Learning from Failure: The Role of Disclosure on Innovation[☆]

Menghan Wang[a]

[a] *University of Hong Kong, Faculty of Business and Economics*

Abstract

This paper examines how the disclosure of failures shape innovation activities. I exploit an expansion in mandatory disclosure requirements for clinical trial results—from only trials on approved products to all applicable trials, including those on unapproved ones—as a positive shock to the availability of failure information. There is a significant increase in the number of new trials initiated, which becomes evident one year after the policy change and persists in the following years. This increase is driven by trials on existing drugs rather than novel drugs, suggesting a "fast-follow" strategy. Textual analysis of trial summaries further supports the learning effect and the importance of disclosing trial results. Consistent with the knowledge spillover channel, sponsors benefits more in medical areas where they had less internal expertise prior to the policy change. However, mandatory disclosure of failures carries proprietary costs, especially for innovative ones whose private knowledge becomes accessible to competitors. Sponsors with a higher risk of losing informational advantages are less inclined to initiate new trials under such disclosure requirements. This paper contributes to the discussion on mandatory disclosure of innovation outcomes, under the trade-off between the social benefits from knowledge spillover and the proprietary costs borne by the innovating entities.

Keywords: disclosure, innovation, knowledge spillover, proprietary cost, clinical trials
JEL: D22, D83, L65, O31, O32

Email address: menghanw@connect.hku.hk (Menghan Wang)

July 31, 2024

1. Introduction

In academic research and the realm of intellectual property in general, the phenomenon of publication bias, or file drawer problem, presents a significant challenge. Empirical results that reject null hypothesis are disproportionately more likely to get published, and there exists an under-representation of marginally insignificant statistics relatively to significant ones (Sterling, 1959; Brodeur et al., 2016; Andrews and Kasy, 2019). Similar bias extends to the patent system, where only those successful inventions get filed and those failed ones left invisible.

From individual perspective, this selective disclosure of success is not surprising due to high proprietary cost associated with disclosure (Bhattacharya and Ritter, 1983; Verrecchia, 1983; Anton and Yao, 1994; Glaeser, 2018; Griffin, Hong and Ryou, 2022). Once the finding of an attempt is made public, competitors can learn from the information disclosed and potentially benefit from it without paying any research costs themselves, leading to a free-riding problem. Even though this attempt fails, competitors may succeed with modifications to the original approach. Moreover, the market may perceive the disclosure of failure as a negative signal regarding the firm's or inventor's R&D capability, which could harm their reputation and financing ability. Hence, there is little incentives for individuals to voluntarily disclose their failures.

However, from social perspective, the importance of cumulative innovation in scientific and technical advancements has been widely recognized, as suggested by the famous quote from Isaac Newton about standing on the shoulders of giants. Those underreported null results and failed experiments could hold substantial value (Abadie, 2020). Each inventor hiding their failures may be socially suboptimal, and such information asymmetry could lead to duplicated research and development investments (Loury, 1979; Reinganum, 1985; Scotchmer, 1991). The loss would be even greater when ideas are harder to find, and more investments in R&D are needed to keep research productivity at prior level (Bloom et al., 2020). A more complete understanding of what has been done and found could save duplicated and redundant efforts, and potentially provide insights for future research and innovation.

The main research question this paper seeks to answer is whether knowledge of past failure could stimulate future innovation. Two opposing forces may be at play. The positive force comes from the spillover effect of knowledge sharing, which could save efforts on duplicative research, stimulate new idea generation, and reduce uncertainty inherent in R&D investments (Scotchmer, 1991; Furman and Stern, 2011; Badertscher, Shroff and White, 2013; Furman, Nagler and Watzinger, 2021; Kim and Valentine, 2021; Hegde, Herkenhoff and Zhu, 2023; Dyer et al., 2023). The downside could be the deterrence effect due to fear of failure, or afraid of losing R&D race after observing others progress (Glaeser and Landsman, 2021; Krieger, 2021; Zhang, 2024). And the proprietary cost associated with the disclosure, such as the risk of competitors appropriating or expropriating their findings, may deter firms from starting new trials to avoid mandatory disclosure (Aoki and Spiegel, 2009; Kim and Valentine, 2021).

Existing empirical evidence is scare due to the lack of counterfactuals. In reality, while it is relatively easy to learn others' success, failures often remain hidden due to fear of copycats and reputation loss. Even for those observable ones, there could be endogeneity concerns that both the decision of voluntary disclosure and future innovation activities may be influenced by other factors, such as investment opportunities.

In this paper, I use an expansion in disclosure requirements for clinical trial results as a positive shock to the availability of failure information. Prior to the policy change, Section 801 of the Food and Drug Administration Amendments Act of 2007 (FDAAA 801) mandated applicable clinical trials of drug, biological, or device products that have been *approved, licensed, or cleared* by the FDA to submit their summary results. Effective as of January 18, 2017, the Final Rule of FDAAA 801 expanded the result submission requirements to *all* applicable trials, including those of unapproved products. The result disclosure encompasses detailed narratives and statistics on participant flow, demographic and baseline characteristics, primary and secondary outcomes, and adverse events, beyond mere status of success or failure. This change in the mandatory disclosure rule for trials on unapproved products presents a quasi-natural experiment, allowing for assessing

the impact of failure disclosure. The treated group consists of medical conditions with lower disclosure rate of applicable trials prior to the rule change, as they are more exposed to the positive shock. I find an 11.9% increase in the number of new trials initiated in the treated group compared to the control group after the policy change. The parallel trend assumption holds, and the positive effect became significant one year after the event. The effect is not attributable to the abnormal surge in trial initiations due to COVID-19. The results remain robust after excluding trials, medical conditions, or sample period related to COVID-19.

To investigate whether the increase in new trials comes from explorative or exploitative innovation, I distinguish interventional clinical trials that involves at least one new drug that has neither been used in previous trials nor approved by FDA, and those involve existing drugs only. The positive effect dominates in trials on existing drugs. While there is no significant change in the number of trials on new drugs after the policy change, and the coefficient estimate is negative. The findings suggest that learning from others failure primarily stimulates incremental improvements instead of novel drug compounds.

Exploring the heterogeneity effect across different phases, I find the positive effect shows up quicker for early-phase trials, while it takes longer time to became evident for late-phase trials. It is consistent with the prediction as the initiations of late-phase trials are contingent on the success of early-phase ones. This timing patterns across phases further support the causality that the increase is driven by the disclosure policy change.

To provide further evidence that the new trials initiated after the rule change draw upon previous trials, I compare the textual similarity between the study summaries of the new trials and previous trials in the same medical condition conducted by other sponsors. A greater similarity indicates a higher level of learning from others. I further differentiate past trials based on their result availability. After the policy change, I observe a significant increase in the similarity to prior trials with disclosed results. In contrast, there is little change in the similarity to prior trials without results. This evidence supports that the availability of trial outcomes enhances the learning effect.

Subsequent trials could assimilate and incorporate those external knowledge into their research design.

Finally, I test two mechanisms that could potentially explain the effect of failure disclosure on future innovation. The first one is the knowledge spillover effect that researchers could learn from others failures and stimulate their own idea generation. If this channel holds, the positive effect on innovation is expected to be stronger in those medical fields where sponsors process less in-house knowledge prior to the policy change. While for areas where they already have substantial internal expertise or hold a dominant position, the potential for learning from others diminishes, as well as the associated spillover benefits. Using the sponsor-MeSH-year sample to capture the heterogeneity across sponsor-MeSH pairs, I find that the increase in trial initiation concentrates in medical conditions where sponsors initiated fewer trials prior to the event, supporting the knowledge spillover channel. Moreover, this evidence is inconsistent with the deterrence effect, which posits that the perceived risk of failure increases after observing others' failures. Sponsors are expected to avoid areas where they have limited in-house knowledge, if the perceived risk of failure is high.

The second mechanism considers the proprietary cost of disclosure. The main concern is the risk of competitors expropriating disclosed results, thereby placing the original sponsor at a competitive disadvantage. Such concern may deter the initiation of new trials in the first place, particularly those under the mandatory result disclosure requirement. To test this hypothesis, I split the sponsors into high and low group based on their level of proprietary cost. The proprietary cost is proxied by the number of trials subjected to the disclosure requirement and initiated prior to the issuance of the Final Rule. It measures sponsor's burden of disclosure under the new rule, and ensures that trial initiation decisions were made ex-ante, thus, not endogenously influenced by the shock. Consistent with the prediction, sponsors with higher proprietary cost initiate less trials subjected to the mandatory disclosure rule compared to those with lower proprietary cost after the shock. While there is no difference in the total number of new trials initiated between sponsors

with high and low proprietary cost.

