From Free Rider to Innovator: The Rise of China's Drug Development

Panle Jia Barwick Hongyuan Xia Tianli Xia*

January 2026

PRELIMINARY: PLEASE DO NOT CIRCULATE

Abstract

This paper examines the transition of developing economies from pharmaceutical "free riders" to innovators, using China's recent ascent as a case study. Since 2020, China has surpassed the US in its annual volume of registered clinical trials. By linking global clinical trial registries with granular drug sales data and scholarly publication data, we provide evidence that this shift was primarily driven by the National Reimbursement Drug List (NRDL) reform. This policy shock increased the effective market size for innovative drugs by combining price negotiations with massive insurance expansion. We document a sharp rise in both the quantity (94% increase post-reform) and novelty of drug trials, with growth concentrated in reform-exposed disease categories, first or best in class drugs, and among domestic firms. A decomposition exercise shows that the NRDL reform accounts for 40% of the growth in trial activity in the Oncology sector, outweighing upstream knowledge accumulation and talent flows (23%), while other government policies seem minor. Our results suggest that strategic public healthcare purchasing can simultaneously improve patient access and achieve dynamic gains by incentivizing indigenous innovation.

---

*Barwick: Department of Economics, UW-Madison, NBER, and CEPR, pbarwick@wisc.edu. H. Xia: Department of Economics, Cornell University, hx276@cornell.edu. T. Xia: Simon Business School, University of Rochester, tianli.xia@rochester.edu. We are grateful to SinoHealth for generously providing the data. We thank Shizhe Yu and Congrong Pan for providing stellar research support and Rebekah Dix, Chris Snyder, Fiona Scott Morton, and Pierre Azouley for helpful comments.

1 Introduction

Since World War II, the global landscape of pharmaceutical innovation has been defined by a clear hierarchy. The United States has served as the undisputed leader in biomedical research, generating the vast majority of novel drug concepts and "first-in-class" therapies, with Europe and Japan serving as a strong second pillar of this ecosystem. In this traditional division of labor, developing nations have played a peripheral role, functioning primarily as consumers of existing medicines or producers of generic copies rather than creators of novel technologies.

This hierarchy aligns with the conventional economic wisdom of the "free-rider" hypothesis, which suggests that smaller or developing nations optimally rely on the diffusion of innovation from wealthy nations rather than investing in indigenous creation (Olson and Zeckhauser, 1966; Kyle, 2007; Lakdawalla, 2018; Garthwaite, 2025). Pharmaceutical R&D is characterized by high risks of failure and exorbitant fixed costs, with a Phase III trial costing, at the median, $21.4 million (Martin et al., 2017). For developing countries, where budgets are constrained and technological frontiers are distant, the expectation has been that the "innovation gap" between the frontier and the developing world would remain stable or even widen over time.

This paper challenges this conventional wisdom by documenting a structural shift in the global supply of medical innovation, where China has transitioned from a peripheral actor to a leading force in pharmaceutical and biomedical research over the past decade. This transformation is punctuated by a dramatic expansion in clinical trial activity: by 2020, China surpassed the US in the total number of annual registered clinical trials (Chen et al., 2024). Critically, this progress is not merely a quantity-driven expansion of low-value generic testing. Leveraging scientific novelty metrics, we document a sharp rise in the investigation of new molecular entities and novel biological mechanisms, particularly in high-burden therapeutic areas such as oncology. These patterns suggest that China is rapidly narrowing the "innovation gap" with global leaders and is poised to shape the future of global biopharmaceutical R&D.

To analyze this transition, we leverage the Citeline TrialTrove database, which provides a comprehensive census of global drug development projects. This dataset allows us to track

the lifecycle of clinical R&D with granular detail across therapeutic areas and firm types. Our identification strategy employs a difference-in-differences framework that benchmarks China's trajectory against the U.S., the global frontier in biomedical research, while utilizing Europe as an alternative counterfactual to ensure the robustness of our estimates. To disentangle the mechanisms underlying this rise, we integrate the trial registries with four complementary data sources: (a) hospital-level pharmaceutical sales from SinoHealth; (b) comprehensive records of scholarly publications and author affiliations from OpenAlex; (c) application backlog durations from the National Medical Products Administration; and (d) a complete archive of central policy documents issued by China's State Council.

All analyses are conducted at the disease–country–year level. Our econometric specifications include disease-by-country fixed effects to absorb time-invariant comparative advantages and year fixed effects to account for global temporal shocks. Furthermore, we incorporate disease-specific time trends and disease-by-year fixed effects in robustness analyses, which allow us to isolate country-specific deviations from global innovations in specific therapeutic areas.

Our analysis documents several stylized facts that challenge the "free-rider" hypothesis in global pharmaceutical R&D. First, we identify a sharp structural break in China's clinical trial activity beginning in 2017. By 2020, China's annual trial initiations surpassed those of the U.S. Second, this expansion is not merely a quantitative surge in low-value, "me-too" drug testing. Rather, we document a significant shift toward the global quality frontier, characterized by a sharp rise in high-novelty trials and new molecular entities (NMEs). Third, we demonstrate that this growth is driven almost entirely by indigenous innovative capacity. The surge is led by domestic Chinese firms rather than the geographic reallocation of multinational corporations (MNCs) pipelines. By 2024, China's annual trial volume achieved a 172% relative expansion compared to the U.S. and the pre-2016 baseline. Finally, the expansion of indigenous clinical activity extends well beyond trial initiation, manifesting in a sharp increase in global licensing transactions and foreign regulatory approvals for drugs developed by Chinese

firms. These findings suggest that China has transitioned from a passive consumer of global medical inventions to a major producer of pharma innovation, underpinned by a fast-growing R&D ecosystem shaped by domestic institutional incentives.

We argue that this structural transformation was not an incidental byproduct of aggregate economic growth, but rather the consequence of a specific policy shock: the National Reimbursement Drug List (NRDL) reform. By combining centralized price negotiations with a large-scale expansion in insurance coverage, the reform fundamentally altered the "effective market size" facing innovative pharmaceutical products. The NRDL reform has three distinctive features. First, unlike the broad insurance expansions commonly studied in the literature, NRDL inclusion is narrowly targeted at innovative drugs, and the scale of the reform is substantial: the cumulative government spending on innovative drugs reached ¥460 billion in 2024, while total public spending on pharmaceuticals exceeded 1.4 trillion RMB, equivalent to 2.2% of China's GDP. Second, the reform was implemented in a developing economy that had historically been a consumer, rather than a producer, of frontier pharmaceutical innovation. Third, the NRDL negotiation mechanism simultaneously compresses prices through centralized bargaining while expanding quantities via universal insurance coverage, rendering the ex-ante effect on expected revenues—and thus innovation incentives—theoretically ambiguous.

Our empirical strategy exploits the staggered implementation of this reform across disease categories to demonstrate that (1) overall, the quantity expansion dominates the price contraction and leads to revenue increase by more than 600% for covered drugs, and (2) therapeutic areas with greater exposure to quantity and revenue expansion experienced disproportionately larger increases in clinical trial activity. Crucially, in a placebo test of disease categories that never experienced NRDL expansion during our sample period, we show no systematic increase in trial activity after 2016. Furthermore, we document significant heterogeneity in firm responsiveness: domestic Chinese firms exhibit a much stronger response to the reform than their international counterparts. This divergence may be attributed to lower institu-

tional frictions or a greater ability to navigate the local regulatory and negotiation landscape (Javorcik and Wei, 2009). Our finding of a positive impact on both drug access and clinical activity suggests that a well-designed insurance framework can reconcile the tension between affordability and R&D incentives. This provides evidence that a country (including developing nations) can utilize public purchasing power to simultaneously improve patient access and strengthen domestic innovation. A back-of-the-envelope exercise shows that the induced R&D activity can more than double the consumer surplus gained via improved access in the short run.

Beyond demand-side incentives unleashed by NRDL, we account for alternative supply-side drivers, including upstream knowledge accumulation, international talent flows (specifically the return/migration of scientists from the U.S. and Europe to China), sector-specific industrial policies, and regulatory reforms that reduced clinical-trial approval latencies. A decomposition exercise reveals that the NRDL reform accounts for approximately 40% of the observed surge in trials in Oncology, while knowledge creation in the upstream contributes 23%. Other industrial and regulatory policies appear to play a comparatively modest role.

China's experience offers a valuable case study. It demonstrates that through deliberate institutional design, specifically the leverage of state monopsony power to expand market access while maintaining innovation incentives, developing nations can escape the "free-rider trap" and move beyond passive consumption to become significant contributors to the global stock of scientific knowledge, potentially generating positive international spillovers in biomedical research.

Related Literature Our analysis contributes to several strands of literature. First, it relates to the literature on market size and innovation. A large body of work based on U.S. data has established that expected revenue is a key determinant of pharmaceutical R&D, providing empirical support for Schmookler (1966)'s hypothesis. Acemoglu et al. (2006) demonstrate that exogenous market expansions driven by demographic shifts stimulate drug innovation,

followed by Blume-Kohout and Sood (2013)'s study on Medicare expansion, Yin (2008)'s study on orphan drug act, and Dubois et al. (2015)'s work on comparing market size and innovation in OECD countries. Our study provides the first evidence in a developing country context and highlights the importance of "effective market size" and the effects of a well-designed insurance policy.

Second, our study contributes to the literature on firm incentives in conducting pharmaceutical R&D. Consistent with prior studies (e.g., Budish et al. (2015), Hermosilla (2021), Agha et al. (2022), Alsan et al. (2024)), we find that firms respond strategically to market-based incentives by reorienting research efforts toward therapeutic areas with higher expected profitability. However, while some research suggests that demand-side shocks primarily pull incremental innovation (Dranove et al., 2022), the trials induced by the NRDL reform are frequently characterized by high scientific novelty. This likely reflects the specific criteria of NRDL, which places a premium on novelty/originality and unmet clinical needs.

Third, our paper contributes to the literature on industrial policy and its impact on innovation. Much of the existing research focuses on direct supply-side interventions, such as government subsidies and tax credits (Branstetter and Sakakibara, 2002; Bronzini and Piselli, 2016; Howell, 2017; Dechezleprêtre et al., 2023; Melnik and Smyth, 2024). In the pharmaceutical sector specifically, Azoulay et al. (2019) and Ward and Dranove (1995) show that public funding (such as NIH grants) can increase private-sector patenting and R&D investment, though Wallsten (2000) cautions that such grants may crowd out firm-financed research. Our analyses focus on demand-side policies, specifically public purchasing and insurance programs. The most notable example is Finkelstein (2004), which finds that government vaccine recommendation and policies have dynamic effects that stimulate R&D. While nearly all the evidence is from U.S., we extend this literature by studying a large-scale reimbursement reform implemented outside the U.S.. Our contribution is to show that state-led insurance expansion can not only stimulate pharmaceutical innovation at scale, but also systematically redirect innovative effort toward previously unmet clinical needs, rather than merely creating

"wasteful innovations".

Lastly, our study contributes to a growing literature documenting China's rapid technological ascent, while speaking directly to debates about the quality of this progress.[1] A number of recent studies emphasize the scale and breadth of China's advance: Leung et al. (2024) shows that China has attained leadership in nearly 90% of critical technologies as measured by high-impact scholarly output. At the same time, a complementary strand of the literature raises concerns about institutional constraints and mismeasurement. Branstetter et al. (2023) argues that narrow patent protection may limit the pace of Chinese innovation, while Chen et al. (2021) cautions that reported R&D growth may partly reflect relabeling responses to tax incentives rather than real inventive activity. Similarly, Wei et al. (2023) finds that China's flagship innovation policy (InnoCom) generated negative returns by encouraging a proliferation of low-quality patents. We contribute to this debate by providing the most comprehensive evidence to date on China's biopharmaceutical research through the lens of clinical trial activity. Leveraging granular novelty indices and decomposing the sources of growth, we show that the surge in clinical development reflects a genuine reorientation toward the global quality frontier, rather than an administrative artifact or a purely quantitative expansion.

The remainder of the paper is organized as follows. Section 2 details the institutional background of the Chinese healthcare system and describes the data. Section 3 analyzes both the quantity and quality of clinical trials in China relative to the U.S., as well as downstream business development and drug licensing activities. Section 4 investigates the underlying mechanisms and discusses the policy implications. Section 5 concludes.

[1]For recent debates in the public media, see NYT (Aug 17,2025), Washington Post (Sep 22, 2025), and Bloomberg (June 10, 2025).

2 Institutional Background and Data

2.1 Drug Development Process

The regulatory process governing drug development and market authorization in China and the US is similar. In both jurisdictions, a drug sponsor must demonstrate that a product is safe and effective through a series of randomized clinical trials, typically benchmarking the investigational drug against a placebo or the prevailing standard of care (National Institutes of Health, 2017).

The drug development lifecycle begins with the identification of a lead compound. To protect the resulting intellectual property, sponsors generally file patent applications prior to the commencement of human testing. Before clinical trials begin, sponsors must submit an Investigational New Drug (IND) application, which provides comprehensive pre-clinical evidence on the compound's toxicity and pharmacological profile.

Clinical evaluation proceeds through three main stages before sponsors may apply for marketing approval. Phase I trials assess basic safety, dosage ranges, and potential side effects, typically enrolling a small group of healthy volunteers. Phase II and Phase III evaluate efficacy, continue to monitor safety, and benchmark the drug against existing therapies. Phase III studies, often involving thousands of subjects, and conducted in a blinded and randomized setting, serve as the "pivotal" trials required for regulatory approval. After approval, sponsors (and other public organizations) may conduct Phase IV studies, which are usually observational and focus on longer-term safety and real-world effectiveness. These trials represent the most capital-intensive phase of the R&D pipeline. On average, Phase I trials cost approximately $3.4 million, and Phase III trials cost roughly $21.4 million (Martin et al., 2017). In addition, trials involve high risks of failure. In our data, each new drug discovery requires an average 123 trials, and only 4 percent of the projects lead to a successful drug. Therefore, the trials and the related financial incentives are key considerations for pharmaceutical innovation.

2.2 The Chinese Healthcare System

The Chinese healthcare system (CHS) provides nearly universal coverage to over 95% of the population, making it the world's largest public health insurance program (Yu, 2015). Established in 1999, it has since evolved into two main schemes: the urban employee basic medical insurance (UEBMI), which covers formal-sector employees and retirees, and the urban and rural resident basic medical insurance (URRBMI), which covers the remaining population, including children and the elderly.

As the dominant purchaser in the domestic market, the CHS has the ability to exert significant monopsony power. Prescription drug spending constitutes approximately 40% of total CHS expenditures. In 2022, the system spent over ¥1.4 trillion ($200 billion) on pharmaceuticals (National Health Commission of the PRC, 2023). In contrast, the private commercial insurance market remains underdeveloped, accounting for less than 7% of total health expenditures. Consequently, inclusion in the National Reimbursement Drug List (NRDL) is the primary gateway to the Chinese pharmaceutical market. Historically, the NRDL prioritized low-cost generics and off-patent molecules due to budget considerations, effectively excluding innovative, high-priced therapies and forcing patients to bear the full cost out-of-pocket.[2]

The NRDL Reform Starting in 2016, the National Healthcare Security Administration (NHSA) launched a reform to integrate innovative therapies into the insurance catalog through centralized price negotiations. This ongoing reform represents a fundamental shift from a rigid, budget-capped list to a dynamic, value-based reimbursement model.

The negotiation process follows an annual cycle. Each spring, NHSA defines an "eligible list" of drugs, targeting innovative drugs (obtaining market approval within the last five years) that address significant unmet clinical needs.[3] Firms with eligible drugs and NHSA exchange information on costs and medical effectiveness during the pre-negotiation stage, which can last

[2]For the 2017 NRDL list, see: NRDL (2017 version).

[3]For the 2024 eligibility criteria, see Inclusion Criteria: 2024 version.

for months. On the day of negotiation, government representatives negotiate with delegates from each drug company simultaneously and independently. Successful negotiations result in inclusion in the NRDL, typically involving average price reductions of 50–60% (Barwick et al., 2025). If negotiations fail, the drug remains on the private market with significantly lower volume. The reform has a huge impact on the drug market and has successfully included 949 drugs in NRDL by 2025, and the accrued spending on innovative drugs has reached ¥460 billion ($66 billion).[4]

From an economics perspective, the reform introduces a price-quantity trade-off. On one hand, firms face deep, centralized price cuts that compress profit margins. On the other hand, successful inclusion triggers a massive expansion in effective market size through nearly universal insurance coverage and low coinsurance rates. Whether this reform encourages or deters innovation depends on whether the quantity effect outweighs the price compression. In analyses below, we exploit the staggered entry of therapeutic categories into the negotiation rounds to identify the impact of this policy shock on the direction and quality of R&D.

2.3 Data

Our empirical analysis integrates four primary data sources with several auxiliary datasets to track the lifecycle of pharmaceutical innovation.

Citeline TrialTrove Database Clinical trials constitute the most critical and costly stage of drug development and directly determine whether a drug project can advance toward regulatory approval. We utilize the Citeline TrialTrove database (2010–2024) to construct a comprehensive census of clinical trials initiated globally. TrialTrove aggregates data from official national registries (e.g., ClinicalTrials.gov in the U.S. and the ChinCTR in China) alongside public disclosures. For each trial, we observe the tested drug, sponsor (private firm vs. public institution), therapeutic area, disease/indication, trial start date, completion sta-

[4]Source: 2025 National Forum on Medical Insurance & Innovative Drug Development.

tus, trial phase, objective, methodology, duration, and trial population. Table 2 reports the number of trials across disease categories and phases in China and the United States.

Citeline Pharmaprojects Database To capture the commercial realization of R&D, we link trials to the Citeline Pharmaprojects database, which tracks key development milestones including patenting, out-licensing, and regulatory approvals(Cunningham et al., 2021). These events serve as proxies for the "quality" of trials, as licensing deals, particularly acquired by the multinational firms, reflect a market recognition of the underlying innovation.

