firms strategically adjust their investment in clinical trials based on the expected future availability of information from other trials. Active and recruiting 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.

6 Dynamic Model of Trial Investment

6.1 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 therapeutical treatment 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 μ_M is the component common to all drugs in M and λ_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 μ_M. We assume that both the individual drug-specific component and the group-level productivity are normally distributed with the following distributions.

$$\mu_G \sim \mathcal{N}(0, \sigma_M) \qquad \text{group-level productivity}$$
$$\lambda_j^o \sim \mathcal{N}(, \sigma_\lambda) \qquad \text{drug-specific productivity}$$

Thus, observing λ_k where $k \neq j$ is informative about group-level productivity μ_M which in turn is informative about the drug effectiveness λ_j. Firms have the same identities as drugs, denoted by j, and invest in trial sites to obtain signals about λ_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 μ_M. $\tilde{\lambda}_j^o$ and $\tilde{\sigma}_j$ denote the beliefs about the mean and variance of drug-specific effectiveness λ_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.

Drug trials are started, completed, and eventually reported at different rates. We define these rates as follows. First, trials of other firms are started (i.e., they enter O) with probability α_S which denotes the trial start rate. Second, α_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 α_D. Thus, α_D is the disclosure rate of completed trials which governs at what pace completed trials of other firms are publicly disclosed such that their individual productivity λ_k can be observed. Recall that the disclosed trial result is the combination of group- and individual effectiveness components, $\lambda_k = \mu_M + \lambda_k^o$. To summarize, these transition rates are

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

Given the firm's own signal s_{jt} from its drug trials and the observation of λ_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:

$$V(S) = \max\left\{\beta E\left[V(S')\right] - cA, \max_{A^* \in [0,\bar{A}]}\left\{\beta E\left[V(S')|A^*\right] - cA - \delta\right\}, \pi(S)\right\} \tag{4}$$

where β 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 δ 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 (4) 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 β. 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 δ. This choice also allows the firm to learn from the outcomes of other trials.[11] 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.

6.2 Solving the Model

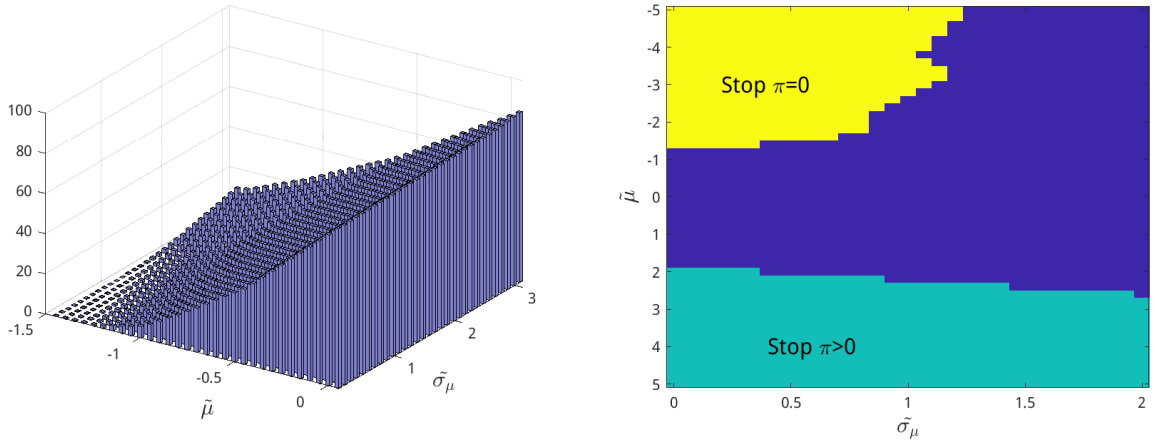
We solve the model using value function iteration, where the value function is approximated by a neural network on each iteration of the Bellman equation.[12] This approach allows us to handle the high-dimensional state space while maintaining sufficient flexibility to capture complex patterns in the decision-making environment.