This paper contributes to the literature in several ways. First, it complements the literature on the effects of disclosure regulations on innovation. Many empirical studies provide supporting evidence for the positive role of disclosure requirements through knowledge spillover and technology diffusion. Using the American Inventor's Protection Act of 1999 (AIPA) as a positive shock that accelerated patent disclosure, prior research documents increased patent citations and follow-on inventions, along with a reduction in duplicated research efforts and improved efficiency in ideas commercialization post-AIPA (Hegde and Luo, 2018; Lück et al., 2020; Baruffaldi and Simeth, 2020; Kim and Valentine, 2021; Hegde, Herkenhoff and Zhu, 2023). Additionally, Furman, Nagler and Watzinger (2021) find that local patenting activity increases after the opening of patent libraries. Dyer et al. (2023) show that high quality patent disclosures spur follow-on innovation, exploiting the exogenous variation of disclosure quality from the assignment to patent examiners. In contrast, mandated secrecy, such as the Invention Secrecy Act, the Uniform Trade Secrets Act (UTSA) and the Inevitable Disclosure Doctrine (IDD), could reduce and delay follow-on inventions (Contigiani, Hsu and Barankay, 2018; Rassenfosse, Pellegrino and Raiteri, 2024; Gross, 2019; Ganglmair and Reimers, 2022). However, some studies also suggest the adverse effect of mandated disclosure (Bessen, 2005; Aoki and Spiegel, 2009; Kim and Valentine, 2021). Aghion, Bergeaud and Van Reenen (2023) shows theoretically and empirically that regulation burden, including but not limited to disclosure requirements, discourage firms from innovation. While most of the prior studies focus on the disclosure of successful inventions, such as patent publications, this paper examines the disclosure of failure information and highlights its importance on future innovations.[1]

Second, this paper broadly relates to the literature on the interaction between disclosure decision and innovation strategies in the context of R&D competition. On the one hand, when deciding

[1] Krieger (2021) documents that failure news of competitors lead to a 23% jump in the exit rate of focal firm's projects in the same technology area. While his study examines the continuation decisions of ongoing parallel projects, my analysis focused on subsequent trial initiations.

whether to disclose their innovation outcomes, firms trade off the benefits of sending out positive signals on their innovation prospects to attract investors or deter competitors, against the potential risk of losing informational advantages to their rivals (Bhattacharya and Ritter, 1983; Anton and Yao, 1994; Jones, 2007; Hughes and Pae, 2015; Glaeser and Landsman, 2021; Capkun et al., 2023; Chu et al., 2024). On the other hand, when setting innovation strategies in light of their peers' disclosure, firms trade off the encouragement effect from the resolved uncertainty and knowledge spillover due to greater information available, against the deterrence effect from the fear of losing the race, especially when the competitors are strong rivals or already claim in-term success (Harris and Vickers, 1987; Lippman and McCardle, 1987; Choi, 1991; Doraszelski, 2003; Kim and Valentine, 2021; Hegde, Herkenhoff and Zhu, 2023; Zhang, 2024; Hsu et al., 2019). When studying the impact of disclosure on innovation, such interactions impose endogeneity concern that voluntary disclosure decisions can be influenced by innovation opportunities. To address the concern, this paper examines a policy change in mandatory disclosure requirements for clinical trial results as a quasi-natural experiment, in an effort to provide more causal evidence.

Lastly, this paper connects to the literature on drug development in the context of finance and accounting.[2] The drug development process is characterized by substantial investments, protracted timelines, and high uncertainty with a low probability of success (Lo and Thakor, 2022). Despite a significant demand for financing, Thakor and Lo (2017) argue theoretically that the low success rate and the specialized expertise required to evaluate project potential lead to insufficient investments in R&D. Empirical evidence support that firms underinvest in novel drugs, and short-term stock market downturns lead to discontinuation of drug developments (Krieger, Li and Papanikolaou, 2022; Mace, 2023). This paper reveals the potential positive impact that disclosure requirements could have on promoting clinical research.

The rest of the paper proceeds as follows. In Section 2, I explain the institutional background and research design. Section 3 describes the data. I examine the impact of failure disclosure on

[2] A comprehensive literature review on financing biomedical innovation is provided by Lo and Thakor (2022)

future trial initiation in Section 4, and provide supporting evidence for the learning effect through textual analysis in Section 5. Further, I investigate two underlying mechanisms, the knowledge spillover effect and the proprietary cost of disclosure, in Section 6.

2. Institutional Background and Empirical Design

The setting of clinical trials is one of the few areas that one can find failure disclosure. As the disclosure of trial failures are of great importance, beyond the benefits of knowledge spillover and reduction of repeated efforts. First, safety is always the top priority in clinical trials that involve human subjects. Any results or indication of adverse events should be made public to enable more informed decision-making by doctors and patients, and to prevent similar incidents in the future. Second, participants of clinical trials have the right to know their condition. Transparency in trial design and results helps to build trust between participants and researchers.

As one of the attempts towards greater transparency, the Food and Drug Administration Modernization Act (FDAMA) of 1997 required National Institutes of Health (NIH) to establish a publicly accessible website for clinical studies. This act resulted in the launch of ClinicalTrials.gov, which was first made accessible to public in Feburary, 2000. It allows sponsors and investigators to submit and update information about clinical research studies and their results, aiming to facilitate information sharing among researchers, healthcare professionals, patients and the general public. Today it is the world's largest database of clinical trials, including studies from all over the globe and studies of all different types of interventions.

In September 2007, congress passed FDAAA (Food and Drug Administration Amendments Act of 2007) which expanded the submission requirements to ClinicalTrials.gov, and regulated the trial result submission for the first time. Section 801 of FDAAA (FDAAA 801) required the submission of summary results information for applicable clinical trials (ACTs) of drugs, biological products, or devices that have been *approved, licensed, or cleared* by the FDA. The submission should be made no later than 12 months after primary completion date, or within 30 days of ap-

proval, licensure or clearance of the drug or device.[3]

In November 2014, the U.S. Department of Health and Human Services issued a Notice of Proposed Rulemaking (NPRM) for public comment. One major change of the proposed rule is to expand the scope of results submission requirements. The Final Rule of FDAAA 801 was issued in September 2016, and came into effective on January 18, 2017. It mandated that *all* ACTs should submit their summary results within 12 months after its primary completion date, despite their approval status.[4] The result disclosure requirements cover a comprehensive set of data points, including participant flow, demographic and baseline characteristics, primary and secondary outcomes, and adverse event information. Both narrative and numeric information are provided. Delayed submissions are allowed in certain circumstances for up to two additional years.[5]

Information submitted to ClinicalTrials.gov undergoes a quality control review to ensure there is no apparent errors, deficiencies, or inconsistencies.[6] The Final Rule also outlines the potential civil or criminal actions, including civil monetary penalty actions, and grant funding actions that may be taken if responsible parties fail to comply with the requirements.[7] However, the actual enforcement has been limited. According to the estimation of the Bennett Institute at the University of Oxford, the US government could have imposed fines of over 60 trillion USD on non-compliance, yet no fines have been claimed to date.[8]

[3]Primary completion date is the date of final data collection for the primary outcome measure.

[4]For each trial initiated before January 18, 2017, it is considered as an ACT if it satisfies the following conditions as defined in section 402(j) of the Public Health Service (PHS) Act: (1) it is a non-phase 1 interventional trial of drugs, medical devices, or biologics; (2) it has at least one U.S. research site, or are conducted under an investigational new drug (IND) application or an investigational device exemption (IDE). For each trial initiated on or after January 18, 2017, it is considered as an ACT if it satisfies the following conditions as defined in 42 CFR Part 11: (1) the study is interventional; (2) it has at least one study facility located in the U.S. or a U.S. territory, or it is conducted under an IND or IDE, or it involves a drug, biological, or device product that is manufactured in and exported from the U.S. or a U.S. territory for study in another country; (3) it evaluates at least one drug, biological, or device product regulated by FDA; (4) it is not a Phase 1 trial of a drug and/or biological product, and it is not a device feasibility study.

[5]It is also permitted for responsible parties to request extensions to the result submission deadlines for "good cause" as well as a permanent waiver of results submission under extraordinary circumstances.

[6]Sponsors and investigators themselves are responsible for ensuring the accuracy, safety and scientific validity.

[7]More details on FDAAA 801 and the Final Rule can be found at https://clinicaltrials.gov/policy/fdaaa-801-final-rule, and clinical trial reporting requirements at https://clinicaltrials.gov/policy/reporting-requirements.

[8]More information can be found on their website "FDAAA TrialsTracker" at https://fdaaa.trialstracker.

I use this expansion in the disclosure requirements on trial results, from trials on approved products to all applicable trials irrespective of their approval status, as a positive shock to the availability of failure information. The regression model is shown below.

$$Num\ of\ trials\ initiated_{mt} = \beta_0 + \beta_1 Treat_m \times Post_t + \vec{\gamma} \cdot \vec{V}_{mt} + \phi_m + \tau_t + \epsilon_{mt} \tag{1}$$

where m represents medical condition, t represents year.

The dependent variable, $Num\ of\ trials\ initiated_{mt}$, is a count-like variable that measures the number of new trials related to a medical condition, initiated in a given year. I use it as a proxy for innovation attempts. The treated group are those medical conditions with lower cumulative disclosure rate of ACTs' results prior to the event, as they are more exposed to the positive disclosure shock.[9] The cumulative disclosure rate is computed as the number of ACTs completed and with results submitted on or before 2016, scaled by the number of ACTs completed on or before 2016. $Post_t$ dummy equals 1 for years on and greater than 2017, when the Final Rule of FDAAA801 took effect. Control variables include the percentage of phase 1, 2, 3, 4 trials started, and the percentage of FDA, NIH, industry-sponsored trials per year per medical condition, accounting for the time-varying characteristics across medical conditions. I also include medical condition fixed effect and year fixed effect to control for unobservable characteristics across medical conditions and time trends. Standard errors are clustered at medical condition level. The Poisson model is used as the distribution of the outcome variable is right-skewed (Cohn, Liu and Wardlaw, 2022; Chen and Roth, 2024).

net/trials/.

[9] A higher results disclosure rate mainly reflects the tendency to voluntarily disclose failures, rather than the mandatory disclosure requirements of approved drugs. In my sample, less than 5% of ACTs receive approval according to the Orange and Purple book from FDA or BioMedTracker. For comparison, Wong, Siah and Lo (2019) estimate an average probability of successful approval from phase 1 of 6.9% using a comprehensive sample of over 400,000 entries of clinical trial data for over 21,143 compounds from Informa Pharma Intelligence's Trialtrove and Pharmaprojects databases from January 1, 2000 to October 31, 2015.

