Supplementary Material for "Estimating and Predicting Clinical Trial Outcomes in China"

Chaoyi Zhao, Timothy Pang, and Andrew W. Lo

A Literature Review

A well established body of research has characterized the average risks of drug development using historical success rates, often summarized as PoSs for phase transitions and for the overall Phase I-to-Approval pathway. Prior work has emphasized the high attrition rates in clinical development and argued that better termination and selection rules could materially improve R&D productivity.[1] Subsequent large-sample studies have provided widely used phase-transition benchmarks across therapeutic areas and time periods.[2,3] Methodological advances have also addressed practical complications that arise in registry-based settings—censoring, missing transitions, and outcome lags—leading to PoS estimators that can be applied to large observational datasets.[4] These PoS estimates provide critical aggregate benchmarks, but they do not directly answer the operational question that sponsors and investors face: *which* specific ongoing trials are more likely to succeed versus fail?

This gap has motivated a fast-growing literature on trial-level predictive modeling. Using inputs ranging from drug and target properties to subject-level early-phase data and structured trial characteristics, prior studies have shown that machine learning classifiers can achieve meaningful discrimination between trials that eventually succeed versus fail.[5–7]

A particularly important development has been the use of unstructured text—trial titles, objectives, and eligibility criteria—because such narrative fields often encode scientific rationale, patient selection logic, endpoint choice, and operational complexity issues that are difficult to represent with a small set of structured features. Recent work has therefore applied natural language processing methods and pretrained language models to clinical-trial information, reporting strong predictive performance by exploiting complex structures and textual semantics.[8,9]

Most of this predictive modeling literature relies on ClinicalTrials.gov or global trial databases that aggregate multiple registries.[5,10] ClinicalTrials.gov, in particular, has been a foundational resource because it offers large-scale standardized fields and increasingly comprehensive results reporting over time. Building on such data, researchers have reported performance metrics such as AUC and F1 scores that provide useful reference points for what is achievable in trial-outcome prediction.[5,6,8,9]

Taken together, prior work has produced refined aggregate PoS estimates and trial-level prediction models that incorporate increasingly diverse registry information. However, both strands have developed primarily with ClinicalTrials.gov or global commercial data. Our study connects them in a China-specific setting by estimating phase-transition benchmarks and evaluating whether structured and Chinese-language registry information can rank pending trials. This design complements global evidence while allowing the assumptions and predictive limits of registry-based analysis to be examined in a distinct institutional and reporting environment.

B Data and Descriptive Statistics

This appendix describes the data underlying our analysis. Section B.1 introduces the datasets and the variables we construct from them, and Section B.2 reports descriptive statistics for drug development projects and clinical trials.

B.1 Datasets and Variables

Our dataset was downloaded from the DXY Insight database; see `https://db.dxy.cn/`. This database includes all clinical trials registered at China's Drug Clinical Trial Registration and Information Disclosure Platform (hereinafter, the Platform) since 2013, the year of the foundation of the Platform. This Platform was established by the Center for Drug Evaluation (CDE) at the former China Food and Drug Administration (CFDA)—now the National Medical Products Administration (NMPA). The Platform was first announced on November 1, 2012, marking the beginning of its pilot phase. Following a series of official notices and implementation guidelines issued in late 2012 and in 2013, the independent platform website was officially launched on November 25, 2013. Since the release of CFDA Announcement No. 28 of 2013, registration on the Platform of all drug clinical trials conducted in China has been mandatory. The Platform serves as the official national registry for drug clinical trials, providing comprehensive, authoritative, and publicly accessible information on clinical trial approvals and progress in China. See `http://www.chinadrugtrials.org.cn/snipet/434.html` for a history of the Platform.

The DXY Insight database provides two datasets: the dataset of drug development projects (the Project Dataset) and the dataset of clinical trials (the Trial Dataset). The Project Dataset and the Trial Dataset are closely related and complementary. Each drug development project represents a combination of all clinical trials at different phases but conducted by the same sponsor for the same drug in the same indication. In other words, the Project Dataset aggregates information at the project level, while the Trial Dataset provides trial-level details for each phase in that project. Together, these two datasets form a hierarchical structure that enables linking of overall drug development progress with underlying clinical evidence from relevant clinical trials.

Project Dataset. The Project Dataset includes 5,648 drug development projects for non-generic drugs from 2013 to 2025. It summarizes information at the project level, covering the composition and characteristics of each drug, participating companies, and development progress. This dataset includes variables such as First Disclosure Date, Sponsor, Biologic/Chemical, Domestic/Imported, and Global Launch Status (whether the active ingredient is already marketed somewhere in the world), and links each project to its associated trials via Trial IDs. It also tracks Current Project Status (approved or not approved).

Trial Dataset. The Trial Dataset includes 14,769 clinical trials for non-generic drugs from 2013 to 2025. It records detailed trial-level information registered on the Platform. It includes Trial ID, First Disclosure Date, Completion Date, Public Title, Official Title, Phase,

Therapeutic Area, Indication, and Scope. It also includes information related to the sponsor, such as the Sponsor Name, Parent Group, and Number of Participating Institutions. Regarding the therapy tested, it includes information about Drug Name, Drug Ingredient, Target, Modality, Dosage Form, and Line. It also includes design information such as Objectives, Trial Design, Inclusion Criteria, Exclusion Criteria, Target Domestic Enrollment, Target International Enrollment, Minimum Enrollment Age, Maximum Enrollment Age, Combination Therapy, and Control Drugs. Table B.1 lists the value types for all variables in the Trial Dataset, and Table B.2 provides all values that categorical variables can take.

Table B.1: Variables in the Trial Dataset.

Name	Type	Name	Type
Trial ID	Text	Target	Text
First Disclosure Date	Date	Modality	Categorical
Completion Date	Date	Dosage Form	Categorical
Public Title	Text (in Chinese)	Line	Categorical
Official Title	Text (in Chinese)	Objectives	Text (in Chinese)
Phase	Categorical	Trial Design	Text (in Chinese)
Therapeutic Area	Categorical	Inclusion Criteria	Text (in Chinese)
Indication	Text (in Chinese)	Exclusion Criteria	Text (in Chinese)
Scope	Categorical	Target Domestic Enrollment	Integer
Sponsor Name	Text (in Chinese)	Target International Enrollment	Integer
Parent Group	Text (in Chinese)	Minimum Enrollment Age	Number
Number of Participating Institutions	Integer	Maximum Enrollment Age	Number
Drug Name	Text (in Chinese)	Combination Therapy	Text (in Chinese)
Drug Ingredient	Text (in Chinese)	Control Drugs	Text (in Chinese)

Table B.2: Categorical variables in the Trial Dataset.

Name	Values
Phase	I; II; III; IV; I/II; II/III; III/IV; BE
Therapeutic Area	Immunology; Endocrinology; Otolaryngology; Respiratory; Cardiovascular; Infectious; Urology; Gastroenterology; Reproductive; Dermatology; Ophthalmology; Neurology; Psychiatry; Rare; Oncology; Hematology; Musculoskeletal; Other
Scope	Domestic; International
Modality	ASO, CAR-T, PROTAC, siRNA, Other Protein, Other Diagnostic Reagent, Chemical, Monospecific Antibody, Bispecific Antibody, Trispecific Antibody, Gene Therapy, Peptide, Microbiome-Based, Antibody–Drug Conjugate, Antibody Fusion Protein, Undetermined, Therapeutic Vaccine, Therapeutic Radiopharmaceutical, Mixed Antibody, Oncolytic Virus, Cyclic Peptide, Cytokine, Diagnostic Radiopharmaceutical, Enzyme, Mesenchymal Stem Cell, Non-Antibody Fusion Protein, Non-Degrading Molecular Glue, Molecular Glue Degrader, Prophylactic Vaccine; Other
Dosage Form	Dermatological Preparation; Diagnostic and Radioactive Reagent; Implant; Inhaled Preparation; Injectable Preparation; Nasal Preparation; Ophthalmic Preparation; Oral Liquid Dose Preparation; Oral Patch; Oral Solid Dose Preparation; Otic Preparation; Patch; Suppository; Vaginal Preparation; Other
Line	First; Second; Third; Last; Adjuvant; Neoadjuvant; Maintenance/Consolidation; Other

Among all of these fields, the Dosage Form field in the Trial Dataset is recorded at a very granular level. To ensure that each category used in our analysis contains a sufficient number of observations, we group these detailed Dosage Forms into broader classes as summarized in Table B.3.

Of the 14,769 clinical trials in the Trial Dataset, 13,339 clinical trials appear in the Project Dataset, covering 90.32% of all trials. In our analysis, we match both datasets and focus on the 13,339 trials that appear in both datasets only.

Table B.3: Trial Dataset dosage form grouping details.

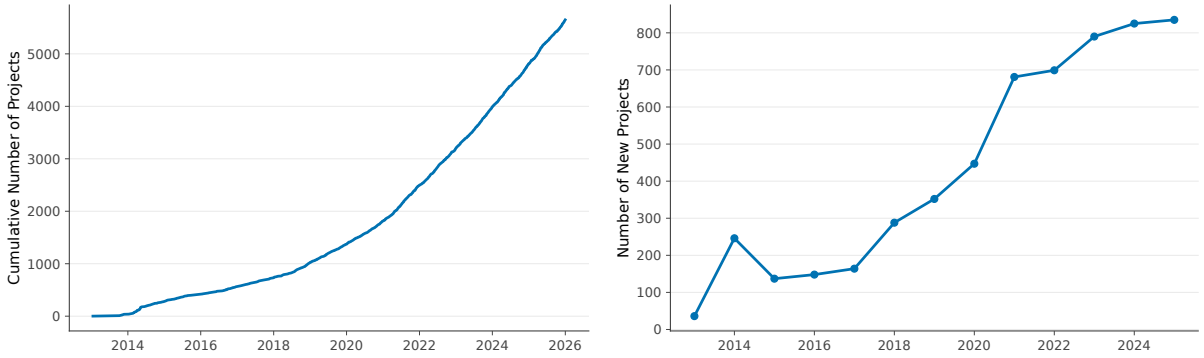
Class	Dosage Form
Dermatological Preparation	Gel; Ointment/Cream/Plaster; Topical Liquid Preparation
Diagnostic and Radioactive Reagent	Diagnostic and Radioactive Reagent
Implant	Implant
Inhaled Preparation	Aerosol/Spray/Powder Inhaler; Inhaler
Injectable Preparation	Emulsion for Injection; Injection; Liposomes for Injection; Microspheres for Injection
Nasal Preparation	Nasal Preparation
Ophthalmic Preparation	Ophthalmic Preparation
Oral Liquid Dose Preparation	Drop Preparation; Emulsion; Oral Liquid Preparation; Suspension
Oral Patch	Oral Patch
Oral Solid Dose Preparation	Chewable Tablet; Dispersible Tablet; Enteric Coated Capsule; Enteric Coated Tablet; Granule; Immediate Release Tablet; Lozenge; Orally Disintegrating Film; Oral Disintegrating Tablet; Powder; Regular Capsule; Regular Tablet; Soft Capsule; Sublingual Tablet; Sustained Release/Controlled Release Capsule; Sustained Release/Controlled Release Tablet
Otic Preparation	Otic Preparation
Patch	Patch
Suppository	Suppository
Vaginal Preparation	Vaginal Preparation
Other	Other

B.2 Descriptive Statistics

We report descriptive statistics for the two units of analysis in turn: drug development projects (Section B.2.1) and individual clinical trials (Section B.2.2).

B.2.1 Drug Development Projects

Figure B.1 provides an overview of the evolution of the 5,648 drug development projects registered on the Platform. Throughout this paper, the year of each project is defined based on its First Disclosure Date. Figure B.1a plots the cumulative number of projects over time, while Figure B.1b reports the annual number of newly disclosed projects.



(a) Cumulative numbers.

(b) Annual increments.

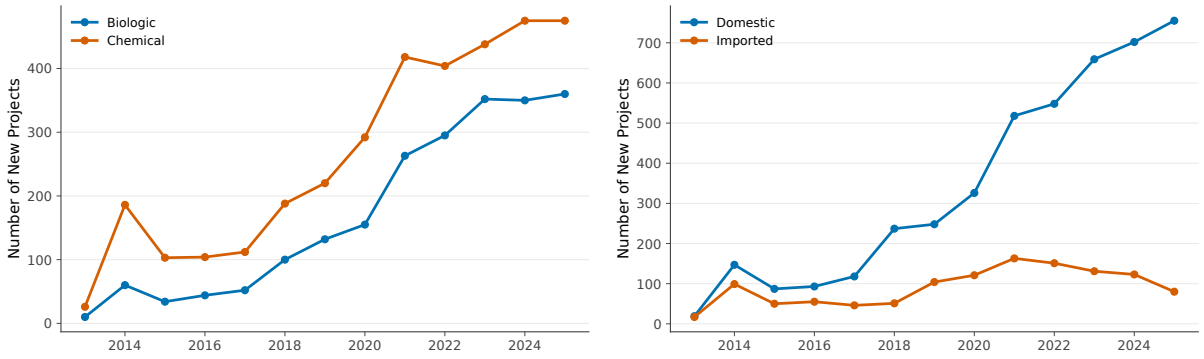
Figure B.1: Cumulative numbers (a) and annual increments (b) of drug development projects in our dataset.

Both panels reveal a clear and persistent upward trajectory in China's drug development activity since 2013. The total number of projects has expanded rapidly, reflecting the

substantial growth of the domestic pharmaceutical and biotechnology sectors as well as the increasing emphasis on clinical research following a series of regulatory reforms.

The noticeable acceleration after 2017 coincides with major policy changes introduced by the NMPA, including reforms to review and approval procedures and the establishment of more innovation-oriented development pathways. The October 8, 2017 *Opinions on Deepening the Reform of the Review & Approval System to Encourage Innovation in Drugs and Medical Devices* marked the deregulation of certain aspects of clinical-trial applications and expanded priority review and approval for innovative and imported drugs. These trends highlight the ongoing expansion and diversification of China's clinical development landscape. While in the first decade of this century, biopharmaceutical innovation remained centered in the U.S., Europe, and Japan,[11] these policy changes and the attendant maturation of the Chinese biopharmaceutical industry coincided with the consistent upward trajectory in China's drug development activity shown in Figure B.1.

Figure B.2 further characterizes the composition of new drug development projects over time. We classify projects into biologic versus chemical drugs (Figure B.2a) and domestic versus imported drugs (Figure B.2b) based on the Biologic/Chemical and Domestic/Imported variables, respectively.



(a) Biologic versus chemical drugs.

(b) Domestic versus imported drugs.

Figure B.2: Numbers of projects of biologic versus chemical drugs (a) and domestic versus imported drugs (b) in the new drug development projects in each year.

Figure B.2 shows two clear patterns. First, in Figure B.2a, the numbers of both chemical and biologic projects increase over time, with a pronounced acceleration after around 2017, consistent with the overall expansion of new drug development activity shown in Figure B.1. Chemical projects remain more numerous than biologic projects throughout the sample period, which is consistent with the fact that small-molecule development is generally more established and scalable.

Second, in Figure B.2b, domestic projects increase substantially and dominate the project counts in later years, whereas imported projects remain comparatively flat. This divergence suggests that the observed growth is primarily driven by domestically developed pipelines rather than by an increase in imported programs, which may reflect greater domestic R&D capacity and stronger incentives for local development during the later period.

Figure B.3 shows the top 15 companies sponsoring the highest number of drug develop-

ment projects in our dataset. Jiangsu Hengrui, Novartis, Chia Tai TianQing Group, Qilu, and AstraZeneca are the most prolific sponsors, sponsoring 92, 70, 65, 60, and 59 projects, respectively. It is worth noting that the top-ranked firms include both major domestic pharmaceutical companies and multinational companies, such as Novartis, AstraZeneca, Pfizer, Roche, Sanofi, Eli Lilly, Merck & Co., and Bayer. This pattern suggests that the drug development projects registered on the Platform reflect participation from both China-based and global R&D players, rather than being dominated by a single market or sponsor type.

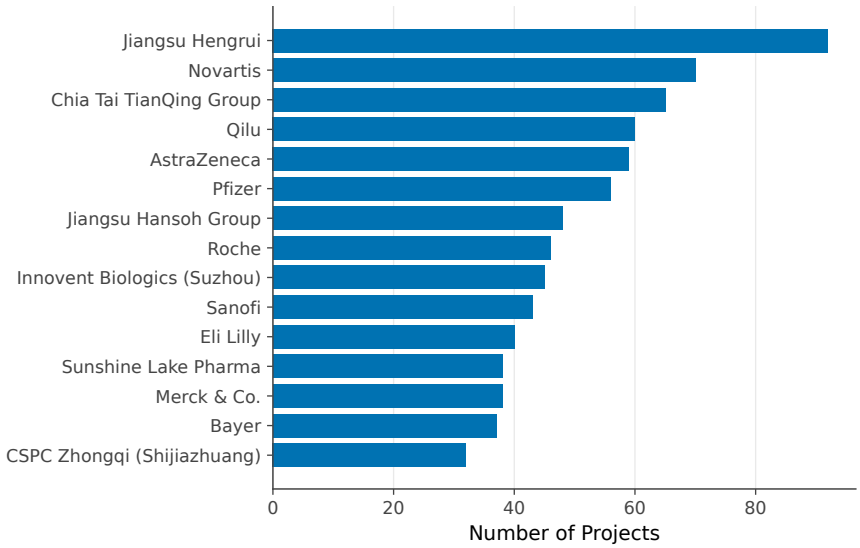
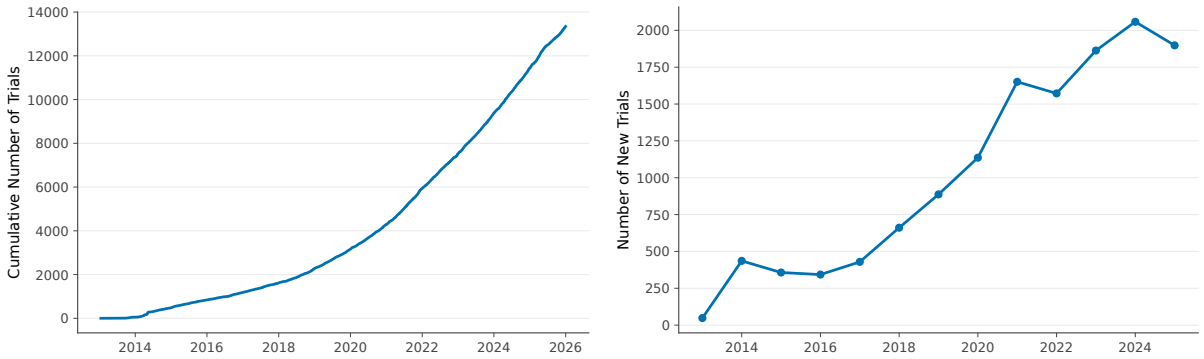


Figure B.3: Top 15 companies with the highest number of drug development projects.

B.2.2 Clinical Trials

After characterizing these drug development projects, we turn to study the 13,339 clinical trials in the Platform in China. Figure B.4 shows the cumulative numbers (Figure B.4a) and annual increments (Figure B.4b) of clinical trials. Figure B.4a shows a steadily accelerating accumulation of registered clinical trials over time. Figure B.4b confirms that this acceleration is driven by rising annual inflows: yearly new trials grow from a few hundred in the mid-2010s to well above 1,500 in the early 2020s, peaking above 2,000 by 2024. Overall, Figure B.4 suggests sustained growth in clinical trial activity in China, with a clear step-up in momentum in the last several years.