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

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

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

Figure 7: Value Function and Optimal Stopping Rule



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

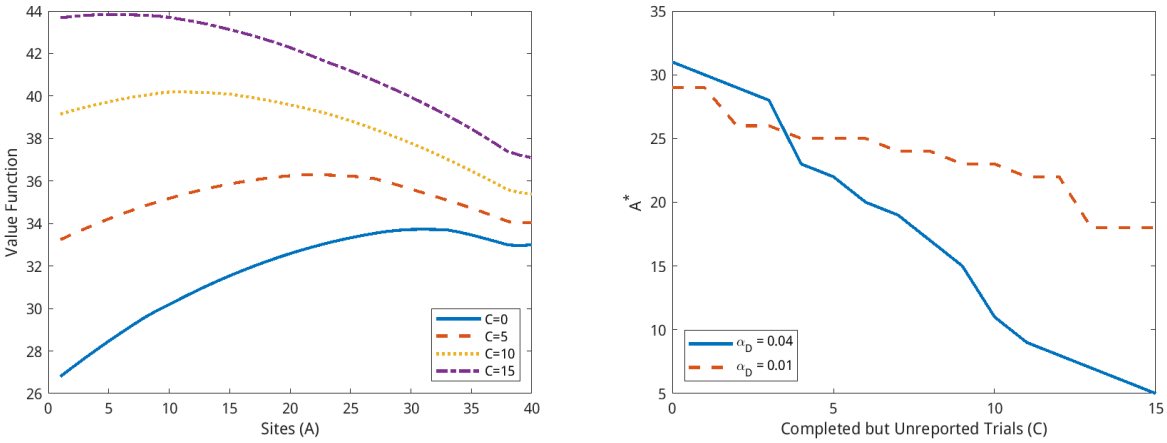
Notes: Panel (i) presents a 3D bar plot of the value function over the space of belief means ($\tilde{\mu}_M$) and standard deviations ($\tilde{\sigma}_M$), representing the continuation value of a trial given current beliefs. The value function increases with both $\tilde{\mu}_M$ and $\tilde{\sigma}_M$, 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.

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 α_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.

We first plot the value function and the optimal stopping rule in $(\tilde{\mu}_M; \tilde{\sigma}_M)$ space which are shown in Figure 7. It consists of two complementary panels. In the left panel, a 3D bar plot shows

Figure 8: Value Function and the Role of Disclosure



(i) Value Function over A and C **(ii)** Optimal A^* for different α_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 α_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 α_D in panel (ii).

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 8 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 α_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 sits. Panel (ii) formalizes this relationship by plotting the optimal number of active sites A^* as a function of C, for two different values of α_D. In both cases, A^* declines monotonically with C, and the drop is markedly steeper when the disclosure probability α_D is higher.

The relationships depicted in Figure 8 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.

7 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.

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 Groups

This section details the definition of Mechanism-of-Action (MOA) groups and the steps to construct them from clinical trial data.

MOA groups are non-overlapping technology groups composed of phase 2 clinical trials of all diseases. Each MOA group is characterized by a unique combination of up to two MOAs shared by all trials in the group. For a trial with a single MOA, it is uniquely categorized into the MOA group characterized by its own MOA. For a trial with two or more MOAs that can be categorized into multiple MOA groups, we give priority to the two-MOA group with the minimum variance in its trials' outcomes.

The steps of constructing MOA groups are as follows:

1. Identify and rank potential MOA groups

   We begin with the sample of reported U.S. industry-sponsored phase 2 trials of all diseases. From reported trials with two and more MOAs, we enumerate all existing two-MOA combinations and obtain a comprehensive set of potential two-MOA groups. Each potential group contains all reported trials that use the two-MOA combination exclusively or with other MOAs. We calculate the group-level sample variances of trial outcomes (positive or negative) and rank the groups by variance, from minimum to maximum. For groups with only one reported trial and hence no well-defined sample variance, we rank them the lowest due to lack of learning opportunity from other trials in the same group.

2. Categorize reported and unreported trials

   We categorize trials with one single MOA into their corresponding one-MOA groups. For trials involving two or more MOAs, we enumerate all possible two-MOA combinations from the trial's MOA profile and retain only those combinations existing in reported trials. The trial is then categorized into the MOA group with the smallest variance in its trial outcomes among these potential groups.

3. Create new groups for trials unmatched to existing groups

   For remaining trials not categorized to any existing MOA groups in step 2, we create new MOA groups using each trial's original MOA profile. Categorization is finished.