SinoHealth Drug Sales Data We obtain granular pharmaceutical sales data from SinoHealth (2016–2024), a leading health-data provider in China (Xia, 2024; Barwick et al., 2025). Unlike other databases, SinoHealth provides superior coverage of both urban and rural hospital channels, through which nearly all innovative drugs in China are dispensed. The data are at the province-quarter-SKU level and report quantities dispensed (years of supply) and retail prices at the SKU (stock keeping unit) level.

OpenAlex Publications To measure upstream knowledge creation, we utilize OpenAlex, an open, large-scale scholarly database that systematically tracks global scientific publications and the related metadata. It provides comprehensive information about authors, citations, institutions, journals, and concepts (i.e., keywords). We focus on biomedical publications with valid PubMed IDs, where the first and last authors are affiliated with either the Chinese or U.S. institutions.

Auxiliary Dataset: IND Application and Approval Dates The 2015 clinical trial application reform in China streamlined regulatory procedures by clearing the backlog of pending IND (investigational new drug) applications and is therefore likely to have boosted clinical trial activity. To measure the backlog days between application and approval, we collect the application and approval dates for all IND applications from the Center for Drug Evaluation

of China's NMPA. We further assign each IND to a disease so that we can calculate backlog days for each disease subcategory in our data. The data scraping process is documented in Appendix A.5.2.

Auxiliary Dataset: Policy Documents We have scraped all biopharma-related policies in China from the State Council Policy Document Library. We identify 830 policy documents targeting the biopharmaceutical sector. We use GPT to classify these documents into four categories: (a) policies that promote advanced manufacturing; (b) innovation policies that reduce reliance on foreign technologies; (c) capital market reform; and (d) fiscal subsidies. The data scraping process is documented in Appendix A.5.3.

Data Contributions A significant contribution of our data work is linking these disparate sources systematically. Appendix A.4.1 details the procedures undertaken to link drug names in the SinoHealth sales data with the disease categories in the Citeline databases, including GPT-based semantic mapping and manual verification. Appendix A.4.2 explains the graph embedding approach we took to link keywords in OpenAlex scientific studies and disease categories in Citeline. By representing keywords and disease categories as points in a high-dimensional vector space, we identify semantically similar clusters of keywords and diseases. This allows us to systematically identify the keywords in OpenAlex publications that are most closely related to each Citeline disease category.

Another significant contribution of our data work is the development of four distinct metrics to quantify clinical trial novelty, addressing a central challenge in the economics of innovation: how to distinguish between frontier R&D and incremental ("me-too") development. Following industry standards, we first identify the biological "target", e.g., the specific protein, gene, or enzyme the drug is designed to modulate. We calculate a novelty score based on the target's historical prevalence in the global R&D pipeline; trials targeting previously unexploited biological pathways are classified as higher-novelty. Second, we employ a fine-tuned Large Language Model to categorize trials based on their functional objectives, distinguishing

between New Molecular Entities (NMEs), the highest tier of innovation, and other trials (such as trials that test for novel drug combinations or new indications of existing therapies). Third, we extract information on the trial's control group. Trials that use innovative drugs as controls test against the "best-in-class" standard and therefore reflect a higher quality benchmark than trials that use placebos or generics. Finally, we highlight trial speed as an important dimension of quality. Faster trials indicate greater efficiency, shorten the time from development to commercialization, and ultimately improve consumer welfare. These measures allow us to more precisely track China's movement toward the global technological frontier. Appendix A.3 provides details about how these novelty measures are constructed.

3 Empirical Results

Our empirical analysis begins by comparing clinical trial and drug development trajectories in China with those in the U.S.. We utilize the U.S. as our primary comparison group for three reasons. First, the U.S. serves as a proxy for the time-varying global innovation frontier. Since scientific knowledge is a global public good, the supply of new drug targets is simultaneously available to all countries. Using the U.S. as a control allows us to isolate China-specific changes in innovation activities from global technological shifts. Second, the U.S. provides a direct test of the "free-rider" hypothesis. If China were merely a follower, we would expect a stable or widening gap between the two nations. A narrowing gap, or a structural break, suggests a transition toward indigenous innovation. Third, the U.S. represents a stable baseline of a mature market, as its regulatory and reimbursement environments remained comparatively constant during our sample period. In robustness checks, we also use Europe and US + Europe as alternative control groups.

3.1 Time Trend of Trials and Drugs

We begin our empirical analysis by documenting the aggregate trends in clinical trial initiations in China relative to the U.S.. Beyond simple counts, we investigate whether this growth represents a move toward the global technological frontier or merely a quantitative expansion. We further decompose these trends by sponsor nationality to separate indigenous capacity building from the geographic reallocation of multinational (MNC) activity.

R&D Volume Figure 1 illustrates the time trends in clinical trial activities in China and the U.S. from 2010 to 2024. Panel (a) shows that while trial activities in the U.S. remained relatively stable, China underwent a dramatic structural shift. Starting from a baseline of fewer than 1,000 annual trials in 2010, China's volume surged post-2016, ultimately surpassing the U.S. in 2020. By 2024, China initiated more than 5,000 clinical trials annually.

R&D Novelty Because a volume-based comparison may be confounded by routine or low-value R&D, Panels (b) and (c) utilize two complementary measures to evaluate trial novelty. First, we employ a novelty index that quantifies how often the tested drug's mechanism has been explored (i.e., "target") following Dranove et al. (2022). We classify trials above median of this index as "high-novelty" projects.[5]

Second, we utilize a fine-tuned GPT-4o mini model to perform a functional classification of trial objectives based on textual registry data. This classification distinguishes between (i) new molecular entities (NMEs), (ii) novel drug combinations, (iii) new indications for existing molecules, and (iv) bioequivalence tests for generic drugs.[6] This dual-measurement approach

[5]To be more specific, Dranove et al. (2022) compute novelty of a trial i with mechanism $m(i)$ at stage $j(i)$ as:

$$Novelty_i = \frac{1}{1 + \text{Number of prior trials at stage } j(i) \text{ using mechanism } m(i)}.$$

This measure distinguishes drugs with mechanisms rarely explored before from those relying on well-established pathways. We calculate the novelty median within each disease subcategory and year, pooling observations across all countries to benchmark novelty against the global clinical development frontier rather than country-specific baselines.

[6]We manually labeled a training set of trials and used the model to classify the remaining trials. Fine-

ensures that our findings are benchmarked against the global clinical development frontier.

The results in Panel (b) of Figure 1 demonstrate that China's expansion is not limited to incremental projects. The volume of high-novelty trials rose from fewer than 400 in 2016 to over 2,400 by 2021. Furthermore, Panel (c) confirms that the surge is primarily driven by trials investigating novel mechanisms, new combinations, or new indications, while generic bioequivalence testing accounts for a diminishing share of the total growth. Collectively, these patterns indicate that China's rapid catch-up encompasses both the quantity and the scientific quality of innovative clinical development.

Domestically-led R&D A natural question that follows from these aggregate trends is whether the rapid acceleration in clinical activity is driven by domestic Chinese firms or by multinational (MNC) incumbents operating within China. Distinguishing between these two groups is essential because growth led by Chinese firms reflects an expansion of China's indigenous innovative capacity, whereas a surge led by international firms might simply represent a geographic reallocation of global R&D activities without local capability building.[7]

Figure 2 compares the trial initiations by sponsor nationality. Panel (a) focuses on trials conducted in China and shows that the post-2016 expansion is driven almost exclusively by domestic firms. In contrast, international firms account for only a small fraction of total trials in China, and their activities remained relatively stable over the sample period.

Panel (b) examines the presence of Chinese sponsors in trials conducted in the U.S.. While their initial volume was limited relative to MNCs, we document a steady increase in Chinese-sponsored trials in the U.S. after 2016, peaking around 2021. This trajectory suggests a nascent "internationalization" of Chinese pharmaceutical firms, though the aggregate magnitude of their overseas activity remains modest. Overall, Figure 2 shows that the surge in

tuning reduces hallucination risk and enables consistent large-scale categorization of trial novelty. Details on the construction of these two novelty measures can be found in Appendix A.3.

[7]The divergence between domestic and foreign firms may be attributed to lower institutional frictions or a greater ability to navigate the local regulatory and negotiation landscape (Javorcik and Wei, 2009).

China's clinical trial volume is a domestic phenomenon led by indigenous firms, rather than a byproduct of shifting multinational R&D strategies. Note that the relatively modest increase in U.S.-based trials initiated by Chinese firms does not imply a lack of innovativeness. Conducting clinical trials in the United States, particularly at early exploratory stages, is costly, and Chinese firms face additional institutional and regulatory frictions when operating overseas. As a result, Chinese firms commonly enter the U.S. market through out-licensing arrangements or collaborations with multinational pharmaceutical companies rather than directly sponsoring trials.

R&D outcomes Whether this increased investment in trial activity translates into progress at later stages of the drug development lifecycle is a non-trivial question. Given that the majority of observed trials were initiated within the last decade and the average pharmaceutical development cycle spans 8–12 years, many long-run product-market outcomes remain unobserved. To partially bridge this gap, we examine intermediate drug-level milestones recorded in the Citeline Pharmaprojects database. In particular, we analyze license-out transactions, which serve as a market-based valuation of a drug's promise, and new drug approvals (NDAs), which represent the final regulatory validation of clinical success and the prerequisite for market entry.

Figure A.2 compares these downstream milestones for China and the U.S.. Panel (a) shows a gradual rise in license-out events for Chinese-originated drugs starting several years after 2016, a pattern consistent with the expected lag between early-stage trials and subsequent commercialization activity. Panel (b) reveals a more pronounced acceleration in new drug approvals of Chinese-originated drugs, which rose from fewer than five prior to 2016 to over forty by 2024. Panel (c) displays a sharp increase in the number of new drugs approved in China, rising from fewer than 20 before 2016 to around 100 by 2024, surpassing the corresponding number in the U.S.[8] These findings are corroborated by an industry report indicating that

[8]For each drug, we only count the first approval anywhere in the world.

China has overtaken the EU4+UK in total new drug approvals over the past five years, trailing only the U.S. (IQVIA Institute for Human Data Science, 2024).[9] If the trend continues, China may surpass the U.S. in new drug access in the near future. These trends indicate that China's catch-up is not confined to trial initiations. The rise in licensing and approvals indicates that the increased R&D effort is beginning to reshape both the scientific pipeline and the commercialization landscape of drug development.

3.2 Structural Break Test

The descriptive evidence suggests a notable inflection point in China's clinical trial activity relative to the U.S. during the mid-2010s. To identify the timing of this shift, we employ a structural break test following the framework established by Quandt (1960) and Andrews (1993). Specifically, we implement a supremum Wald (Sup-W) test, which identifies an unknown break date by maximizing the Wald statistic across a series of Chow tests for all possible partitions of the sample period.

The test yields a clear global maximum in 2017 (Wald statistic = 23.57, p-value = 0.00), providing robust statistical evidence of a regime shift in China's trial output. Since a break identified in 2017 reflects a divergence in the data-generating process beginning that year, we designate 2016, the final year of the pre-break regime, as the reference period ($t = 0$). This data-driven identification provides the empirical basis for the treatment period used in the event-study and difference-in-differences specifications presented in the following section.

3.3 DID and Event Study

The descriptive evidence suggests a sharp expansion in China's clinical trial activity relative to the U.S. post 2016. To formally quantify the magnitude of this shift, we exploit the granular nature of our micro-data by estimating regressions at the disease–country–year level. This

[9]Over the past five years, the U.S. approved 267 new drugs, China approved 192 drugs, and EU4+UK approved 187 drugs. And many drugs approved China are first-in-world and first-in-class.

high-frequency, disaggregated approach is motivated by three considerations. First, drug development is frequently catalyzed by global scientific breakthroughs that are disease-specific, such as the emergence of checkpoint inhibitors in oncology. Our framework accounts for these time-varying, disease-specific global shocks that might otherwise confound the estimates. Second, a disaggregated analysis addresses potential compositional differences. If China's growth were merely driven by "low-hanging fruit", such as me-too drugs in the crowded antibiotics sector, a country-level aggregate would not distinguish between increases in trial volume from changes in the R&D mix. Finally, the therapeutic area serves as the economic unit of decision-making, as pharmaceutical firms typically allocate R&D budgets and manage pipelines along these specific disease classifications.

Aggregate the data to disease-country-year level, we estimate the following difference-in-differences specification, using 2016 as the baseline year:

$$Y_{d,c,t} = \beta_c \mathbf{1}(\text{China})_c \times \text{Post2016}_t + \phi_{d,c} + \xi_t + \epsilon_{d,c,t}, \tag{1}$$

where the dependent variable is the log number of trials (or drug projects) for disease d in country c and year t. Consistent with the results of our structural break test, Post2016_t is an indicator equal to one for all years from 2017 onward. We include disease-by-country fixed effects ($\phi_{d,c}$) to control for time-invariant factors, such as country-specific comparative advantages in certain therapeutic areas. Year fixed effects (ξ_t) absorb global temporal trends in drug development. These fixed effects ensure that identification comes from relative changes within disease categories after 2016 rather than shifts in disease composition or global trends. Under the assumption of parallel trends, which we test formally in the following section, β identifies the differential growth in China's clinical output.

Other specifications are implemented as robustness checks. To account for the count nature of the data and the prevalence of zeros, we complement the OLS estimates with Poisson Pseudo–Maximum Likelihood (PPML) regressions. As a more demanding robustness exercise,

we further augment this specification with disease-specific linear trends and disease-year fixed effects. These specifications ensure that our estimates are not confounded by time-varying, disease-specific global shocks—such as surges in worldwide investment in immunology—that may coincide with shifts in China's domestic trial activity.

Table 3 reports the baseline difference-in-differences estimates. Across all specifications, the coefficients are positive and statistically significant at the 1% level, indicating a substantial acceleration in China's clinical trial output relative to the U.S.. The estimate in Column 1 implies that the total number of trials in China increased by approximately 94% ($e^{0.66} - 1$) post 2016.

Beyond volume, we find evidence of a corresponding surge in the quality and novelty of these trials. Columns (2) and (3) show that high-novelty trials (defined by the above median trials according to Dranove et al. (2022)'s measure) and trials investigating non-generic compounds rose by 65% and 67%, respectively. These results indicate that China's post-2016 growth reflects a broad-based strengthening of clinical development capabilities across therapeutic areas, rather than a concentrated shift toward low-value, imitative drug projects..

Figure 3 plots the event-study coefficients underlying the DID results in Table 3. It provides a visual test of the parallel-trends assumption and quantifies the dynamic effects following the structural break. The estimated coefficients prior to 2016 are small in magnitude and statistically indistinguishable from zero, suggesting that China and the U.S. followed comparable growth trajectories in clinical trial activity during the pre-treatment period, as shown in Panel (a). Moreover, we assess the sensitivity of the parallel-trends assumption using the approach proposed by Roth (2022). Figure A.1 shows that the post-treatment effect remains statistically significant for moderate deviations from parallel pre-trends.

Beginning in 2017, we observe a clear divergence. By 2024, the relative increase reaches approximately 1.0 log point, representing a 172% expansion in China's trial volume relative to the U.S.. Panels (b) and (c) reveal analogous patterns for high-novelty and non-generic trials:

both series exhibit flat pre-trends followed by significant post-break growth of 0.8 log points. These patterns suggest that the expansion in China's clinical activity is not a transitory shock but a shift to a new steady state with higher intensity and more innovative R&D.

To investigate the drivers of this expansion, Figure 4 decomposes China's trial activity by firm origin. This analysis allows us to distinguish between a surge driven by the growth of domestic innovative capacity and one driven by multinational corporations (MNCs) offshoring clinical trials to the Chinese market.[10] The event-study estimates for both groups exhibit stable, near-zero pre-trends, further validating our empirical design. However, the post-2016 trajectories diverge sharply: while the relative growth for domestic Chinese firms climbs to approximately 1.2 log points by 2024, the estimates for international firms remain relatively flat. This asymmetry suggests that the post-2016 acceleration is primarily an endogenous phenomenon driven by domestic firms, rather than a result of global firms reallocating their R&D pipelines to China.

Finally, Figure A.3 examines the transmission of China's clinical surge to downstream drug development outcomes, specifically license-out transactions and new drug approvals. As shown in Panels (a) and (b), both indicators exhibit a marked upturn beginning around 2021. This timing is consistent with the standard gestation period in the pharmaceutical industry, reflecting the necessary lag as the surge in early-phase trial activity matures into late-stage pipeline milestones. The coefficients display greater variance than the trial-level estimates, given the relative sparsity of licensing and approval events at the disease-year level. Nevertheless, the distinct post-2016 upward trajectory provides evidence that China's clinical expansion is yielding tangible innovative outputs, suggesting that the initial increase in R&D effort is successfully translating into commercial and regulatory success.

[10] The estimation equation is: $Y_{d,c,t} = \beta_{c_1} \mathbf{1}(\text{ChineseFirmInChina})_{d,c} \times Post2016_t + \beta_{c_2} \mathbf{1}(\text{InternationalFirmInChina})_{d,c} \times Post2016_t + \phi_{d,c} + \xi_t + \epsilon_{d,c,t}$.

3.4 Heterogeneity and Robustness

In this subsection, we investigate heterogeneity across trial stages and whether a trial is sponsored by private or public entities, followed by a series of robustness checks. These results help inform the mechanisms analysis in the subsequent section.

Figure A.4 plots event-study estimates for clinical trials by development phase following 2016. We observe an immediate and statistically significant increase in Phase I trials, whereas the rise in Phase II and Phase III trials emerges only after a lag of approximately two to three years. This timing pattern is consistent with new drug development timelines, in which early-stage discoveries require several years to progress to later phases. The delayed response in advanced-phase trials suggests that the post-2016 expansion reflects the entry of genuinely novel projects rather than the sheer resumption of previously suspended or terminated pipelines, which are likely to be more marginal.

Table A.1 splits trials by stages and sponsor types. Trial activity increased across all stages, with Phase I trials registering the highest increase ($e^{0.703} - 1 = 102\%$ annual growth), followed by Phase II ($e^{0.353} - 1 = 42\%$) and Phase III ($e^{0.177} - 1 = 19\%$). This ordering is consistent with the high risk of failure and attrition inherent in the trial process.