Figure B.5 summarizes the therapeutic-area composition of all clinical trials in our sample as of the end of 2025. In the Trial Dataset, a single clinical trial may be tagged with multiple therapeutic areas. In such cases, we use a non-mutually-exclusive counting scheme: the trial is counted once within each associated area. For example, if a trial is classified as both Oncology and Cardiovascular, it is included in the counts for both Oncology and Cardiovascular. We use this principle throughout the analysis in this article. Oncology dominates the landscape, accounting for more than two-fifths of trials (42.7%), far exceeding any other single area. This suggests that, in China's clinical trial landscape, cancer-related



(a) Cumulative numbers.

(b) Annual increments.

Figure B.4: Cumulative numbers and annual increments of clinical trials.

R&D has been a particularly prominent focus.[12] The next largest categories are Immunology (11.4%), Infectious Disease (9.6%), and Endocrinology (9.4%).

Following the cross-sectional snapshot in Figure B.5, we next examine how the therapeutic-area composition evolves over time. Figure B.6 reports the annual numbers of newly registered clinical trials by therapeutic area, separating the top five areas (Figure B.6a) from the remaining areas (Figure B.6b); the top five are defined by ranking areas according to their total trial counts aggregated across all years in the full dataset.

The time-series patterns are broadly consistent with Figure B.5. Oncology dominates throughout the period and exhibits a pronounced expansion after the late 2010s, followed by a high plateau from 2021–2025. In contrast, most non-top-five areas contribute smaller but steadily increasing numbers of new trials, with many showing accelerated growth after around 2018–2020, indicating a broad-based expansion of clinical trial activity beyond the largest therapeutic areas.

Figure B.7 further breaks down the clinical-trial landscape by showing the phase composition in each therapeutic area. The stacked bars indicate how trials in each area are distributed across Phase I, Phase II, Phase III, Phase IV, and combined-phase categories (I/II, II/III, and III/IV), providing a more granular view of where development activity is concentrated along the clinical pipeline. BE refers to bioequivalence studies, which test whether a candidate product has equivalent exposure to a reference drug; such studies are rare in the Trial Dataset for non-generic drugs.

Across most therapeutic areas, Phase I trials account for the larger share, with Phase II and Phase III contributing substantial additional volumes. In contrast, Phase IV trials are rare in the dataset. Combined-phase trials appear more frequently in the Oncology area than in other therapeutic areas, which is broadly consistent with international development practices in oncology that often employ seamless or adaptive phase transitions to accelerate clinical development.

Figure B.8 shows a histogram of the number of participating institutions per clinical trial. Although most trials involve only a small number of participating institutions, the distribution has a pronounced right tail: many trials involve dozens of participating institutions. In particular, over 1,500 trials are conducted by more than 50 institutions. This pattern

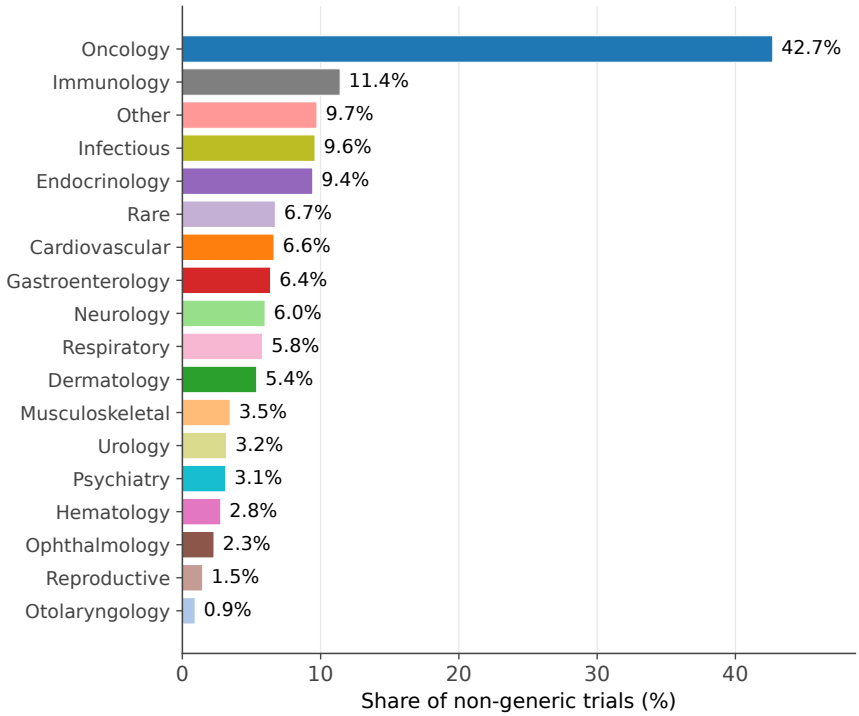
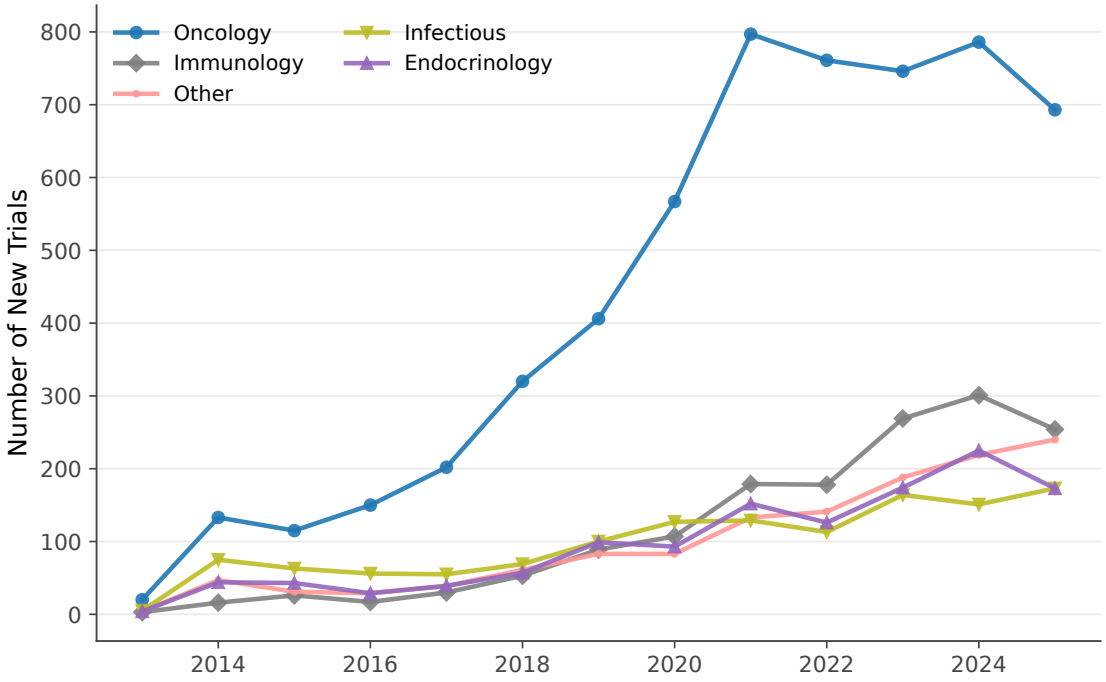
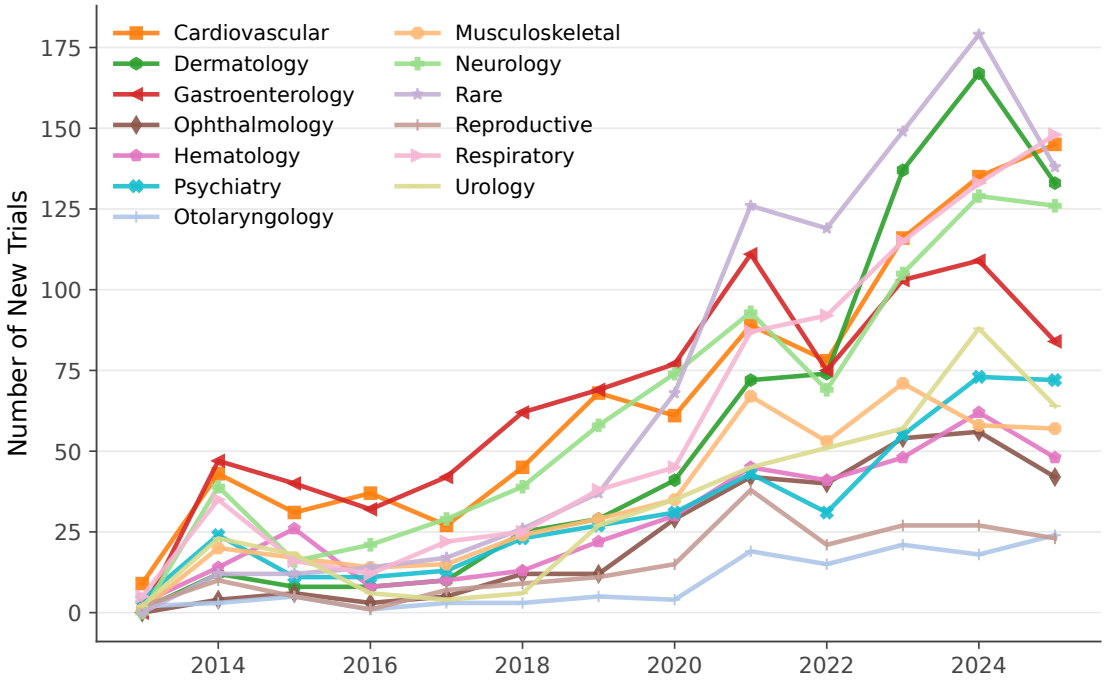


Figure B.5: Therapeutic area distribution of all clinical trials as of the end of 2025. Areas are non-mutually exclusive (a trial may be tagged with multiple areas), so shares need not sum to 100%.

highlights the fact that large-scale collaborative trials are common in our Trial Dataset and likely reflect the need for broader recruitment and coordination in more resource-intensive study designs.



(a) Top 5 therapeutic areas.



(b) Other therapeutic areas.

Figure B.6: Numbers of new clinical trials in the top five therapeutic areas (a) and the remaining therapeutic areas (b) in each year. The top five therapeutic areas are ranked by the total number of trials across all years in the full dataset.

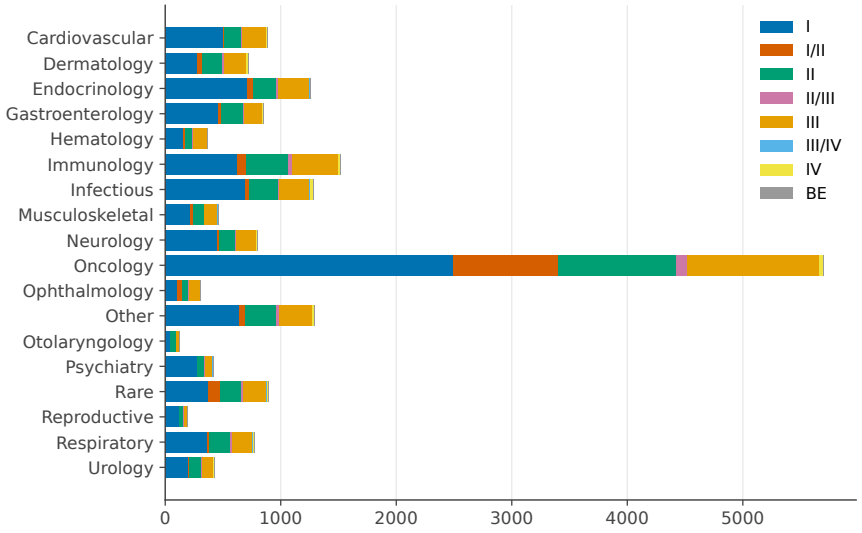


Figure B.7: Numbers of clinical trials by phase in each therapeutic area.

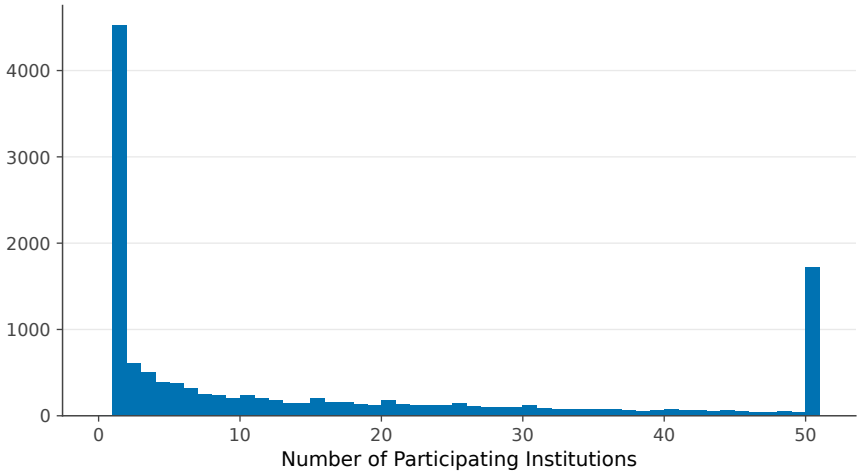


Figure B.8: Histogram of the number of participating institutions per clinical trial. Trials with over 50 participating institutions are truncated at 50.

C PoS Estimation

This appendix provides details and additional results for our PoS estimation. Section C.1 describes the estimation method; Section C.2 examines missing phase records and registry coverage; Section C.3 reports the underlying trial-outcome counts and the estimated PoS by therapeutic area; Section C.4 compares our Chinese estimates with global benchmarks; and Section C.5 assesses robustness to the labeling assumptions and across novelty strata.

C.1 Estimation Method

Our procedure for estimating the PoS of Chinese clinical trials consists of four steps: (1) preprocessing and phase harmonization, (2) duration estimation and completion-date imputation, (3) trial outcome labeling, and (4) PoS estimation.

Step 1: Preprocessing and Phase Harmonization. We first preprocess the trial records and retain only Phases I–III trials for subsequent analyses. In particular, we exclude Phase IV trials and BE trials because they account for very small proportions of all trials (see Figure B.7), and do not reflect the core development-stage transitions. Therefore, we explicitly exclude analysis of post-approval development, which, in the form of drug lifecycle management, plays an increasingly important role in the industry.[13] After this restriction, our PoS estimation covers the 5,616 non-generic projects—of the 5,648 in the Project Dataset—that retain at least one Phase I–III trial; the remaining projects have no phase-classified trial and therefore cannot enter any phase transition.

We further harmonize combined phases into single normalized phase labels using a "lower" rule that maps combined phases to their respective earlier phases. Specifically, we treat Phases I/II, II/III, and III/IV trials as Phases I, II, and III trials, respectively. This is because assigning a combined-phase trial to the later phase (e.g., mapping I/II to Phase II) can lead to overly optimistic inferences about development maturity and downstream outcomes. By adopting the conservative "lower" rule, we avoid inflating the apparent level of advancement of a project and ensure that subsequent duration imputation and PoS estimation are anchored to the minimum level of evidence implied by the recorded phase label.

Step 2: Duration Estimation and Completion-date Imputation. To obtain the status of a trial at a specific time, we need to bound each trial's timeline. We anchor each trial's timeline at its First Disclosure Date and its end at its Completion Date (Table B.1). The trial's duration for labeling purposes is defined as their difference—the elapsed time from first disclosure to recorded completion, in days.

However, the Completion Date is missing for a proportion of the trials in the Trial Dataset. This is also common in other datasets, such as data for the U.S. Therefore, we follow the literature[4] and impute missing completion dates using empirically estimated median phase durations based on trials in the same therapeutic area with observed start and end dates. Some trials with observed disclosure and completion dates have negative intervals, i.e., the recorded completion date precedes the first disclosure date. This can arise from recording errors in the Trial Dataset, but also from retrospective or delayed registration, in

which case neither date is necessarily wrong. When calculating the median phase durations, we discard these trials but do not remove them in later steps. If a trial is associated with multiple therapeutic areas, we impute its duration as the maximum median duration across those areas. This choice is intentionally conservative: using the longest duration reduces the risk of prematurely treating the trial as having had sufficient time to progress.

Step 3: Trial Outcome Labeling. We evaluate the status of each trial at the end of each year from 2013–2025. We consider four different statuses: successful, failed, in progress, and not available. For each year, if a trial has not yet been disclosed at the end of that year, we label it not available. If it has been disclosed but has not yet reached its completion date (or imputed completion date from Step 2), we label it in progress.

Once the end of that year passes the trial completion date, we determine whether the trial should be considered successful or a failure, using project-level information from the Project Database. For a Phase I or Phase II trial, we treat "success" as meaning that the project has progressed to a later clinical phase or has reached approval. Specifically, a Phase I trial is labeled successful if the project has any Phase II or Phase III trial, or if the project has been approved; a Phase II trial is labeled successful if the project has any Phase III trial, or if the project has been approved. Phase III trials are labeled successful only if the project has been approved.

If none of these success conditions is met, we label a completed trial as failed only after an additional grace period. In particular, we consider a project as failed only if the end of the year exceeds the completion date by more than half the median trial duration of that therapeutic area and that phase. Similarly, as in Step 2, if a trial belongs to multiple areas, we use the maximum of the median trial durations for different therapeutic areas. Before that buffer window elapses, we keep the trial labeled as in progress, reflecting the possibility of administrative lag or delayed disclosure of subsequent development steps. This buffering mechanism prevents prematurely classifying trials as failed immediately upon completion.

This grace period method is also adopted by the literature when analyzing global clinical trials.[4] However, our choice of grace period is different. In particular, the literature[4] set grace periods of 360, 540, and 900 days for Phases I, II, and III trials, respectively. This is because, for example, the U.S. Food and Drug Administration (FDA) has a 6-month period to decide if it wishes to follow-up on a filing, and an additional 18 months to deliver a verdict; together with roughly six months to prepare and submit the New Drug Application after Phase III, this implies an overall time between Phase III and Approval of around 30 months, or 900 days. However, the Chinese drug development industry has a shorter history, and the Platform only includes trials since 2013. Therefore, the grace period values used in the literature[4] may be too long for China and would be likely to introduce more biases to our analysis. Therefore, we use a simple heuristic and set the grace periods as half the median trial durations.

We assess the robustness of our PoS estimates to these labeling choices—the combined-phase harmonization rule, the grace period, completion-date imputation, and the follow-up window—in a sensitivity analysis reported in Appendix C.5.1.

Step 4: PoS Estimation. At any observation cutoff, many programs are still ongoing, so we know only that they have not yet succeeded or failed, but not their eventual outcome—a form of *right-censoring*. A survival or competing-risks framework would therefore in principle be natural. We instead follow the ratio estimators of Wong et al.,[4] which accommodate this censoring by explicitly counting in-progress and missing transitions (below) and keep our estimates directly comparable to the global benchmark. We adopt their two definitions of PoS—the path-by-path PoS and the phase-by-phase PoS. For the completeness of this article, we include their definitions as follows.