3. Data and Sample

This section introduces the data sources, outlines the sample composition, and presents the descriptive statistics of the sample used in the empirical analysis.

3.1. Clinical Trials Data

The trial level data is obtained from ClinicalTrials.gov.[10] For each clinical trial, it provides detailed information on the disease studied, timeline and status, sponsors and collaborators, study design and interventions.[11] I restrict the sample to trials with start date between 2000 and 2022, and primary completion date on or after 2008, when the result database is first released. The final sample consisted of 328,177 clinical trials. The details of sample selection process is reported in Table 1 Panel A.

Among these trials, 52,457 (16.0%) posted their results on ClinicalTrials.gov. One concern is that trials with no results submitted to ClinicalTrials.gov may disclose their results through journal publication or other avenues, including medical conferences, press releases, company websites, earning conference calls, etc. To ensure the accuracy and completeness of the result disclosure status, I supplement ClinicalTrials.gov using the data from PubMed and BioMedTracker. Two methods are used to find journal publications, which are related to clinical trials, from PubMed. First, I check if there is any self-reported results reference available on ClinicalTrials.gov. Second, I search for publications in PubMed using the NCT number of each trial.[12] To ensure the publications are related to trial results, I exclude journal articles that are published prior to the primary completion date, or those mention "study protocol" in their titles. If at least one publication satisfy

[10]The data is downloaded on March 2, 2024 from https://classic.clinicaltrials.gov/AllPublicXML.zip.

[11]The definitions of all data elements on ClinicalTrials.gov can be found at https://clinicaltrials.gov/policy/protocol-definitions.

[12]In September 2004, the International Committee of Medical Journal Editors (ICMJE) issued a policy that requires trial registration as a condition of publication. The members of ICMJE will only consider a trial for publication if it has been registered in a comprehensive and publicly available database, such as ClinicalTrials.gov. The policy applies to any clinical trial starting enrollment after July 1, 2005 (De Angelis et al., 2004).

the above criteria, I mark the trial as result available. I find 17,484 trials with results on PubMed, around 59.6% of them does not have results reported on ClinicalTrials.gov. To account for results disclosure from other sources, I use BioMedTracker data, which tracks the development progress of drugs. If it records events indicating that results have been made public, I mark the trial as having results available.[13] BioMedTracker identified 1,248 trials with disclosed results, around 34.1% of them does not have results reported on ClinicalTrials.gov. Overall, 63,246 trials have disclosed their results according to at least one of these three data sources, representing 19.3% of the trial sample.

In the trial sample, 15.8% of them are ACTs under FDA regulation. Among them, 58.6% have results disclosed. Categorized by study type, the majority of the trials, accounting for 76.3%, are interventional studies, with the remaining 23.7% as observational studies. By the agency class of lead sponsors, a small fraction of the trials are sponsored by government agencies, 0.9% by FDA and 1.4% by NIH. 22.5% of the trials are sponsored by industry entities, such as pharmaceutical firms and biotechnology companies. The remaining 75.2% are sponsored by other organizations, including hospitals and research institutes. Regarding the distribution by trial phases, 8.3% are for phase 1, 16.0% for phase 2, 9.5% for phase 3, and 6.8% for phase 4.[14]

3.2. Medical Condition-Year Data

To construct the medical condition-year level data for analysis, I first define medical condition based on the tree structure of Medical Subject Headings (MeSH). The MeSH vocabulary serves as a thesaurus that facilitates searching and indexing literature in the life sciences. It is used by the trials registered on ClinicalTrials.gov, as well as PubMed and NLM's catalog of book holdings. Each MeSH term is located in one or more trees. The tree structure goes from more general terms to more specific terms. I define medical conditions based on MeSH tree structure at Level 3

[13] An event is considered as results disclosure if it mentions "Results" in its "Event Type".

[14] The remaining trials are either without phases (for example, studies of devices or behavioral interventions), or with missing value of "Study Phase".

(Zhang, 2024).[15] Then I count the number of trials initiated, completed, with results available per MeSH per year.[16] If one trial is associated with N different MeSH, I count the number as 1/N for each MeSH so that the total number of new trials won't be inflated by higher number of associated MeSHs.

In the sample selection process, I exclude MeSHs with less than 50 trials completed over the period from 2008 to 2022. Since few trials have been conducted in those areas, leaving little prior knowledge for future investigators to learn. Moreover, these relatively small (niche) areas may not be comparable to those large ones. Less than 1% of the clinical trials fall exclusively into those MeSH.[17] I further exclude years before 2013 for DiD analysis, as the disclosure rate is relatively low in early years, and it leaves five-year of prior trials for content analysis in Section 5. The final MeSH-year sample comprises 4,780 observations, spanning years 2013 to 2022. The details of the MeSH-year sample is reported in Table 1 Panel A, all continuous variables are winsorized at 1%.

Table 2 Panel A presents the summary statistics. In the final MeSH-year sample, around 42.8 trials are initiated per MeSH per year on average, with a standard deviation of 58.8. The average cumulative disclosure rate of ACTs is 51.8%. For MeSHs belonging to the treated group, the average cumulative disclosure rate of ACTs prior to the event is 36.2%. While the control group exhibits a higher average disclosure rate of 58.6%. Following the implementation of the Final Rule, the treated group experiences a greater increase in the average ACT disclosure rate, rising to 66.2% in 2022. The average rate of the control group also increases, though with a lower growth rate, reaching 76.3% in 2022.[18]

[15] More details of MeSH tree structure can be found on https://meshb-prev.nlm.nih.gov/treeView.

[16] The initiation status of a trial is determined by its "Study Start Date". A trial is considered as completed in a given year if its "Primary Completion Date" falls within that year, and its "Overall Recruitment Status" is marked as "Completed" or "Terminated" according to ClinicalTrials.gov. The definition of those variables are provided in Appendix A.

[17] The main results remain robust under alternative thresholds. For example, the results hold when excluding MeSH terms with fewer than 10, 20, or 30 completed trials.

[18] The cumulative disclosure rate of ACTs is not the same as the compliance rate to the Final Rule, as it does not account for the one-year period for results submission after the primary completion date, or the delayed submissions under certain circumstances. Moreover, it does not exclude the early trials which are not subjected to the Final Rule.

3.3. Sponsor-Medical Condition-Year Data

To test the underlying mechanisms, I construct a sponsor-MeSH-year sample to account for the heterogeneity across sponsors and sponsor-MeSH pairs. Only the lead sponsor of each trial is considered to ensure a one-to-one match between trials and their sponsors.[19] Similar to the construction of MeSH-year sample, I count the number of trials initiated per sponsor per MeSH per year. To exclude inactive sponsors and under-represented diseases, I restrict the sample to sponsors with at least 10 trials completed, and MeSHs with at least 50 trials completed. This results in 15,869,600 observations from 3,320 sponsors across 478 MeSHs and over the sample period from 2013 to 2022.

Then I exclude sponsor-MeSH pairs with zero trials initiated over the sample period, which indicates that the sponsor did not conduct research in those fields. To ensure the comparability between the treated and control groups, I also require sponsors included in the sample to have similar number of trials in the treated and control MeSH groups prior to the event. Specifically, I exclude sponsors whose ratio of trials in treated MeSH to trials in control MeSH falls outside the range of 0.5 to 1.5. The final sample consists of 842,880 observations from 1,181 sponsors and 478 MeSH. On average, a sponsor initiate 0.089 trials per year for a given medical condition. The sample composition of the sponsor-MeSH-year data is reported in Table 1 Panel C and its descriptive statistics in Table 1 Panel C.

4. Failure Disclosure and New Trial Initiation

4.1. Baseline Analysis

Following the regression model in Equation (1), I find a positive effect of the failure disclosure on future trial initiation, as shown Table 3. Column (1) presents the baseline DiD analysis. On

The actual compliance rate should be higher than the cumulative disclosure rate, though there remains room for improvement in current compliance status.

[19] In most of the clinical trials, the lead sponsor is also the responsible party.

average, the number of trials initiated in the treated MeSHs after the event is 13.2% ($e^{0.124} - 1$) more compared to the control MeSHs. As shown in Column (2), there is no significant difference in trial initiations between the treated and control group before the policy change, satisfying the parallel trend hypothesis. This positive effect became evident one year after the implementation of the Final Rule, and the effect persists over time. Figure 1 plots dynamic trends of the differential effect between treated and control MeSH groups around the rule change. To ensure the treated and control MeSH group are comparable, I present the DiD results after entropy balance using means, variances, and skewness for all control variables in Column (3) and (4). The positive effect remains statistically significant, with a slightly reduced magnitude of 11.9% ($e^{0.112} - 1$). These results are consistent with the learning effect that researchers learn from past failures and generate more new ideas.

To alleviate the concern that the effect is driven by the abnormal increase in the amount of clinical trials due to COVID-19, I exclude COVID-related trials, COVID-related medical conditions, or COVID period from 2020 to 2022.[20] As shown in Appendix B, the positive effect remains statistically significant, though its magnitude drops to around 6-7%.