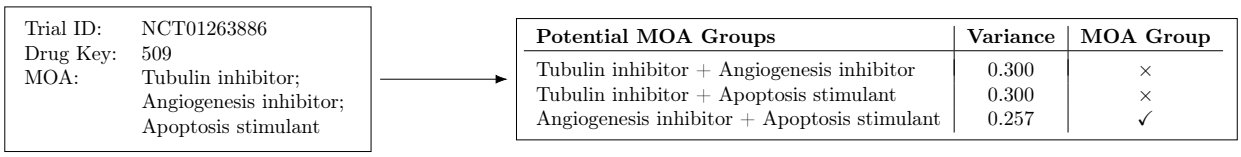


Figure A1: Example of MOA Group Selection Process

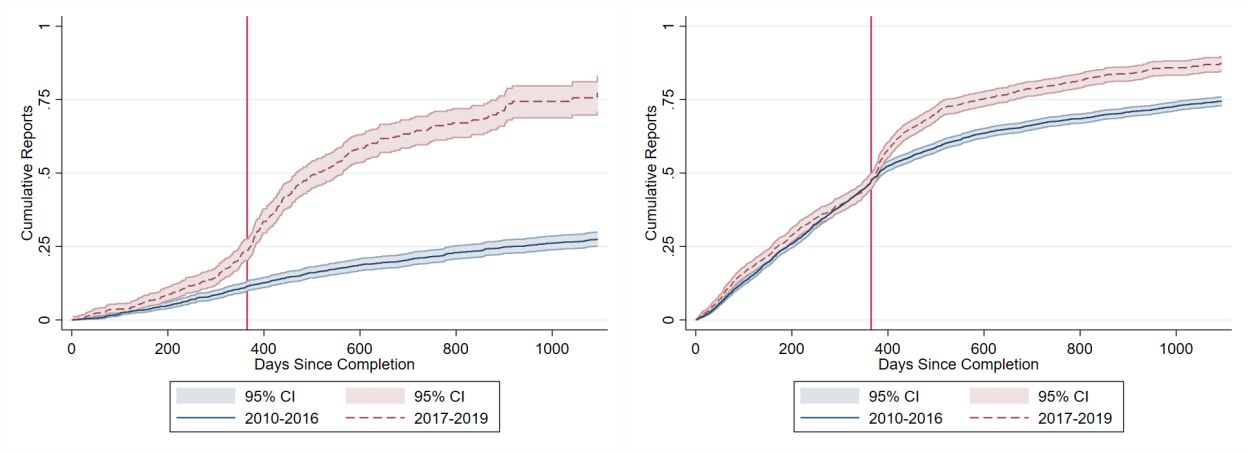
B Additional Figures and Tables

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.

Figure A2: Cumulative Results Disclosure: Non-Industry and All Trials



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

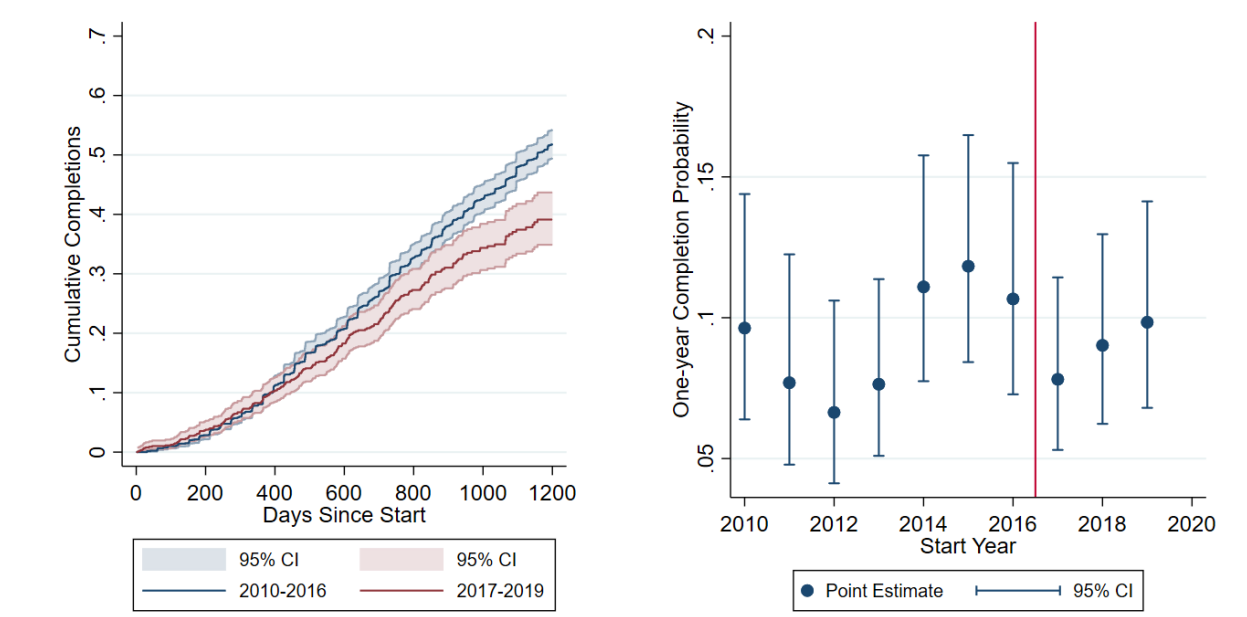
Notes: Kaplan-Meier failure curves recording the cumulative probability of results reporting for Phase 2 trials with at least one site in the United States. The two time periods refer to the year of trial completion which is relevant for reporting.