Path-by-Path PoS. The path-by-path PoS is estimated based on the assumption that every drug development program passes through Phases I, II, and III trials in order. The path-by-path PoS is the probability that a project eventually succeeds along an implied development path from the earliest observed phase through later phases to approval. In this formulation, projects that have already reached a later stage (e.g., have a Phase III trial) are treated as having necessarily succeeded in earlier transitions (e.g., Phase I to Phase II and Phase II to Phase III). If we observe data for Phases I and III but not Phase II for a project, the path-by-path estimator treats the intermediate transition as successful for estimation; the registry cannot determine whether Phase II was truly skipped or simply unobserved in the linked data. Similarly, approved projects are treated as successful for all transitions.

More formally, for $j = \text{I, II, and III}$, let n^j be the number of drug development paths with observed Phase j trials, and n^j_s be the number of drug development paths where we observe phase transitions of state s of Phase j, with

$$s = \begin{cases} \text{ip}, & \text{if there are still Phase } j \text{ trials in progress,} \\ \text{t}, & \text{if all Phase } j \text{ trials have failed (i.e., terminated),} \\ \text{m}, & \text{if the phase transition can be inferred to be missing.} \end{cases}$$

The path-by-path PoS from Phase j to Phase $j+1$ is[4]

$$\text{PoS}_{j,j+1}(\text{Path-By-Path}) = \frac{n^{j+1}}{n^j + n^j_{\text{m}} - n^j_{\text{ip}}},$$

and the path-by-path PoS from Phase I to Approval, $\text{PoS}_{\text{I,Approval}}$, is

$$\text{PoS}_{\text{I,Approval}}(\text{Path-By-Path}) = \frac{n^{\text{Approval}}}{n^1 + n^1_{\text{m}} - n^1_{\text{ip}} - n^2_{\text{ip}} - n^3_{\text{ip}}},$$

where n^{Approval} is the number of approved drug development paths.

Phase-by-Phase PoS. Under the phase-by-phase definition, we compute PoS by conditioning directly on whether a project has reached the following phase in the dataset, without considering the missing data. We then obtain an overall Phase I-to-Approval PoS by multi-

plying these phase-by-phase transition probabilities. More formally, we have

$$\text{PoS}_{j,j+1}(\text{Phase-By-Phase}) = \frac{n^{j+1} - n^j_{\text{m}}}{n^j - n^j_{\text{ip}}},$$
$$\text{PoS}_{\text{I,Approval}}(\text{Phase-By-Phase}) = \prod_{j=\text{I,II,III}} \text{PoS}_{j,j+1}(\text{Phase-By-Phase}).$$

It is worth noting that, if no phase transitions are missing, the path-by-path and phase-by-phase methods should produce the same results. The phase-by-phase method is more widely used in the literature.[2,14] However, if phase transitions are missing, the path-by-path method will be more representative of actual approval rates, while the phase-by-phase method will underestimate the individual transition PoS.[4] In this article, for completeness, we report results obtained by both methods.

C.2 Missing Phase Records and Registry Coverage

Missing intermediate phases are a known feature of global clinical-development databases, rather than an anomaly specific to China. Wong et al. developed the path-by-path estimator expressly to accommodate paths in which a higher phase is observed but an intermediate phase is not.[4] Under their algorithm, observing Phase III or approval implies successful completion of the earlier transitions for path reconstruction, whether or not a standalone Phase II record is observed.

We therefore define a missing Phase II transition in the same way: a project has reached Phase III or approval but has no observed Phase II record. In the Chinese sample, 983 of 1,664 such projects meet this definition, or 59.1%. For comparison, 501 of these same 1,664 projects (30.1%) had no observed Phase I record, and 83 of 593 approved projects (14.0%) had no observed Phase III record. Thus, although missing records occur at more than one point in the development path, missing Phase II is the largest missing-phase category in our data. Wong et al. did not report the corresponding percentage directly, but it can be recovered from the path-by-path and phase-by-phase estimates published for the same global all-indications, industry-sponsored sample. Let p denote the path-by-path Phase II-to-III estimate, q the phase-by-phase estimate, H the number of paths reaching Phase III or approval, and m the number with a missing Phase II transition. Their equations imply

$$\frac{m}{H} = \frac{1 - q/p}{1 - q}.$$

Using their reported $p = 48.6\%$ and $q = 38.2\%$ gives an implied missing-transition proportion of approximately 34.6%.

We next examined whether missing Phase II records became less common as registry coverage matured. We grouped the same 1,664 projects that reached Phase III or approval by first disclosure year. The missing-Phase-II proportions were 57.7% (165/286), 49.2% (164/333), and 54.9% (322/586) for projects first disclosed in 2013–2015, 2016–2018, and 2019–2021, respectively. The absence of a sustained decline across these three more mature cohorts provides little evidence that incomplete coverage during the registry's early years is

the principal explanation. The proportion was 72.3% (332/459) in 2022–2025, but this most recent cohort includes only programs that progressed rapidly enough to reach Phase III or approval by the end of 2025 and is therefore not directly comparable with the earlier cohorts.

Company disclosures indicate that at least some missing Phase II records reflect deliberate progression from Phase I to Phase III rather than incomplete registry coverage. The first example is Hinova Pharmaceuticals' deuterated enzalutamide (HC-1119), a Class 1 innovative drug (a new molecular entity that has not been marketed elsewhere before its approval in China) for metastatic castration-resistant prostate cancer that has since been approved. Its linked registry records contain Phase I trials (the earliest being CTR20170042) and a Phase III trial (CTR20190199), with no Phase II record, and Hinova's Shanghai Stock Exchange listing prospectus states that, following prior communications, the FDA allowed the compound to skip Phases I and II in the U.S. and enter Phase III directly, and that the NMPA likewise agreed to direct entry into Phase III in China. The second example is Shanghai Yizhong's paclitaxel polymeric micelle for non-small-cell lung cancer, approved in October 2021. Its linked registry records contain a Phase I trial (CTR20130637) and a Phase III trial (CTR20150217), with no Phase II record, and the company's Shanghai Stock Exchange offering prospectus states that the product completed exactly these two trials—Phase I and the Phase III in non-small-cell lung cancer, cited by the same registration numbers—and notes that additional indications may proceed directly to Phase III under the prevailing review-and-approval rules.

The third example is Jiangsu Recbio's nine-valent HPV vaccine REC603. The linked registry records contain a Phase I trial (CTR20182557) and a Phase III trial (CTR20210947), with no Phase II record. More importantly, Recbio's 2021 annual report filed with the Hong Kong Stock Exchange states that, following communication with the CDE, the NMPA had no objection to the company proceeding directly to Phase III and that no Phase II clinical trial was conducted for REC603. This disclosure directly confirms that deliberate Phase II omission occurs. The fourth example is Chengdu Olymvax's AC–Hib combination vaccine, which provides corroborating evidence. Its linked registry records likewise contain a Phase I trial (CTR20190550) and a Phase III trial (CTR20191736), but no Phase II record, while the company's formal Shanghai Stock Exchange filing refers to the relevant research agreement as a Phase I/III clinical study. This documentation implies deliberate Phase II omission. At the same time, these two vaccine examples reflect the less rigid separation between distinct trial phases in vaccine development than in other therapeutic areas.

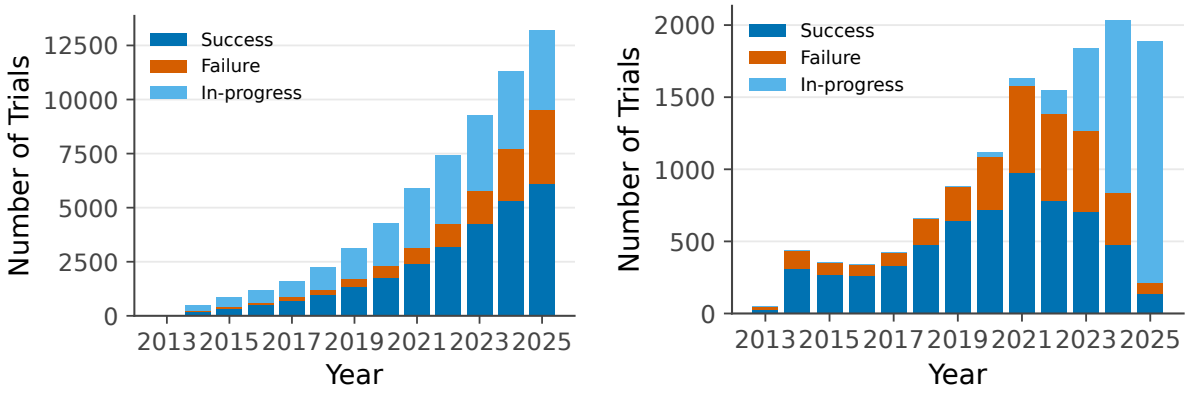
C.3 Trial Outcomes and PoS Estimates

We first summarize the trial-outcome counts that enter the estimator (Section C.3.1) and then report the resulting PoS estimates by therapeutic area (Section C.3.2).

C.3.1 Trial-Outcome Counts

Based on the method discussed in Section C.1, in this section we summarize the outcomes of Chinese clinical trials registered on the Platform. Figure C.1 reports the numbers of trials classified as successful, failed, and in progress. Figure C.1a presents the cumulative number of trials first disclosed up to the end of each year, with each bar decomposed by outcome

status measured at the end of that year. For example, the 2020 bar shows the outcome composition, as of the end of 2020, among all trials registered on or before 2020. Overall, successful trials account for the largest share, while a substantial portion of trials remain in progress, and a smaller fraction are classified as failed.



(a) Cumulative numbers by year.

(b) Numbers started in each year.

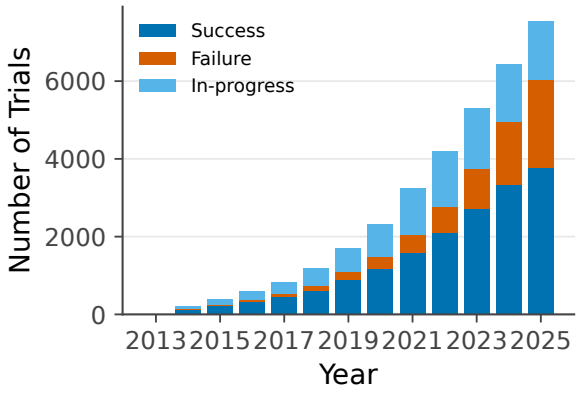
Figure C.1: Numbers of trials by year, shown as cumulative counts of trials first started up to each year with outcomes as of the end of that year (a), and counts of trials first started in each year with outcomes as of 2025 (b).

Figure C.1b adopts another perspective: for each year, it counts trials first disclosed in that year and then breaks them down by their outcomes observed by the end of 2025. For example, the 2020 bar shows the outcome composition, as of the end of 2025, among all trials first disclosed in 2020. This view makes clear that earlier cohorts have had more time to resolve, so a larger fraction of their trials are classified as successful or failed, whereas more recent cohorts are dominated by in-progress trials, consistent with the fact that many trials first disclosed closer to 2025 have not yet reached a definitive outcome by the end of the observation window.

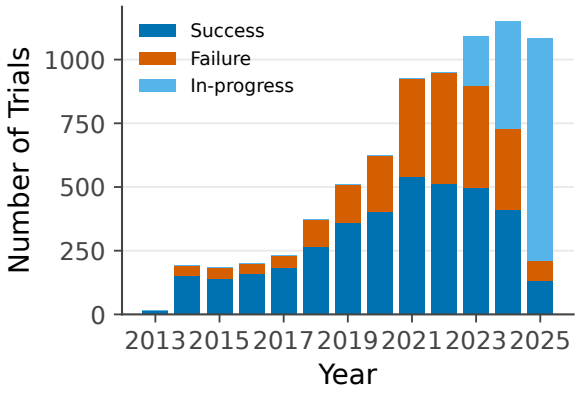
Figure C.2 further stratifies the trial outcomes shown in Figure C.1 by different phases. Figures C.2a (Phase I), C.2c (Phase II), and C.2e (Phase III) use the same method as in Figure C.1a, and Figures C.2b (Phase I), C.2d (Phase II), and C.2f (Phase III) use the same method as in Figure C.1b. The phase-specific patterns are broadly consistent with the aggregate figures in Figure C.1: in the cumulative panels, successful trials account for the largest share, while in the cohort-by-disclosure-year panels, recent cohorts are dominated by in-progress trials.

C.3.2 Estimated PoS by Therapeutic Area

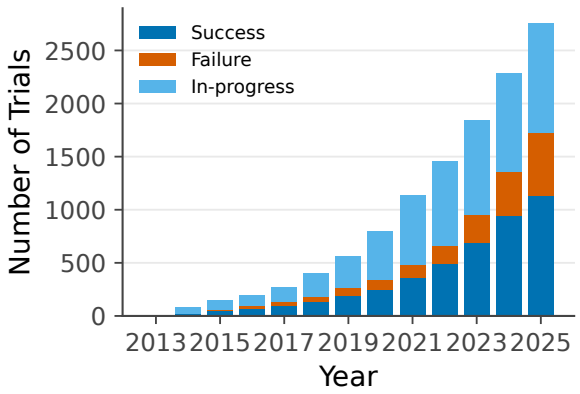
Table C.1 reports the numbers of trials, the median trial durations, and the PoS estimates for different phases and therapeutic areas based on all trials from 2013–2025. PoS results obtained by both the phase-by-phase method and the path-by-path method discussed in Appendix C.1 are provided.



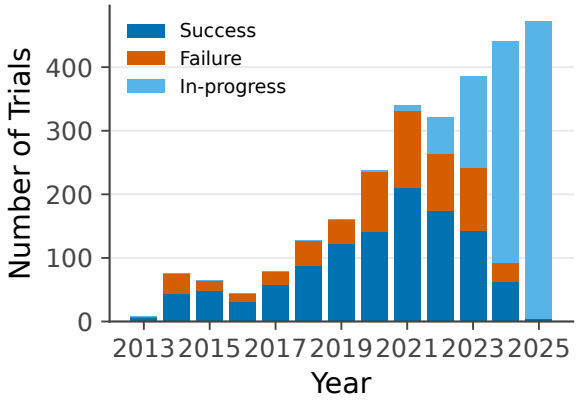
(a) Cumulative numbers by year (Phase I).



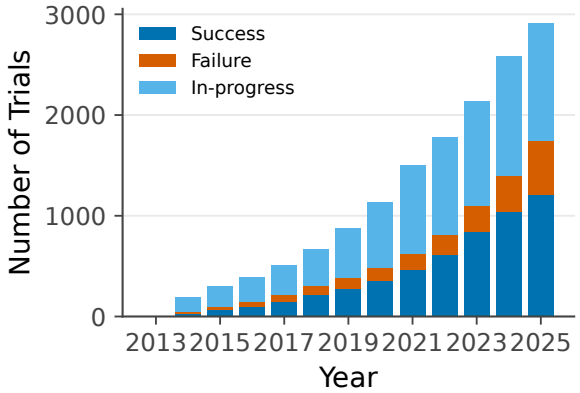
(b) Numbers started in each year (Phase I).



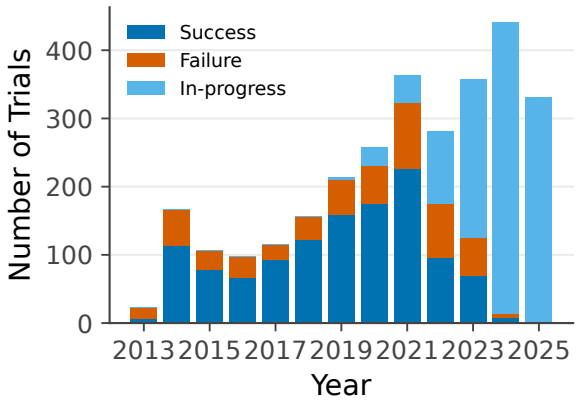
(c) Cumulative numbers by year (Phase II).



(d) Numbers started in each year (Phase II).



(e) Cumulative numbers by year (Phase III).



(f) Numbers started in each year (Phase III).

Figure C.2: Numbers of trials by year for different phases, shown as cumulative counts of trials first started up to each year with outcomes as of the end of that year (a, c, and e for Phases I, II, and III, respectively), and counts of trials first started in each year with outcomes as of 2025 (b, d, and f for Phases I, II, and III, respectively).

Table C.1: Trial counts, median durations, and estimated PoS by therapeutic area.

Therapeutic Area	Number of Trials				Median Duration (Year)			
	Phase I	Phase II	Phase III	Overall	Phase I	Phase II	Phase III	Sum I–III
Oncology	3,401	1,120	1,144	5,665	1.73	2.73	3.77	8.23
Cardiovascular	512	161	201	874	0.41	1.47	2.08	3.95
Dermatology	320	189	198	707	0.69	1.19	1.65	3.53
Gastroenterology	491	200	148	839	0.44	1.65	1.24	3.34
Endocrinology	763	212	276	1,251	0.49	1.09	1.90	3.48
Ophthalmology	143	69	92	304	0.61	1.10	2.25	3.95
Hematology	176	64	123	363	0.54	1.16	1.74	3.44
Immunology	704	399	396	1,499	0.64	1.51	2.13	4.28
Infectious	731	261	261	1,253	0.54	1.40	1.51	3.46
Psychiatry	279	65	66	410	0.41	1.46	1.34	3.22
Otolaryngology	47	50	25	122	0.44	1.05	1.73	3.23
Musculoskeletal	241	98	117	456	0.64	1.48	2.24	4.36
Neurology	472	145	171	788	0.41	1.53	1.71	3.65
Other	694	290	292	1,276	0.47	0.83	1.18	2.48
Rare	480	200	201	881	0.86	1.93	2.08	4.87
Reproductive	126	35	34	195	0.34	1.27	1.40	3.01
Respiratory	385	197	179	761	0.52	1.43	2.34	4.30
Urology	209	109	100	418	0.35	1.66	1.98	3.99

Therapeutic Area	Phase-By-Phase PoS (%)				Path-By-Path PoS (%)			
	I → II	II → III	III → A	Overall	I → II	II → III	III → A	Overall
Oncology	47.07 (1.35)	76.39 (2.24)	74.33 (2.70)	26.72 (1.47)	54.36 (1.25)	86.68 (1.35)	76.66 (2.50)	20.11 (1.21)
Cardiovascular	50.83 (2.88)	53.66 (5.51)	54.17 (5.87)	14.77 (2.36)	60.64 (2.52)	78.16 (3.13)	59.76 (5.42)	18.28 (2.36)
Dermatology	61.95 (3.39)	64.44 (5.05)	60.29 (5.93)	24.07 (3.30)	74.43 (2.50)	81.50 (2.95)	63.01 (5.65)	25.14 (3.21)
Gastroenterology	51.03 (2.94)	48.57 (4.88)	61.76 (5.89)	15.31 (2.30)	61.62 (2.53)	67.47 (3.64)	69.77 (4.95)	21.28 (2.44)
Endocrinology	54.79 (2.57)	53.85 (4.89)	57.52 (4.65)	16.97 (2.21)	63.12 (2.25)	79.49 (2.64)	61.90 (4.33)	22.67 (2.26)
Ophthalmology	55.75 (4.67)	48.57 (8.45)	33.33 (7.55)	9.03 (2.69)	69.14 (3.63)	80.43 (4.14)	39.53 (7.46)	15.32 (3.42)
Hematology	52.94 (5.41)	70.00 (7.25)	65.31 (6.80)	24.20 (4.33)	73.15 (3.63)	87.50 (3.38)	67.92 (6.41)	34.29 (4.63)
Immunology	56.25 (2.43)	57.99 (3.80)	74.12 (4.75)	24.18 (2.45)	67.90 (1.96)	75.09 (2.56)	77.78 (4.18)	21.88 (2.20)
Infectious	60.96 (2.28)	56.86 (4.00)	43.62 (4.06)	15.12 (1.85)	66.79 (2.03)	77.08 (2.48)	48.47 (3.91)	19.41 (1.96)
Psychiatry	45.27 (4.09)	41.38 (9.15)	48.65 (8.22)	9.11 (2.67)	56.45 (3.64)	79.01 (4.52)	62.00 (6.86)	20.95 (3.34)
Otolaryngology	60.00 (8.28)	40.00 (9.80)	61.54 (13.49)	14.77 (5.27)	77.78 (5.24)	59.46 (8.07)	64.29 (12.81)	20.93 (6.20)
Musculoskeletal	45.14 (4.15)	58.82 (6.89)	59.57 (7.16)	15.82 (3.03)	64.57 (3.20)	81.74 (3.60)	67.24 (6.16)	24.68 (3.43)
Neurology	44.73 (3.00)	50.70 (5.93)	45.68 (5.53)	10.36 (1.88)	59.36 (2.54)	79.89 (3.04)	52.69 (5.18)	17.50 (2.27)
Other	55.88 (2.46)	62.20 (4.30)	61.16 (4.43)	21.26 (2.33)	66.17 (2.05)	82.61 (2.28)	66.67 (3.97)	25.47 (2.27)
Rare	39.87 (2.82)	66.33 (4.77)	69.86 (5.37)	18.47 (2.34)	62.68 (2.20)	85.20 (2.38)	80.70 (3.70)	28.05 (2.48)
Reproductive	50.70 (5.93)	64.00 (9.60)	52.63 (11.45)	17.08 (4.94)	63.16 (4.95)	81.25 (5.63)	66.67 (9.07)	25.35 (5.16)
Respiratory	55.43 (3.09)	58.02 (5.48)	62.90 (6.13)	20.23 (2.97)	67.42 (2.49)	79.39 (3.15)	69.33 (5.32)	23.21 (2.82)
Urology	58.62 (4.57)	40.35 (6.50)	60.98 (7.62)	14.42 (3.15)	74.87 (3.14)	68.81 (4.44)	67.35 (6.70)	25.19 (3.79)

Notes: Bernoulli standard errors in parentheses for the PoS estimates. Sum I–III is the sum of the three phase-specific median durations, not the median of total development time. I → II, II → III, III → A, and Overall represent the transitions from Phase I to Phase II, Phase II to Phase III, Phase III to Approval, and Phase I to Approval, respectively.