The patterns reveal significant variation in magnitudes that align with the comparative advantages of different institutions. Private sponsors drove the largest increases in Phase I and Phase III trials, stages most critical for establishing safety, efficacy, and securing regulatory approval. In contrast, public sponsors exhibited their strongest growth in Phase IV trials. Phase IV trials focus on post-approval observational studies and real-world evidence generation, which are less commercially driven. These patterns reflect a structural division of labor: private firms prioritize high-stakes, market-entry R&D in early and pivotal phases, while public institutions focus more heavily on post-market surveillance and broader clinical evidence generation following a drug's commercial launch.

We further evaluate the sensitivity of our baseline estimates to alternative specifications and sample definitions. Table A.2 assesses whether the results are invariant to different fixed-

effect structures and functional form choices. In Column (2), we augment the baseline model with disease-by-year fixed effects, which control for time-varying, disease-specific shocks at the global level. To ensure that our results are not driven by the log transformation of the dependent variable, Columns (3) and (4) report estimates using the raw count of trials as the outcome, employing both OLS and Poisson Pseudo–Maximum Likelihood (PPML) estimators. Across all specifications, the post-2016 upstick remains robust to functional form assumptions and fixed-effect choices.

In Table A.3, we address potential concerns regarding the DID design. Because the globalization of R&D can lead to multicenter trials, to avoid any spillovers across the treated and control groups, Column (2) restricts the treatment group to trials conducted exclusively in China, ensuring the effect is not driven by increasing numbers of global trials including a Chinese site. Furthermore, Columns (3) and (4) verify that our results are not sensitive to the choice of the U.S. as the control group by using Europe (and U.S. + Europe) as alternative control groups. The consistency of the coefficients across these samples suggests that our findings reflect a distinct shift in the treatment group (China) rather than idiosyncratic trends in the control group (U.S.).

Collectively, these results provide the first empirical evidence of a developing economy transitioning into a global frontier in pharmaceutical innovation. In the next section, we turn to the institutional and policy drivers behind this growth, assessing the extent to which this transition was catalyzed by deliberate regulatory reforms versus broader economic and demographic shifts in a growing and aging society.

4 Mechanisms

In this section, we investigate the institutional and economic drivers behind China's rapid ascent in pharmaceutical R&D. First, we examine the National Reimbursement Drug List (NRDL) reform in 2016, which dramatically expanded the market for innovative drugs. We

trace the causal link from this demand-side shock to the subsequent surge in clinical trial activity, documenting how increased expected returns catalyzed R&D investment. Second, we evaluate the role of knowledge creation and human capital, analyzing how talent inflows and domestic knowledge accumulation served as essential upstream inputs for innovation. Third, we study the 2015 clinical trial application (CTA) reform that cleared a substantial backlog of pending applications and accelerated review timelines (Jia et al., 2024). Finally, we assess complementary industrial policies to determine the extent to which broader state-led initiatives and institutional changes account for the R&D growth. We conclude this section with a decomposition exercise to quantify the relative importance of each driver.

4.1 The NRDL Reform

The NRDL Reform and Market Expansion The National Reimbursement Drug List (NRDL) reform possesses three features central to our analyses. First, the staggered roll-out of reimbursement coverage across drug classes provides rich temporal and cross-sectional variation. This gradual expansion allowed firms to update their expectations regarding the long-term profitability of the Chinese market and enables us to construct a treatment intensity measure. Second, the scale of the commitment is both massive and selective. With over ¥100 billion allocated to negotiated drugs in 2024 alone, the reform represents one of the largest public expansions of pharmaceutical coverage globally, comparable in fiscal magnitude to the introduction of Medicare Part D in the U.S.. Unlike Medicare expansion, which is widespread across drug categories, NRDL reform has strict eligibilty criteria targeting novel drugs that recently gained market approval, and favor high-quality drugs with large overall survival gains (Barwick et al., 2025).[11]

Third, the policy represents a sophisticated exercise of government monopsony power.

[11] Anecdotal evidence and government townhalls indicate that the government is aware of NRDL's potential (advantageous) effect on innovation: "In 2024 alone, fund spending on negotiated drugs during the agreement period has already exceeded ¥100 billion, which shows that the healthcare system is providing real, tangible financial support for innovation." Source: Press Conference on the National Healthcare Security Administration's Measures to Support the High-Quality Development of Innovative Drugs, July 2025.

Rather than relying on a laissez-faire approach or direct R&D subsidies, the Chinese government utilizes "price-for-volume" negotiations. This institutional design seeks to resolve a classic market failure: the trade-off between static efficiency (patient access through lower prices) and dynamic efficiency (innovation incentives through firm profits). By leveraging its buyer power to guarantee volume in exchange for price concessions, the state aims to find a "sweet spot" that expands access while preserving, or even enhancing, the net present value (NPV) of innovative drug projects.Therefore, although the demographics and income follow a smooth trend, the reform can create a jump in an "effective market size" via the group purchasing power.

To evaluate the strength of this shock, we first estimate the direct effect of NRDL inclusion on drug sales using drug-quarter level observations. We follow Callaway and Sant'Anna (2021) to address the staggered-rollout nature of the reform. Figure 5 decomposes the impact on the logarithm of retail prices, the logarithm of quantities, and the logarithm of revenue. Following successful negotiation, retail prices drop by an average of 50% (and 66% for oncology drugs). However, this price compression is more than offset by a massive market expansion: quantities sold increase by 350% on average, with cancer drugs seeing a nearly ten-fold (900%) increase in volume.

The net effect on producer revenue (and profit) is overwhelmingly positive. Panels (e) and (f) show a 100% average increase in revenue post-negotiation, with oncology drugs experiencing a 500% surge. Given that marginal production costs for drugs are typically low (Barwick et al. (2025) estimates an approximately 18% of the pre-negotiated price), these revenue gains translate into a significant increase in producer surplus. These results confirm that the NRDL reform fundamentally altered the profitability of the Chinese pharmaceutical market, providing the financial impetus for the clinical trial surge documented in Section 3.

NRDL's Effect on Clinical Trials To formally test whether the NRDL-induced market expansion catalyzed the surge in clinical activity, we construct disease-specific measures of

policy exposure. By merging the Citeline TrialTrove dataset with granular hospital-level sales data from SinoHealth (2017–2024), we exploit variation in the intensity of reimbursement coverage across 312 distinct disease categories. This allows us to quantify the effective "market size" effect on innovation, a central theme in the economics of R&D (Finkelstein, 2004; Acemoglu et al., 2006; Kremer and Snyder, 2018).

We define four complementary measures of disease-level NRDL exposure to proxy for the size of effective market expansion: (a) the log number of covered drugs, (b) the log revenue of covered drugs, (c) the share of covered drugs within the disease category, and (d) the revenue share of covered drugs. We estimate the relationship between reimbursement exposure and clinical activities using the following specification. We assume the number of trials in disease category d in country c at time t is given by:

$$log(\text{Trials})_{d,c,t} = \beta \text{NRDL Exposure}_{d,c,t} + \phi_{d,c} + \xi_t + \epsilon_{d,c,t} \tag{2}$$

where the coefficient β identifies the semi-elasticity of trial activity with respect to NRDL-driven market expansion.

Table 4 reports these results with two key findings. First, disease categories with greater NRDL exposure experienced significantly larger increases in clinical trial activity across all four measures. Both the extensive margin (coverage or not) and the intensive margin (the scale of coverage) boost R&D activity. Second, the magnitudes are economically significant. The coefficient in Column (3) implies that a 100% increase in NRDL-covered drug revenue is associated with a 17.8% increase in clinical trials ($e^{0.164} - 1$). This estimated elasticity is comparable to the findings of Dubois et al. (2015), which document the responsiveness of drug development to market size using variations in market sizes across diseases in OECDs.[12]

To mitigate concerns that our results reflect broad national trends rather than the NRDL specifically, Figure 6 presents event-study estimates stratified by treatment status. Panel

[12]Note the elasticity is much smaller than documented in Acemoglu et al. (2006) (≈ 4.0) using demographic shifters in U.S., likely because the huge absolute value of U.S. market size compared to China or Europe.

(a) conducts a falsification test on disease categories that never received NRDL coverage during our sample period. To ensure a clean control group, we exclude oncology and infectious diseases, as the former is subject to high expectations of future coverage and the latter was confounded by the COVID-19 pandemic. As expected, these "never-treated" categories show no systematic increase in trial activity after 2016. In contrast, Panel (b) reveals a sharp post-2016 divergence for disease categories that experienced NRDL expansion, mirroring the baseline results in Figure 3. This bifurcated pattern—showing minimal expansion in non-reimbursed therapeutic areas alongside a pronounced surge in reimbursed ones—strongly suggests that the NRDL reform played a central role in reorienting China's pharmaceutical R&D toward innovative development.

There is considerable heterogeneity in the scale of NRDL expansion across therapeutic areas, which allows for a more granular examination of how shifts in market size influence the direction of R&D. Figure 7 visualizes this relationship by plotting the change in clinical trial activity against the expansion of NRDL coverage (the number of drugs included in NRDL) for each disease subcategory between 2016 and 2024. Panel (a) displays the raw data, while Panel (b) plots the estimated increase in trials attributable to NRDL exposure, using the elasticity of 0.164 from Table 4: $\Delta\text{Trials}_d = \sum_{t=2016}^{2023} \text{Trials}_{d,t} \times 0.164 \times \frac{\text{NRDL}_{d,t+1} - \text{NRDL}_{d,t}}{\text{NRDL}_{d,t}}$.

The disease categories with the most pronounced expansion in NRDL exposure include non–small cell lung cancer, liver cancer, and renal cancer, diseases that also witnessed some of the sharpest increases in trial volume (Panel (a)). As shown in Panel (b), the upward-sloping fitted line reveals a strong positive association between reimbursement expansion and R&D response. The fact that the regression-based estimates closely track the raw data suggests that our baseline results are not driven by functional form assumptions or outliers, but rather reflect a robust cross-sectional relationship where innovation follows the expansion of the effective market size.

These patterns underscore the role of government reimbursement in overcoming underinvestment in R&D. When demand for innovative drugs is relatively inelastic among high-

income consumers but elastic among the broader population, private market pricing may lead to a sub-optimal volume of innovation. In this context, institutional intervention through "price-for-volume" negotiations can effectively expand the market, shifting the demand curve outward and restoring incentives for innovation (Kremer and Snyder, 2018).

Importantly, our results suggest that the NRDL reform influences not only the scale but also the direction of pharmaceutical innovation. By disproportionately incentivizing R&D in therapeutic areas with high unmet needs and large patient populations, the reform has steered clinical activity toward novel drug development. This reorientation represents a critical second-order benefit of the policy: while the immediate impact of the NRDL is to increase static efficiency through expanded patient access and consumer surplus, its long-run impact is to enhance dynamic efficiency by catalyzing the development of the next generation of therapies. We quantify the size of the long-run benefit in Section 4.5.

4.2 Knowledge Creation and Talent Flow

A competing hypothesis is that China's clinical surge was driven by a supply-side expansion in innovation inputs, specifically the accumulation of scientific knowledge and the return of high-skilled talent from abroad. Existing literature suggests that the stock of public science provides the "fertile soil" for private-sector R&D (Azoulay et al., 2019; Bloom et al., 2013), and that clinical trial activity is highly responsive to new scientific opportunities (Budish et al., 2015). Furthermore, high-skill scientist mobility has been shown to enhance invention, with immigration, diaspora networks, and return migration facilitating cross-border diffusion of ideas (Kerr, 2008). In the Chinese context, global integration has been linked to the country's surge in patenting (Hu and Jefferson, 2009).

To test this supply-side mechanism, we leverage the OpenAlex database to construct disease–year–level measures of knowledge production and talent flows. We proxy for knowledge quantity using total PubMed-indexed publications and for knowledge quality using the count of highly cited papers (top 5%). We also measure talent flows by identifying "returnees":

scientists with at least three publications affiliated with U.S. institutions who subsequently published under a Chinese affiliation.

Figure 8 presents the temporal trends of these indicators for China and the U.S.. While China's scientific output has undeniably matured, the timing does not align with the sudden inflection point observed in clinical trials. Panel (a) shows a steady, secular increase in publication volume since 2010, with no detectable structural break around 2016. Similarly, Panels (b) and (c) demonstrate that the production of "frontier" knowledge (the number of top 5% most-cited papers authored by researchers with Chinese affiliations) and the inflow of U.S.-trained talent have followed a smooth, long-term upward trajectory. These results suggest that while China's scientific capacity is deepening, these supply-side factors lack the discontinuity required to explain the post-2016 acceleration in trial activity.

We formally evaluate the relative contribution of these factors in Table 5 by estimating a "horse-race" specification that includes knowledge and talent measures alongside our NRDL exposure variables. Column (2) shows that knowledge volume is positively correlated with trial activity, every additional 1000 papers contributes to 18% more trials in the disease category, and Column (3) confirms that "high-quality" knowledge (the number of highly cited papers) has an even stronger effect. Column (4) finds a significant positive relationship between returnee talent and trial growth, consistent with the role of international expertise in fostering domestic innovation.

Crucially, the coefficient on NRDL exposure remains large, positive, and statistically significant across all specifications. The elasticity of trial activity with respect to reimbursement revenue is remarkably stable, even when controlling for the full suite of supply-side indicators. This leads to an important conclusion: while knowledge and talent provided the necessary infrastructure and innovative capacity, the market-size expansion triggered by the NRDL reform acted as the primary catalyst. Without the demand-side shock to expected returns, the accumulated supply of knowledge and talent would likely have translated into a much more gradual increase in clinical activity rather than the dramatic shift documented in this paper.

4.3 Application Reform

There are several other policies that may have complemented the increase in drug innovation but are unlikely to be the primary driving forces. For instance, the clinical trial application reform in 2015, studied in Jia et al. (2024), streamlined regulatory procedures by clearing the backlog of pending IND (investigational new drug) applications, which are required prior to initiating clinical trials for new drugs under investigation.

We collect the application and approval dates for all IND applications from the Center for Drug Evaluation of China's NMPA to calculate the backlog time, and obtain two main findings. First, while this reform provided an immediate boost by alleviating administrative bottlenecks, its effects appear transitory: the backlog that accumulated prior to 2014 was resolved rapidly through the reform, and approval days flattened after 2019, as shown in Figure 9. Moreover, the reform alone cannot account for the persistent upward trajectory in trial activity observed after 2016, nor can it explain the heterogeneous increases across firm nationality, therapeutic categories, and trial novelty. Second, as shown in column (5) of Table 5 that controls for both NRDL and application backlog, for each disease category, an additional 100 backlog days are associated with a 4.4% reduction in the number of clinical trials. Taken together, these results indicate that the application reform, by reducing application backlog days, may have complemented the increase in trial activity, but it is unlikely to be the main driver behind it.

4.4 Other Policy Reforms

We also examine whether other government policies may have contributed to the rise in clinical trial activity. To do so, we scrape all policy documents from the State Council Policy Document Library and use GPT-based classification to identify those relevant to biopharma innovation. These policy documents are grouped into four categories: (a) advanced manufacturing policies, which aim to strengthen domestic production capacity; (b) innovation independence policies, which seek to reduce reliance on technology from other countries; (c)

capital market reform policies, which improve firms' access to financing; and (d) tax or subsidy policies, which provide direct incentives for R&D and innovation.

Figure 10 plots the number of policies in each category from 2010 to 2024. Innovation-related policies are relatively rare before 2016, but beginning around 2017, policy activity increases across multiple categories. Notably, "innovation independence" policies rise sharply after 2018, consistent with the heightened geopolitical tensions and the Chinese government's emphasis on technological self-sufficiency. Patterns for manufacturing, capital market reform, and tax/subsidy policies also show a moderate upward trend. Column (6) of Table 5 quantifies the relationship between these policies and clinical trial activity. Although the number of biopharma-related policies is positively and significantly correlated with trial growth, the coefficient on NRDL exposure remains large and stable even after controlling for these policy measures. This suggests that while broader innovation policies may have contributed at the margin, they do not account for the sharp post-2016 increase in clinical trials—reinforcing the central role of the NRDL reform in expanding effective market size and shaping firms' innovation incentives.

4.5 Decomposition of Different Mechanisms

The preceding analysis identifies several empirically important drivers of China's clinical expansion. Because supply-side and demand-side changes occurred simultaneously, we adopt a unified framework to quantify their relative importance. Specifically, we conduct a decomposition that attributes the observed growth in clinical trials to each underlying mechanism.

We proceed in two steps. First, we estimate semi-elasticities of clinical trial activity with respect to each mechanism. Column (7) of Table 5 reports the coefficients from regressions of trial activity on NRDL exposure, measures of knowledge quantity and quality, talent flows, and other policies, absorbing disease-by-country and year fixed effects. Second, we combine these estimates with observed disease-level changes from 2016–2023 to translate coefficients into implied changes in trial counts. For disease subcategory d, the contribution of mechanism

$M \in \{\text{NRDL, talent flow, application reform, other policies}\}$ is calculated as:

$$\begin{aligned}\Delta Y_d(M) &= \sum_{t=2016}^{2023} (\text{TrialNumber}_{t+1} - \text{TrialNumber}_t) \\ &= \text{Coef}_{\text{M}} \times \sum_{t=2016}^{2023} \sum_{j \in d} (\text{M}_{j,t+1} - \text{M}_{j,t}) \times \text{TrialNumber}_{j,t}.\end{aligned}$$

This decomposition quantifies the relative contributions of demand-pull (NRDL) and supply-push forces, yielding a clear ranking of the drivers behind China's post-2016 surge in pharmaceutical innovation.

Table 6 presents the results of the decomposition exercise. To illustrate the economic magnitudes, we highlight the results for Oncology, the largest therapeutic area in our sample. Of the 1,004 additional oncology trials initiated in China in 2024 relative to the 2016 baseline, the NRDL mechanism accounts for approximately 40 percent of the expansion. In comparison, the combined supply-side factors of knowledge accumulation and talent flows explain 23 percent of the growth. Other sector-specific industrial policies account for a negligible share (less than 1 percent), with the remaining variation attributed to unobserved factors. Across all disease categories, the NRDL consistently emerges as the most influential driver, explaining at least 20 percent of the aggregate increase in clinical trial activity. This finding suggests that while secular trends in scientific knowledge and international talent flows provided a critical foundation for China's innovative capacity, their quantitative contributions to the post-2016 acceleration are secondary to the NRDL reform.