4.2. Trials on New vs Existing Drugs

To investigate whether the increased innovation activities are explorative or exploitative in nature, I differentiate trials on new drugs and existing drugs within the sub-sample of interventional clinical trials that only involve drug interventions and with intervention name disclosed. I classify a clinical trial as a trial on new drug if at least one of its interventions cannot be found neither in the past trials with start date earlier than the current trial, nor in the FDA Orange Book with approval date earlier than the start date of the current trial. If all of its interventions can be found in either past trials or the Orange Book, I classify it as a trial on existing drugs.[21] To alleviate the

[20] A trial is classified as COVID-related if it mentions "COVID-19" in its condition MeSH terms. COVID-related MeSHs are "C01.748.610", "C01.925.705", "C01.925.782", "C08.381.677", and "C08.730.610".

[21] Orange book data is downloaded from https://www.fda.gov/drugs/drug-approvals-and-databases/orange-book-data-files.

matching error due to non-standardized intervention name, I remove dosage, frequency, routes of administration, brand name, and only keep the name of the drugs using GPT-3.5 model. Among 93,080 interventional drug trials, 33.3% of them involves new drugs, while the remainder only involves existing drugs.

As shown in Table 4, I find the increase in trial initiations dominates in trials on existing drugs (Column (3) and (4)). The number of trials on new drugs shows little change, even a slight decrease (Column (1) and (2)). This evidence suggests that the innovation activities spurred by past failures are primarily exploitative or incremental innovations, focusing on modifications and improvements within existing drug compounds (intensive margin), such as to explore optimal dosage or to test the combination of multiple drugs. However, it did not inspire more exploratory or disruptive innovations that venture into new territories (extensive margin), searching for new drug candidates. This finding aligns with the learning effect that the availability of past trial results enables better evaluation of follow-on trials built upon prior studies and existing drugs. It is much more time-consuming and expertise-intensive to develop a new drug, which usually starts from pre-clinical research on the toxicity and mechanism of action of drug compounds. This process, taking up to several years of laboratory and animal studies, remains difficult to accelerate, despite access to past failure information.

4.3. Trials by Phases

The drug development process follows several phases. Phase 1 trial aims to test the safety and appropriate dosage of the drug, usually involving a small number of participants and lasting for several months. If successful, phase 2 trial investigates the efficacy and potential side effects of the drug. It involves more participants and takes up longer time. In order to determine whether a drug provides a net treatment benefits, phase 3 trial performs more extensive and rigorous testing on the drug efficacy, based on a much larger sample and lasting for up to several years. After the success of phase 3 trial, the developer can file an application to FDA. Phase 4 trial, or post-marketing surveillance study, is usually conducted after the drug has been approved to provide

further evidence on its long-term efficacy and potential side effects in a boarder patient pool.

As shown in Table 5, there exists variations in the timing of the effect across different phases of clinical trials. The positive effect shows up quicker in early phase trials compared to later phases. The increase in number of phase 1 trials initiated became significant one year after the shock, while it takes around two years for phase 2 and 3 trials, and around three years for phase 4 trials. The delayed response in late-phase trials is not surprising, as the initiations of late-phase trials are contingent on the completion and success of early-phase trials. This evidence provides further support that the changes in the number of new trials initiated are likely driven by the change in disclosure policy.

5. Learning Effect: Text-Based Evidence

To provide more evidence that researchers learn from past trial results, I quantify the level of learning by the textual similarity between current trials and related past trials. For each trial initiated on or after 2013, I compute the similarity between its summary and those of related past trials. To qualify as a related past trial, a trial need to meet all the following criteria: (1) medical relevance: it shares at least one common MeSH with the current trial; (2) information availability: it is posted before the start year of the current trial; (3) information recency: its primary completion year is no more than 5 years before the start year of the current trial; (4) external knowledge: it does not have the same primary sponsor as the current trial. The similarity between a pair of related trials is calculated as the cosine similarity of their TF-IDF vectors, rescaled to the range of 0 to 100. When computing the TF-IDF vectors, I remove stopwords from trial summaries, and drop infrequent words from the corpus if they appear in less than 100 trials.[22] If the trial has multiple related past trials that satisfy the above criteria, the average similarity is used. If the trial has no related past trial, the similarity is defined as zero.

[22] I obtain the generic stopwords list ("StopWords_Generic.txt") from the Notre Dame Software Repository for Accounting and Finance (SRAF) at https://sraf.nd.edu/textual-analysis/stopwords/.

Further, I distinguish related past trials with results available, and those without results available before the year when the current trial started. The difference between these two measures, similarity to past related trials with result available versus similarity to those without results unavailable, is used to proxy the level of learning from past results disclosure. The regression model is presented below.

$$Similarity_i = \beta_0 + \beta_1 Treat_i \times Post_t + \vec{\gamma} \cdot \vec{V}_{it} + \sigma_s + \tau_t + \epsilon_i \tag{2}$$

where i represents clinical trial, s represents the lead sponsor of the trial, t represents the year when the trial started.

The sample for this analysis consists of trials started on or after 2013 and with at least one condition MeSH within the 478 MeSHs used in the baseline analysis. A clinical trial is assigned to the treated group if its MeSH is in the treated MeSH group as defined in Equation (1). If a trial has multiple MeSHs, it is assigned to the treated group if more than half of its MeSHs are in the treated group. Control variables include the percentage of phase 1, 2, 3, 4 trials started, the percentage of FDA, NIH, industry-sponsored trials based on the trial's MeSH and its initiation year. For trials with multiple MeSHs, the average values are taken. I also add sponsor and year fixed effects. Standard errors are clustered by sponsor. The summary statistics are reported in Table 2 Panel B.

Table 6 presents the results. After the event, trials associated with MeSHs from the treated group became more similar in study summary to prior trials with results available. In terms of magnitude, the increase in similarity is around 3.2% compared to the sample mean, when sponsor fixed effect and year fixed effect are added in Column (1), and 3.0% when sponsor-year fixed effect is included in Column (3). In contrast, the similarity between treated trials and past related trials without results available remains unchanged before and after the policy change, as shown in Column (2) and (4). The differences in the effect between similarity to past trials with results and

those without results are statistically significant. As the incremental result disclosure after the rule change comes from failed trials, the increased similarity provide further support for more learning from past failures. Results are consistent if $Treat_i$ is defined as a continuous variable, measured by the proportion of MeSHs belonging to treatment group for a given trial.

6. Mechanism Tests

This section presents the mechanism tests. Two main mechanisms are: (1) the knowledge spillover effect that the disclosure of failures stimulates idea generation, thus, leads to the increase in future trial initiations; (2) the proprietary cost of disclosure that competitors could leverage the disclosed results at minimal expense. To test the two hypotheses, I construct the sample at sponsor-MeSH-year level to capture the heterogeneity across sponsor-MeSH pairs. If the knowledge spillover effect is in place, sponsors would benefit more in those research areas (MeSHs) where they have less in-house knowledge. If the proprietary cost of disclosure plays a role, sponsors facing a higher risk of information leakage to their competitors would be discouraged to initiate new trials after the policy change, especially for those trials subjected to mandatory disclosure rule, in order to avoid results disclosure. The regression model is shown below.

$$Num\ of\ trials\ initiated_{smt} = \beta_0 + \beta_1 Treat_m \times Post_t + \vec{\gamma} \cdot \vec{V}_{mt} + \phi_m + \lambda_{st} + \epsilon_{smt} \tag{3}$$

where s represents sponsor, m represents medical condition, t represents year.

The dependent variable is the number of new trials initiated by a sponsor related to a medical condition in a given year. The treated and control MeSH groups follow the same definition as in Equation (1). I control for sponsor-year fixed effect and MeSH fixed effect to capture unobservable characteristics. I also add time-varying MeSH characteristics as control variables.

In untabulated results, there is no significant increase in the number of trials initiated using the full sample of sponsor-MeSH-year data. The null results are consistent across various specifica-

tions, including controlling for sponsor-year fixed effect or sponsor and year fixed effect separately, and applying entropy balance.

6.1. Knowledge Spillover Effect

Knowledge spillover refers to the diffusion of insights and findings from one entity to others in the field. When more trial results are disclosed publicly, they could contribute to a collective pool of knowledge and enhances the innovation potential of the entire industry. To test this knowledge spillover effect, I use the cumulative number of trials initiated prior to the policy change as a proxy for in-house knowledge. The high group comprises sponsor-MeSH pairs whose cumulative number falls above the median compared to other MeSH within the same sponsor. The more trials that the sponsor has conducted in a medical field, the more experienced and knowledgeable it might be, thus, the potential benefits from knowledge spillover would be limited. Comparison within a sponsor help mitigate the concerns of other unobservable and confounding sponsors characteristics. For example, sponsors with fewer trials initiated could be those with less in-house knowledge, but may also be those more financially constrained, thus, less capable of investing in future trials. By comparing across MeSHs within a single sponsors, the results are not affected by those common sponsor characteristics.

Table 7 presents the results. In areas where sponsors have lower levels of in-house knowledge accumulation, there is a significant increase of 24.6% ($e^{0.220} - 1$) in the number of trials initiated after the policy change. In contrast, no effect is observed in areas where sponsors process higher levels of internal expertise. The difference between the coefficient of the high and low group is statistically significant at 1% level. This finding is consistent with the prediction of knowledge spillover that sponsors who possess limited internal knowledge rely more on external information sources, and benefit more from learning from others. Moreover, the results contradict with the prediction of deterrence effect that the fear of failure, amplified by the knowledge of others' failures, would deter innovation. If the fear of failure plays a dominant role, sponsors are expected to

shy away from medical fields where they possess less in-house knowledge, as the perceived risk of failure are likely higher.