From Table C.1, we make the following observations. First, consistent with Figures B.5–B.7, trials for oncology dominate in volume, with 5,665 trials in total across Phases I–III. Second, oncology also exhibits markedly longer development times, with phase-specific median durations summing to 8.23 years, whereas the corresponding sums for most other areas are approximately 2.48–4.87 years. This contrast likely reflects the greater complexity of oncology development, where trial endpoints, patient heterogeneity, and the frequent use of complex trial designs can prolong the overall timeline.

Third, for most therapeutic areas, the overall PoS remains below 30% under both the phase-by-phase and path-by-path estimates, a level in line with that reported for Chinese drug development projects in the literature.[3] Among all therapeutic areas, trials in hematology show the highest overall path-by-path PoS (34.29%), while trials in ophthalmology have the lowest (15.32%). Trials in oncology have a path-by-path PoS of 20.11%.

Fourth, comparing the two estimation approaches, we find that, for all phase transitions, the phase-by-phase PoS is lower (more conservative) than the path-by-path PoS. This is consistent with the literature.[4] For example, for oncology trials, the phase-by-phase PoS and the path-by-path PoS from Phase I to Phase II are 47.14% and 54.42%, respectively. This implies that when there are missing data the phase-by-phase method may underestimate the PoS.

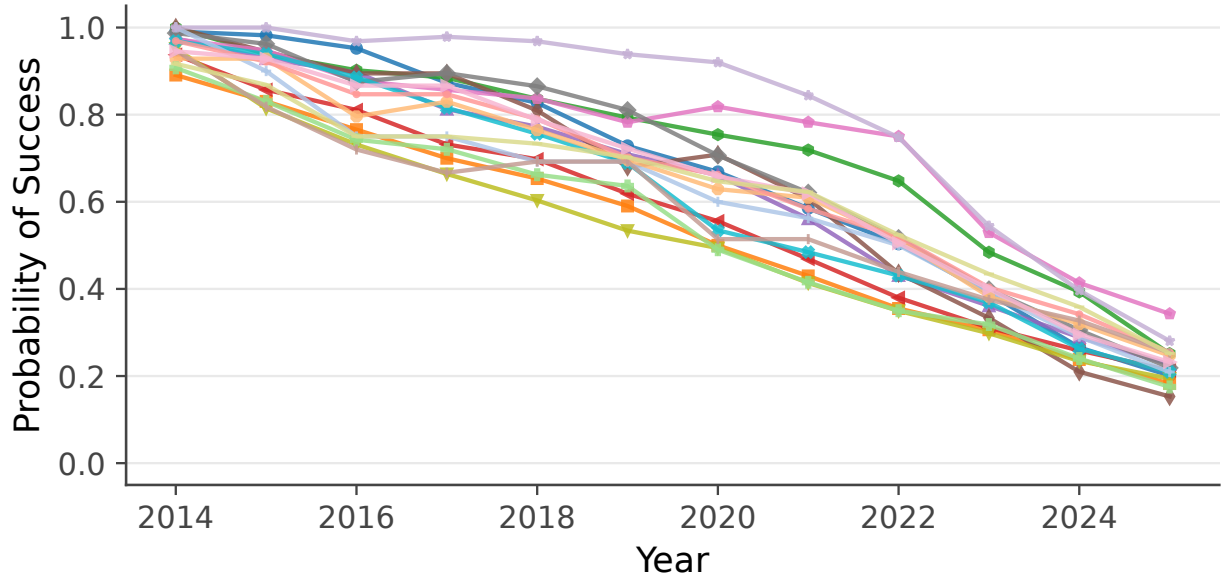
We next examine the time-series evolution of PoS estimates. We recompute the PoS year by year, where the estimate for year Y is obtained using all trials available up to the end of year Y. Figure C.3 reports the resulting overall PoS from Phase I to Approval as a function of time. Figure C.3a and Figure C.3b show the path-by-path PoS and phase-by-phase PoS, respectively.

From Figure C.3, we observe that both the path-by-path PoS and the phase-by-phase PoS exhibit a clear downward trend over time across most therapeutic areas, indicating that the implied probability of progressing from Phase I to Approval has declined as the sample has expanded. Because each yearly value reuses all earlier projects and adds newly disclosed ones, this decline mixes the entry of young, largely unresolved programs with additional follow-up on older ones, and cannot by itself separate cohort effects from vintage effects.

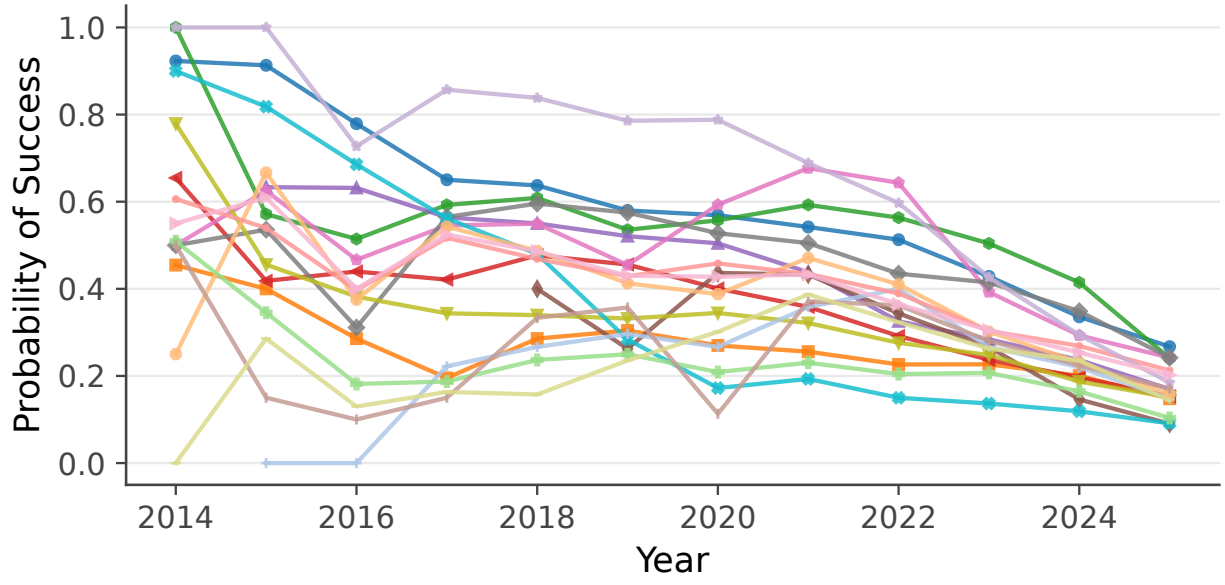
We also note that some therapeutic areas display flat trajectories in certain periods. For example, in Figure C.3b, the phase-by-phase PoS for drugs in the otolaryngology area is estimated to be zero before 2016. This is because there were not enough trials in this area during that period (see Figure B.6b and Table C.1). We still report these trajectories because the lack of data is a fact that cannot be changed. However, they should be interpreted with caution as they primarily reflect limited sample support rather than substantive changes in underlying success rates.

We further decompose the overall PoS time series into different phase transitions—Phase I to Phase II, Phase II to Phase III, and Phase III to Approval, where each year's estimate is computed using all trials available up to that year. Figure C.4 and Figure C.5 show the trajectories for path-by-path and phase-by-phase PoS, respectively.

Similar to Figure C.3, we observe that the PoS tends to decrease over time in most therapeutic areas, with the most pronounced downward movement typically appearing in the final transition (from Phase III to Approval). Meanwhile, the transitions from Phase I to Phase II and from Phase II to Phase III remain comparatively high for many areas



(a) Path-by-path PoS.



(b) Phase-by-phase PoS.

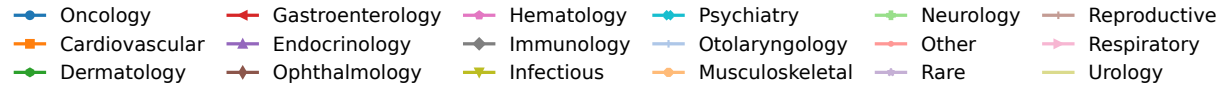
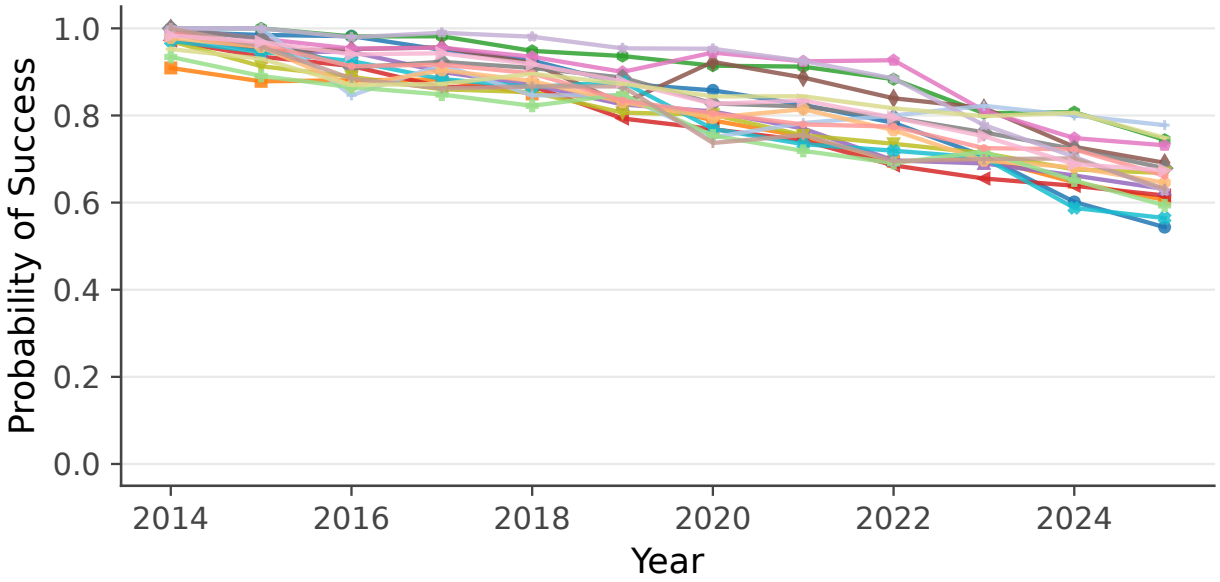
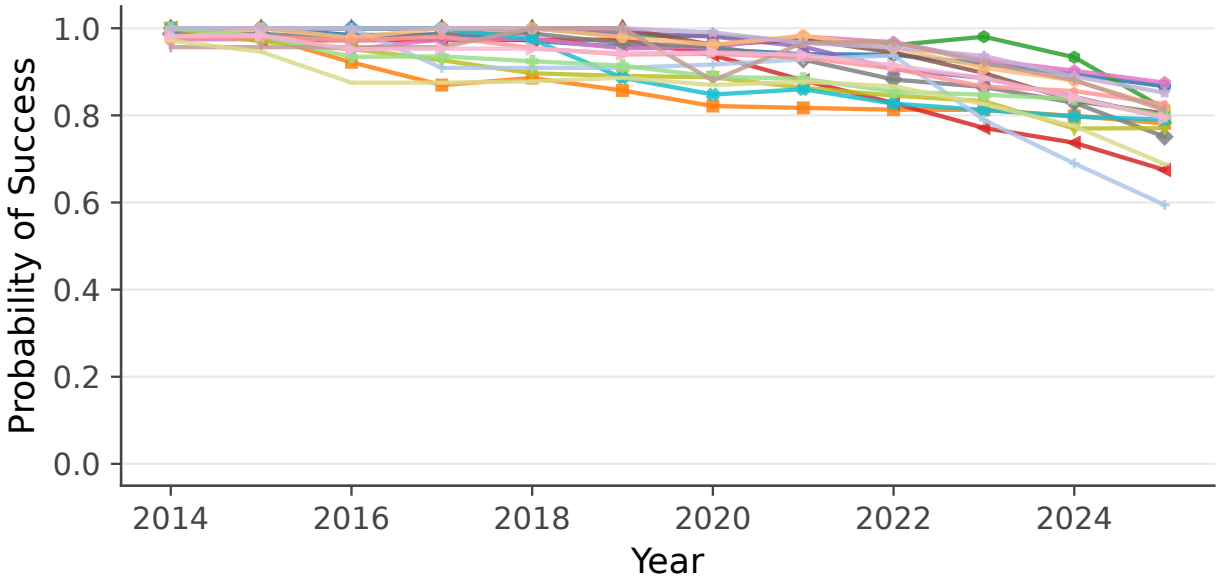


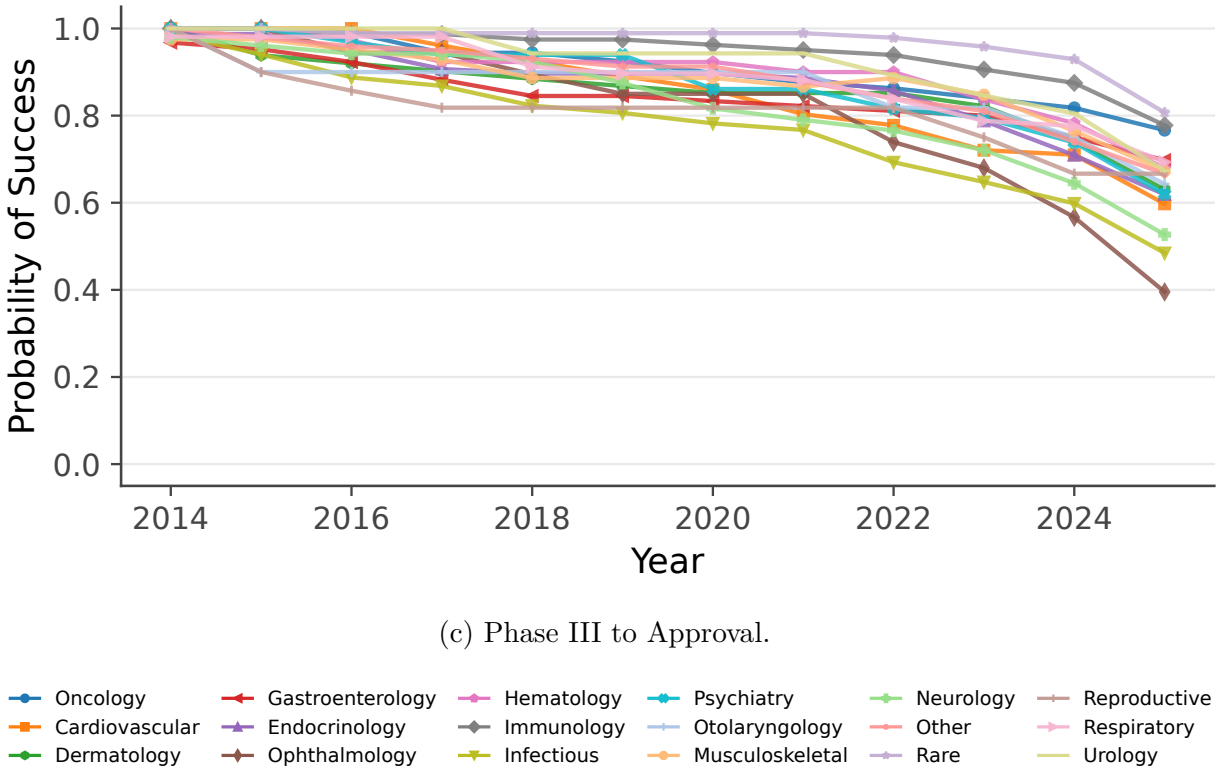
Figure C.3: Overall PoS from Phase I to Approval by therapeutic area over time, where each year's estimate is computed using all trials available up to that year, using the path-by-path PoS (a) and the phase-by-phase PoS (b).



(a) Phase I to Phase II.