4.6 Welfare Implications

Before we conclude, we conduct a back-of-the-envelope exercise to interpret the improvement in clinical trial activity in terms of potential new drug discovery and, ultimately, social surplus, and benchmark it against the short-term gains from increased market access.

The short-run welfare gain comes from improved access and lower prices for existing in-

novative drugs. Without NRDL expansion, these drugs are typically sold only in the private market, without insurance coverage, and at high prices. This reflects strong market power: patented drugs face no close competitors, so manufacturers can charge prices far above marginal cost. When a drug is added to the NRDL, two forces reduce the effective price faced by patients. First, insurance coverage acts like a price subsidy, sharply lowering out-of-pocket payments. Second, the government uses its monopsony bargaining power to negotiate substantial price cuts with manufacturers. Together, these channels generate immediate welfare gains by expanding utilization among patients who were previously priced out of the market.

The long-run gain operates through innovation incentives. By expanding coverage and increasing the expected effective market size, the NRDL raises the expected returns to R&D in related therapeutic areas. This stimulates firms to start or accelerate drug-development pipelines, which may eventually lead to the discovery of new drugs. The social value of these future drugs is unknown ex ante. However, we can form expectations by looking at the realized welfare generated by new drug launches in similar disease categories in recent years, and use those observed benefits to proxy for the expected value of innovations induced by the NRDL expansion.

Based on this intuition, and focusing on cancer drugs alone, the short-run gain from expanding access is estimated to be around ¥40 billion ($5.6 billion) per year from 2016 to 2023 (Barwick et al., 2025). The cumulative increase in oncology trials is 7084 from 2016 to 2024. Given that 40% of this increase is attributable to the pull policy from the NRDL, and that the average conversion rate is 123 trials per successful new drug, this implies approximately 23 new oncology drugs over the next five to ten years, likely yielding a welfare gain of at least ¥21.16 billion per year in the near future. This highlights that the dynamic gains of boosted innovation from the NRDL reform can be comparable to the static gains of expanding access.[13]

[13]We assume the revenue for each new oncology drug is ¥0.46 billion per year (which corresponds to the

5 Conclusion

This paper documents a historic shift in the global landscape of pharmaceutical innovation. Using granular data on clinical trials with an event study framework, we provide the first empirical evidence of China's transition from a follower in generic manufacturing to a global frontier in clinical R&D. We show that by 2024, China's clinical trial activity has surged by over 170% relative to the U.S., a trend driven primarily by domestic firms and increasingly focused on high-novelty, non-generic compounds.

Our analysis suggests that this transition was not a passive consequence of China's economic growth. Instead, we identify the 2016 National Reimbursement Drug List (NRDL) reform as the primary catalyst. By leveraging government monopsony power to guarantee market volume, the NRDL reform not only increases access, but also altered the expected returns to innovation. Our growth-accounting decomposition reveals that this demand-pull effect explains a substantially larger share of the R&D surge than supply-side factors such as knowledge accumulation or talent return. In addition, our welfare exercise indicates the long-run gains from induced innovation can be as sizable as short-run gains from improved access.

These findings offer a new perspective on the policy studied here. While the pharmaceutical literature often frames the trade-off between static efficiency (patient access) and dynamic efficiency (innovation) as close to a zero-sum game, the Chinese experience suggests that a sufficiently large market expansion can mitigate this tension. China's experience indicates that a developing economy can successfully "pull" frontier innovation into its domestic market.

---

average sales per new cancer drug originated from China.) Assuming constant-elasticity demand with $\varepsilon = -1.5$, consumer surplus under monopoly equals $CS = PQ \cdot \frac{-1/\varepsilon}{1+1/\varepsilon}$. Substituting $\varepsilon = -1.5$ gives $CS = 2PQ$. With total revenue $PQ = 0.46 \times 23 = 10.58$ billion, $CS = 21.16$ billion. The CS gains under monopoly is a lower bound as insurance expansion is likely to increase access and CS per drugs.

References

Acemoglu, Daron, David Cutler, Amy Finkelstein, and Joshua Linn, "Did medicare induce pharmaceutical innovation?," *American Economic Review*, 2006, *96* (2), 103–107.

Agha, Leila, Soomi Kim, and Danielle Li, "Insurance Design and Pharmaceutical Innovation," *American Economic Review: Insights*, June 2022, *4* (2), 191–208.

Alsan, Marcella, Maya Durvasula, Harsh Gupta, Joshua Schwartzstein, and Heidi Williams, "Representation and Extrapolation: Evidence from Clinical Trials," *The Quarterly Journal of Economics*, February 2024, *139* (1), 575–635.

Andrews, Donald WK, "Tests for parameter instability and structural change with unknown change point," *Econometrica: Journal of the Econometric Society*, 1993, pp. 821–856.

Azoulay, Pierre, Joshua S Graff Zivin, Danielle Li, and Bhaven N Sampat, "Public R&D investments and private-sector patenting: evidence from NIH funding rules," *The Review of economic studies*, 2019, *86* (1), 117–152.

Barwick, Panle, Ashley Swanson, and Tianli Xia, "A Double Dose of Reform: Insurance and Centralized Negotiation in Drug Markets," March 2025. Working paper.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen, "Identifying technology spillovers and product market rivalry," *Econometrica*, 2013, *81* (4), 1347–1393.

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.

Branstetter, Lee G and Mariko Sakakibara, "When do research consortia work well and why? Evidence from Japanese panel data," *American Economic Review*, 2002, *92* (1), 143–159.

Branstetter, Lee G., Douglas Hanley, and Huiyan Zhang, "Unleashing the Dragon: The Case for Patent Reform in China," *Carnegie Mellon Working Paper Series*, 2023.

Bronzini, Raffaello and Paolo Piselli, "The impact of R&D subsidies on firm innovation," *Research Policy*, 2016, *45* (2), 442–457.

Budish, Eric, Benjamin N Roin, and Heidi Williams, "Do firms underinvest in long-term research? Evidence from cancer clinical trials," *American Economic Review*, 2015, *105* (7), 2044–2085.

Callaway, Brantly and Pedro H.C. Sant'Anna, "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230. Themed Issue: Treatment Effect 1.

Chen, Zhao, Zhikuo Liu, Juan Carlos Suárez Serrato, and Daniel Yi Xu, "Notching R&D Investment with Corporate Income Tax Cuts in China," *American Economic Review*, July 2021, *111* (7), 2065–2100.

Chen, Zimeng, Hui Zhong, Hongxi Hu, Fanpu Kong, Wannian Liang, and Guanqiao Li, "Chinese innovative drug R&D trends in 2024," *Nature Reviews Drug Discovery*, July 2024, *23* (11), 810–811. Bandiera_abtest: a Cg_type: From The Analyst's Couch Publisher: Nature Publishing Group.

Cunningham, Colleen, Florian Ederer, and Song Ma, "Killer Acquisitions," *Journal of Political Economy*, March 2021, *129* (3), 649–702. Publisher: The University of Chicago Press.

Dechezleprêtre, Antoine, Elias Einiö, Ralf Martin, Kieu-Trang Nguyen, and John Van Reenen, "Do tax incentives increase firm innovation? An RD design for R&D, patents, and spillovers," *American Economic Journal: Economic Policy*, 2023, *15* (4), 486–521.

Dranove, David, Craig Garthwaite, and Manuel Hermosilla, "Does consumer demand pull scientifically novel drug innovation?," *The RAND Journal of Economics*, 2022, *53* (3), 590–638.

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. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1756-2171.12113.

Finkelstein, Amy, "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry," *Quarterly Journal of Economics*, 2004, *119* (2), 527–564.

Garthwaite, Craig, "Economic Markets and Pharmaceutical Innovation," *Journal of Economic Perspectives*, May 2025, *39* (2), 3–26.

Grover, Aditya and Jure Leskovec, "node2vec: Scalable feature learning for networks," in "Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining" 2016, pp. 855–864.

Hermosilla, Manuel, "Rushed Innovation: Evidence from Drug Licensing," *Management Science*, January 2021, *67* (1), 257–278. Publisher: INFORMS.

Howell, Sabrina T, "Financing innovation: Evidence from R&D grants," *American economic review*, 2017, *107* (4), 1136–1164.

Hu, Albert Guangzhou and Gary H Jefferson, "A great wall of patents: What is behind China's recent patent explosion?," *Journal of Development Economics*, 2009, *90* (1), 57–68.

IQVIA Institute for Human Data Science, "Global Trends in R&D 2024: Activity, Productivity, and Enablers," 2024. Accessed 2025.

Javorcik, Beata Smarzynska and Shang-Jin Wei, "Corruption and cross-border investment in emerging markets: Firm-level evidence," *Journal of International Economics*, 2009, *78* (2), 432–443.

Jia, Ruixue, Xiao Ma, Jianan Yang, and Yiran Zhang, "Improving Regulation for Innovation: Evidence from China's Pharmaceutical Industry," *Working Paper*, 2024.

Kerr, William R, "Ethnic scientific communities and international technology diffusion," *The Review of Economics and Statistics*, 2008, *90* (3), 518–537.

Kremer, Michael and Christopher M. Snyder, "Worst-Case Bounds on R&D and Pricing Distortions: Theory with an Application Assuming Consumer Values Follow the World Income Distribution," October 2018.

Kyle, Margaret, "Pharmaceutical Price Controls and Entry Strategy," *Review of Economics and Statistics*, 2007.

Lakdawalla, Darius N., "Economics of the Pharmaceutical Industry," *Journal of Economic Literature*, June 2018, *56* (2), 397–449.

Leung, Jennifer Wong, Stephan Robin, and Danielle Cave, "ASPI's two-decade Critical Technology Tracker: The rewards of long-term research investment," *Australian Strategic Policy Institute*, August 2024.

Martin, Linda, Melissa Hutchens, Conrad Hawkins, and Alaina Radnov, "How much do clinical trials cost?," *Nature Reviews Drug Discovery*, 2017, *16* (6), 381–382.

Melnik, Walter and Andrew Smyth, "R&D tax credits and innovation," *Journal of Public Economics*, 2024, *236*, 105157.

National Health Commission of the PRC, "2022 Health Statistics Yearbook," 2023. Accessed: 2024-07-30.

Olson, M. and R. Zeckhauser, "An Economic Theory of Alliances," *Review of Economics and Statistics*, 1966.

Priem, Jason, Heather Piwowar, and Richard Orr, "OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts," *arXiv preprint arXiv:2205.01833*, 2022.

Quandt, Richard E, "Tests of the hypothesis that a linear regression system obeys two separate regimes," *Journal of the American statistical Association*, 1960, *55* (290), 324–330.

Roth, Jonathan, "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends," *American Economic Review: Insights*, 2022, *4* (3), 305–322.

Schmookler, Jacob, *Invention and Economic Growth*, Harvard University Press, 1966. Google-Books-ID: eQozAAAAMAAJ.

Wallsten, Scott J., "The Effects of Government-Industry R&D Programs on Private R&D: The Case of the Small Business Innovation Research Program," *RAND Journal of Economics*, 2000, *31* (1), 82–100.

Ward, Michael R. and David Dranove, "The Vertical Chain of Research and Development in the Pharmaceutical Industry," *Economic Inquiry*, 1995, *33* (1), 70–87.

Wei, Shang-Jin, Jianhuan Xu, Ge Yin, and Xiaobo Zhang, "Mild Government Failure," NBER Working Papers 31178, National Bureau of Economic Research, Inc 2023.

Xia, Tianli, "Welfare Effects of Resale Price Maintenance: Evidence from the Chinese Pharmaceutical Industry," April 2024. Working paper.

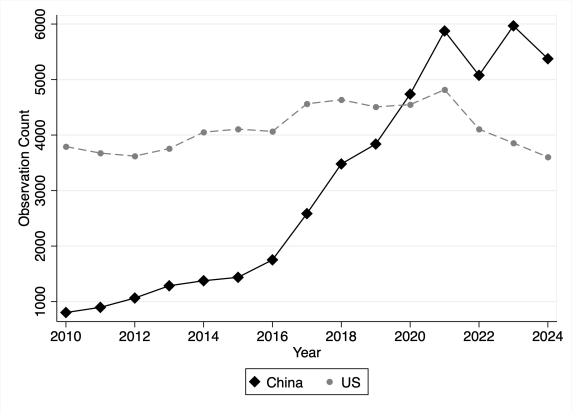
Yin, Wesley, "Market incentives and pharmaceutical innovation," *Journal of Health Economics*, July 2008, *27* (4), 1060–1077.

Yu, Hao, "Universal health insurance coverage for 1.3 billion people: What accounts for China's success?," *Health Policy*, 2015, *119* (9), 1145–1152.

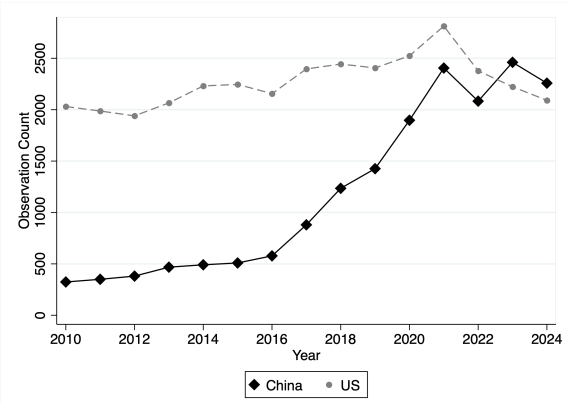
Figures

Figure 1: Time Trend - Number of Trials (Quantity and Novelty)

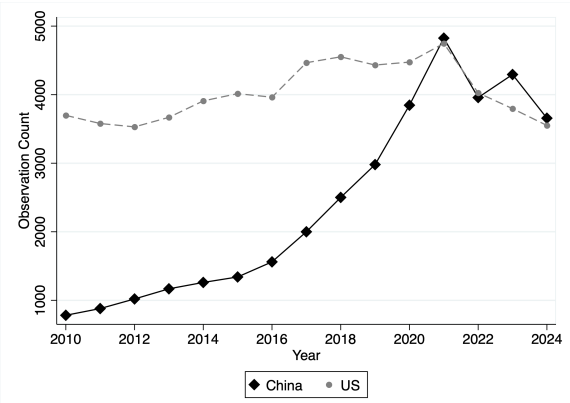
(a) All Trials
China vs US



(b) Novelty Score > Median
China vs US



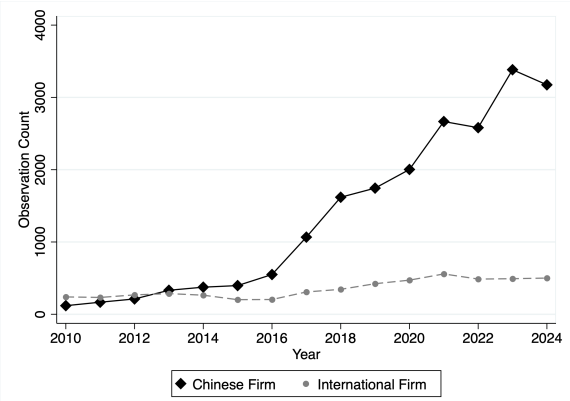
(c) Non-Generic Trials
China vs US



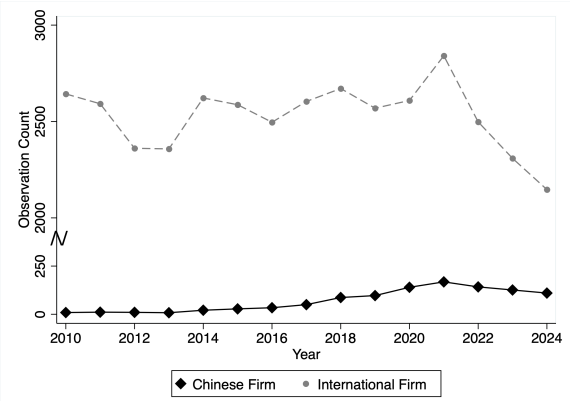
Notes. Data source is Citeline TrialTrove database. The figures plot the number of clinical trial observations by year. The vertical axis in each panel measures the observation count, while the horizontal axis indicates calendar year (2010–2024). Panel (a) compares all trials conducted in China and the United States. Panel (b) compares trials with novelty score (calculated based on Dranove et al. (2022)) above median, and Panel (c) compare trials investigating non-generic drugs.

Figure 2: Time Trend - Number of Trials (Chinese Firm or International Firm)

(a) Trials in China
Chinese Firm vs International Firm



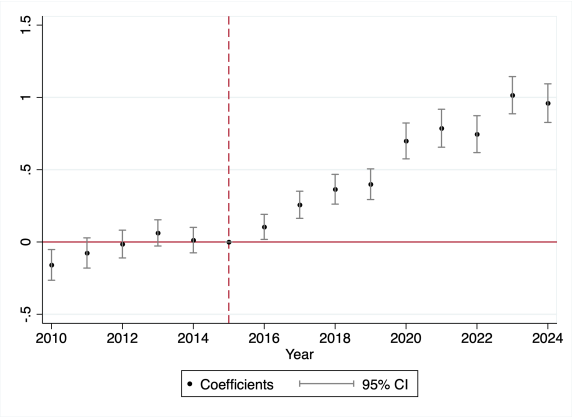
(b) Trials in the US
Chinese Firm vs International Firm



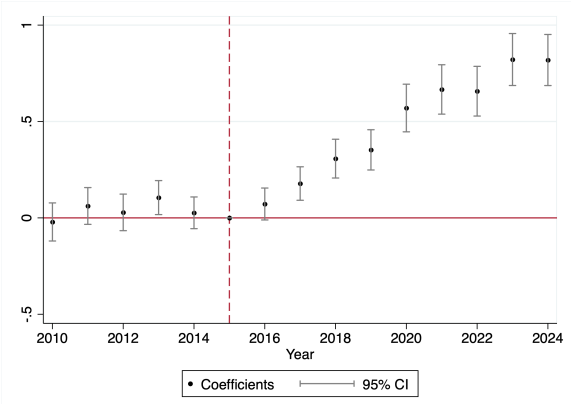
Notes. Data source is Citeline TrialTrove database. The figures plot the number of clinical trial observations by year. The vertical axis in each panel measures the observation count, while the horizontal axis indicates calendar year (2010–2024). Panel (a) restricts to trials conducted in China, distinguishing between Chinese firms and international firms. Panel (b) restricts to trials conducted in the United States, distinguishing between Chinese firms and international firms.