6.2. Proprietary Cost of Disclosure

Proprietary cost mainly reflects the value of trial results as exclusive assets to its sponsors. Once these results are mandated to disclose publicly, they can no longer enjoy the benefits of exclusivity. Their competitors could free ride without incurring the original research cost, undermining the sponsors' competitiveness or first mover advantages. If proprietary cost serve as one of the decisive factors of trial initiations, sponsors with higher proprietary cost are expected to initiate fewer trials after the policy change to avoid results disclosure. To examine this mechanism, I use the number of ACTs started before September 21, 2016 when the Final Rule was issued and completed after January 18, 2017 when the rule came into effective, as a proxy for proprietary cost. The high proprietary cost group consists of sponsors whose number of trials meeting the above criteria fall above the median of all sponsors.[23] Those trials are subjected to mandatory results disclosure requirement, but their sponsors did not know this ex-ante when they make the decision to start the trials, which ensures the grouping is not endogenously affected by the shock.

When examining the impact of policy change on the number of new trials initiated, there is no differential effect between sponsors with higher and lower proprietary cost. As shown in Table 8 Panel A Column (1) and (2), in the full sample, both groups of sponsors do not experience change in trials initiation after the shock. In Column (3) and (4) based on the sub-sample of sponsor-MeSH pairs with lower in-house knowledge prior to the event, both groups of sponsors experience significant increase of over 20% in the number of trials initiated. The coefficient differences between the high and low group are statistically insignificant.

Restricting the analysis to trials that are mandated to disclose their summary results irrespective of FDA approval status after the rule change, sponsors with lower proprietary cost initiate

[23] Majority of the sponsors have zero trials satisfying the above criteria, namely, the median value is zero.

significantly more trials following the policy change.[24] While there is no change in the number of new trials subjected to the disclosure rule among sponsors with high proprietary cost. The results hold under both the full sample (Column (1) and (2)) and the sub-sample of sponsor-MeSH pairs with low in-house knowledge (Column (3) and (4)). The coefficient difference between the high and low group is statistically significant at 1% level in the full sample, and insignificant with a p-value of 0.148 in the sub-sample. These findings suggest that when sponsors already suffer from high proprietary cost and potential information leakage, they are less willing to start new trials that are subjected to the mandatory disclosure requirement compared to those sponsors with lower proprietary cost. However, this does not affect the initiation of non-applicable trials, which are not bounded by the disclosure requirement and do not incur proprietary information cost.

7. Conclusion

In this study, I explore the effect of failure disclosure on future innovation activities in the setting of clinical trials. Using an extension in the result submission requirement from ACTs of drugs, biologics, or devices that are approved, licensed, or cleared by FDA to all ACTs despite their approval status, I document an increase in trials initiations following this positive shock to the availability of failure information. The positive effect dominates in trials on existing drugs, suggesting that the new innovations are mainly incremental modifications based on prior studies rather than disruptive innovations of new compounds. Textual analysis on trial summaries provides further support for learning from past trials, especially those with disclosed results. Consistent with the knowledge spillover channel, sponsors benefits more from the disclosure of others' failures in medical fields where they process less in-house knowledge. The proprietary cost associated with disclosure also influences the decision to initiate trials. Sponsors facing higher proprietary cost are less willing to start new trials subject to mandatory disclosure requirements.

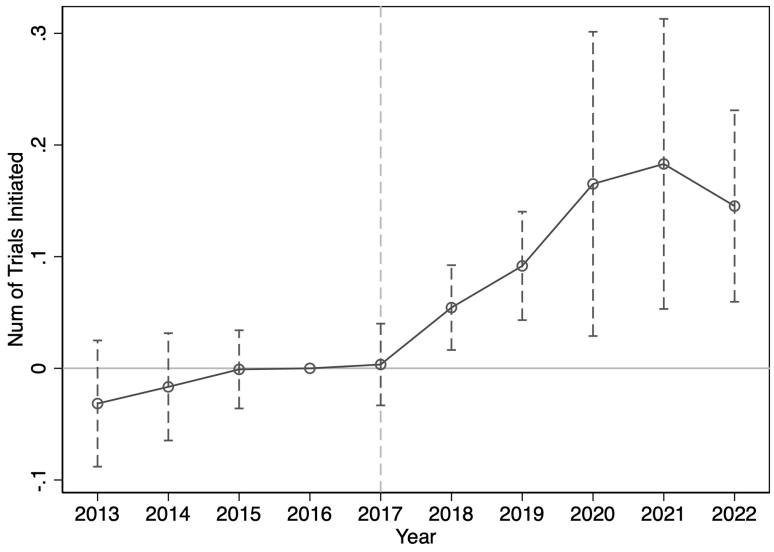
[24] In September 2016, together with the issuance of Final Rule of FDAAA 801, the final NIH policy also mandated disclosure of trial results. The NIH policy applies to all clinical trials funded in whole or in part by NIH and initiated after January 18, 2017. To compute the number of trials subjected to mandatory results disclosure, I sum up the number of ACTs under FDA regulation and NIH-funded trials.

This paper contributes to the discussion on whether and how failure disclosure can stimulate innovation activities, which despite its importance, lacks empirical evidence. The findings support mandated disclosure policies, when disclosure is privately costly but socially beneficial. It also echoes the ongoing efforts to promote greater research transparency. However, the optimal level of disclosure requirements remains an open question, given the trade-off between the social benefits from knowledge spillover and the proprietary cost borne by the innovating entities.

Figure 1: Dynamic Trends

This figures show the dynamic effect of the expansion in disclosure requirements on new trials initiation over the period from 2013 to 2022. Figure (a) presents the dynamic plot of the DiD analysis, and Figure (b) presents the plot after entropy balancing. The confidence interval of 5% significance level is presented. Control variables include the percentage of phase 1, 2, 3, 4 trials started, percentage of FDA, NIH, industry-sponsored trials per year per medical condition. MeSH fixed effect and year fixed effect are included. The regression results are reported in Table 3. The sample composition is presented in Table 1 and variable definition in Appendix A.

(a) DID



(b) DID+Entropy balance

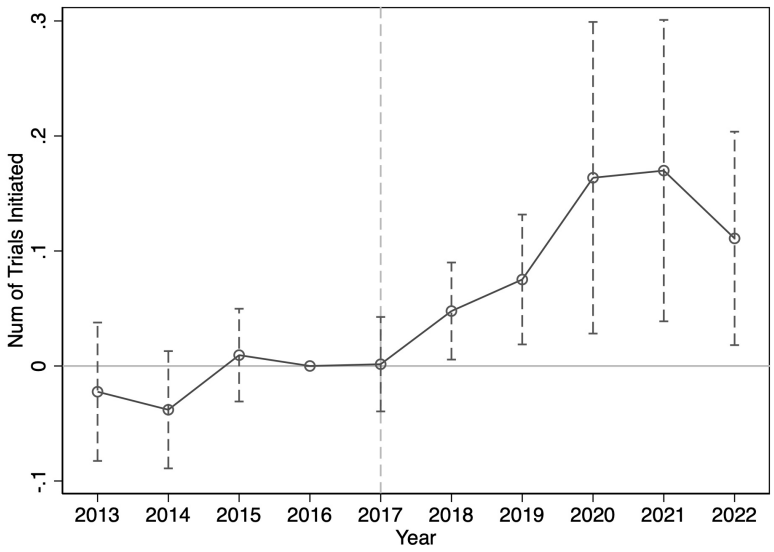


Table 1: Sample composition

Panel A. Trial-level data

	Number of Trials
Clinical trials data downloaded on March 2, 2024	485,171
minus: initiation year before 2000, or after 2022	(56,915)
minus: primary completion year before 2008	(30,342)
minus: missing condition MeSH	(69,737)
Clinical trials sample	328,177

Panel B. MeSH-year-level data

	MeSH-Year Obs
MeSH-year sample from 328,177 trials	31,890
minus: MeSH with less than 50 trials completed	(24,720)
minus: years on or before 2012	(2,390)
MeSH-year sample	4,780

Panel C. Sponsor-MeSH-year-level data

	Sponsor-MeSH-Year Obs
Sponsor-MeSH-year sample from sponsors with at least 10 trials completed, and MeSHs with at least 50 trials completed	15,869,600
minus: sponsor-MeSH pairs with zero trial started	(14,169,520)
minus: sponsors with an incomparable number of trials in treated and control MeSH groups	(858,870)
Sponsor-MeSH-year sample	841,210

This table shows the sample composition. Panel A presents the sample selection of clinical trials from ClinicalTrials.gov. Panel B presents the MeSH-year sample used in Section 4. Panel C presents the Sponsor-MeSH-year sample used in Section 6.