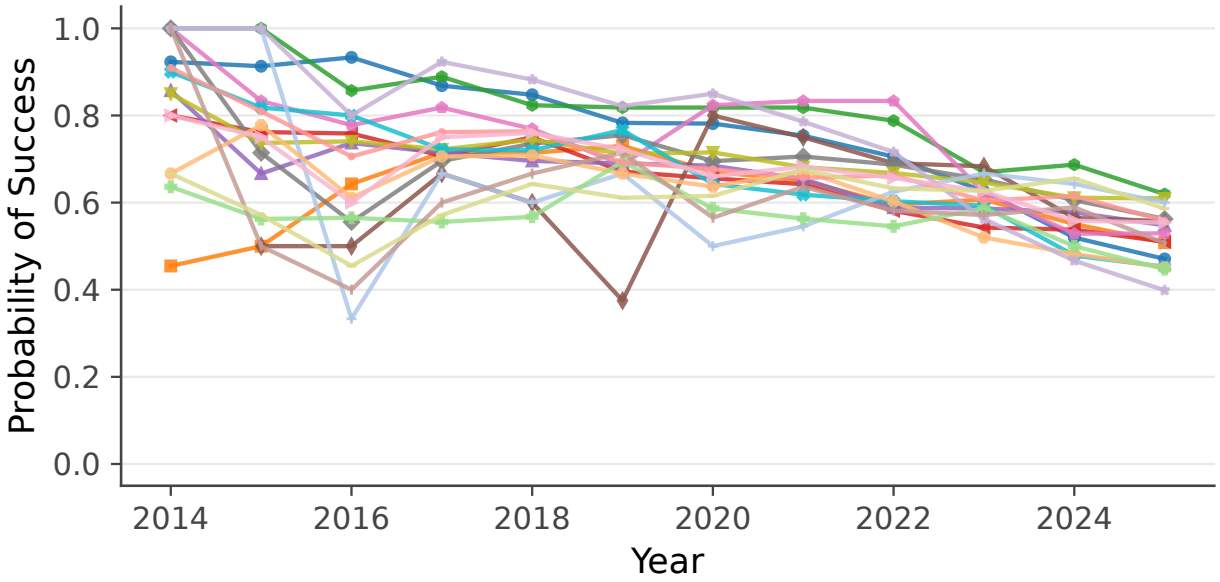


(b) Phase II to Phase III.

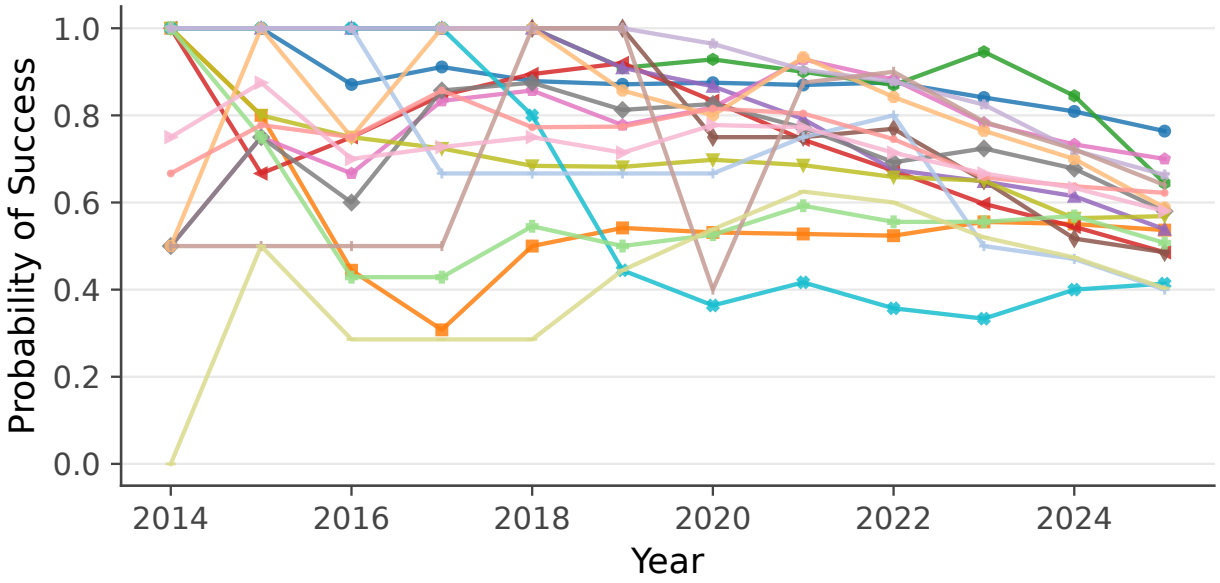


(c) Phase III to Approval.

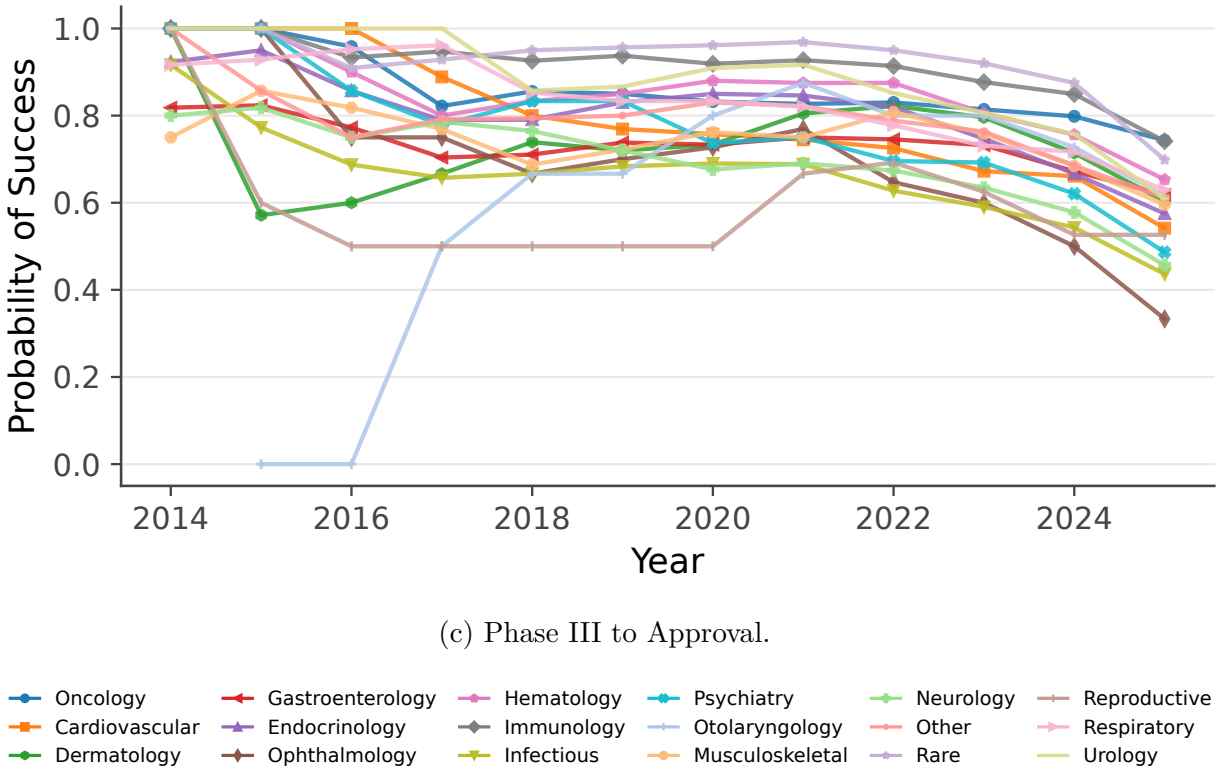
Figure C.4: Path-by-path PoS by therapeutic area over time for Phase I to Phase II (a), Phase II to Phase III (b), and Phase III to Approval (c), where each year's estimate is computed using all trials available up to that year.



(a) Phase I to Phase II.



(b) Phase II to Phase III.



(c) Phase III to Approval.

Figure C.5: Phase-by-phase PoS by therapeutic area over time for Phase I to Phase II (a), Phase II to Phase III (b), and Phase III to Approval (c), where each year's estimate is computed using all trials available up to that year.

but still exhibit gradual deterioration as the sample expands over time. In addition, compared to the path-by-path estimates (Figure C.4), the phase-by-phase estimates (Figure C.5) show greater year-to-year variability—particularly in earlier years and in smaller therapeutic areas—reflecting the higher sensitivity of phase-by-phase PoS to sparse observations and missing data.

C.4 Comparison with Global Benchmarks

In our main article, we compare the PoS estimates from China with those reported for the global market from Project ALPHA in Table 1. Project ALPHA (`https://projectalpha.mit.edu/pos/`), maintained by the MIT Laboratory for Financial Engineering, reports path-by-path PoS estimates computed with the same estimator we use,[4] drawing on Citeline drug-development data. We use the 2025Q4 release, and compute the reported global standard errors as Bernoulli standard errors from Project ALPHA's path denominators. Project ALPHA adopts a classification scheme that does not perfectly align with ours, so we construct an explicit crosswalk that maps each therapeutic area in the Chinese dataset to the closest corresponding category (or categories) in the global dataset. Table C.2 summarizes this mapping.

In some cases, multiple Chinese therapeutic areas map to the same global category. Four global categories each subsume two Chinese areas—Alimentary/Metabolic (Gastroenterology and Endocrinology), Sensory (Ophthalmology and Otolaryngology), Neurological (Psychiatry and Neurology), and Genitourinary (Reproductive and Urology)—so 8 of the 18 Chinese areas are benchmarked against a broader global aggregate rather than a one-to-one counterpart, while the remaining 10 map one-to-one. Comparisons for these eight aggregated areas should therefore be read with more caution, as the global benchmark may represent a broader category than the corresponding Chinese subgroup.

Table C.2: Mapping of therapeutic areas between our Chinese dataset and the global dataset from Project ALPHA.

Chinese Dataset	Global Dataset (Project ALPHA)	Chinese Dataset	Global Dataset (Project ALPHA)
Oncology	Anticancer	Psychiatry	Neurological
Cardiovascular	Cardiovascular	Otolaryngology	Sensory
Dermatology	Dermatological	Musculoskeletal	Musculoskeletal
Gastroenterology	Alimentary/Metabolic	Neurology	Neurological
Endocrinology	Alimentary/Metabolic	Other	Miscellaneous
Ophthalmology	Sensory	Rare	Rare
Hematology	Blood and Clotting	Reproductive	Genitourinary (Including Sex Hormones)
Immunology	Immunological	Respiratory	Respiratory
Infectious	Anti-infective; Antiparasitic	Urology	Genitourinary (Including Sex Hormones)

Beyond classification, one might also ask whether the two sides label trial outcomes differently. Project ALPHA applies the same outcome-labeling and estimation methodology of Wong et al.[4] that we adopt. We further probe the labeling choices on the Chinese side—the combined-phase rule, the grace period, and completion-date imputation—in the sensitivity analysis of Appendix C.5.1, and find that the main China–global patterns persist across these variations.

C.5 Robustness

We assess robustness along two dimensions: sensitivity to the trial-outcome labeling assumptions (Section C.5.1) and stratification of the estimates by drug novelty (Section C.5.2).

C.5.1 Sensitivity to Labeling Assumptions

We re-estimate PoS under five alternatives to our baseline labeling, all defined in Appendix C.1: (i) mapping combined-phase trials to the *higher* rather than the lower phase (*Upper Phase Rule*); (ii) a full-median rather than half-median grace period before an unresolved trial is treated as failed (*Full-Median Grace*); (iii) the longer 360/540/900-day grace periods of Wong et al.[4] (*WSL Grace*); (iv) imputing missing completion dates from the median duration pooled across all therapeutic areas, rather than the area-specific median (*Pooled-Duration Imputation*); and (v) restricting to the more mature cohort of trials first disclosed by 2019, whose outcomes are more fully resolved (*Disclosure* $\leq$ *2019*). Table C.3 reports the resulting estimates, pooled across all therapeutic areas and by area, each with its Bernoulli standard error. Its final panel also reports success, failure, and in-progress counts for each pooled phase transition in the baseline and mature cohorts.

The headline contrast—China's Phase II-to-Phase III PoS far exceeding the global benchmark—holds under every variation. The pooled Phase II-to-Phase III estimate stays between 77.26% and 86.62% across the five alternatives (versus a global benchmark of 43.77%), and it remains above the global benchmark in every therapeutic area.

The two variations most relevant to in-progress transitions leave the estimate well above the global benchmark: the longer WSL grace periods raise the pooled estimate to 85.20%, and restricting to the mature 2019 cohort still leaves it at 77.26%. In the baseline cohort, 22.20%, 24.16%, and 42.07% of eligible Phase I-to-Phase II, Phase II-to-Phase III, and Phase III-to-approval transitions remain in progress, respectively; in the mature 2019 cohort, these proportions fall to 0.07%, 0%, and 1.79%. Because the mature 2019 cohort has no in-progress Phase II-to-Phase III transitions, its 77.26% estimate suggests that the elevated Phase II-to-Phase III transition is not driven by unresolved recent transitions. It is thus a stable feature of the Chinese data rather than a consequence of any single labeling choice.

Table C.3: Sensitivity of China path-by-path PoS estimates (%) to alternative outcome-labeling assumptions, pooled across all therapeutic areas and by area, with a pooled resolution-status breakdown.

	Phase-Transition PoS			Overall
Assumption	I → II	II → III	III → A	I → A
Panel A: Oncology				
Baseline	54.36 (1.25)	86.68 (1.35)	76.66 (2.50)	20.11 (1.21)
Upper Phase Rule	66.94 (1.06)	81.36 (1.45)	75.86 (2.51)	20.50 (1.23)
Full-Median Grace	62.45 (1.30)	92.17 (1.10)	83.65 (2.28)	26.60 (1.54)
WSL Grace	55.41 (1.26)	86.95 (1.34)	79.42 (2.43)	20.91 (1.25)
Pooled-Duration Imputation	40.72 (1.07)	79.11 (1.54)	61.97 (2.58)	12.55 (0.79)
Disclosure $\leq$ 2019	51.04 (2.08)	79.25 (2.36)	73.21 (2.96)	28.92 (1.90)
Panel B: Cardiovascular				
Baseline	60.64 (2.52)	78.16 (3.13)	59.76 (5.42)	18.28 (2.36)

Continued on next page

Table C.3: Sensitivity of China path-by-path PoS estimates and resolution-status breakdown (continued).

	Phase-Transition PoS			Overall
Assumption	I → II	II → III	III → A	I → A
Upper Phase Rule	62.50 (2.47)	78.02 (3.07)	59.76 (5.42)	18.42 (2.38)
Full-Median Grace	62.47 (2.53)	82.42 (2.96)	72.06 (5.44)	20.94 (2.66)
WSL Grace	68.88 (2.54)	82.42 (2.96)	72.06 (5.44)	24.50 (3.04)
Pooled-Duration Imputation	62.81 (2.54)	79.53 (3.09)	57.65 (5.36)	19.22 (2.47)
Disclosure ≤ 2019	68.63 (4.59)	67.14 (5.61)	63.83 (7.01)	29.41 (4.51)
Panel C: Dermatology				
Baseline	74.43 (2.50)	81.50 (2.95)	63.01 (5.65)	25.14 (3.21)
Upper Phase Rule	76.56 (2.37)	80.45 (2.96)	63.01 (5.65)	25.14 (3.21)
Full-Median Grace	79.93 (2.38)	90.97 (2.30)	74.19 (5.56)	34.59 (4.12)
WSL Grace	82.55 (2.29)	93.38 (2.02)	75.41 (5.51)	38.66 (4.46)
Pooled-Duration Imputation	73.46 (2.51)	83.93 (2.83)	66.67 (5.68)	25.84 (3.28)
Disclosure ≤ 2019	72.22 (6.10)	89.74 (4.86)	74.29 (7.39)	48.15 (6.80)
Panel D: Gastroenterology				
Baseline	61.62 (2.53)	67.47 (3.64)	69.77 (4.95)	21.28 (2.44)
Upper Phase Rule	64.97 (2.47)	63.68 (3.49)	68.18 (4.97)	20.83 (2.39)
Full-Median Grace	63.16 (2.54)	74.17 (3.56)	74.07 (4.87)	23.72 (2.67)
WSL Grace	65.33 (2.55)	73.20 (3.58)	77.92 (4.73)	25.10 (2.80)
Pooled-Duration Imputation	62.13 (2.53)	64.74 (3.63)	72.29 (4.91)	21.20 (2.43)
Disclosure ≤ 2019	62.71 (4.45)	77.03 (4.89)	73.68 (5.83)	35.59 (4.41)
Panel E: Endocrinology				
Baseline	63.12 (2.25)	79.49 (2.64)	61.90 (4.33)	22.67 (2.26)
Upper Phase Rule	67.09 (2.16)	77.60 (2.64)	60.00 (4.30)	22.81 (2.27)
Full-Median Grace	65.10 (2.25)	84.55 (2.44)	72.22 (4.31)	26.17 (2.55)
WSL Grace	68.79 (2.25)	85.32 (2.40)	75.00 (4.25)	29.10 (2.77)
Pooled-Duration Imputation	62.85 (2.25)	79.83 (2.63)	61.90 (4.33)	22.61 (2.25)
Disclosure ≤ 2019	58.45 (4.14)	87.95 (3.57)	78.08 (4.84)	40.14 (4.11)
Panel F: Ophthalmology				
Baseline	69.14 (3.63)	80.43 (4.14)	39.53 (7.46)	15.32 (3.42)
Upper Phase Rule	81.87 (2.95)	77.88 (4.07)	34.69 (6.80)	16.50 (3.66)
Full-Median Grace	71.79 (3.60)	85.06 (3.82)	56.67 (9.05)	19.54 (4.25)
WSL Grace	73.68 (3.57)	85.06 (3.82)	60.71 (9.23)	20.99 (4.52)
Pooled-Duration Imputation	69.14 (3.63)	81.32 (4.09)	36.96 (7.12)	15.04 (3.36)
Disclosure ≤ 2019	48.48 (8.70)	81.25 (9.76)	69.23 (12.80)	27.27 (7.75)
Panel G: Hematology				
Baseline	73.15 (3.63)	87.50 (3.38)	67.92 (6.41)	34.29 (4.63)
Upper Phase Rule	82.69 (3.03)	84.31 (3.60)	64.29 (6.40)	36.36 (4.83)
Full-Median Grace	75.69 (3.57)	90.32 (3.07)	78.26 (6.08)	40.00 (5.16)
WSL Grace	77.30 (3.53)	89.36 (3.18)	80.00 (5.96)	41.38 (5.28)
Pooled-Duration Imputation	71.71 (3.65)	86.60 (3.46)	64.29 (6.40)	32.14 (4.41)
Disclosure ≤ 2019	70.73 (7.11)	86.21 (6.40)	80.00 (8.00)	48.78 (7.81)
Panel H: Immunology				
Baseline	67.90 (1.96)	75.09 (2.56)	77.78 (4.18)	21.88 (2.20)
Upper Phase Rule	72.74 (1.82)	72.73 (2.54)	76.24 (4.24)	22.13 (2.23)
Full-Median Grace	71.56 (1.94)	84.92 (2.25)	87.50 (3.53)	27.60 (2.68)
WSL Grace	74.47 (1.92)	84.25 (2.29)	90.59 (3.17)	29.96 (2.86)
Pooled-Duration Imputation	66.38 (1.96)	74.05 (2.58)	77.78 (4.18)	20.87 (2.12)
Disclosure ≤ 2019	57.14 (4.54)	75.00 (5.25)	86.00 (4.91)	36.44 (4.43)
Panel I: Infectious				
Baseline	66.79 (2.03)	77.08 (2.48)	48.47 (3.91)	19.41 (1.96)
Upper Phase Rule	69.64 (1.96)	76.00 (2.47)	47.02 (3.85)	19.41 (1.96)
Full-Median Grace	69.25 (2.03)	79.86 (2.41)	57.66 (4.22)	22.44 (2.22)
WSL Grace	72.18 (2.01)	79.86 (2.41)	64.23 (4.32)	24.92 (2.43)
Pooled-Duration Imputation	66.67 (2.03)	77.08 (2.48)	52.32 (4.06)	19.95 (2.01)
Disclosure ≤ 2019	64.85 (3.36)	72.52 (3.90)	63.16 (4.95)	29.70 (3.22)

Continued on next page

Table C.3: Sensitivity of China path-by-path PoS estimates and resolution-status breakdown (continued).