Figure 3: Event Study - Trials (Quantity and Novelty)

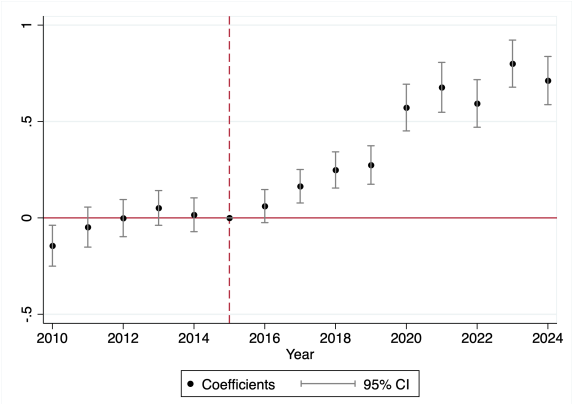
(a) All Trials
China vs US



(b) Novelty Score > Median
China vs US



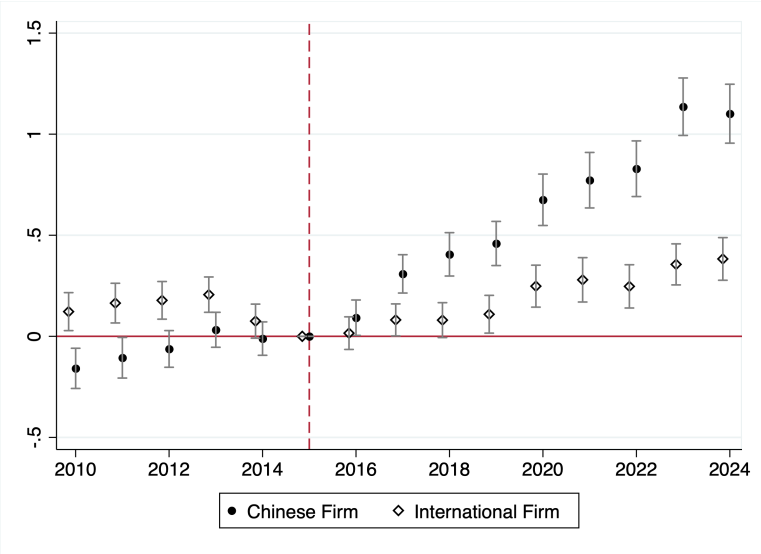
(c) Non-Generic Trials
China vs US



Notes. Data source is Citeline TrialTrove Database. The figure presents an event-study analysis of disease–country level clinical trial activity, with 2015 as the base year. The vertical axis in each panel plots the estimated coefficients on the interaction term from the difference-in-differences specification, along with 95% confidence intervals. The horizontal axis indicates calendar year (2010–2024). Panel (a) compares all trials conducted in China and the United States. Panel (b) restricts to trials of novelty score above median. Panel (c) restricts to trials investigating non-generic drugs. Robust standard errors are clustered at the disease × country level.

Figure 4: Event Study - Trials (Chinese Firm vs International Firm)

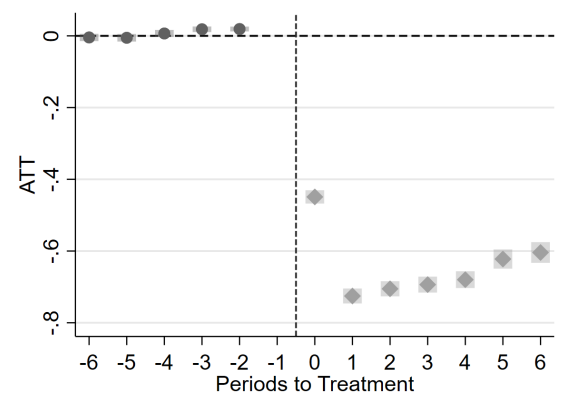
(a) All Trials
Chinese Firm vs International Firm



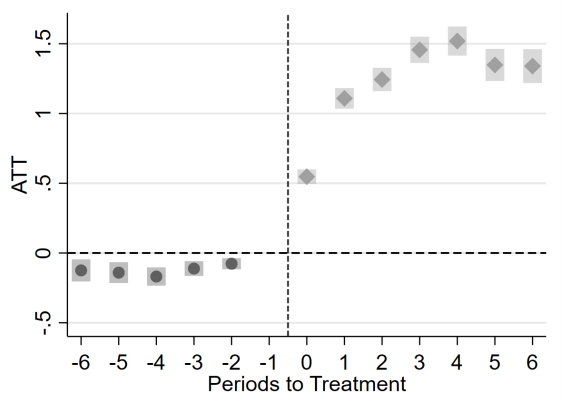
Notes. Data source is Citeline TrialTrove Database. The figure presents an event-study analysis of disease–country level clinical trial activity, with 2015 as the base year. The vertical axis plots the estimated coefficients on the interaction term from the difference-in-differences specification, along with 95% confidence intervals. The horizontal axis indicates calendar year (2010–2024). The black line compares trials conducted in China by Chinese firms and those conducted in the United States. The gray line compares trials conducted in China by international firms and those conducted in the United States. Robust standard errors are clustered at the disease × country level.

Figure 5: Effects of the NRDL Reform on Price and Quantity

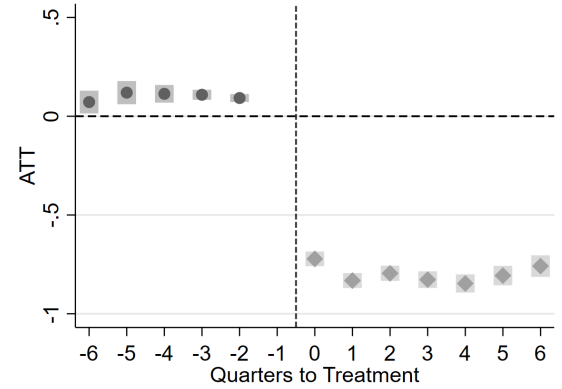
(a) Retail Price: All Categories



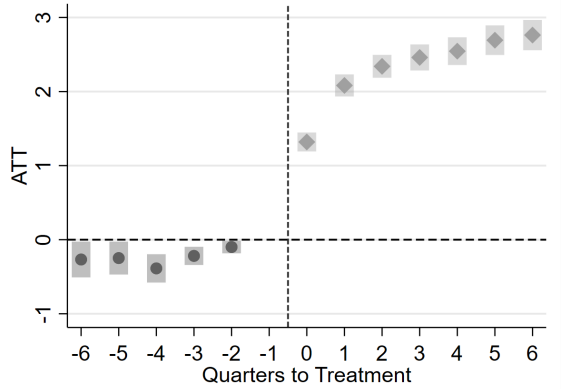
(b) Quantity: All Categories



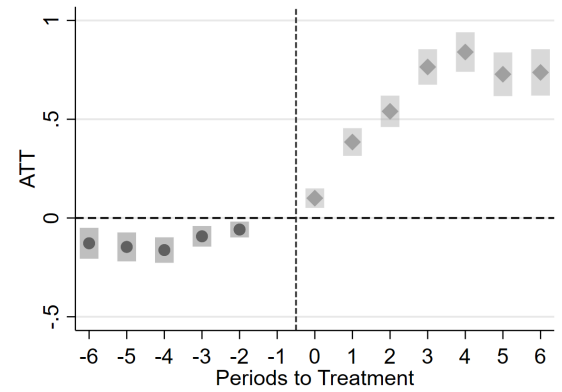
(c) Retail Price: Cancer Drugs



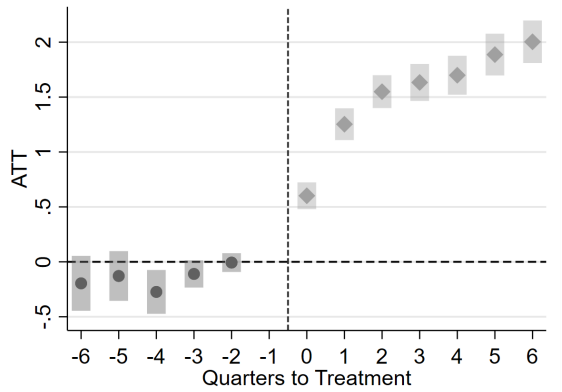
(d) Quantity: Cancer Drugs



(e) Revenue: All Drugs



(f) Revenue: Cancer Drugs

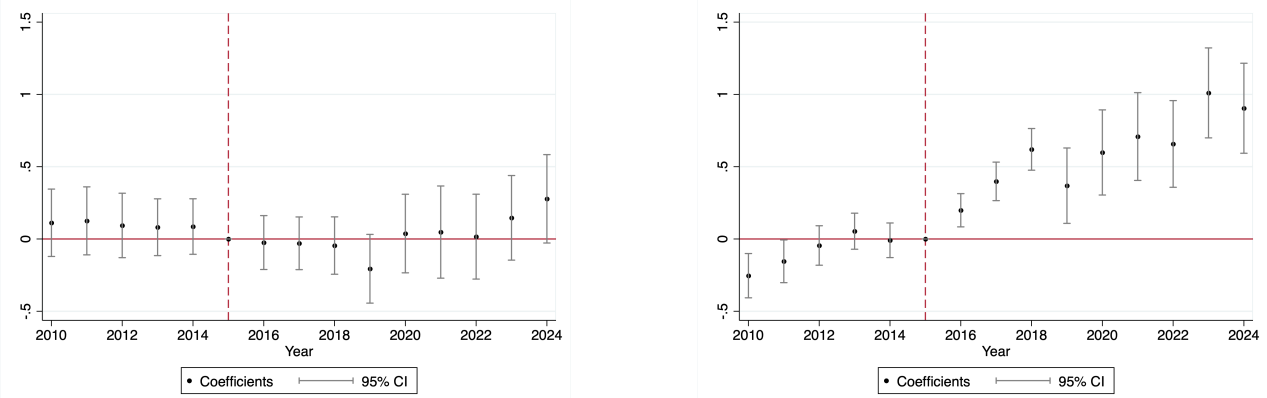


Notes. The data source is Sinohealth drug sales. This figure reports the effects of the NRDL Reform on log(retail prices) (a and c), log(quantities) (b and d), and log(revenue) (e and f) of successfully negotiated drugs. The horizontal axis denotes quarters relative to the negotiation period. The vertical axis reports estimated dynamic treatment effects and 95% confidence intervals using the CSDID method from Callaway and Sant'Anna (2021).

Figure 6: Mechanism: NRDL Effects across Disease Categories (Event Study)

(a) No NRDL Exposure (Placebo)

(b) NRDL Exposure

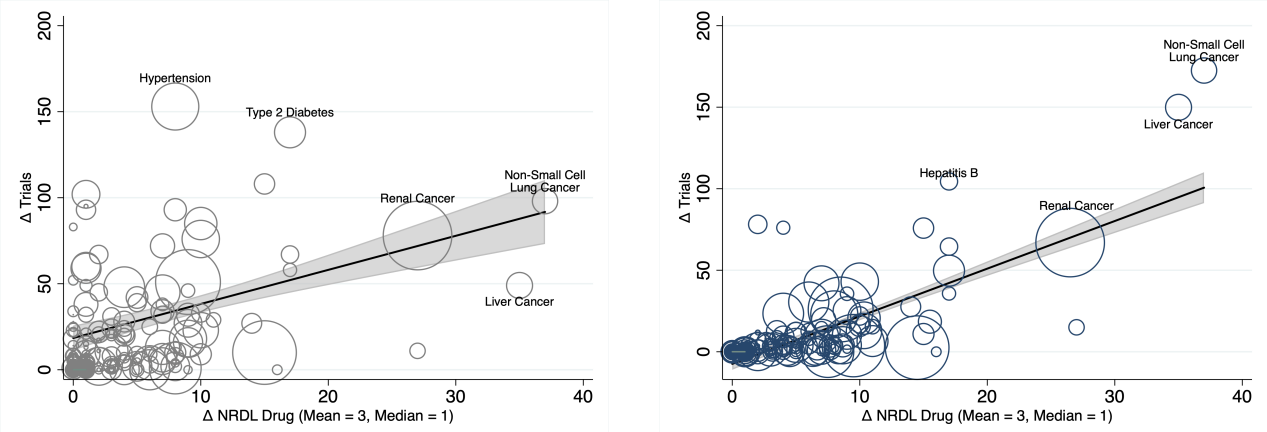


Notes. Data source is Citeline TrialTrove Database. The figure presents an event-study analysis of disease–country level clinical trial activity, with 2015 as the base year. The vertical axis in each panel plots the estimated coefficients on the interaction term from the difference-in-differences specification, along with 95% confidence intervals. The horizontal axis indicates calendar year (2010–2024). Panel (a) only include disease categories without any drug ever covered by NRDL and are not Oncology or Infectious disease. Panel (b) include disease categories with drug that are ever covered by NRDL. Robust standard errors are clustered at the disease × country level.

Figure 7: Mechanism: NRDL Effects across Disease Categories (Magnitude)

(a) Raw Data

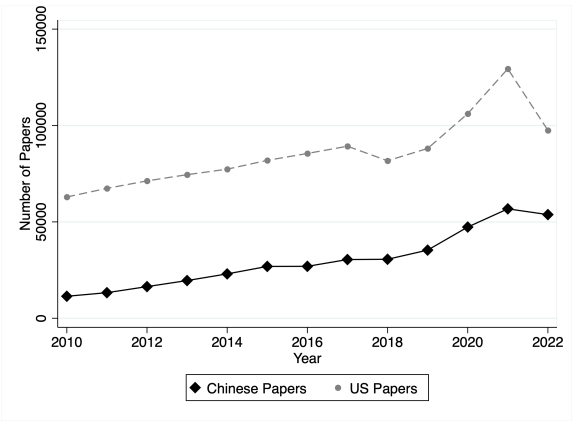
(b) Reg. Results:
Increased Trials vs NRDL Exposure



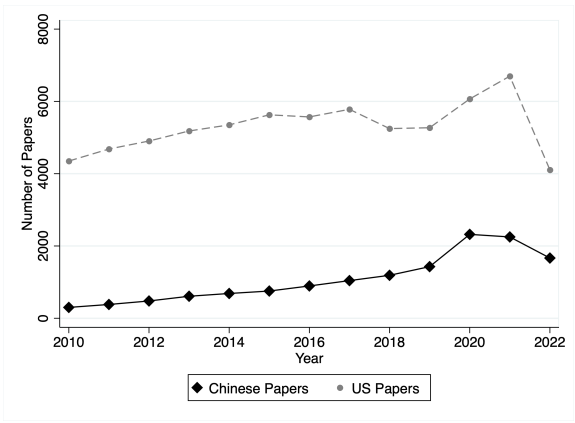
Notes. Data source is Citeline TrialTrove Database. The figure reports, for each disease subcategory, the change in clinical trial activity and the change in the number of NRDL-included drugs between 2016 and 2024. Panel (a) displays the raw data. Panel (b) presents the estimated increase in trials associated with NRDL exposure for each disease subcategory. The size of each circle is proportional to the number of NRDL-included drugs in that disease subcategory.

Figure 8: Mechanism: Time Trend - Papers and Scientists

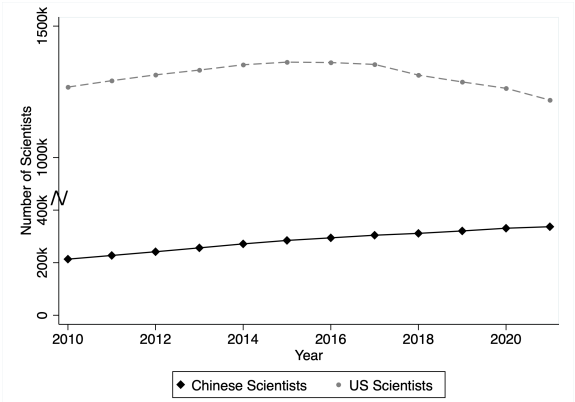
(a) Number of Papers



(b) Number of Top 5% Cited Papers



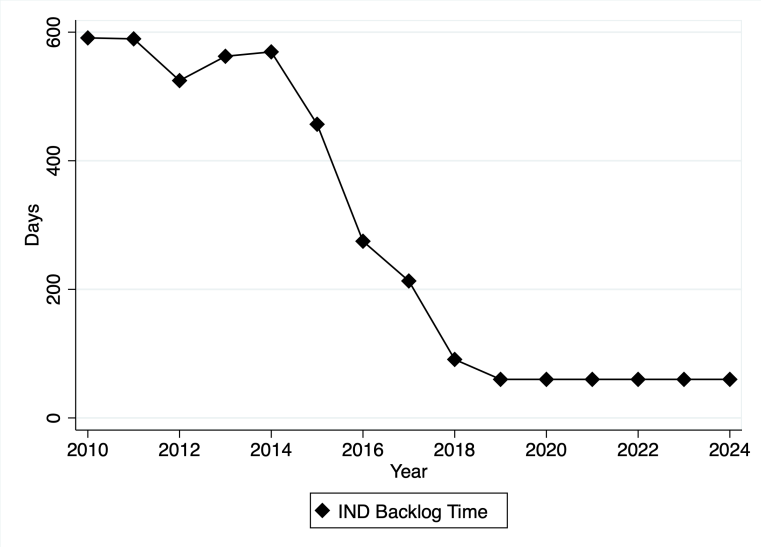
(c) Number of Scientists w. US Experience



Notes. Data source is OpenAlex publication database, and only papers with PubMed id are included. The figure plots time trends in knowledge production and international talent flows for China and the United States. The vertical axes report the number of papers (Panel (a)), the number of Top 5% cited papers (Panel (b)), and the number of scientists with U.S. experience (Panel (c)), while the horizontal axis indicates calendar year. Top 5% cited papers are defined within the disease category and year, and scientists with U.S. experience are those publish at least three papers with a US affiliation.