Table 2: Summary statistics

Panel A. MeSH-year level observations

	Obs	Mean	Std. Dev.	P10	P25	P50	P75	P90
Num of trials initiated	4,780	42.770	58.827	6.415	10.340	21.029	44.432	111.979
Num of trials initiated (new drug)	4,780	3.082	4.892	0.000	0.433	1.275	3.354	8.239
Num of trials initiated (old drug)	4,780	7.752	12.625	0.474	1.242	3.167	7.805	20.930
Num of trials initiated (phase 1)	4,780	3.283	5.614	0.000	0.325	1.167	3.500	8.370
Num of trials initiated (phase 2)	4,780	6.088	11.162	0.167	0.733	2.183	5.837	14.710
Num of trials initiated (phase 3)	4,780	3.336	5.286	0.000	0.500	1.457	3.634	8.478
Num of trials initiated (phase 4)	4,780	2.411	4.013	0.000	0.367	1.000	2.385	6.237
ACT cumulative disclosure rate	4,780	0.518	0.175	0.302	0.416	0.529	0.647	0.732
Treat	4,780	0.500	0.500	0.000	0.000	0.500	1.000	1.000
Post	4,780	0.600	0.490	0.000	0.000	1.000	1.000	1.000
% Phase 1	4,780	0.073	0.071	0.000	0.020	0.054	0.106	0.167
% Phase 2	4,780	0.132	0.112	0.013	0.047	0.102	0.192	0.299
% Phase 3	4,780	0.080	0.068	0.000	0.031	0.067	0.112	0.169
% Phase 4	4,780	0.060	0.055	0.000	0.020	0.047	0.087	0.133
% FDA sponsored	4,780	0.007	0.015	0.000	0.000	0.000	0.006	0.022
% NIH sponsored	4,780	0.009	0.019	0.000	0.000	0.000	0.011	0.029
% Industry sponsored	4,780	0.214	0.153	0.036	0.095	0.190	0.304	0.426

Panel B. Trial level observations

	Obs	Mean	Std. Dev.	P10	P25	P50	P75	P90
Similarity (with results)	228,108	4.662	3.281	1.656	2.430	3.718	5.836	8.947
Similarity (without results)	228,108	4.591	3.001	1.797	2.537	3.733	5.710	8.569
Treat	228,108	0.424	0.494	0.000	0.000	0.000	1.000	1.000
Post	228,108	0.674	0.469	0.000	0.000	1.000	1.000	1.000
% Phase 1	228,108	0.077	0.064	0.020	0.033	0.060	0.101	0.152
% Phase 2	228,108	0.140	0.104	0.037	0.062	0.108	0.189	0.303
% Phase 3	228,108	0.077	0.042	0.028	0.047	0.072	0.099	0.130
% Phase 4	228,108	0.056	0.038	0.015	0.028	0.049	0.077	0.109
% FDA sponsored	228,108	0.007	0.010	0.000	0.000	0.003	0.009	0.017
% NIH sponsored	228,108	0.008	0.012	0.000	0.001	0.004	0.011	0.021
% Industry sponsored	228,108	0.205	0.112	0.080	0.129	0.189	0.259	0.347

Panel C. Sponsor-MeSH-year level observations

	Obs	Mean	Std. Dev.	P10	P25	P50	P75	P90
Num of trials initiated	842,880	0.089	0.226	0.000	0.000	0.000	0.000	0.333
Treat	842,880	0.470	0.499	0.000	0.000	0.000	1.000	1.000
Post	842,880	0.600	0.490	0.000	0.000	1.000	1.000	1.000
In-house knowledge[25]	842,880	0.569	1.080	0.000	0.000	0.200	0.600	1.518
Proprietary cost[26]	842,880	0.015	0.077	0.000	0.000	0.000	0.000	0.000

[25] In-house knowledge is measured by the number of clinical trials initiated on or before 2016 per sponsor-MeSH

[26] Proprietary cost is measured by the number of clinical trials initiated before Sep 21, 2016 and completed after Jan 18, 2017 per sponsor

% Phase 1	842,880	0.072	0.062	0.009	0.027	0.057	0.100	0.153
% Phase 2	842,880	0.132	0.105	0.028	0.054	0.103	0.177	0.292
% Phase 3	842,880	0.078	0.053	0.018	0.041	0.071	0.106	0.144
% Phase 4	842,880	0.059	0.046	0.010	0.025	0.049	0.083	0.121
% FDA sponsored	842,880	0.007	0.013	0.000	0.000	0.001	0.009	0.021
% NIH sponsored	842,880	0.009	0.015	0.000	0.000	0.003	0.011	0.025
% Industry sponsored	842,880	0.203	0.127	0.055	0.109	0.184	0.274	0.374

This table shows the number of observations, the mean, the standard deviation, and 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentiles for the variables and observations used in my empirical tests. Panel A presents the summary statistics for the MeSH-year sample used in Section 4. Panel B comprises the sub-sample of clinical trials used for the textual analysis in Section 5. Panel C comprises the Sponsor-MeSH-year sample used in Section 6. Continuous variables are winsorized at 1%.

Table 3: New Trials Initiation

	Num of trials initiated			
	DID		DID+Entropy balance	
	(1)	(2)	(3)	(4)
Treat × Post	0.124***		0.112***	
	(0.041)		(0.041)	
Treat × Pre4		-0.032		-0.022
		(0.029)		(0.031)
Treat × Pre3		-0.017		-0.038
		(0.025)		(0.026)
Treat × Pre2		-0.001		0.009
		(0.018)		(0.021)
Treat × Post0		0.003		0.002
		(0.019)		(0.021)
Treat × Post1		0.054***		0.048**
		(0.019)		(0.022)
Treat × Post2		0.092***		0.075***
		(0.025)		(0.029)
Treat × Post3		0.165**		0.164**
		(0.070)		(0.069)
Treat × Post4		0.183***		0.170**
		(0.066)		(0.067)
Treat × Post5		0.145***		0.111**
		(0.044)		(0.047)
Controls	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Obs	4,780	4,780	4,780	4,780

This table shows the impact of the expansion in disclosure requirements on new trial initiations using Poisson regression model. The dependent variable is the number of new trials related to a medical condition (MeSH) initiated in a given year. The treated group are those MeSHs with lower cumulative disclosure rate of ACTs' results prior to the event. Column (1) and (2) present the DiD analysis and its dynamic trends. In Column (3) and (4), the matching is based on entropy balancing using means, variances, and skewness for all control variables. Control variables include the percentage of phase 1, 2, 3, 4 trials started, percentage of FDA, NIH, industry-sponsored trials per year per MeSH. MeSH fixed effect and year fixed effect are added. The sample composition is presented in Table 1 and variable definition in Appendix A. Standard errors are clustered at MeSH level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

Table 4: New Trials Initiation: New vs Existing Drug

	Num of trials initiated			
	New Drug		Existing Drug	
	(1)	(2)	(3)	(4)
Treat × Post	-0.032		0.130***	
	(0.050)		(0.044)	
Treat × Pre4		-0.049		-0.033
		(0.058)		(0.040)
Treat × Pre3		-0.082		-0.027
		(0.055)		(0.034)
Treat × Pre2		0.031		-0.012
		(0.057)		(0.035)
Treat × Post0		-0.152***		0.013
		(0.055)		(0.036)
Treat × Post1		-0.024		0.058
		(0.055)		(0.037)
Treat × Post2		-0.157***		0.121***
		(0.060)		(0.039)
Treat × Post3		0.049		0.203**
		(0.092)		(0.097)
Treat × Post4		-0.040		0.166***
		(0.083)		(0.063)
Treat × Post5		-0.047		0.096**
		(0.069)		(0.045)
Controls	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Obs	4,780	4,780	4,780	4,780

This table compares the impact of expansion in disclosure requirements on new trial initiations between trials on new drugs and trials on existing drugs, using Poisson regression model. The dependent variable of Column (1) and (2) is the number of drug trials initiated related to at least one new drug per medical condition (MeSH) per year. A drug is considered new if it has never used in previous trials or included in FDA Orange Book. The dependent variable of Column (3) and (4) is the number of drug trials initiated related to existing drugs per MeSH per year. The treated group are those MeSHs with lower cumulative disclosure rate of ACTs' results prior to the event. The matching is based on entropy balancing using means, variances, and skewness for all control variables. Control variables include the percentage of phase 1, 2, 3, 4 trials started, percentage of FDA, NIH, industry-sponsored trials per year per MeSH. MeSH fixed effect and year fixed effect are added. The sample composition is presented in Table 1 and variable definition in Appendix A. Standard errors are clustered at MeSH level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

Table 5: New Trials Initiation: By Phases

	Num of trials initiated (by phases)			
	Phase 1	Phase 2	Phase 3	Phase 4
	(1)	(2)	(3)	(4)
Treat × Pre4	0.013	-0.060	-0.021	-0.008
	(0.047)	(0.040)	(0.065)	(0.053)
Treat × Pre3	0.064	-0.019	0.044	-0.074
	(0.057)	(0.032)	(0.058)	(0.061)
Treat × Pre2	0.005	-0.038	0.032	0.061
	(0.040)	(0.035)	(0.063)	(0.039)
Treat × Post0	0.023	0.024	0.033	0.010
	(0.041)	(0.029)	(0.056)	(0.043)
Treat × Post1	0.066*	0.053	0.077	0.052
	(0.040)	(0.033)	(0.050)	(0.043)
Treat × Post2	0.112**	0.100***	0.124**	0.044
	(0.044)	(0.033)	(0.061)	(0.046)
Treat × Post3	0.261**	0.215**	0.241**	0.219**
	(0.102)	(0.095)	(0.108)	(0.088)
Treat × Post4	0.203***	0.229***	0.231**	0.241***
	(0.073)	(0.079)	(0.103)	(0.085)
Treat × Post5	0.220***	0.124***	0.204**	0.160***
	(0.058)	(0.047)	(0.088)	(0.056)
Controls	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Obs	4,750	4,770	4,750	4,750

This table compares the impact of expansion in disclosure requirements on new trial initiations by phases using Poisson regression model. The dependent variable is the number of trials related to a medical condition (MeSH) initiated in a given year: Column (1) for phase-1 trials, Column (2) for phase-2 trials, Column (3) for phase-3 trials, and Column (4) for phase-4 trials. The treated group are those MeSHs with lower cumulative disclosure rate of ACTs' results prior to the event. The matching is based on entropy balancing using means, variances, and skewness for all control variables. Control variables include the percentage of phase 1, 2, 3, 4 trials started, percentage of FDA, NIH, industry-sponsored trials per year per MeSH. MeSH fixed effect and year fixed effect are added. The sample composition is presented in Table 1 and variable definition in Appendix A. Standard errors are clustered at MeSH level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