	Phase-Transition PoS			Overall
Assumption	I → II	II → III	III → A	I → A
Panel J: Psychiatry				
Baseline	56.45 (3.64)	79.01 (4.52)	62.00 (6.86)	20.95 (3.34)
Upper Phase Rule	58.29 (3.61)	80.95 (4.28)	59.62 (6.80)	21.23 (3.38)
Full-Median Grace	58.66 (3.68)	81.01 (4.41)	73.81 (6.78)	23.66 (3.71)
WSL Grace	65.22 (3.75)	80.00 (4.47)	77.50 (6.60)	27.68 (4.23)
Pooled-Duration Imputation	59.66 (3.70)	79.01 (4.52)	72.09 (6.84)	23.66 (3.71)
Disclosure $\leq$ 2019	69.81 (6.31)	70.27 (7.51)	69.23 (9.05)	33.96 (6.51)
Panel K: Otolaryngology				
Baseline	77.78 (5.24)	59.46 (8.07)	64.29 (12.81)	20.93 (6.20)
Upper Phase Rule	79.69 (5.03)	62.50 (7.65)	60.00 (12.65)	20.93 (6.20)
Full-Median Grace	83.05 (4.88)	70.97 (8.15)	75.00 (12.50)	29.03 (8.15)
WSL Grace	84.48 (4.75)	70.97 (8.15)	81.82 (11.63)	31.03 (8.59)
Pooled-Duration Imputation	77.78 (5.24)	61.11 (8.12)	60.00 (12.65)	20.93 (6.20)
Disclosure $\leq$ 2019	55.56 (16.56)	80.00 (17.89)	75.00 (21.65)	33.33 (15.71)
Panel L: Musculoskeletal				
Baseline	64.57 (3.20)	81.74 (3.60)	67.24 (6.16)	24.68 (3.43)
Upper Phase Rule	66.81 (3.09)	79.34 (3.68)	67.24 (6.16)	24.38 (3.39)
Full-Median Grace	66.98 (3.21)	85.45 (3.36)	79.59 (5.76)	28.68 (3.88)
WSL Grace	70.24 (3.19)	83.19 (3.52)	79.59 (5.76)	30.23 (4.04)
Pooled-Duration Imputation	63.72 (3.20)	81.03 (3.64)	65.00 (6.16)	23.78 (3.32)
Disclosure $\leq$ 2019	54.55 (6.13)	86.11 (5.76)	70.00 (8.37)	32.31 (5.80)
Panel M: Neurology				
Baseline	59.36 (2.54)	79.89 (3.04)	52.69 (5.18)	17.50 (2.27)
Upper Phase Rule	63.78 (2.46)	79.26 (2.96)	51.58 (5.13)	18.01 (2.33)
Full-Median Grace	62.71 (2.57)	85.80 (2.74)	65.33 (5.50)	21.30 (2.70)
WSL Grace	67.68 (2.58)	85.28 (2.78)	68.06 (5.49)	24.26 (3.02)
Pooled-Duration Imputation	61.33 (2.56)	79.43 (3.06)	53.26 (5.20)	18.28 (2.36)
Disclosure $\leq$ 2019	63.00 (4.83)	68.25 (5.86)	65.12 (7.27)	28.00 (4.49)
Panel N: Other				
Baseline	66.17 (2.05)	82.61 (2.28)	66.67 (3.97)	25.47 (2.27)
Upper Phase Rule	70.40 (1.94)	81.57 (2.27)	65.73 (3.97)	26.04 (2.31)
Full-Median Grace	69.02 (2.05)	87.02 (2.08)	73.44 (3.90)	29.38 (2.55)
WSL Grace	72.88 (2.02)	87.36 (2.06)	80.34 (3.67)	33.45 (2.81)
Pooled-Duration Imputation	66.29 (2.05)	82.91 (2.27)	66.67 (3.97)	25.61 (2.28)
Disclosure $\leq$ 2019	64.49 (4.07)	86.52 (3.62)	80.52 (4.51)	44.93 (4.23)
Panel O: Rare				
Baseline	62.68 (2.20)	85.20 (2.38)	80.70 (3.70)	28.05 (2.48)
Upper Phase Rule	72.61 (1.93)	80.80 (2.49)	78.63 (3.79)	29.58 (2.59)
Full-Median Grace	69.72 (2.20)	91.35 (1.95)	93.88 (2.42)	37.10 (3.07)
WSL Grace	66.23 (2.21)	87.96 (2.21)	94.85 (2.25)	33.09 (2.82)
Pooled-Duration Imputation	54.38 (2.11)	78.51 (2.64)	76.67 (3.86)	21.55 (1.99)
Disclosure $\leq$ 2019	68.89 (4.88)	88.71 (4.02)	90.91 (3.88)	55.56 (5.24)
Panel P: Reproductive				
Baseline	63.16 (4.95)	81.25 (5.63)	66.67 (9.07)	25.35 (5.16)
Upper Phase Rule	64.21 (4.92)	79.59 (5.76)	66.67 (9.07)	25.35 (5.16)
Full-Median Grace	66.67 (4.97)	88.64 (4.78)	69.23 (9.05)	29.51 (5.84)
WSL Grace	68.97 (4.96)	90.70 (4.43)	75.00 (8.84)	32.73 (6.33)
Pooled-Duration Imputation	64.52 (4.96)	82.98 (5.48)	66.67 (9.07)	26.47 (5.35)
Disclosure $\leq$ 2019	56.67 (9.05)	76.47 (10.29)	69.23 (12.80)	30.00 (8.37)
Panel Q: Respiratory				
Baseline	67.42 (2.49)	79.39 (3.15)	69.33 (5.32)	23.21 (2.82)
Upper Phase Rule	70.22 (2.39)	78.86 (3.09)	67.53 (5.34)	23.32 (2.83)
Full-Median Grace	70.00 (2.49)	86.75 (2.76)	77.61 (5.09)	27.51 (3.25)
WSL Grace	73.23 (2.46)	86.75 (2.76)	80.00 (4.96)	30.23 (3.50)

Continued on next page

Table C.3: Sensitivity of China path-by-path PoS estimates and resolution-status breakdown (continued).

	Phase-Transition PoS			Overall
Assumption	I → II	II → III	III → A	I → A
Pooled-Duration Imputation	67.23 (2.49)	78.92 (3.17)	64.20 (5.33)	22.41 (2.74)
Disclosure ≤ 2019	72.50 (4.99)	87.93 (4.28)	76.00 (6.04)	48.10 (5.62)
Panel R: Urology				
Baseline	74.87 (3.14)	68.81 (4.44)	67.35 (6.70)	25.19 (3.79)
Upper Phase Rule	77.32 (3.01)	68.75 (4.38)	66.00 (6.70)	25.58 (3.84)
Full-Median Grace	76.88 (3.09)	76.53 (4.28)	80.49 (6.19)	30.84 (4.46)
WSL Grace	81.25 (2.94)	75.76 (4.31)	82.50 (6.01)	34.02 (4.81)
Pooled-Duration Imputation	74.48 (3.15)	68.18 (4.44)	66.00 (6.70)	24.63 (3.72)
Disclosure ≤ 2019	65.96 (6.91)	70.97 (8.15)	72.73 (9.50)	34.04 (6.91)
Panel S: All Therapeutic Areas (Pooled)				
Baseline	61.78 (0.74)	81.29 (0.86)	61.51 (1.57)	19.66 (0.72)
Upper Phase Rule	68.49 (0.67)	78.90 (0.87)	60.33 (1.56)	19.94 (0.73)
Full-Median Grace	66.51 (0.74)	86.62 (0.78)	71.02 (1.57)	24.19 (0.87)
WSL Grace	66.25 (0.74)	85.20 (0.80)	73.39 (1.55)	23.99 (0.86)
Pooled-Duration Imputation	55.62 (0.71)	79.39 (0.88)	58.54 (1.55)	16.48 (0.62)
Disclosure ≤ 2019	58.57 (1.34)	77.26 (1.49)	72.02 (1.83)	32.27 (1.27)
Panel T: Resolution Status, All Therapeutic Areas (Pooled); S/F/P Counts (In-Progress %)				
Baseline	2,699/1,670/1,247 (22.20%)	1,664/383/652 (24.16%)	593/371/700 (42.07%)	—
Disclosure ≤ 2019	796/563/1 (0.07%)	615/181/0 (0.00%)	435/169/11 (1.79%)	—

Notes: Each cell reports the estimate with its Bernoulli standard error in parentheses. "A" denotes approval; the endpoint is registry-inferred progression or approval. In Panel T, cells report success/failure/in-progress (S/F/P) counts, with the in-progress share of all eligible transitions in parentheses. The assumptions are defined in Appendix C.1.

C.5.2 PoS by Novelty Stratum

One possible contributor to China's high Phase II-to-Phase III transition is program mix: a pipeline tilted toward low-risk follow-on ("me too") drugs could raise the observed transition rate. To assess this, we stratify the non-generic cohort using registry-derived proxies for novelty and recompute path-by-path PoS within each stratum (Table C.4). The proxies come directly from database fields. We use *Global Launch Status*—whether the active ingredient is already marketed somewhere in the world (*follow-on*) or *not yet launched*; *Domestic/Imported*; and *Biologic/Chemical*. Within oncology—where mechanism crowding is a recognized concern—we additionally split trials by *Target*, separating PD-1 or PD-L1 (a heavily pursued, "saturated" mechanism) from other targets.

Novelty accounts for part of the gap: the Phase II-to-Phase III transition is highest for follow-on and imported programs (96.37% and 90.44%, respectively), consistent with their lower scientific risk. But it does not explain the gap away—the pooled Phase II-to-Phase III PoS is 69.27% even for not-yet-launched programs and 74.27% for domestic ones, both far above the global benchmark of 43.77%. Because this comparison rests on the Phase II-to-Phase III transition, it is unaffected by the construction that leaves Phase III-to-approval and overall PoSs undefined for the not-yet-launched stratum (Table C.4). The elevated transition is thus a broad feature of Chinese development, not merely an artifact of follow-on activity. Within oncology, mechanism saturation makes little difference: PD-1/PD-L1 programs transition at 83.33%, close to the 87.02% for other oncology targets.

Table C.4: Path-by-path PoS by novelty stratum (%, standard errors in parentheses).

Stratum	*N*	I→II	II→III	III→A	Overall
All Non-Generic (Pooled)					
All Non-Generic	5,616	61.78 (0.74)	81.29 (0.86)	61.51 (1.57)	19.66 (0.72)
Launched Globally (Follow-On)	1,067	89.01 (0.97)	96.37 (0.62)	80.24 (1.46)	66.93 (1.58)
Not Yet Launched (Novelty Proxy)	4,549	53.30 (0.86)	69.27 (1.37)	—	—
Imported	1,185	87.34 (1.00)	90.44 (0.99)	61.69 (2.09)	43.56 (1.79)
Domestic	4,431	53.03 (0.87)	74.27 (1.28)	61.28 (2.37)	11.48 (0.67)
Chemical	3,414	61.47 (0.92)	79.47 (1.11)	61.76 (1.87)	20.47 (0.90)
Biologic	2,202	62.33 (1.23)	84.66 (1.35)	60.96 (2.85)	17.98 (1.22)
Oncology: Mechanism Saturation					
All Oncology Projects	2,444	54.36 (1.25)	86.68 (1.35)	76.66 (2.50)	20.11 (1.21)
PD-1/PD-L1	218	58.17 (3.99)	83.33 (4.81)	92.31 (5.23)	24.00 (4.27)
Other Targets	2,226	53.95 (1.32)	87.02 (1.40)	75.10 (2.68)	19.72 (1.26)

Notes: *N* is the number of projects in each stratum that enter the estimator. Novelty is proxied by Global Launch Status (already launched globally, i.e., follow-on, versus not yet launched); *Domestic/Imported*; *Biologic/Chemical*; and, within oncology, *Target* (PD-1/PD-L1 versus other targets). For the not-yet-launched stratum, Phase III-to-approval and overall PoSs are undefined by construction (—), because approval changes launch status.

D Trial Outcome Prediction

This appendix provides details and additional results for our trial-outcome prediction models. Section D.1 describes the features and the models we train; Section D.2 reports the full set of performance, coefficient, and feature-importance results; and Section D.3 examines robustness through feature ablation and stratification by novelty.

D.1 Features and Models

We describe how the features are constructed (Section D.1.1), the training and evaluation procedure (Section D.1.2), and the additional models we consider beyond those in the main text (Section D.1.3).

D.1.1 Feature Construction

For a given calendar year and each trial, we compute various features using only information that is available up to the end of that year. Our goal is to use these features to predict the outcome of the trial obtained as in Appendix C. The features we construct are as follows, with a summary in Table D.1.

Table D.1: Features used to predict trial outcomes.

Class	Name	Type
Text	BERT PoS	Value in [0,1]
Sponsor	Past PoS	Value in [0,1]
	Number of Past Sponsored Trials	Integer
Phase	I; II	One-hot
Information	Days Since First Disclosure	Integer
	Number of Participating Institutions	Integer
	International Scope	One-hot
	Imported Drug	One-hot
Design	Target Domestic Enrollment	Integer
	Target International Enrollment	Integer
	With Active Control	One-hot
	With Combination Therapy	One-hot
	With Upper Age Limit	One-hot
	Minimum Enrollment Age	Numeric
	Maximum Enrollment Age	Numeric
	Number of Inclusion Criteria	Integer
	Number of Exclusion Criteria	Integer
Therapeutic Area	See Table B.2	One-hot
Line	See Table B.2	One-hot
Modality	See Table B.2	One-hot
Dosage Form	See Table B.2	One-hot

Text (BERT). As shown in Table B.1, the Trial Dataset provides a large amount of unstructured text information related to the trials. However, most machine learning models

can only deal with numerical variables and cannot deal with text information directly. To incorporate text information into our machine learning models, we use BERT.[15]

For each trial, we concatenate its key text fields into a single string, with bracketed field markers:

> [靶点] {Target} [目的] {Objectives} [标题] {Official Title} [SEP] {Public Title}
> [适应症] {Indication} [设计] {Trial Design}

All fields are provided in the Trial Dataset; we omit highly templated or sparsely populated fields (such as drug name, ingredient, and inclusion/exclusion criteria), which we found added little.

We feed this text to the publicly available `bert-base-chinese` model (`https://huggingface.co/google-bert/bert-base-chinese`). Because a trial's text can exceed BERT's maximum input length, we split its token sequence into consecutive windows of 510 tokens—corresponding to a maximum input length of 512 after adding the boundary tokens `[CLS]` and `[SEP]`—advancing by a stride of 446 tokens, so that adjacent windows overlap by 64 tokens and preserve contextual continuity. We encode each window independently, take BERT's output vector at the `[CLS]` position as the window representation, and average these across windows to obtain a single 768-dimensional document embedding for the trial. This embedding step uses only text and no outcome labels.

To turn the embedding into a scalar feature, we fit a lightweight linear (logistic) head—with class-weighted loss and early stopping—that maps the 768-dimensional embedding to a success probability. This head is the only label-trained component, and we fit it in a leakage-safe manner: for each calendar year's forward-looking prediction, the head is trained only on that year's training fold, never on validation or test trials. Its predicted probability is the *BERT PoS* feature in Table D.1. (For the full-sample feature-importance analysis reported for 2025, we instead fit the head on all labeled trials.)

Sponsor. We use the Parent Group in the Trial Dataset as the sponsors of the trials. For each year, we compute the historical PoS of trials with known results sponsored by each group. Because some sponsors may have limited history, when estimating the historical PoS, we apply a Beta(1,1) prior distribution to stabilize the estimates. When a trial has multiple listed sponsors, we aggregate sponsor-level statistics by averaging across the involved groups. This historical PoS is named Past PoS. We also calculate the number of previously disclosed trials for that sponsor, which is named "Number of Past Sponsored Trials."

Phase. We use Phase in the Trial Dataset as a categorical feature and encode it via one-hot indicators. Phase III is treated as the baseline category, and therefore we only include indicators for Phase I and Phase II as features.

Trial Information. For each year, we include Days Since First Disclosure as a feature, computed as the number of days between the end of that year and the First Disclosure Date in the Trial Dataset. We also include the Number of Participating Institutions in the Trial Dataset as a feature. In addition, we include indicator variables for whether the trial

is "International" in Scope in the Trial Dataset, and whether the drug is imported in the Project Dataset.

Design. Recruitment features include Target Domestic Enrollment and Target International Enrollment in the Trial Dataset. Eligibility features include Minimum Enrollment Age and Maximum Enrollment Age in the Trial Dataset, and an additional indicator for whether an upper age limit is explicitly present (With Upper Age Limit). We also include indicators that capture whether the trial involves combination therapy (With Combination Therapy, based on Combination Therapy in the Trial Dataset) and whether it uses an active control (With Active Control, based on Control Drugs in the Trial Dataset). Finally, we summarize the eligibility criteria by counting the Number of Inclusion Criteria and the Number of Exclusion Criteria, based on the Inclusion Criteria and the Exclusion Criteria in the Trial Dataset.

Other Categorical Information. The Therapeutic Area, Line, Modality, and Dosage Form in the Trial Dataset are categorical variables, and we encode them as one-hot indicator variables. Table B.2 provides all potential labels for these variables. To ensure there are enough trials for each label, for each year, we categorize labels that appear fewer than 20 times in all disclosed trials as "Other." We use "Other" as the baseline category and therefore exclude it from the feature variables.

For features with missing data, for each year, we use medians to impute all numeric features, and use the most frequent categories to impute all categorical variables. If there are features with zero variances because of lack of data for a specific year, we remove these features before training to ensure that the feature set used for prediction is well-defined and numerically stable. Finally, we standardize all features by subtracting by their means and then dividing by their standard deviations across different trials, ensuring all features have the same magnitudes.

Example D.1 (Trial CTR20180975)**.** *To make the time-stratified feature construction concrete, consider trial CTR20180975, a Phase III trial of sintilimab—an anti-PD-1 monoclonal antibody developed by Innovent Biologics—combined with chemotherapy for non-squamous non-small-cell lung cancer, first disclosed on 2018-07-23. Registration details are publicly available on the Platform (*`http://www.chinadrugtrials.org.cn/`*) under registration number CTR20180975. If we evaluate this trial at the end of 2019, the features known at that point include: Phase = III; Therapeutic Area = Oncology; Modality = Monospecific Antibody; Target = PD-1; Scope = Domestic; Target Domestic Enrollment = 378; Number of Participating Institutions = 48; and Days Since First Disclosure = 526 (from 2018-07-23 to the end of 2019). The* Sponsor: Past PoS *feature is Innovent Biologics' historical success rate computed only from its trials whose outcomes were already known by the end of 2019, and the* Text: BERT PoS *feature is the probability the BERT head assigns to the trial's concatenated text fields. Because the trial had not yet resolved by the end of 2019, it enters the 2019 out-of-sample test set; sintilimab was subsequently approved for this indication, so the project—and hence this Phase III trial—is ultimately labeled a success.*

D.1.2 Training and Evaluation Procedure

For each calendar year Y, we construct a prediction model at the end of year Y using only trials that have known outcomes by the end of that year. Among these labeled trials, we perform a chronological split: the earliest 80% are used for training, and the remaining 20% are used for validation, mimicking a realistic forward-looking setting. For each model and year we tune a small hyperparameter grid, fitting each candidate configuration on the training fold and keeping the one with the highest validation AUC; the grids cover regularization strength and class weighting (logistic regression and the support vector machine), the number of neighbors and the weighting scheme (k-nearest neighbors), tree depth, leaf size, and class weighting (random forest), the number of trees, learning rate, and depth (gradient boosting and XGBoost), and the hidden-layer sizes and regularization (multilayer perceptron). We also use the validation fold to choose each model's classification threshold, selecting the value that maximizes Youden's J statistic (sensitivity plus specificity minus one) so that the threshold balances identifying successes and failures rather than defaulting to the majority class.