Table A2: Cox Model Estimates

Start Year	Compete	Complete	Start	Start
2011	1.023	1.043		
	(0.063)	(0.063)		
2012	0.967	0.935	1.003	1.025
	(0.064)	(0.064)	(0.180)	(0.181)
2013	0.952	0.997	0.948	0.975
	(0.065)	(0.065)	(0.178)	(0.180)
2014	0.937	0.912	0.842	0.851
	(0.064)	(0.064)	(0.177)	(0.179)
2015	0.903	0.903	0.827	0.831
	(0.063)	(0.063)	(0.172)	(0.173)
2016	0.714***	0.761***	0.679*	0.668*
	(0.066)	(0.066)	(0.176)	(0.178)
2017	0.524***	0.613***	0.483***	0.487***
	(0.069)	(0.069)	(0.174)	(0.176)
2018	0.599***	0.640***	0.466***	0.497***
	(0.074)	(0.074)	(0.180)	(0.183)
2019	0.524***	0.564***	0.430***	0.429***
	(0.092)	(0.093)	(0.181)	(0.183)
Disease FE	NO	YES	NO	YES

Notes: Estimates from a cox proportional hazards model. The outcome in the first two columns in phase 2 trial completion, and year year refers to the start year of the phase two trial. The outcome of the second two columns is phase 2 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 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.

Table A3: Response of Trial Start to Results Disclosure: Robustness

		Phase II Start			
Same Drug	Successful Phase 2	0.0104	0.0068	0.0104	0.0069
		(0.0080)	(0.0055)	(0.0080)	(0.0055)
	Unsuccessful Phase 2	-0.0167***	-0.0058	-0.0167***	-0.0058
		(0.0052)	(0.0051)	(0.0052)	(0.0051)
Same Therapeutic Class	Successful Phase 2	-0.0106*	0.0072	-0.0109*	0.0071
		(0.0064)	(0.0065)	(0.0064)	(0.0065)
	Unsuccessful Phase 2	0.0182**	0.0017	0.0193***	0.0022
		(0.0072)	(0.0069)	(0.0075)	(0.0070)
Same Mechanism of Action	Successful Phase 2	-0.0043	0.0263***	-0.0155	0.0181***
		(0.0071)	(0.0092)	(0.0122)	(0.0066)
	Unsuccessful Phase 2	-0.0182***	-0.0078	0.0012	-0.0134**
		(0.0047)	(0.0055)	(0.0043)	(0.0053)
	Completed Suc. Phase 2	0.0140	0.0022		
		(0.0106)	(0.0071)		
	Completed Unsuc. Phase 2	0.0174***	-0.0139		
		(0.0041)	(0.0094)		
	Started Phase 3			0.0256**	0.0055
				(0.0126)	(0.0065)
Time FE		YES	YES	YES	YES
Drug FE		YES	NO	YES	NO
Drug-Year FE		NO	YES	NO	YES
N		1,617,366	1,617,366	1,617,366	1,617,366
R^2		0.1893	0.3397	0.1895	0.3397