Figure 9: Mechanism: Application Backlog Time

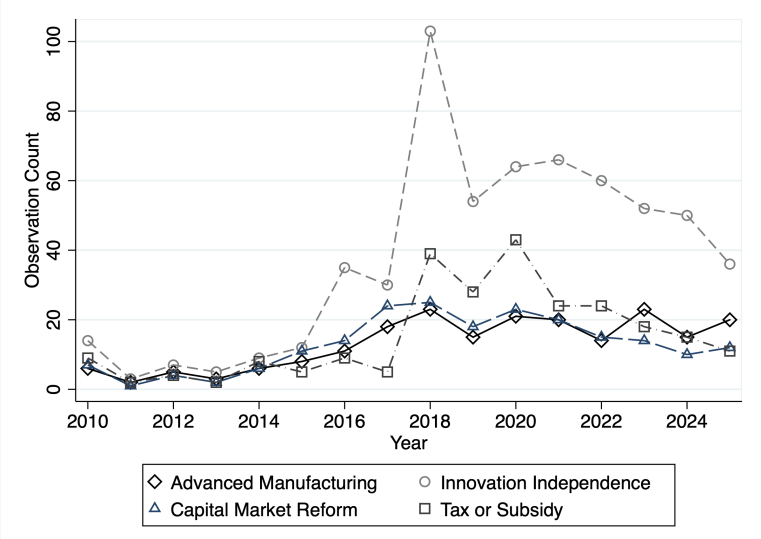
(a) Average Application Backlog Days



Notes. Data source is the National Medical Products Administration (NMPA). The figures plot the mean of application backlog days (i.e. approval date - application date) by year.

Figure 10: Mechanism: Time Trend - Other Policies

(a) Number of Different Policies



Notes. Data source is the State Council Policy Document Library. The figures plot the number of policy documents by year. The vertical axis in each panel measures the observation count, while the horizontal axis indicates calendar year (2010–2024). Policies are classified into 4 types: (a) advanced manufacturing policies to promote industry production capacity (b) innovation independence policies to reduce reliance on western countries (c) capital market reform policies to alleviate barriers of raising funding, and (d) tax or subsidy to incentivize innovation.

Tables

Table 1: Summary of Each Round of Negotiation 2016-2022

Round	2016	2017	2018	2019	2020	2021	2022	All
# Negotiated Products	5	44	18	150	162	117	147	643
# Successful Negotiations	3	36	17	97	119	94	121	487

Notes. Data source: NHSA website. Overall success rate: 75.7%.

Table 2: Descriptive Statistics - Number of Trials by Disease and Country (2010-2024)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	China				US			
	Phase I	Phase II	Phase III	Phase IV	Phase I	Phase II	Phase III	Phase IV
Autoimmune/Inflammation	2061	1154	731	920	1529	2730	1758	1060
Central Nervous System (CNS)	2327	508	441	2029	2091	3004	2030	2746
Cardiovascular	2726	343	534	1572	797	1053	854	1005
Genitourinary	416	102	96	428	136	188	184	288
Infectious Disease	2198	853	618	1306	1735	1853	1145	1256
Metabolic/Endocrinology	2291	462	572	1294	1296	1589	1154	905
Oncology	4293	7711	2528	1463	6579	10241	1892	348
Ophthalmology	85	92	90	142	130	499	311	204
Vaccines (Infectious Disease)	33	13	53	42	36	48	34	12

Notes. Data source is Citeline TrialTrove Database. The table summarize the number of clinical trials in each disease category across year in each country.

Table 3: Difference in Difference - Trials (Quantity and Novelty)

	(1)	(2)	(3)
	(Log) All Trials	(Log) Novelty > Median Trials	(Log) Non-Generic Trials
China × (Year > 2016)	0.663***	0.508***	0.515***
	(0.046)	(0.046)	(0.043)
Constant	1.259***	0.926***	1.238***
	(0.011)	(0.011)	(0.011)
Disease × Country FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Adjusted R-squared	0.867	0.819	0.864
Observations	9030	9030	9030

Notes. Data source is Citeline TrialTrove Database. Dependent variable is log(number of trials + 1) for Column (1), log(number of trials of novelty score above median + 1) for Column (2), and log(number of trials investigating non-generic drugs + 1) for Column (3). The table reports difference-in-differences estimates. The key independent variable is an indicator for post-2016 years interacted with an indicator for China relative to the United States. Specifications include disease-by-country fixed effects and year fixed effects. Robust standard errors are clustered at the disease × country level.

Table 4: Mechanism: NRDL Intensity and Number of Trials

	(1)	(2)	(3)	(4)	(5)
	Mean of Independent Var	(Log) All Trials			
Log(Number of NRDL-Included Drugs)	0.428	0.164***			
		(0.038)			
Log(Revenue of NRDL-Included Drugs)	0.107		0.280***		
			(0.084)		
Ratio of NRDL-Included Drugs	0.059			0.454***	
				(0.165)	
Revenue Ratio of NRDL-Included Drugs	0.104				0.512***
					(0.143)
Constant		1.305***	1.303***	1.300***	1.299***
		(0.008)	(0.008)	(0.008)	(0.008)
Disease × Country FE		Yes	Yes	Yes	Yes
Year FE		Yes	Yes	Yes	Yes
Adjusted R-squared		0.870	0.869	0.869	0.870
Observations		9030	9030	9030	9030

Notes. Data source is Citeline TrialTrove database and Sinohealth database. Dependent variable is log(number of trials + 1). The table reports regressions of the number of clinical trials at the disease-country-year level on measures of exposure to the NRDL reform. NRDL exposure is measured in four ways: (1) log number of NRDL-included drugs, (2) log revenue of NRDL-included drugs, (3) ratio of NRDL-included drugs to all drugs in the disease area, and (4) revenue ratio of NRDL-included drugs to total disease-level drug revenues. All specifications include disease-by-country and year fixed effects. Robust standard errors are clustered at the disease × country level.

Table 5: Mechanisms: Summary

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mechanisms	NRDL	Paper Quantity	Paper Quality	Talent Flow	Application Backlog	Policy	All
D.V.	(Log) All Trials						
Revenue Ratio of NRDL-Included Drugs	0.512***	0.495***	0.474***	0.501***	0.459***	0.502***	0.433***
	(0.143)	(0.140)	(0.139)	(0.143)	(0.142)	(0.143)	(0.138)
Number of Papers (Thousands)		0.182**					-0.017
		(0.082)					(0.056)
Number of Papers w. High Citations (Thousands)			5.104***				7.336***
			(1.494)				(1.917)
Number of Scientists w. US Experience (Thousands)				0.009***			-0.019**
				(0.003)			(0.008)
Application Backlog Days (Hundreds)					-0.044***		-0.036***
					(0.008)		(0.008)
Number of Policies						0.005***	0.002***
						(0.001)	(0.001)
Constant	1.299***	1.271***	1.257***	1.280***	1.388***	1.251***	1.336***
	(0.008)	(0.015)	(0.015)	(0.010)	(0.017)	(0.013)	(0.026)
Disease × Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	0.870	0.872	0.873	0.870	0.863	0.871	0.875
Observations	9030	9030	9030	9030	9030	9030	9030

Notes. Data source is Citeline TrialTrove database, Sinohealth database, OpenAlex database, and State Council Policy Document Library. Dependent variable is log(number of trials + 1). The table reports regressions of the number of clinical trials at the disease-country-year level on different mechanisms. All specifications include disease-by-country and year fixed effects. Robust standard errors are clustered at the disease × country level.

Table 6: Quantification of Different Mechanisms

	All Increased Trials	NRDL	Knowledge & Talent	Application Reform	Other Policies	Unexplained
Oncology	1004.00	398.53	232.34	2.20	0.46	370.47
Autoimmune/Inflammation	609.00	33.83	22.92	0.08	0.12	552.06
Central Nervous System	572.00	43.03	55.60	0.50	0.13	472.73
Metabolic/Endocrinology	496.00	47.92	36.25	0.03	0.12	411.68
Cardiovascular	467.00	26.43	22.74	0.36	0.16	417.30
Infectious Disease	152.00	44.11	84.23	0.44	0.13	23.09
Genitourinary	58.00	1.40	4.43	0.00	0.03	52.14
Ophthalmology	22.00	1.72	0.63	0.03	0.01	19.62
Vaccines (Infectious Disease)	-14.00	0.00	0.08	0.00	0.00	-14.08
Total	3366.00	596.98	459.22	3.64	1.16	2305.00

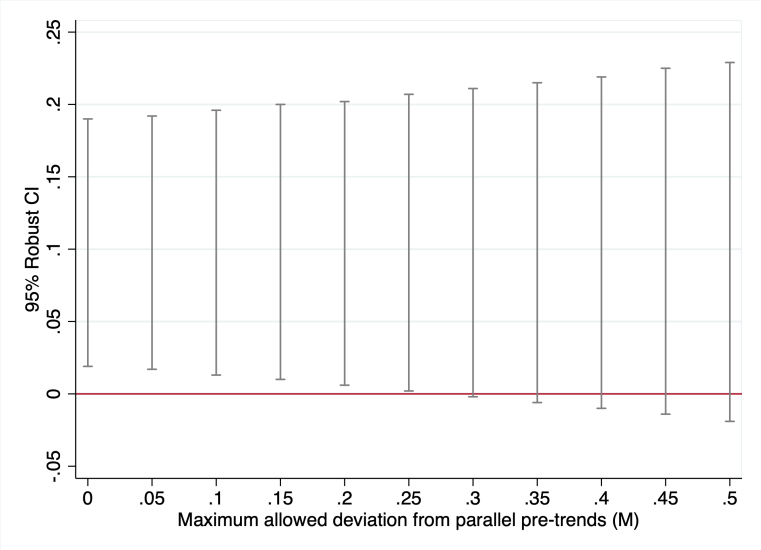
Notes. This table reports quantification of the contribution of different mechanisms to the increase of clinical trials.

Appendix

A.1 Figures

Figure A.1: Sensitivity of Event Study Estimates to Violations of Parallel Trends

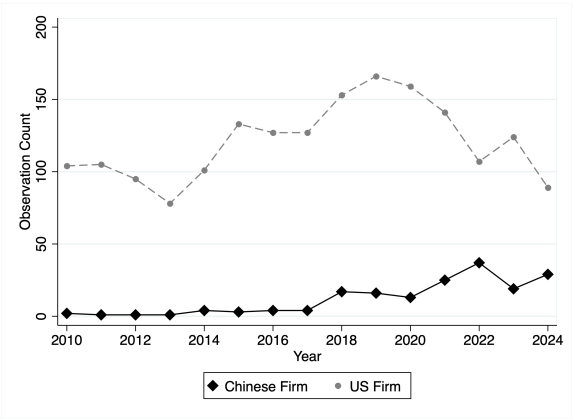
(a) HonestDiD Sensitivity Analysis



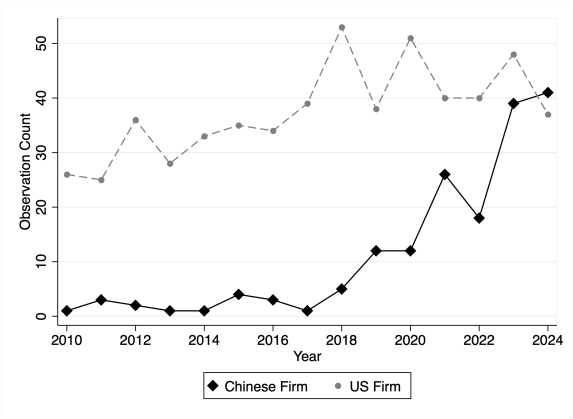
Notes. Data source is Citeline TrialTrove Database. This figure reports a sensitivity analysis of the baseline event study estimates using the HonestDiD approach. The horizontal axis shows the maximum allowed deviation from parallel pre-treatment trends (M), while the vertical axis displays the corresponding 95% robust confidence interval for the post-treatment effect. Confidence intervals are constructed following the C-LF method with $\alpha = 0.05$.

Figure A.2: Time Trend - Number of Drugs

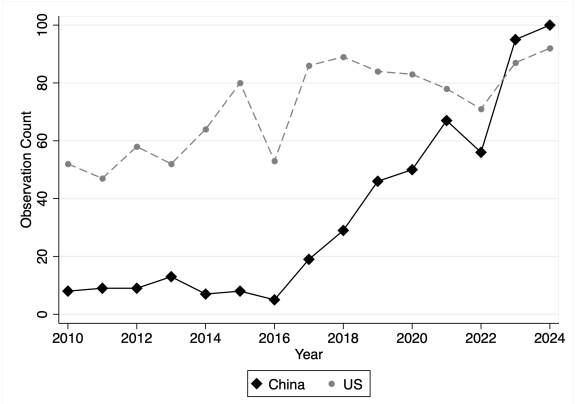
(a) Number of Licensed-Out Drugs
Chinese Firm vs US Firm



(b) Number of Newly Approved Drugs
Chinese Firm vs US Firm



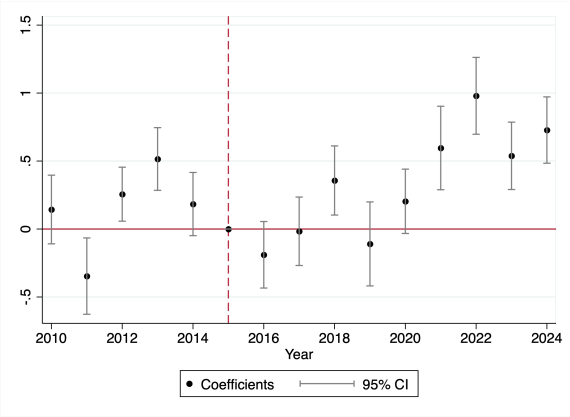
(c) Number of Newly Approved Drugs
China vs US



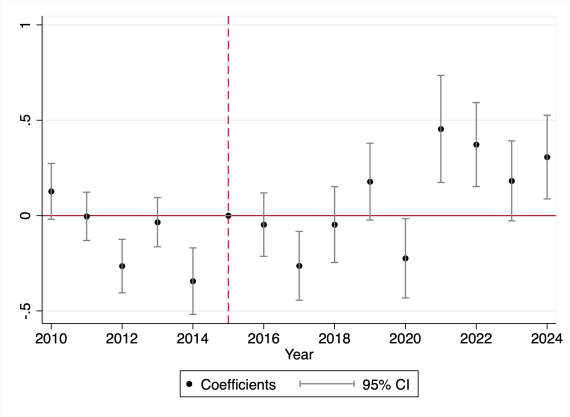
Notes. Data source is Citeline Pharmaprojects database. The figures plot the number of clinical trial observations by year. The vertical axis in each panel measures the observation count, while the horizontal axis indicates calendar year (2010–2024). Panel (a) compares license-out drugs originated by firms with headquarter in China and the United States. Panel (b) compares newly approved drugs originated by firms with headquarter in China and the United States. Panel (c) compares newly approved drugs in China and the United States.

Figure A.3: Event Study - Drug Pipelines

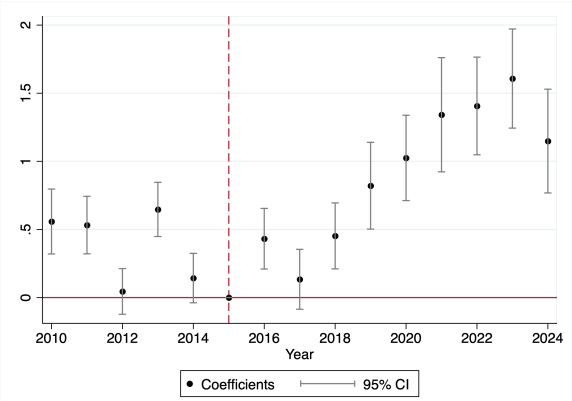
(a) License-Out Drugs
Originated from China vs US



(b) Newly Approved Drugs
Originated from China vs US



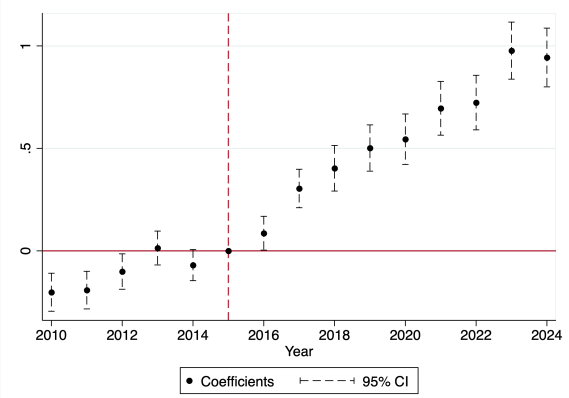
(c) Newly Approved Drugs
in China vs US



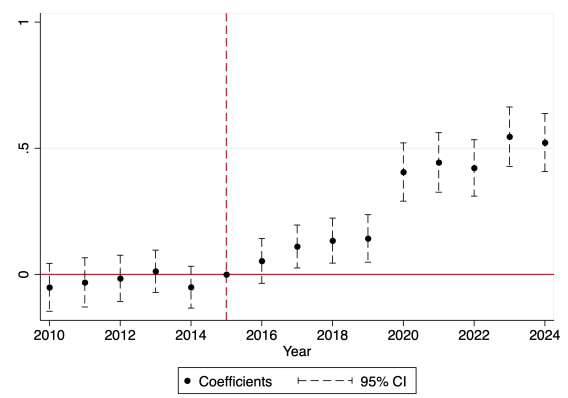
Notes. Data source is Citeline Pharmaprojects Database. The figure presents an event-study analysis of disease–country level clinical trial activity, with 2015 as the base year. The vertical axis in each panel plots the estimated coefficients on the interaction term from the difference-in-differences specification, along with 95% confidence intervals. The horizontal axis indicates calendar year (2010–2024). Panel (a) compares license-out drugs originated by firms with headquarter in China and the United States. Panel (b) compares newly approved drugs originated by firms with headquarter in China and the United States. Panel (c) compares newly approved drugs in China and the United States. Robust standard errors are clustered at the disease × country level.