After training and validation, the out-of-sample test set consists of all trials whose outcome is not yet known at the end of year Y. We use the trained model to predict the outcomes of these trials, and compare the predictions with what subsequently happens: by the end of our sample (2025), a test trial is labeled a success or a failure if its status has resolved, and is excluded from the evaluation if it is still in progress. The number of test trials (# Test in Table 2) therefore counts only those that resolve by 2025. Of the 13,339 matched trials, the 13,205 with Phase I–III labels (Appendix C.1) form the prediction universe; 9,526 of them resolve by the end of 2025 and constitute all training, validation, and realized test observations across the annual exercises, while the remaining 3,679 are still pending at the end of 2025 and never enter any reported metric. We repeat this process for each year from 2015 through 2024; 2013 and 2014 are omitted because too few outcomes had resolved to train on, and 2025 leaves no room for outcomes to have resolved. Because every feature is computed using only information available up to the end of year Y, and every fitted preprocessing step—missing-value imputation, standardization, category collapsing, the sponsor-history statistics, and the BERT head (Appendix D.1.1)—is estimated on the training period alone, the design is strictly forward-looking and avoids look-ahead bias (target leakage).

Beyond discrimination, we evaluate probability calibration using the Brier score and the expected calibration error (ECE); the ECE is the sample-weighted absolute gap between mean predicted probability and observed success frequency across ten equal-width probability bins. As a feasible no-skill benchmark for calibration, we use a predictor that makes the same forecast for every pending trial: the overall success rate among trials whose outcomes are already known by the end of year Y.

The most recent years do call for some care. A trial pending at the end of year Y enters the metrics only once it resolves, by 2025 at the latest, so for years close to 2025 only the trials that resolve within a short window are evaluated (for example, 1,799 of those pending at the end of 2024), and these may not be representative of all pending trials. The AUC nevertheless rises gradually across the whole sample period rather than jumping only in the

final year, so the improvement is unlikely to be merely an artifact of this shortening window; it is the exact level for the latest year that should be read with the most caution.

D.1.3 Additional Models

To complement our main models, we consider several widely used classifiers that can flexibly capture nonlinear relationships between trial characteristics and outcomes. These models serve as robustness checks.

Gradient Boosting Tree. Gradient boosting builds an additive ensemble of decision trees sequentially, where each new tree is fitted to reduce the errors of the current ensemble.

XGBoost. XGBoost is an optimized implementation of gradient boosting with additional regularization and engineering improvements. We use it as a strong tree-based baseline that is widely adopted in applied machine learning.

Support Vector Machine. The support vector machine can learn nonlinear decision boundaries by implicitly mapping features into a higher-dimensional space. We use a support vector machine with a radial basis function kernel.

K-Nearest Neighbors. The k-nearest neighbors classifier predicts outcomes based on the labels of the most similar observations under a distance metric in the feature space. It provides a simple nonparametric benchmark.

Multilayer Perceptron. The multilayer perceptron is a feedforward neural network that learns nonlinear transformations of the input features through stacked hidden layers, enabling flexible decision boundaries on tabular data.

D.2 Performance, Coefficients, and Feature Importance

We report out-of-sample performance for every model and evaluation year (Section D.2.1), the full set of logistic-regression coefficients (Section D.2.2), and feature-importance rankings (Section D.2.3).

D.2.1 Performance by Year and Model

Table D.2 reports the full out-of-sample evaluation: for every model and year it gives discrimination (AUC and PR-AUC), precision, recall, and F1 separately for the success and the failure class, calibration (Brier score and ECE), and decision-oriented metrics (the lift and precision in the top predicted-success decile and the realized failure rate in the bottom decile); a final panel summarizes each model as an average across evaluation years, weighted by the number of realized outcomes. A no-skill classifier would have an AUC of 50%, a PR-AUC equal to the success share, and a lift of 1.00; the models exceed these, with the

tree-based models (random forest, gradient boosting, XGBoost) strongest on AUC and PR-AUC, while the support vector machine and k-nearest neighbors are the weakest and do not improve on logistic regression or random forest. Across models, the Brier scores and ECE values range from somewhat better than the no-skill benchmark to somewhat worse, while the largest gains over the benchmark come in discrimination. The decision-oriented metrics are informative in both directions: the top predicted-success decile is enriched for successes (its precision exceeds the base rate) and the bottom decile concentrates failures (its failure rate exceeds the base rate), so the scores help both to prioritize promising trials and to deprioritize likely failures.

Table D.2: Out-of-sample model performance by year for all models, at the validation-selected (Youden-optimal) classification threshold.

Year	Model	# Test	Discrimination		Success Class			Failure Class			Calibration		Decision Value		
			AUC	PR-AUC	P	R	F1	P	R	F1	Brier	ECE	TD-Lift	TD-Prec	BD-Fail
2015	RF	9,115	55.87	68.60	66.00	79.89	72.29	41.55	25.78	31.81	0.274	0.213	1.13	72.92	43.53
2015	LR	9,115	54.04	67.52	65.15	88.93	75.20	41.53	14.18	21.14	0.316	0.288	1.12	71.82	42.65
2015	XGB	9,115	54.27	68.11	64.45	97.92	77.74	40.78	2.58	4.86	0.313	0.284	1.11	71.38	38.49
2015	GBT	9,115	54.38	68.10	64.37	99.32	78.11	40.30	0.83	1.63	0.324	0.304	1.13	72.59	40.35
2015	MLP	9,115	54.06	67.13	66.57	60.30	63.28	38.79	45.37	41.82	0.233	0.054	1.10	70.72	40.90
2015	SVM	9,115	52.83	65.66	66.07	73.74	69.69	40.08	31.68	35.39	0.235	0.055	1.01	65.02	40.35
2015	KNN	9,115	52.49	66.27	64.49	96.88	77.43	40.20	3.78	6.92	0.281	0.217	1.06	67.98	40.02
2015	*No-Skill*	9,115	50.00	64.33	64.33	81.27	71.81	35.67	18.73	24.57	0.258	0.169	1.00	64.33	35.67
2016	RF	8,906	55.53	69.85	65.19	83.18	73.09	40.42	20.43	27.14	0.249	0.142	1.19	76.66	40.63
2016	LR	8,906	53.67	67.30	64.90	81.47	72.25	38.86	21.09	27.34	0.295	0.241	1.10	70.82	36.14
2016	XGB	8,906	55.85	69.66	66.76	66.18	66.47	40.36	40.99	40.67	0.282	0.222	1.20	76.88	41.30
2016	GBT	8,906	52.95	66.72	64.70	91.86	75.92	41.21	10.22	16.37	0.286	0.205	1.06	68.01	41.30
2016	MLP	8,906	48.24	62.82	64.19	74.59	69.00	35.89	25.48	29.80	0.244	0.097	0.96	61.50	34.68
2016	SVM	8,906	52.04	67.05	64.34	88.77	74.60	37.12	11.88	18.00	0.239	0.097	1.12	71.83	36.03
2016	KNN	8,906	50.06	64.80	64.09	89.10	74.55	35.17	10.59	16.28	0.279	0.210	1.02	65.43	35.13
2016	*No-Skill*	8,906	50.00	64.17	64.17	77.90	70.37	35.83	22.10	27.34	0.249	0.137	1.00	64.17	35.83
2017	RF	8,634	54.14	68.21	64.91	65.40	65.15	38.12	37.62	37.87	0.255	0.150	1.16	74.28	37.08
2017	LR	8,634	52.92	66.25	64.60	65.11	64.85	37.56	37.05	37.30	0.297	0.248	1.05	67.21	37.31
2017	XGB	8,634	54.29	67.43	68.78	42.66	52.66	39.42	65.83	49.31	0.288	0.236	1.08	68.83	37.89
2017	GBT	8,634	54.61	68.21	70.34	43.55	53.79	40.43	67.60	50.59	0.278	0.216	1.15	73.70	34.07
2017	MLP	8,634	45.84	60.83	55.66	4.46	8.26	35.73	93.72	51.74	0.244	0.081	0.89	56.78	30.13
2017	SVM	8,634	51.37	65.80	64.70	50.92	56.99	37.05	50.98	42.91	0.241	0.099	1.08	69.18	32.91
2017	KNN	8,634	47.17	61.61	63.27	44.67	52.37	35.72	54.24	43.07	0.299	0.242	0.92	58.52	30.82
2017	*No-Skill*	8,634	50.00	63.83	63.83	77.02	69.81	36.17	22.98	28.11	0.248	0.132	1.00	63.83	36.17
2018	RF	8,299	59.30	70.71	69.35	47.78	56.58	41.58	63.78	50.34	0.231	0.050	1.21	76.14	51.69
2018	LR	8,299	57.21	68.52	67.18	64.52	65.82	43.00	45.91	44.41	0.269	0.172	1.14	72.17	44.34
2018	XGB	8,299	58.41	69.41	64.71	88.61	74.80	46.65	17.08	25.01	0.310	0.284	1.19	75.06	45.54
2018	GBT	8,299	58.02	69.24	71.04	39.58	50.83	41.09	72.32	52.41	0.268	0.198	1.18	74.34	44.10
2018	MLP	8,299	48.66	61.87	62.03	48.08	54.17	35.73	49.51	41.50	0.243	0.077	0.95	60.12	35.90
2018	SVM	8,299	56.51	68.39	64.84	77.07	70.43	41.85	28.30	33.77	0.255	0.159	1.11	70.00	43.37
2018	KNN	8,299	48.93	62.24	63.29	86.15	72.97	37.52	14.27	20.67	0.308	0.257	0.95	60.00	39.40
2018	*No-Skill*	8,299	50.00	63.18	63.18	77.83	69.74	36.82	22.17	27.68	0.254	0.147	1.00	63.18	36.82
2019	RF	7,786	61.24	71.57	64.18	87.73	74.13	49.19	19.52	27.95	0.249	0.146	1.27	78.82	49.81
2019	LR	7,786	57.43	67.89	64.91	74.34	69.30	44.58	33.92	38.53	0.292	0.245	1.15	71.37	46.85
2019	XGB	7,786	61.06	69.82	68.31	66.85	67.57	47.36	49.03	48.18	0.286	0.240	1.17	72.66	52.25
2019	GBT	7,786	60.42	69.54	67.40	67.90	67.65	46.58	46.01	46.29	0.279	0.225	1.18	73.56	51.60
2019	MLP	7,786	59.66	70.01	65.68	79.18	71.80	48.31	31.99	38.49	0.308	0.276	1.20	74.84	51.73
2019	SVM	7,786	57.85	67.76	63.17	93.12	75.28	48.77	10.76	17.64	0.260	0.169	1.11	69.32	48.52
2019	KNN	7,786	52.20	63.42	63.22	79.24	70.33	41.50	24.21	30.58	0.301	0.242	1.02	63.29	40.69
2019	*No-Skill*	7,786	50.00	62.18	62.18	77.99	69.19	37.82	22.01	27.83	0.260	0.158	1.00	62.18	37.82
2020	RF	7,208	62.40	71.55	63.78	87.63	73.83	53.26	22.07	31.21	0.247	0.140	1.28	78.09	56.59

Continued on next page

Table D.2: Out-of-sample model performance by year for all models (continued).

Year	Model	# Test	Discrimination		Success Class			Failure Class			Calibration		Decision Value		
			AUC	PR-AUC	P	R	F1	P	R	F1	Brier	ECE	TD-Lift	TD-Prec	BD-Fail
2020	LR	7,208	60.16	68.70	66.54	67.31	66.92	47.86	46.99	47.42	0.257	0.139	1.15	70.04	54.65
2020	XGB	7,208	58.44	68.32	62.13	90.23	73.59	47.56	13.88	21.49	0.341	0.323	1.20	73.23	47.43
2020	GBT	7,208	57.96	66.56	63.10	85.31	72.54	48.73	21.86	30.18	0.330	0.307	1.14	69.63	49.79
2020	MLP	7,208	50.60	61.87	83.33	0.11	0.23	38.99	99.96	56.10	0.241	0.042	1.05	64.22	39.94
2020	SVM	7,208	60.02	69.50	67.82	61.42	64.46	47.36	54.36	50.62	0.240	0.080	1.24	75.87	49.51
2020	KNN	7,208	54.76	64.73	63.38	74.13	68.34	44.84	32.93	37.97	0.288	0.224	1.12	68.52	48.68
2020	*No-Skill*	7,208	50.00	61.03	61.03	77.61	68.33	38.97	22.39	28.44	0.265	0.166	1.00	61.03	38.97
2021	RF	6,362	64.77	71.82	63.04	86.97	73.10	57.82	25.94	35.82	0.287	0.243	1.37	81.29	61.79
2021	LR	6,362	61.32	67.99	63.22	78.16	69.90	51.70	33.96	41.00	0.280	0.217	1.19	70.44	56.45
2021	XGB	6,362	60.75	68.24	61.23	91.00	73.21	55.51	16.31	25.21	0.348	0.333	1.25	73.90	57.70
2021	GBT	6,362	61.41	68.12	62.22	87.61	72.77	55.82	22.74	32.32	0.333	0.314	1.23	72.80	59.28
2021	MLP	6,362	63.41	70.02	65.82	73.30	69.36	53.55	44.72	48.74	0.283	0.230	1.28	75.94	58.33
2021	SVM	6,362	63.87	71.37	68.48	65.31	66.86	52.78	56.32	54.49	0.229	0.034	1.35	80.03	55.82
2021	KNN	6,362	57.23	66.26	60.89	82.78	70.17	47.66	22.78	30.83	0.292	0.225	1.27	75.47	50.00
2021	*No-Skill*	6,362	50.00	59.23	59.23	76.80	66.88	40.77	23.20	29.57	0.272	0.176	1.00	59.23	40.77
2022	RF	5,263	70.01	74.27	67.42	75.19	71.09	62.50	53.24	57.50	0.283	0.243	1.48	83.08	66.16
2022	LR	5,263	66.18	69.79	63.08	78.97	70.13	59.94	40.50	48.34	0.275	0.215	1.38	77.95	66.92
2022	XGB	5,263	67.16	71.28	68.97	64.45	66.63	57.80	62.67	60.13	0.315	0.292	1.37	77.19	63.12
2022	GBT	5,263	66.86	70.48	69.74	60.77	64.95	56.67	66.06	61.01	0.297	0.261	1.38	77.76	63.69
2022	MLP	5,263	67.45	72.36	65.26	73.94	69.33	59.52	49.33	53.94	0.304	0.273	1.47	82.70	65.21
2022	SVM	5,263	67.50	71.96	66.55	72.15	69.24	59.80	53.32	56.37	0.224	0.022	1.42	79.85	66.16
2022	KNN	5,263	63.08	67.43	71.57	27.11	39.32	47.86	86.14	61.53	0.288	0.230	1.36	76.43	58.94
2022	*No-Skill*	5,263	50.00	56.28	56.28	75.91	64.64	43.72	24.09	31.06	0.285	0.196	1.00	56.28	43.72
2023	RF	3,717	72.61	73.17	59.02	88.05	70.67	73.28	34.89	47.27	0.284	0.259	1.58	81.72	82.53
2023	LR	3,717	67.06	66.84	59.33	75.95	66.62	63.50	44.56	52.37	0.273	0.209	1.47	75.81	69.62
2023	XGB	3,717	69.48	70.24	58.81	85.65	69.74	70.27	36.11	47.71	0.278	0.231	1.56	80.65	74.46
2023	GBT	3,717	70.84	70.77	57.34	90.51	70.20	73.66	28.28	40.87	0.330	0.325	1.57	81.18	75.27
2023	MLP	3,717	70.47	71.23	58.63	86.33	69.83	70.69	35.11	46.92	0.277	0.236	1.61	83.06	75.00
2023	SVM	3,717	68.43	69.40	60.73	74.70	66.99	64.31	48.56	55.33	0.225	0.023	1.53	78.76	69.89
2023	KNN	3,717	69.67	70.47	61.37	78.40	68.85	67.35	47.44	55.67	0.276	0.228	1.54	79.57	73.92
2023	*No-Skill*	3,717	50.00	51.57	51.57	73.70	60.68	48.43	26.30	34.09	0.299	0.221	1.00	51.57	48.43
2024	RF	1,799	78.59	71.74	63.52	78.12	70.07	78.53	64.06	70.56	0.270	0.270	1.87	83.33	91.67
2024	LR	1,799	70.80	65.58	53.50	79.38	63.92	73.04	44.74	55.49	0.217	0.047	1.71	76.11	81.11
2024	XGB	1,799	74.42	68.43	56.01	86.75	68.07	81.07	45.45	58.24	0.239	0.186	1.79	79.44	90.56
2024	GBT	1,799	73.78	68.81	56.35	81.00	66.46	76.58	49.75	60.32	0.308	0.311	1.84	81.67	87.78
2024	MLP	1,799	75.28	70.88	58.98	79.25	67.63	77.07	55.86	64.77	0.259	0.229	1.91	85.00	89.44
2024	SVM	1,799	71.83	67.39	54.44	78.12	64.17	73.12	47.65	57.70	0.215	0.054	1.71	76.11	77.78
2024	KNN	1,799	69.94	60.96	56.95	72.25	63.69	71.68	56.26	63.04	0.283	0.243	1.40	62.22	81.67
2024	*No-Skill*	1,799	50.00	44.47	44.47	69.86	54.34	55.53	30.14	39.07	0.311	0.254	1.00	44.47	55.53
Weighted Averages Across Years															
All	RF	67,089	60.84	70.73	65.10	76.94	69.75	49.04	33.87	38.04	0.259	0.169	1.28	77.59	52.41
All	LR	67,089	58.10	67.75	64.44	75.06	69.07	46.35	34.23	38.43	0.284	0.218	1.18	71.58	48.82
All	XGB	67,089	59.24	69.00	65.06	76.75	69.02	48.21	33.24	34.71	0.304	0.268	1.22	74.01	49.36
All	GBT	67,089	58.83	68.43	65.83	73.50	67.21	47.75	37.04	35.74	0.300	0.257	1.21	73.28	49.28
All	MLP	67,089	55.25	65.88	65.17	53.48	51.58	44.58	53.74	45.11	0.260	0.139	1.14	68.67	45.98
All	SVM	67,089	57.92	68.15	65.02	73.38	68.28	46.29	36.52	38.59	0.240	0.089	1.20	72.39	47.26
All	KNN	67,089	53.95	64.66	63.70	74.82	66.95	42.93	30.07	31.36	0.291	0.231	1.10	66.81	44.54
All	*No-Skill*	67,089	50.00	61.14	61.14	77.51	68.31	38.86	22.49	28.47	0.263	0.164	1.00	61.14	38.86

Notes: # Test is the realized-outcome test set; training and validation sizes are as in Table 2 of the main text. P, R, and F1 are precision, recall, and F1 for the success and the failure class; Brier and ECE are the calibration measures defined in Appendix D.1.2; TD-Lift, TD-Prec, and BD-Fail are the top-decile lift, top-decile precision, and bottom-decile failure rate. Higher is better for every column except Brier and ECE, for which lower is better. All metrics are reported as percentages, except TD-Lift (a multiplicative ratio) and Brier and ECE (unit-free scores). The *No-Skill* row is a benchmark that assigns every pending trial in that year the same predicted probability: the success rate among trials whose outcomes were already known by the end of the year; its AUC, PR-AUC, Brier, and ECE follow from this constant forecast. Its precision, recall, and F1 entries are the expected values of a randomized classifier that predicts success with that same probability, so each precision

equals the corresponding realized test-set class share while each recall equals the constant itself; TD-Lift, TD-Prec, and BD-Fail are expected values under random tie-breaking among the tied constant scores. The final panel reports averages across evaluation years, weighted by the number of realized outcomes; its N is the pooled number of realized-outcome test evaluations (trial–years) across the evaluation years—repeated, non-independent evaluations of 9,526 unique trials, each evaluated in every year before its outcome resolves. RF, LR, GBT, XGB, SVM, KNN, and MLP denote random forest, logistic regression, gradient boosting trees, XGBoost, support vector machine, k-nearest neighbors, and multilayer perceptron.