Table 6: Learning Effect: Analysis of Trial Summaries

	Similarity of trial summaries			
	w/ results	w/o results	w/ results	w/o results
	(1)	(2)	(3)	(4)
Treat × Post	0.151***	0.038	0.140***	0.031
	(0.039)	(0.036)	(0.045)	(0.041)
Test of coefficient difference between similarity to past trials with results and without results				
Difference (p-value)	0.113*** (0.000)		0.108*** (0.000)	
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	No	No
Sponsor FE	Yes	Yes	No	No
Year × Sponsor FE	No	No	Yes	Yes
Obs	213,259	213,259	188,136	188,136
Adj. R^2	0.140	0.108	0.136	0.103

This table examines the learning effect based on the textual analysis of trial summaries. In Column (1) and (3), the dependent variables are the similarity of trial summaries between the current trial and its past related trials with disclosed results. In Column (2) and (4), the dependent variables are the similarity of trial summaries between the current trial and its past related trials without disclosed results. The treated group consists trials with medical condition (MeSH) in the treated group as defined in Table 3. If a trial has multiple MeSHs, it is assigned to the treated group if more than half of its MeSHs are in the treated group. Control variables include the percentage of phase 1, 2, 3, 4 trials started, the percentage of FDA, NIH, industry-sponsored trials based on the trial's MeSH and its initiation year. For trials with multiple MeSHs, the average values are taken. Sponsor and year fixed effect are added. The sample comprises trials started on or after 2013 and with at least one condition MeSH within the 478 MeSHs used in Table 3. Standard errors are clustered at sponsor level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

Table 7: New Trials Initiation: Knowledge Spillover

In-house knowledge =	Num of trials initiated Low (1)	 High (2)
Treat × Post	0.220***	-0.007
	(0.064)	(0.060)
Test of coefficient difference between high-knowledge areas and low-knowledge areas		
Difference (p-value)	0.227*** (0.010)	
Controls	Yes	Yes
MeSH FE	Yes	Yes
Sponsor × Year FE	Yes	Yes
Obs	363,397	323,970

This table compares the impact of the expansion in disclosure requirements on new trial initiations between areas where sponsors process high and low in-house knowledge, using Poisson regression model. The dependent variable is the number of new trials per sponsor per medical condition (MeSH) per year. The treated group are those MeSHs with lower cumulative disclosure rate of ACTs' results prior to the event. The matching is based on entropy balancing using means, variances, and skewness for all control variables. Control variables include the percentage of phase 1, 2, 3, 4 trials started, percentage of FDA, NIH, industry-sponsored trials per year per MeSH. MeSH fixed effect and year fixed effect are added. The sample composition is presented in Table 1 and variable definition in Appendix A. Standard errors are clustered at MeSH level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

Table 8: New Trials Initiation: Proprietary Cost of Disclosure

Panel A. Effect on all trials initiation

	Num of trials initiated			
	Full sample		Sub-sample with low in-house knowledge	
Proprietary cost =	High	Low	High	Low
	(1)	(2)	(3)	(4)
Treat × Post	-0.000	0.028	0.202***	0.241***
	(0.040)	(0.042)	(0.073)	(0.069)
Test of coefficient difference between sponsors with high and low proprietary costs				
Difference (p-value)	-0.028 (0.635)		-0.039 (0.696)	
Controls	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes
Sponsor × Year FE	Yes	Yes	Yes	Yes
Obs	237,494	541,546	111,805	251,592

Panel B. Effect on trials subjected to mandatory disclosure requirement

	Num of trials initiated			
	Full sample		Sub-sample with low in-house knowledge	
Proprietary cost =	High	Low	High	Low
	(1)	(2)	(3)	(4)
Treat × Post	-0.029	0.454***	0.063	0.513*
	(0.051)	(0.143)	(0.098)	(0.295)
Test of coefficient difference between sponsors with high and low proprietary costs				
Difference (p-value)	-0.483***(0.001)		-0.450 (0.148)	
Controls	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes
Sponsor × Year FE	Yes	Yes	Yes	Yes
Obs	199,616	63,025	81,187	21,616

This table shows the impact of the expansion in disclosure requirements on new trial initiations between sponsors with high and low proprietary cost, using Poisson regression model. In Panel A, the dependent variable is the number of new trials per sponsor per medical condition (MeSH) per year. In Panel B, the dependent variable is the number of new trials subjected to result disclosure requirement per sponsor per MeSH per year. In both panels, Column (1) and (2) present the results based on the full sample, Column (3) and (4) present the results based on the sub-sample of sponsor-MeSH pairs with low in-house knowledge. The treated group are those MeSHs with lower cumulative disclosure rate of ACTs' results prior to the event. The matching is based on entropy balancing using means, variances, and skewness for all control variables. Control variables include the percentage of phase 1, 2, 3, 4 trials started, percentage of FDA, NIH, industry-sponsored trials per year per MeSH. MeSH fixed effect and year fixed effect are added. The sample composition is presented in Table 1 and variable definition in Appendix A. Standard errors are clustered at MeSH level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

References

Abadie, A., 2020. Statistical nonsignificance in empirical economics. American Economic Review: Insights 2 (2), 193–208.

Aghion, P., Bergeaud, A., Van Reenen, J., 2023. The impact of regulation on innovation. American Economic Review 113 (11), 2894–2936.

Andrews, I., Kasy, M., 2019. Identification of and correction for publication bias. American Economic Review 109 (8), 2766–2794.

Anton, J. J., Yao, D. A., 1994. Expropriation and inventions: Appropriable rents in the absence of property rights. Review of Economic Studies 84 (1), 190–209.

Aoki, R., Spiegel, Y., 2009. Pre-grant patent publication and cumulative innovation. International Journal of Industrial Organization 27 (3), 333–345.

Badertscher, B., Shroff, N., White, H. D., 2013. Externalities of public firm presence: Evidence from private firms' investment decisions. Journal of Financial Economics 109 (3), 682–706.

Baruffaldi, S. H., Simeth, M., 2020. Patents and knowledge diffusion: The effect of early disclosure. Research Policy 49 (4), 03927.

Bessen, J., 2005. Patents and the diffusion of technical information. Economics Letters 86 (1), 121–128.

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

Bloom, N., Jones, C. I., Van Reenen, J., Webb, M., 2020. Are ideas getting harder to find? The American Economic Review 110 (4), 1104–1144.

Brodeur, A., Lé, M., Sangnier, M., Zylberberg, Y., 2016. Star wars: The empirics strike back. American Economic Journal: Applied Economics 8 (1), 1–32.

Capkun, V., Lou, Y., Otto, C. A., Wang, Y., 2023. Do firms respond to peer disclosures? evidence from disclosures of clinical trial results. The Accounting Review 98 (3), 71–108.

Chen, J., Roth, J., 2024. Logs with zeros? some problems and solutions. Quarterly Journal of Economics 139 (2), 891–936.

Choi, J. P., 1991. Dynamic r&d competition under "hazard rate" uncertainty. The RAND Journal of Economics 22 (4), 596–610.

Chu, J., He, Y., Hui, K. W., Lehavy, R., 2024. New product announcements, innovation disclosure, and future firm performance. Review of Accounting Studies, 291–315.

Cohn, J. B., Liu, Z., Wardlaw, M. I., 2022. Count (and count-like) data in finance. Journal of Financial Economics 146 (2), 529–551.

Contigiani, A., Hsu, D. H., Barankay, I., 2018. Trade secrets and innovation: Evidence from the "inevitable disclosure" doctrine. Strategic Management Journal 39 (11), 2921–2942.

De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A. J. P., Schroeder, T. V., Sox, H. C., Van Der Weyden, M. B., 2004. Clinical trial registration: a statement from the international committee of medical journal editors. CMAJ: Canadian Medical Association Journal 171 (6), 606–607.

Doraszelski, U., 2003. An r&d race with knowledge accumulation. The RAND Journal of Economics 34 (1), 20–42.

Dyer, T. A., Glaeser, S., Lang, M. H., Sprecher, C., 2023. The effect of patent disclosure quality on innovation. Journal of Accounting and Economics 29, 101647.

Furman, J. L., Nagler, M., Watzinger, M., 2021. Disclosure and subsequent innovation: Evidence from the patent depository library program. American Economic Journal: Economic Policy 13 (4), 239–270.

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

Ganglmair, B., Reimers, I., 2022. Visibility of technology and cumulative innovation: Evidence from trade secrets laws. Working paper.

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

Glaeser, S. A., Landsman, W. R., 2021. Deterrent disclosure. Accounting Review 96 (5), 291–315.

Griffin, P. A., Hong, H. A., Ryou, J. W., 2022. Proprietary costs: Why do r&d-active firms choose single-lender financing? The Accounting Review 97 (6), 263–296.

Gross, D. P., 2019. The consequences of invention secrecy: Evidence from the uspto patent secrecy program in world war ii. Working paper.

Harris, C., Vickers, J., 1987. Racing with uncertainty. Review of Economic Studies 54 (1), 1–21.

Hegde, D., Herkenhoff, K., Zhu, C., 2023. Patent publication and innovation. Journal of Political Economy 131 (7), 1845–1903.

Hegde, D., Luo, H., 2018. Patent publication and the market for ideas. Management Science 64 (2), 652–672.

Hsu, P.-H., Lee, K., Moon, S. K., Oh, S., 2019. Information disclosure and peer innovation: Evidence from mandatory reporting of clinical trials. Working paper.