Figure A.4: Event Study - Trials of Different Phases

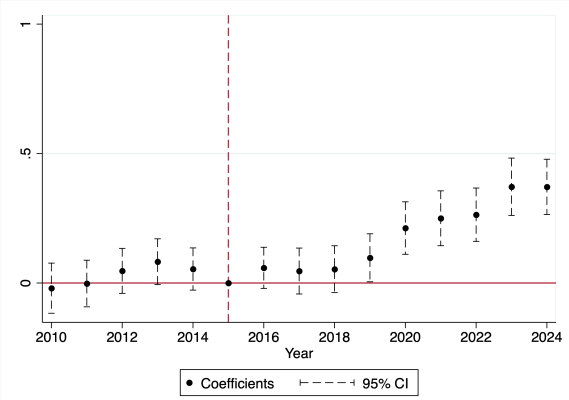
(a) Phase I



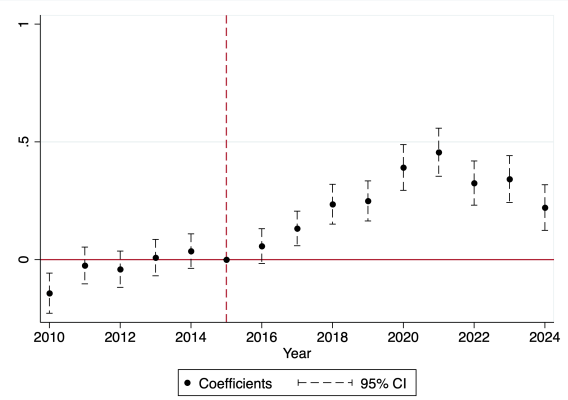
(b) Phase II



(c) Phase III



(d) Phase IV



Notes. Data source is Citeline TrialTrove Database. The figure presents an event-study analysis of disease–country level clinical trial activity, with 2015 as the base year. The vertical axis in each panel plots the estimated coefficients on the interaction term from the difference-in-differences specification, along with 95% confidence intervals. The horizontal axis indicates calendar year (2010–2024). Panels (a) to (d) compares trials of Phase I to IV conducted in China and the United States. Robust standard errors are clustered at the disease × country level.

A.2 Tables

Table A.1: Robustness - Different Sponsors

	(1)	(2)	(3)	(4)
	(Log) Trials by			
	Any Sponsors	Private Sponsors	Public Sponsors	Hybrid Sponsors
All Trials				
China × (Year > 2016)	0.663***	0.619***	0.408***	0.346***
	(0.046)	(0.050)	(0.035)	(0.031)
Phase-I Trials				
China × (Year > 2016)	0.703***	0.655***	0.173***	0.197***
	(0.052)	(0.052)	(0.024)	(0.024)
Phase-II Trials				
China × (Year > 2016)	0.353***	0.213***	0.244***	0.178***
	(0.038)	(0.034)	(0.030)	(0.023)
Phase-III Trials				
China × (Year > 2016)	0.177***	0.152***	0.069***	0.040***
	(0.031)	(0.031)	(0.015)	(0.012)
Phase-IV Trials				
China × (Year > 2016)	0.309***	0.059***	0.269***	0.128***
	(0.028)	(0.017)	(0.026)	(0.016)
Disease × Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	9030	9030	9030	9030

Notes. Data source is Citeline TrialTrove Database. Dependent variable is log(number of trials + 1), number of trials are calculated in different sponsors and phases within the disease-country-year group. The table reports difference-in-differences estimates. The key independent variable is an indicator for post-2016 years interacted with an indicator for China relative to the US. Specifications include disease-by-country fixed effects and year fixed effects. Robust standard errors are clustered at the disease × country level.

Table A.2: Robustness - Different FEs and Models

	(1)	(2)	(3)	(4)
	(Log) All Trials		All Trials	
China × (Year > 2016)	0.663***	0.682***	11.522***	1.301***
	(0.046)	(0.046)	(1.436)	(0.079)
Constant	1.259***	1.365***	7.797***	2.968***
	(0.011)	(0.012)	(0.350)	(0.028)
Percentage Change	94.1%	97.8%	164.9%	167.3%
Disease × Country FE	Yes	Yes	Yes	Yes
Disease × Year FE	No	Yes	No	No
Year FE	Yes	No	Yes	Yes
Model	OLS	OLS	OLS	Poisson
Observations	9030	9030	9030	9030

Notes. Data source is Citeline TrialTrove Database. Dependent variable is log(number of trials + 1) for Columns (1) and (2) and number of trials for Columns (3) and (4). The table reports difference-in-differences estimates. The key independent variable is an indicator for post-2016 years interacted with an indicator for China relative to the United States. Specifications varies with fixed effects combinations and models. Robust standard errors are clustered at the disease × country level.

Table A.3: Robustness - Different Treatment or Control Groups

	(1)	(2)	(3)	(4)
	(Log) All Trials			
China (*vs US*) × (Year > 2016)	0.663***			
	(0.046)			
China (*only*) × (Year > 2016)		0.654***		
		(0.046)		
China (*vs EU*) × (Year > 2016)			0.683***	
			(0.045)	
China (*vs EU+US*) × (Year > 2016)				0.675***
				(0.046)
Constant	1.259***	1.234***	1.194***	1.440***
	(0.011)	(0.011)	(0.011)	(0.011)
Disease × Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Adjusted R-squared	0.867	0.864	0.869	0.884
Observations	9030	9030	9030	9030

Notes. Data source is Citeline TrialTrove Database. Dependent variable is log(number of trials + 1) . The table reports difference-in-differences estimates. The key independent variable is an indicator for post-2016 years interacted with an indicator for China relative to different control group. Specifications include disease-by-country fixed effects and year fixed effects. Robust standard errors are clustered at the disease × country level.

A.3 Trial Novelty Calculation

We construct measures of trial novelty in several ways. First, we calculate a novelty score based on the extent to which a target has not yet been explored, following Dranove et al. (2022). Second, we use GPT to classify a trial's objective into one of four categories: investigation of a new drug, a new combination of existing drugs, a new indication for an existing drug, or a bio-equivalence test for a generic drug. This appendix provides the technical details underlying the construction of these measures.

A.3.1 Novelty Score

The first novelty measure follows the approach proposed by Dranove et al. (2022), adapted to the structure of the Citeline database. Whereas Dranove et al. (2022) construct novelty using "target-based actions" (TBA) from the Cortellis database, our data contain mechanisms of action (MOA), which are conceptually similar. MOAs describe the specific scientific strategy through which a molecule produces a pharmacological effect. Each MOA corresponds to a pair consisting of the biological entity targeted by the molecule (e.g., p38 MAP kinase) and the mechanism used to modulate that entity (e.g., inhibition). In total, there are 4,053 unique MOAs in the data. MOA information is available for 379,993 out of 463,412 trials in the database, covering approximately 82% of all trials. These trials include not only those conducted in China and the United States but also trials conducted in other countries, ensuring that the novelty measure reflects the global frontier of drug development. For the 107,256 trials conducted in China and the United States between 2010 and 2024 (our main analysis sample), the required information is available for 103,959 trials.

We compute translational novelty at the trial level by assessing the extent to which the MOA deployed in a given trial has previously been used. Consider a trial i that deploys a single MOA k. Let $D_{k(i)}$ represent the number of previous deployments of k in stages that are at least as advanced as that of trial i. For example, if i is a Phase 2 trial, $D_{k(i)}$ would

represent the count of other trials associated with k that have been tested in Phases 2 or 3 prior to the initiation of trial i, that is, Phase 1 deployments of k are not counted. This restriction follows Dranove et al. (2022): evidence generated at more advanced stages (e.g., Phase 3) is more informative about translational potential than evidence generated earlier, and knowledge flows are asymmetric (Phase 3 $\rightarrow$ Phase 1, but not vice versa).

The translational novelty score for trial i is computed as:

$$\text{Novelty}_i = \frac{1}{1 + D_{k(i)}} \in (0, 1]$$

Thus, the score equals 1 only if trial i deploys k for the first time ($D_{k(i)} = 0$). If $D_{k(i)} = 1$, the score is 0.5; if $D_{k(i)} = 2$, the score is 0.33; and so on. That is, the score decreases as the number of k's previous trials increases, approaching 0 in the limit. It is important to note that, except for trials with no prior deployments, our inference will not exploit information contained by the specific value of scores and instead relies on the implied ordering.

Many trials deploy multiple MOAs. In such cases, we compute a trial-level novelty score by taking the average of the MOA-specific novelty values for that trial:

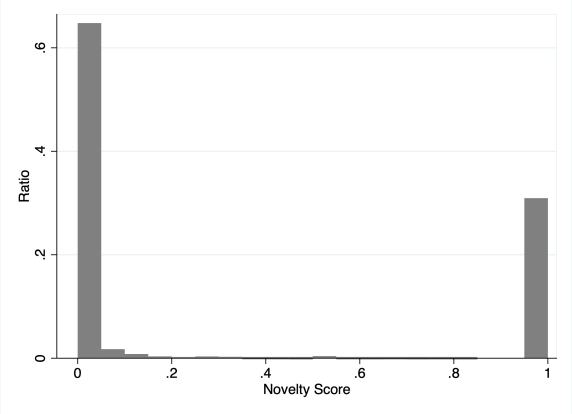
$$\text{Novelty}_i = \frac{1}{K_i} \sum_{k \in \text{MOAs}(i)} \frac{1}{1 + D_{k(i)}}$$

where K_i denotes the number of MOAs associated with trial i. Averaging yields a conservative measure of novelty for multi-target trials, balancing the contribution of highly novel targets and more established ones.

Figure A.5 plots the distribution of the novelty scores, which is highly right-skewed: the mean novelty score is 0.32, while the median is essentially zero, reflecting that most trials rely on previously well-explored mechanisms. This distribution closely matches the pattern documented in Dranove et al. (2022), despite relying on MOAs from Citeline rather than TBAs from Cortellis. To classify novel trials, we compute novelty quantiles within disease categories and years (pooling across countries); trials in the top quartile (Quantile 4) are designated as

"novel," while lower-quantile trials represent more incremental mechanisms.

Figure A.5: Distribution of Novelty Score (Dranove et al. (2022))



Notes. Data source is Citeline TrialTrove Database.

A.3.2 GPT-classified Novelty Objective

We also construct an innovation classifier using GPT to determine whether a clinical trial investigates (i) a new drug, (ii) a new combination of existing drugs, (iii) a new indication for an existing drug, or (iv) a bio-equivalence test for a generic drug. This measure allows us to capture innovative content that is not directly observable from structured fields in Citeline. The classifier takes as input a rich set of trial-level information—including the trial title, phase, primary tested drug, mechanism of action, trial objective, primary endpoint, study design, treatment plan, and investigator notes. An illustrative example of the variables fed into the model is shown as follows.

```
TrialID: 322799

StartDate: 2015-07-06

TrialTitle: Intratumoral Phase Ib Clinical Trial Of IIb, IIIc, And IV M1a,
M1b Malignant Melanoma With Recombinant Human GM-CSF Herpes Simplex Virus
Injection (OrienX-010)

TrialPhase: Phase I (Ib)

PrimaryTestedDrug: OrienX-010

PrimaryTestedDrugMechanism: Immuno-oncology therapy
```

```
TrialObjective: To evaluate the safety and tolerability of continuous
administration of OrienX-010 intratumoral injection in patients with stage
IIIb, IIIc, and IV M1a/M1b malignant melanoma who failed standard therapy,
and to evaluate the biodistribution characteristics and biological effects
of continuous administration of OrienX-010 intratumoral injection in the
same patient population.

PrimaryEndpoint: Adverse events, immunogenicity, neutralizing antibody
response, safety and tolerability

StudyDesign: Single-arm, non-randomized, open-label study

TreatmentPlan: Recombinant human GM-CSF herpes simplex virus injection,
intratumoral administration, specification 2.0-8.0 × 10^7 PFU (1.0 ml) per
injection site, injection every two weeks for 16 consecutive weeks, maximum
dose not exceeding 10 ml (8 × 10^8 PFU), adjusted according to lesion size

TrialNotes: Registry ChinaDrugTrials; registration number CTR20150881;
sponsor Beijing Aoyuan Heli Biotechnology Co., Ltd.; recruitment completed;
first informed consent signed on 2015-07-06; trial end date 2017-05-24
```

However, large language models can hallucinate or infer information that is not explicitly present in the input, which is problematic when the classification must be strictly tied to verifiable trial attributes. To reduce hallucinations and ensure consistency, we develop a supervised fine-tuned model based on GPT-4o-mini. The fine-tuning process proceeds in three steps.

First, we randomly sample 200 clinical trials from the full dataset and manually label each trial's treatment arm, control arm, and innovation category using the detailed textual descriptions described above. Two healthcare domain experts independently label all trials and reconcile discrepancies, ensuring reliability and high inter-annotator agreement. Second, we fine-tune GPT-4o-mini on the OpenAI platform using these 200 manually verified examples as the training set.[14] Fine-tuning forces the model to anchor its reasoning on the specific structure and terminology of Citeline data rather than relying on external assumptions, which substantially reduces hallucination and improves the reproducibility of trial-level classifications. Third, we deploy the fine-tuned classifier across the full set of clinical trials in our

[14] Additional fine-tuning documentation: https://platform.openai.com/docs/guides/supervised-fine-tuning

sample, generating for each trial: (i) a predicted innovation category, and (ii) a confidence score. This produces a scalable measure of whether a trial contains at least some innovative content.

The prompt used for classification is provided below. In total, the model successfully classified all 107,256 trials, of which 35% were labeled as new-drug trials, 23% as new-combination trials, 32% as new-indication trials, and 10% as generic bio-equivalence trials.

```
# GPT prompt
message_content = (
    "We provide the trial titles, targeted drug information, and summary below "
    "in the format of information name: information, seperated by ‘;’. "
    "For each trial, follow a step-by-step process: "
    "Step 1: Answer the question: What drugs are used in the treatment versus "
    "control arms of the clinical trial described below? "
    "Extract treatment versus control arms of the clinical trial, If you do "
    "not find treatment drug or control drop, just say ‘n/a’. "
    "If multiple drugs are in the treatment group, say ‘drugAname+drugBname’. "
    "For durg with both code names and branded names, list both "
    "(branded names in brackets)"
    "Step 2: Answer the question: what type is each trial? Classify the "
    "treatment drugs into one of the four catogories:"
    "‘new drug’, ‘existing drug new indication’, "
    "‘existing drug new combination’, ‘generic drug’. "
    "‘new drug’ is a trial on a drug that is not yet on the market"
    "‘existing drug new indication’ is a investigatory trial to test whether "
    "an exisiting drug can treat a new disease"
    "‘existing drug new combination’ is a trial to test whether drugs can be combined"
    "‘generic drug’ is a trial to test bioequivalence of a drug to an exisiting drug"
    "Put all the information in two steps on one line, in the following format: "
    "’Treatment: [treatment drug]. Control: [control drug]. "
    "Classification: [treatment drug catogeory]’  "
    f"{text}"
)

# Send message_content to the GPT model and retrieve output
completion = call_gpt_api(
    client=client,
    model=MODEL,
    messages=[
        {"role": "system", "content": "You are a healthcare economist"},
        {"role": "user", "content": message_content},
    ],
    temperature=0.5
)

gpt_output = completion.choices[0].message.content.strip()
```

A.4 Linkage across Datasets

A.4.1 Citeline Data and SinoHealth Data

Citeline provides detailed disease information for each clinical trial and drug project, covering nine aggregate disease categories and 312 granular disease subcategories. Disease categories include Oncology, Autoimmune or Inflammation, Central Nervous System (CNS), Metabolic Endocrinology, Cardiovascular, Infectious Disease, Genitourinary, Ophthalmology, and Vaccines. Disease subcategories, for example, are Non–Small Cell Lung Cancer, Type 2 Diabetes, Rheumatoid Arthritis. In contrast, the SinoHealth retail sales data report drug names and sales volumes but do not contain disease classifications. There are 1,377 unique drugs, with different brands of the same drug treated as a single drug. To merge the two datasets, we perform a multi-step mapping procedure that combines structured ATC information with LLM-assisted classification and expert review.

First, we construct candidate disease information for each drug. For every drug in the SinoHealth dataset, we extract (i) the drug name, including all synonyms, and (ii) its ATC (Anatomical Therapeutic Chemical) classification. We also assemble the complete list of 312 disease categories and subcategories from Citeline, which will serve as the candidate set to which each SinoHealth drug can be assigned. Second, we use GPT-4o to generate disease assignments based on drug profiles. We provide GPT-4o with the drug name, its ATC code, and the full set of disease categories. The model is instructed to determine the most likely disease indications for each drug and to output at most five disease categories in ranked order of relevance. This step leverages GPT's ability to interpret pharmaceutical naming conventions, ATC structure, and therapeutic context. Third, we implement manual clinical validation by domain experts. To ensure accuracy and avoid hallucination, every GPT-generated classification is reviewed by a healthcare expert specializing in drug indications. For instance, a typical drug is mapped to three disease categories, reflecting the common presence of multiple approved or off-label indications.

Together, this pipeline enables us to reliably map sales data to disease categories at a level of granularity comparable to Citeline, thereby allowing us to measure disease-specific market expansion and link it to clinical trial activity.