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 A4: Response of Trial Start to Results Disclosure: Phase I

		Phase 2 Start	Phase 1 Start	Phase 3 Start
Same Drug	Successful Phase 1	0.0032	-0.0001	0.0007
		(0.0040)	(0.0022)	(0.0009)
	Unsuccessful Phase 1	-0.0047***	-0.0125***	-0.0035***
		(0.0014)	(0.0009)	(0.0007)
Same Therapeutic Class	Successful Phase 1	0.0019	0.0012	0.0020
		(0.0029)	(0.0021)	(0.0014)
	Unsuccessful Phase 1	0.0011	0.0008	0.0007
		(0.0029)	(0.0021)	(0.0014)
Same Mechanism of Action	Successful Phase 1	0.0009	-0.0002	0.0027***
		(0.0010)	(0.0008)	(0.0009)
	Unsuccessful Phase 1	0.0016	-0.0000	-0.0026***
		(0.0014)	(0.0009)	(0.0010)
Time FE		YES	YES	YES
Drug FE		YES	YES	YES
N		1,617,366	1,617,366	1,617,366
R^2		0.0098	0.0128	0.0061
Mean of Dep. Var.		0.0098	0.0128	0.0061

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 A5: Response of Trial Start to Results Disclosure: Phase III

		Phase 2 Start	Phase 1 Start	Phase 3 Start
Same Drug	Successful Phase 2	0.0017	-0.0033***	-0.0070***
		(0.0015)	(0.0010)	(0.0006)
	Unsuccessful Phase 2	-0.0039**	-0.0008	-0.0046***
		(0.0018)	(0.0009)	(0.0006)
Same Therapeutic Class	Successful Phase 2	-0.0011	-0.0020	0.0020**
		(0.0015)	(0.0012)	(0.0008)
	Unsuccessful Phase 2	0.0045**	0.0019	0.0010
		(0.0018)	(0.0015)	(0.0007)
Same Mechanism of Action	Successful Phase 2	-0.0220***	-0.0056	-0.0055**
		(0.0077)	(0.0040)	(0.0027)
	Unsuccessful Phase 2	0.0212***	0.0060	0.0042*
		(0.0075)	(0.0038)	(0.0025)
Time FE		YES	YES	YES
Drug FE		YES	YES	YES
N		1,617,366	1,617,366	1,617,366
R^2		0.0098	0.0128	0.0061
Mean of Dep. Var.		0.0098	0.0128	0.0061

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 Investment to MOA State by Policy Period

	Trial Completion		Active or Recruiting		Active Sites	
Start Year $\geq$ 2017	0.006*	0.001	0.817	0.399	-1.118***	-1.208***
	(0.003)	(0.003)	(0.840)	(0.957)	(0.364)	(0.276)
Log Unreported	0.003	0.001	0.361	0.132	0.225	0.072
	(0.003)	(0.002)	(0.712)	(0.370)	(0.332)	(0.219)
Log Unreported $\times$ Start $\geq$ 2017	-0.005	-0.008***	-1.725*	-1.879***	-0.887**	-0.791***
	(0.003)	(0.002)	(1.016)	(0.558)	(0.429)	(0.266)
Log Completed	-0.003		-0.511		-0.250	
	(0.002)		(0.608)		(0.268)	
Log Completed $\times$ Start $\geq$ 2017	-0.001		0.039		0.355	
	(0.003)		(0.691)		(0.344)	
Log Started	0.000			-1.593***		-0.661***
	(0.002)			(0.333)		(0.191)
Log Started $\times$ Start $\geq$ 2017	0.002			0.421		0.382**
	(0.002)			(0.314)		(0.161)
Time FE	YES	YES	YES	YES	YES	YES
MOA Group FE	YES	YES	YES	YES	YES	YES
N	64,529	64,529	64,529	64,529	64,529	64,529

Notes: The table reports coefficient estimates of the regressions described in equations (2) and (2) with additional controls for the number of started and completed trials in M_j. The sample includes all trial-months before 2020 for all phase 2 oncology trials started after 2010. Stars indicate statistical significance: * 90%, ** 95%, *** 99%.