Hughes, J. S., Pae, S., 2015. Discretionary disclosure, spillovers, and competition. Review of Accounting Studies 20 (1), 319–342.

Jones, D. A., 2007. Voluntary disclosure in r&d-intensive industries. Contemporary Accounting Research 24 (2), 489–522.

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

Krieger, J. L., 2021. Trials and terminations: Learning from competitors' r&d failures. Management Science 67 (9), 5525–5548.

Krieger, J. L., Li, D., Papanikolaou, D., 2022. Missing novelty in drug development. Review of Financial Studies 35 (2), 636–679.

Lippman, S. A., McCardle, K. F., 1987. Dropout behavior in r&d races with learning. The RAND Journal of Economics 18 (2), 287–295.

Lo, A. W., Thakor, R. T., 2022. Financing biomedical innovation. Annual Review of Financial Economics 14 (1), 231–270.

Loury, G. C., 1979. Market structure and innovation. Quarterly Journal of Economics 93 (3), 395–410.

Lück, S., Balsmeier, B., Seliger, F., Fleming, L., 2020. Early disclosure of invention and reduced duplication: An empirical test. Management Science 66 (6), 2677–2685.

Mace, C., 2023. The real effects of equity markets on innovation. Journal of Financial and Quantitative Analysis 58 (6), 2522–2552.

Rassenfosse, G. d., Pellegrino, G., Raiteri, E., 2024. Do patents enable disclosure? evidence from the invention secrecy act. International Journal of Industrial Organization 92, 103044.

Reinganum, J. F., 1985. Innovation and industry evolution. Quarterly Journal of Economics 100 (1), 81–99.

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

Sterling, T. D., 1959. Publication decisions and their possible effects on inferences drawn from tests of significance–or vice versa. Journal of the American Statistical Association 54 (285), 30–34.

Thakor, R. T., Lo, A. W., 2017. Optimal financing for r&d-intensive firms. Working paper.

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

Wong, C. H., Siah, K. W., Lo, A. W., 2019. Estimation of clinical trial success rates and related parameters. Biostatistics 20 (2), 273–286.

Zhang, Y., 2024. Corporate r&d investments following competitors' voluntary disclosures: Evidence from the drug development process. Journal of Accounting Research 62 (1), 335–373.

Appendix A. Variable Definition

This table shows the definition of all variables and their sources. MeSH-year level and Sponsor-MeSH-year level variables are constructed from trial level data. CT stands for ClinicalTrials.gov. PM stands for PubMed. BMT stands for BioMedTracker. OB stands for the Orange Book from FDA.

Variable	Definition	Source
Trial level variables		
Study type	The nature of the investigation or investigational use for which clinical study information is being submitted. The type falls into one of the following options: Interventional; Observational; Expanded Access.	CT
Overall recruitment status	The recruitment status for the clinical study as a whole, based upon the status of the individual sites. The status falls into one of the following options: Not yet recruiting; Recruiting; Enrolling by invitation; Active, not recruiting; Completed; Suspended; Terminated; Withdrawn.	CT
Study start date	The estimated date on which the clinical study will be open for recruitment of participants, or the actual date on which the first participant was enrolled.	CT
Primary completion date	The date that the final participant was examined or received an intervention for the purposes of final collection of data for the primary outcome, whether the clinical study concluded according to the pre-specified protocol or was terminated.	CT
Results first posted date	The date when the results of the study was first available on ClinicalTrials.gov.	CT
Brief summary	A short description of the clinical study, including a brief statement of the clinical study's hypothesis, written in language intended for the lay public. The length limit is 5,000 characters.	CT
Primary disease (Condition MeSH)	The name(s) of the disease(s) or condition(s) studied in the clinical study, or the focus of the clinical study. Use, if available, appropriate descriptors from NLM's Medical Subject Headings (MeSH)-controlled vocabulary thesaurus.	CT
Study phase	The numerical phase of a clinical trial for a drug or biological product, consistent with terminology in 21 CFR 312.21 and in 21 CFR 312.85 for phase 4 studies. The phase falls into one of the following options: Early Phase 1, Phase 1, Phase 1/Phase 2, Phase 2, Phase 2/Phase 3, Phase 3, Phase 4, N/A. In Section 4.3, I classify trials belonging to "Early Phase 1" and "Phase 1" as phase 1 trials, "Phase 1/Phase 2" and "Phase 2" as phase 2 trials, "Phase 2/Phase 3" and "Phase 3" as phase 3 trials, "Phase 4" as phase 4 trials.	CT

Sponsor name	The name of the entity or the individual who is the sponsor of the clinical study. When a clinical study is conducted under an investigational new drug application (IND) or investigational device exemption (IDE), the IND or IDE holder is considered the sponsor. When a clinical study is not conducted under an IND or IDE, the single person or entity who initiates the study, by preparing and/or planning the study, and who has authority and control over the study, is considered the sponsor.	CT
Results reference	Citations to publications related to results from this clinical study, provided in PubMed Unique Identifier (PMID) and/or bibliographic citation.	CT
Results available	A dummy variable that equals 1 if the results of a clinical trial is available from any of the three sources: ClinicalTrials.gov, PubMed, and BioMedTracker. For ClinicalTrials.gov, the results are available if "Results First Posted Date" is not missing. For PubMed, the results are available if at least one publication related to the trial can be found using "NCT ID" or "PMID", excluding those published prior to trial completion and those with "study protocol" in their titles. For BioMedTracker, the results are available if it records events indicating that results have been made public.	CT, PM, BMT
Trial on new drugs	A dummy variable that equals 1 if the trial has at least one intervention that cannot be found neither in the past trials with start date earlier than the current trial, nor in the Orange Book with approval date earlier than the start date of the current trial.	CT, OB
Trial on existing drugs	A dummy variable that equals 1 if all of the trial's interventions can be found in either the past trials with start date earlier than the current trial, or the FDA Orange Book with approval date earlier than the start date of the current trial.	CT, OB
Similarity	The textual similarity between the "Brief Summary" of a trial and its related past trials. The similarity is calculated as the cosine similarity of their TF-IDF vectors, rescaled to the range of 0 to 100. If the trial has multiple related past trials that satisfy the above criteria, the average similarity is used. If the trial has no related past trial, the similarity is defined as zero. I differentiate related past trials with disclosed results versus those without. The selection criteria of related past trials are explained in Section 5.	CT
MeSH-year level variables		
Num of trials initiated	The number of clinical trials associated with a given MeSH initiated in a given year.	CT

ACT cumulative disclosure rate	The number of ACTs completed and with results available on or before a given year, scaled by the total number of ACTs completed on or before that year.	CT
% Phase 1	Proportion of phase 1 trials associated with a given MeSH initiated in a given year relative to the total number of trials initiated in that MeSH during the same year.	CT
% Phase 2	Proportion of phase 2 trials associated with a given MeSH initiated in a given year relative to the total number of trials initiated in that MeSH during the same year.	CT
% Phase 3	Proportion of phase 3 trials associated with a given MeSH initiated in a given year relative to the total number of trials initiated in that MeSH during the same year.	CT
% Phase 4	Proportion of phase 4 trials associated with a given MeSH initiated in a given year relative to the total number of trials initiated in that MeSH during the same year.	CT
% FDA sponsored	Proportion of trials sponsored by FDA associated with a given MeSH initiated in a given year relative to the total number of trials initiated in that MeSH during the same year.	CT
% NIH sponsored	Proportion of trials sponsored by NIH associated with a given MeSH initiated in a given year relative to the total number of trials initiated in that MeSH during the same year.	CT
% Industry sponsored	Proportion of trials sponsored by industry entities associated with a given MeSH initiated in a given year relative to the total number of trials initiated in that MeSH during the same year.	CT
Sponsor-MeSH-year level variables		
Num of trials initiated	The number of clinical trials led by the sponsor associated with a given MeSH initiated in a given year.	CT
In-house knowledge	The number of clinical trials led by the sponsor initiated on or before 2016 associated with a given MeSH.	CT
Proprietary cost	The number of clinical trials led by the sponsor initiated before Sep 21, 2016 when the Final Rule was issued and completed after Jan 18, 2017 when the Final Rule came into effective.	CT

Appendix B. Robustness: Exclude COVID-19 Effect

This table shows the impact of expansion in disclosure requirements on new trial initiations after excluding the effect of COVID-19, using Poisson regression model. The dependent variable is the number of new trials related to a medical condition (MeSH), initiated in a given year. The treated group are those MeSHs with lower cumulative disclosure rate of ACTs' results prior to the event. Column (1) and (2) present the DiD analysis and DiD after entropy balance (DID+EB) for the sample excluding COVID-related trials. Column (3) and (4) present the results using the sample excluding COVID-related MeSHs. Column (5) and (6) present the results for sample period from 2013 to 2019 before COVID-19. Control variables include the percentage of phase 1, 2, 3, 4 trials started, percentage of FDA, NIH, industry-sponsored trials per year per MeSH. MeSH fixed effect and year fixed effect are added. The sample composition is presented in Table 1 and variable definition in Appendix A. Standard errors are clustered at MeSH level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

	Number of trials initiated					
	Exclude COVID trials		Exclude COVID MeSH		Exclude COVID period	
	DID	DID+EB	DID	DID+EB	DID	DID+EB
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	0.074***	0.062**	0.072***	0.060**	0.062***	0.050**
	(0.027)	(0.029)	(0.028)	(0.029)	(0.022)	(0.023)
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs	4,770	4,770	4,730	4,730	3,346	3,346