The prompt used for GPT-4o is as follows:

```
# GPT prompt
message_content = (
    "We provide information about drugs and diseases. "
    "For each drug, we include details in Chinese, such as the drug name "
    "and its ATC classification. "
    "Please primarily rely on the most granular information available, "
    "such as the fields ‘shiyingzheng_fenlei’ and ‘atc4_fenlei’, "
    "when making your judgment. "
    "For diseases, we provide a list of disease names (i.e., subcategory) "
    "and their category in a format starting with a numeric disease id. "
    "Please determine which disease(s) this drug is likely used to treat, "
    "based on the drug’s information and your knowledge. "
    "Select AT MOST 5 relevant diseases from the disease list, "
    "ranked in descending order of relevance. "
    "Please ONLY output from the disease list in its original format. "
    "Do NOT include any other information. "
    "Return in the format starting with: ‘Answer:’ followed by the diseases "
    "separated by periods. "
    "Example output: "
    "Answer: 1-(disease category: ...); (disease subcategory: ...). "
    "2-(disease category: ...); (disease subcategory: ...)."
    " "
    f"{text}"
)

# Send message_content to the GPT model and retrieve output
completion = call_gpt_api(
    client=client,
    model=MODEL,
    messages=[
        {"role": "system", "content": "You are a healthcare economist"},
        {"role": "user", "content": message_content},
    ],
    temperature=0.5
)

gpt_output = completion.choices[0].message.content.strip()
```

A.4.2 Citeline Data and OpenAlex Data

OpenAlex provides comprehensive, high-quality metadata on scientific publications, including publication year, author names, affiliations, citations, journal identifiers, and concept tags that summarize a paper's scientific content (Priem et al., 2022). To study knowledge production relevant to biopharma innovation, we focus on papers whose first or last authors are affiliated with institutions in China or the United States and restrict the sample to publications with valid PubMed IDs to ensure biomedical relevance. In total, there are 33,809,530 OpenAlex publications with PubMed IDs. Among these, 9,734,217 of them are affiliated with US, and 2,321,944 are affiliated with China.

Linking OpenAlex publications to Citeline disease categories is challenging because the two databases use entirely different classification systems. OpenAlex concepts are algorithm-generated scientific keywords (e.g., "protein folding," "mental health," "immunotherapy"), whereas Citeline assigns each clinical trial or drug project to a structured hierarchy of nine major disease categories and 312 granular subcategories. Moreover, the relationship between OpenAlex concepts and Citeline disease categories is complex. OpenAlex concepts are substantially more granular and overlapping than disease categories, often requiring many-to-many mappings rather than one-to-one matches. In addition, the OpenAlex concept hierarchy is not a tree—concepts may have multiple parents—making it structurally incompatible with the strictly hierarchical disease classification system. Finally, many concepts are semantically ambiguous or not directly tied to a specific disease, so surface-level keyword matching fails to capture their relevance.

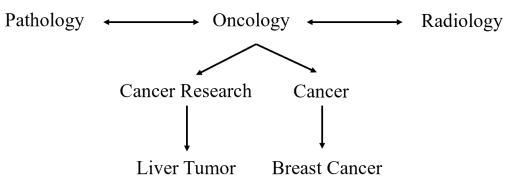
To bridge these systems systematically and avoid ad hoc manual matching, we construct a unified semantic space in which both OpenAlex concepts and Citeline disease categories can be jointly represented and compared using graph embedding.

First, we build concept–disease graphs in OpenAlex and Citeline: We extract all OpenAlex concepts associated with biomedical papers. OpenAlex provides two types of relationships among concepts: "ancestor–descendant" (directed edges): hierarchical scientific topics

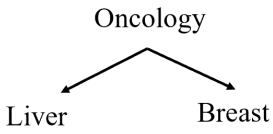
(e.g., "oncology → breast cancer"). "related" (undirected edges): semantically similar topics frequently co-occurring in research. We construct a graph where each concept is a node and each hierarchical or related relationship is an edge. Separately, we build an analogous graph for Citeline diseases, where nodes represent disease categories or subcategories and edges represent their structured parent–child hierarchy (e.g., "oncology → non–small cell lung cancer").

Figure A.6: Example Graph - OpenAlex Concepts and Citeline Disease Categories

(a) OpenAlex Concept



(b) Citeline Disease Category



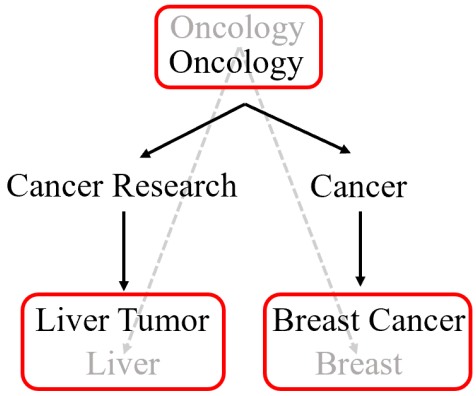
Second, we embed concepts and diseases into vector space using *node2vec* (Grover and Leskovec, 2016). We convert each graph into a high-dimensional vector representation using the node2vec algorithm, which embeds nodes so that: concepts frequently appearing together or closely connected in the graph lie near each other, disease categories in Citeline are similarly embedded based on their hierarchical structure.

This produces two embedding spaces: one for OpenAlex concepts and one for Citeline diseases. Embeddings learned from different graphs of databases (i.e., OpenAlex and Citeline) are not directly comparable because their coordinates are identified only up to an arbitrary rotation. We therefore align the two embedding spaces using a set of anchor pairs, that is concept (OpenAlex) – disease category (Citeline), which provide common reference points across graphs. This alignment preserves within-graph structure while making cross-graph similarity measures interpretable.

One example is shown in Figure A.7. In this example, *Liver Tumor* and *Breast Cancer* are OpenAlex concepts under the broader concept *Oncology*, while *Liver* and *Breast* are Citeline disease subcategories under the disease category *Oncology*. We treat (*Liver Tumor*, *Liver*) and (*Breast Cancer*, *Breast*) as anchor pairs. These anchors provide shared reference points

that allow us to rotate the two embedding spaces into a common coordinate system. After alignment, nodes that are close to these anchors in their respective graphs—illustrated by the local structure around *Oncology* and its neighbors—become comparable across databases, so that cross-graph similarity between a Citeline disease subcategory and an OpenAlex scientific concept has a consistent interpretation.

Figure A.7: Example - Alignment of Two Embedding Spaces



Third, we operationalize this alignment by constructing anchor pairs. To place OpenAlex concepts and Citeline diseases into a shared semantic space, we identify "anchor pairs" — disease (Citeline)–concept (OpenAlex) matches that refer to the same scientific topic. For each Citeline subcategory, we search OpenAlex concepts for lexical matches, synonyms, or closely related terms, yielding 15,958 initial linkages across the 312 subcategories. Each linkage is then evaluated through two layers of validation: manual review by research assistants and a GPT-4o assessment of semantic similarity. After filtering, 12 subcategories had no suitable matches and 23 had only one; for these 35 cases, research assistants identified additional synonyms or clinically equivalent concepts. Following these adjustments, all disease subcategories received valid matches, typically with 2–10 high-quality anchors each. These anchors allow us to align the two embedding spaces and ensure that semantically similar OpenAlex concepts lie close to their corresponding Citeline disease categories, forming the basis for constructing the disease-specific knowledge measures used in the main analysis.

After this process, we retrieve the embeddings of all concepts associated with each publication, align them using the anchor pairs, and compute similarity scores. Specifically, for each publication, we first obtain its concept embeddings and align them via the anchor pairs. We then compute a weighted average of these concept embeddings using the associated confidence scores. The resulting publication-level vector is compared with all disease vectors using cosine similarity, and the five diseases with the highest similarity scores are selected. The method captures the general thematic focus of publications (e.g., anxiety, liver, renal) for about 60% of publications, with similarity scores greater than or equal to 0.5.

```
prompt = f"""
    Here is an example of questions and answers:
    group \t disease_category \t disease_subcategory \t display_name
    90 \t Cardiovascular \t Peripheral Arterial Disease \t Peripheral arterial occlusive
    disease
    90 \t Cardiovascular \t Peripheral Arterial Disease \t Occlusive arterial disease
    90 \t Cardiovascular \t Peripheral Arterial Disease \t Hirschsprung's disease

    Let's review the matchings for Cardiovascular - Peripheral Arterial Disease (PAD).
    I will analyze each entry and mark cases that are acceptable but not precise.

    Group 90 - Peripheral Arterial Disease
    1. PAD -> Peripheral arterial occlusive disease
    Match score: 1
    Reason: Direct concept match, standard term for PAD

    2. PAD -> Occlusive arterial disease
    Match score: 0.5
    Reason: Broader concept, not specific

    3. PAD -> Hirschsprung's disease
    Match score: 0
    Reason: Unrelated, different system

    Now, please evaluate the following entry:
    disease_category: {category}
    disease_subcategory: {subcategory}
    display_name: {concept}

    Please output:
    Match score: [1 / 0.5 / 0]
    Reason: [in Chinese, within 20 words]
    """

prompt = f"""
    You are a medical terminology audit expert. Please evaluate the relationship
```

```
    between a disease subcategory and a medical concept, using the following criteria:

    - If the concept is a broader or more general category (a hypernym) of the
    subcategory → Output: 0.2, Reason: concept is broader
    - If the concept is a narrower or more specific instance (a hyponym) of the
    subcategory → Output: 0.8, Reason: concept is more specific
    - If the concept is at the same level (e.g., treatment method, diagnostic
    tool, body location) → Output: 0.5, Reason: same-level concept

    Output strictly in the following format:
    Match Score: [0.8 / 0.5 / 0.2]
    Reason: [in Chinese, 10-20 characters]

    disease_category: {cat}
    disease_subcategory: {subcat}
    concept: {concept}
"""
```

A.5 Auxilary Data Cleaning and Variable Construction

A.5.1 Papers and Scientists

Once each paper is linked to one or more Citeline disease categories, we construct disease–year level measures of knowledge quantity, knowledge quality, and international talent flows. These measures aggregate paper-level or author-level information to the disease-country-year unit of analysis used in the main text.

Knowledge quantity (number of papers). Knowledge quantity is defined as the total number of PubMed-indexed papers in a given disease–country–year. If a paper maps to multiple diseases, we assign fractional counts based on the number of disease assignments. The distribution of paper counts varies substantially across diseases, with a median of 31 papers and a mean of 199 papers at disease-country-year level.

Knowledge quality (number of highly cited papers). Knowledge quality is measured as the number of "high-impact" papers in each disease–country–year, defined as those in the top 5% of citations among all papers published in the same year. This adjusts for differences in citation windows across publication cohorts. Papers mapped to multiple diseases are again

fractionally allocated.

Talent flows (weighted number of scientists with U.S. experience). Talent flows capture the presence of researchers with international training or collaboration. A scientist is classified as having U.S. experience if they have published at least one paper with a U.S. affiliation on or before year t. We then compute, for each disease–year, the weighted sum of such scientists, where each scientist's weight equals the share of their total publications associated with that disease category. For example, if 50% of a scientist's publications map to Disease A and 20% to Disease B, they contribute weights of 0.5 and 0.2, respectively. There are 7,552,293 unique scientists in OpenAlex, and among them, 2,539,289 have prior experience in the United States.

Together, these measures allow us to quantify how scientific knowledge and international talent evolve across diseases and over time, and to examine whether these upstream factors help explain the rise of clinical trial activity documented in the main analysis.

A.5.2 Application Backlog Time

The clinical trial application reform in 2015, studied in Jia et al. (2024), streamlined regulatory procedures by clearing the backlog of pending IND (investigational new drug) applications, which are required prior to initiating clinical trials for new drugs under investigation. To calculate application backlog days across diseases and over time, we collect the application and approval dates for almost all IND applications from the Center for Drug Evaluation of China's NMPA. The detailed data collection and cleaning process is as follows.

First, we scrape the application and approval dates for all IND applications from the Center for Drug Evaluation of China's NMPA (National Medical Products Administration). The NMPA discloses all drug applications on its website,[15] with information including "Application ID," "Drug Name," "Drug Type," "Application Type," "Registration Category," "Company Name," and "Submission Date." In addition, the NMPA separately lists the approval date for

[15]Website link: https://www.cde.org.cn/main/xxgk/listpage/9f9c74c73e0f8f56a8bfbc646055026d

each application on its website.[16]

Second, we restrict the sample to applications whose fourth character is the letter "L," indicating clinical trial applications, and whose "Application Type" is 1 or 2, which correspond to innovative drugs and are therefore relevant for our focus on IND applications. Moreover, in July 2018, the NMPA introduced the implied approval (silent approval) system, under which applications submitted for more than 60 days are automatically approved. This policy change is reflected in the data: after its introduction, most applications no longer received mailed approval documents, indicating that the silent approval mechanism was effectively in operation. However, because some applications were still rejected, we additionally scrape the database of silent approvals and match entries from this database to the main dataset using the application acceptance number.

Finally, we link these IND drugs to their corresponding disease categories and disease subcategories. We first use large language models to classify diseases based on drug names, company information, and drug types. We then search the drug names in the Citeline Pharmaprojects database, which provides drug name histories, corresponding diseases, and textual descriptions. We feed this additional information back to a large language model to perform reclassification. This procedure yields 15,079 IND applications with classified diseases and calculated application backlog days.

```
# GPT prompt
def classify_direct(drug_name, drug_type, company):
    prompt = f"""
        You are a senior pharmaceutical and clinical indication expert.
        Please determine the disease category and disease subcategory based on
        the drug name I provide, and return the confidence level of this judgment.
        First choose ONE disease_category.
        Then choose ONE disease_subcategory ONLY from that category.

def classify_with_evidence(drug_name, drug_type, company, candidates):
    evidence = "\n\n".join(f"[{i+1}] {c['summary']}" for i, c in enumerate(candidates))
    prompt = f"""
        You are a senior pharmaceutical and clinical indication expert.
        Please determine the disease category and disease subcategory based on
```

[16]Website link: https://www.nmpa.gov.cn/zwfw/zwfwpjfbzs/pjfbzsyp/index.html

```
the drug name and evidence I provide, and return the confidence level
of this judgment.
First choose ONE disease_category.
Then choose ONE disease_subcategory ONLY from that category.
```

A.5.3 Policy Documents and Classification

We also scraped all biopharma-related policies in China from the State Council Policy Document Library. There are 840 policy documents relevant to biopharma innovation. We use GPT to classify these documents into four categories: (a) advanced manufacturing policies that promote the scale-up and modernization of industry production capacity; (b) innovation independence policies that aim to strengthen domestic R&D capabilities and reduce reliance on foreign technologies; (c) capital market reform policies that lower financing barriers and improve firms' access to external funding; and (d) tax and subsidy policies that provide fiscal incentives to support innovation. The detailed steps are as follows:

To construct this dataset, we first scraped all policy documents available in the State Council Policy Document Library, obtaining 18,488 entries. We then implemented a two-stage classification procedure using GPT-4o. In the first stage, we used a broad screening prompt to filter out all policies potentially related to the four categories of innovation policy described above. This step substantially reduced the number of documents requiring manual or fine-grained review. In the second stage, we applied a more detailed prompt to each remaining document, asking the model to assign (i) a relevance score capturing how closely the policy pertains to biopharma innovation and (ii) a brief explanatory note summarizing the policy's purpose and mechanism. The first stage returned 3,474 policy documents relevant to innovation. These documents were further scored in the second stage. Among them, 840 are related to biopharma innovation, and 273 have a relevance score greater than 0.5. These outputs were used both to validate relevance and to facilitate the final four-way policy categorization described above.

The two prompts used in this procedure are provided below:

```
# GPT prompt
SYSTEM_PROMPT = (
    "You are a careful Chinese policy classification expert."
    "You will receive UP TO 10 Chinese policy documents at once."
    "For EACH document, you must extract:"
    "1) the policy date (政策日期) from the text "
    "(any explicit date of issuance/publishing/promulgation; "
    "return as the original string; if none, return \"\")."
    "2) whether it clearly relates to these 4 themes (0/1):"
    "   - high_end_mfg: 高端制造业/高端装备/战略性新兴产业/高端生物医药制造 等。"
    "   - innovation_independence: "
    "     自主创新/科技自立自强/提升原始创新能力/创新平台与人才基地建设 等。"
    "   - capital_market_reforms: 支持通过资本市场/多层次资本市场/股权基金/债券/上市/并
购/"
    "     融资工具来促进上述产业或创新的政策。"
    "   - tax_subsidy: 对医药/药物研发/临床试验/"
    "     创新平台等提供的减税、税收优惠、财政补贴、进口免税、费用减免、贷款贴息等。"
    "Return STRICT JSON ONLY, as a list of objects, in the SAME ORDER as input."
)
```

```
# GPT prompt
SYSTEM_PROMPT = (
    "You are a senior policy evaluation expert "
    "working inside a pharmaceutical company in China."
    "You will be shown Chinese policy documents."
    "Your task is to judge **specifically from the perspective "
    "of a drug/biopharmaceutical R&D enterprise** whether this policy creates "
    "incentives to invest more in drug R&D."
    "When you evaluate, consider the following channels:"
    "1) The policy directly reduces R&D cost "
    "(税收优惠、财政补贴、进口免税、费用加计扣除、贷款贴息等)."
    "2) The policy increases the expected revenue/profit from new drugs "
    "(医保支付、带量采购配套激励、市场准入加速、知识产权保护等)."
    "3) The policy expands or improves the domestic human capital / innovation"
    "ecosystem for biopharma (人才计划、平台建设、国家重点实验室、创新载体、医工结合)."
    "4) The policy explicitly includes ‘高端生物医药制造’, ‘生物医药’, ‘医药产业’ "
    "within broader high-end manufacturing / strategic emerging industries "
    "and provides concrete support measures."
    "Important:"
    "- Some policies talk about ‘high-end manufacturing’, "
    "‘strategic emerging industries’, or ‘innovation-driven development’. "
    "You must judge whether biopharma / drug R&D is "
    "**actually covered or very likely covered** by that scope."
    "- If the policy is very general, very macro, or only about administration"
    "without concrete support instruments, the incentive should be low."
    "- Output a score between 0 and 1 (inclusive). "
    "0 = no meaningful incentive for drug R&D. 1 = very strong and direct incentive. "
    "Values in between (e.g. 0.3, 0.5, 0.8) represent your strength of belief."
    "Return STRICT JSON ONLY, in a list of objects, in the same order as input."
)
```