D.2.2 Logistic Regression Coefficients

In our main article, we report selected coefficients and t-statistics from the logistic regression specification in Table 3. The following Table D.3 reports the full table for all estimated coefficients.

Table D.3: Estimated values of coefficients and t-statistics of the logistic regression for all features.

Feature Name	Coefficient	t-Statistics	Feature Name	Coefficient	t-Statistics
Text: BERT PoS	4.54***	(18.64)	Line: Adjuvant	0.38	(1.07)
Sponsor: Past PoS	6.44***	(35.19)	Line: Neoadjuvant	0.63	(1.10)
Sponsor: Number of Past Sponsored Trials	-0.00	(-0.05)	Line: Maintenance/Consolidation	0.71*	(1.84)
Phase: I	0.60***	(6.17)	Modality: ASO	-1.11**	(-2.12)
Phase: II	0.35***	(3.51)	Modality: CAR-T	0.68*	(1.85)
Information: Days Since First Disclosure	0.21***	(7.22)	Modality: PROTAC	-1.54**	(-2.16)
Information: Number of Participating Institutions	0.05	(1.56)	Modality: siRNA	0.59	(1.56)
Information: International Scope	-0.67***	(-6.05)	Modality: Other Protein	0.06	(0.17)
Information: Imported Drug	0.17*	(1.81)	Modality: Other Diagnostic Reagent	0.71	(1.17)
Design: Target Domestic Enrollment	-0.08**	(-2.34)	Modality: Chemical	0.09	(0.41)
Design: Target International Enrollment	-0.01	(-0.32)	Modality: Monospecific Antibody	0.16	(0.73)
Design: With Active Control	-0.08	(-1.34)	Modality: Bispecific Antibody	-0.10	(-0.39)
Design: With Combination Therapy	0.37***	(4.10)	Modality: Trispecific Antibody	1.82	(1.43)
Design: With Upper Age Limit	-0.07	(-0.97)	Modality: Molecular Glue Degrader	-0.70	(-0.76)
Design: Minimum Enrollment Age	-0.06**	(-2.22)	Modality: Gene Therapy	-1.25***	(-3.04)
Design: Maximum Enrollment Age	-0.07**	(-2.53)	Modality: Peptide	0.20	(0.82)
Design: Number of Inclusion Criteria	0.04	(1.07)	Modality: Microbiome-Based	1.32**	(2.21)
Design: Number of Exclusion Criteria	-0.01	(-0.48)	Modality: Antibody–Drug Conjugate	1.42***	(5.19)
Therapeutic Area: Immunology	0.17	(1.58)	Modality: Antibody Fusion Protein	-0.20	(-0.78)
Therapeutic Area: Endocrinology	-0.01	(-0.10)	Modality: Undetermined	-1.50**	(-2.45)
Therapeutic Area: Otolaryngology	-0.25	(-0.96)	Modality: Therapeutic Vaccine	-0.21	(-0.49)
Therapeutic Area: Respiratory	0.03	(0.24)	Modality: Therapeutic Radiopharmaceutical	0.12	(0.17)
Therapeutic Area: Cardiovascular	-0.00	(-0.02)	Modality: Mixed Antibody	0.72	(1.12)
Therapeutic Area: Infectious	0.13	(1.15)	Modality: Oncolytic Virus	0.24	(0.58)
Therapeutic Area: Urology	-0.05	(-0.32)	Modality: Cyclic Peptide	0.83	(1.51)
Therapeutic Area: Gastroenterology	-0.21*	(-1.86)	Modality: Cytokine	0.26	(0.81)
Therapeutic Area: Reproductive	-0.37*	(-1.86)	Modality: Diagnostic Radiopharmaceutical	-0.39	(-0.92)
Therapeutic Area: Dermatology	0.38**	(2.42)	Modality: Enzyme	-0.32	(-0.87)
Therapeutic Area: Ophthalmology	0.07	(0.25)	Modality: Mesenchymal Stem Cell	0.56	(1.60)

Feature Name	Coefficient	t-Statistics	Feature Name	Coefficient	t-Statistics
Therapeutic Area: Neurology	-0.03	(-0.29)	Modality: Non-Antibody Fusion Protein	0.79**	(2.19)
Therapeutic Area: Psychiatry	-0.38**	(-2.41)	Modality: Non-Degrading Molecular Glue	-1.22*	(-1.76)
Therapeutic Area: Rare	0.13	(1.14)	Modality: Prophylactic Vaccine	0.42	(1.51)
Therapeutic Area: Oncology	-0.32***	(-2.87)	Dosage Form: Dermatological Preparation	-0.86**	(-2.18)
Therapeutic Area: Hematology	0.60***	(3.38)	Dosage Form: Inhaled Preparation	-0.53	(-1.39)
Therapeutic Area: Musculoskeletal	0.09	(0.61)	Dosage Form: Injectable Preparation	-0.93***	(-2.81)
Line: First	0.71***	(4.64)	Dosage Form: Ophthalmic Preparation	-0.02	(-0.05)
Line: Second	0.18	(1.51)	Dosage Form: Oral Liquid Dose Preparation	-1.37***	(-3.47)
Line: Third	0.20	(1.05)	Dosage Form: Oral Solid Dose Preparation	-1.02***	(-3.06)
Line: Last	-0.43**	(-2.31)	Dosage Form: Patch	-1.51***	(-2.81)

Notes: ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

D.2.3 Feature Importance

This appendix summarizes which trial characteristics are most influential for the additional models when trained on the full set of trials with outcomes.

Figure D.1 reports the top 20 in terms of feature importance in the random forest. Several patterns stand out. First, sponsor-level history dominates the model: Sponsor: Past PoS is by far the most important predictor, indicating that the model relies heavily on the sponsor's historical success profile when ranking trial outcomes. Second, textual signals and information-timing are consistently influential. Text: BERT PoS and Information: Days Since First Disclosure appear among the top three features, indicating that the fitted models draw both on the overall positivity signal extracted from the registration text and on how long a trial has been publicly disclosed; their incremental out-of-sample contributions are quantified by the ablations in Appendix D.3. This aligns well with the logistic regression results, where these variables also exhibit strong explanatory power.

Third, multiple trial design and population variables contribute meaningfully, though with smaller marginal importance than sponsor, text, and information. In particular, Design: Target Domestic Enrollment is among the most important design features, and eligibility-related variables—Number of Inclusion/Exclusion Criteria and Minimum/Maximum Enrollment Age—also rank relatively high. This is intuitive: enrollment scale and eligibility stringency/complexity often capture feasibility and patient heterogeneity, and the random forest can exploit nonlinearities and interactions between these features that a linear model may only capture imperfectly.

Finally, many indicators related to therapeutic area, modality, and dosage form appear with only comparatively minor importance in this aggregate model. This does not imply that these factors are irrelevant; rather, it suggests that once sponsor history, text signal, disclosure timing, and basic design and population variables are available, the incremental gain from any single category indicator is limited.

We conclude by pointing out that the random forest feature importance reflects contribution to impurity reduction—it indicates how much a feature is used to improve prediction,

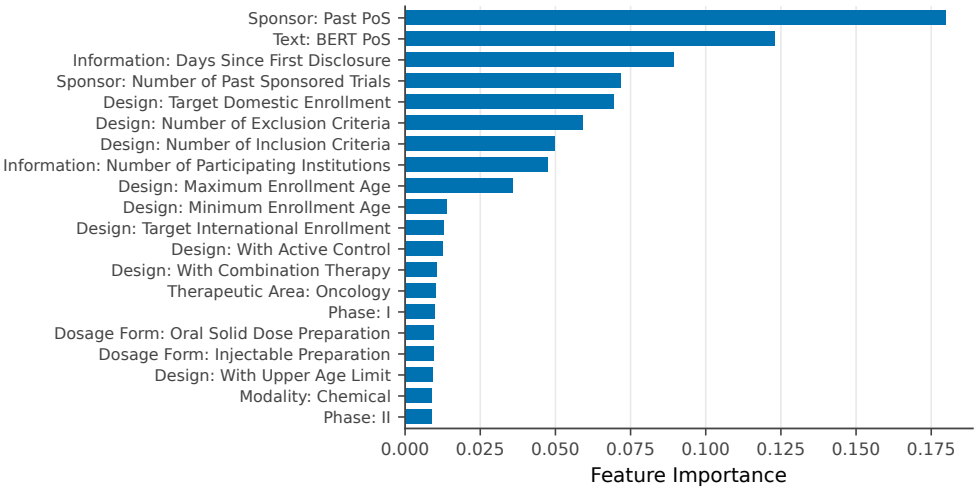


Figure D.1: Feature importance in the random forest model trained on data for all clinical trials.

not whether it increases or decreases PoS. In addition, all of these feature importance aspects should be interpreted in terms of predictive salience rather than causal effect.

Figure D.2 shows the corresponding top-20 feature importance results for other machine learning models.

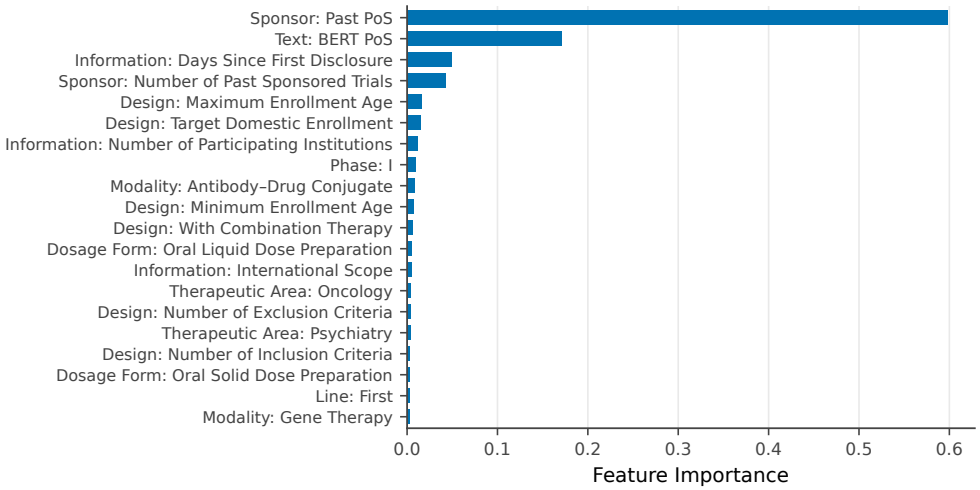
D.3 Robustness

We probe robustness by ablating feature blocks, including an analysis of new sponsors (Section D.3.1), and by evaluating predictive performance within novelty strata (Section D.3.2).

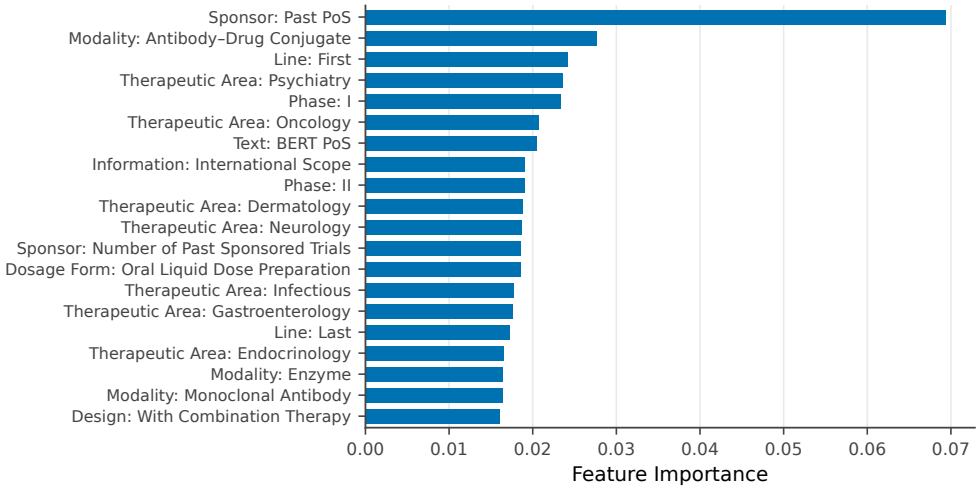
D.3.1 Feature Ablation and New Sponsors

The strong role of sponsor history in our coefficient and feature-importance results (main text and Appendix D.2.3) raises a natural concern: is the model little more than a sponsor-reputation index? To probe this, for each calendar year we re-estimate every model with the sponsor-history block removed, with the BERT text score removed, and with both removed, and recompute discrimination. Table D.4 reports the change in AUC relative to the full model, separately for all trials, for trials whose sponsor has a prior registry record, and for trials from sponsors with no prior record.

Two patterns emerge. First, all three feature groups contribute, so the model is more than a sponsor-reputation index: removing sponsor history lowers AUC the most (by about 0.8 to 3 percentage points across models), removing the BERT text score lowers it modestly (a positive contribution for nearly every model), and even with both removed the AUC stays above 50%—about 57% for the random forest—reflecting independent signal in the design and disclosure features. Second, the clear exception is new sponsors: for trials from sponsors with no prior registry record, full-model AUC is only about 52%, and removing the sponsor-history block barely changes it—a sponsor with no prior record is simply assigned a

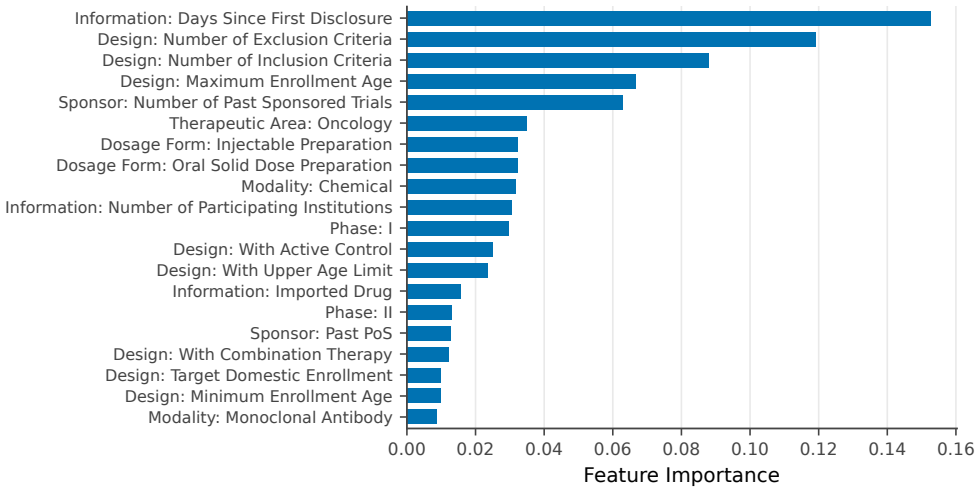


(a) Gradient boosting.

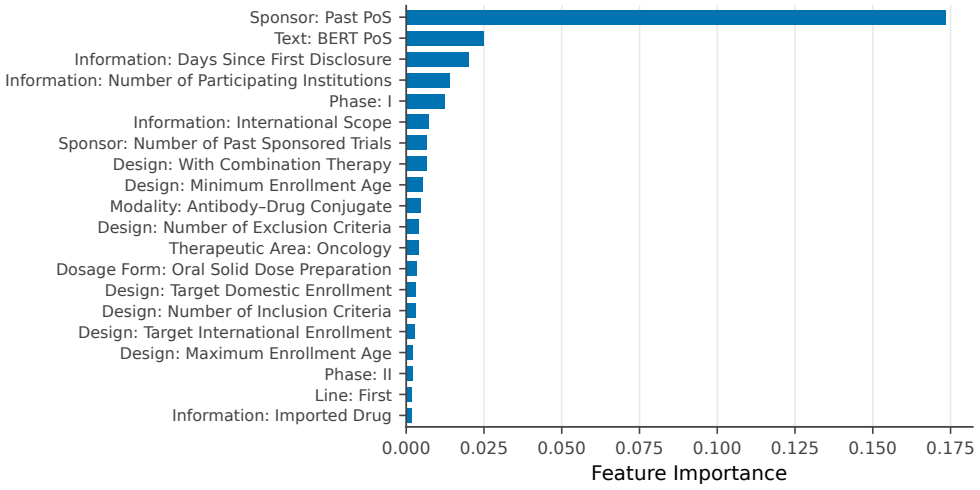


(b) XGBoost.

Figure D.2: Feature importance of other models trained on data for all clinical trials.

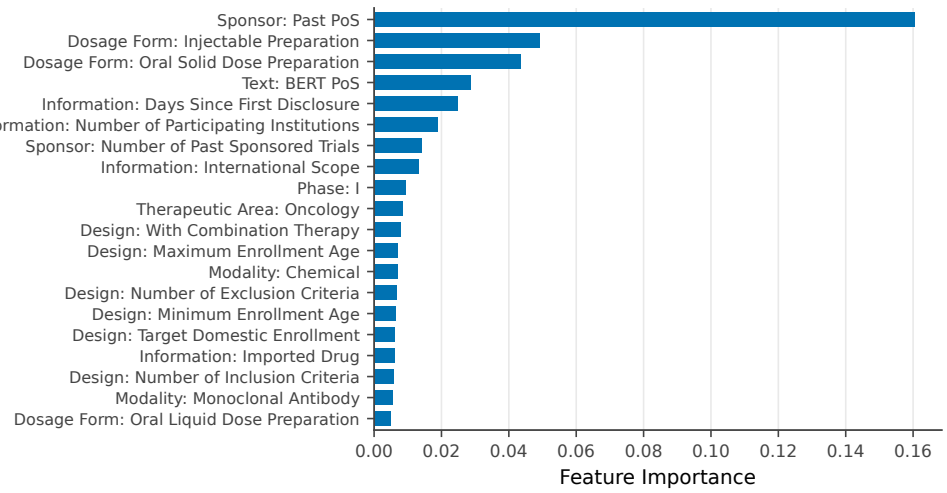


(c) K-nearest neighbors.



(d) Support vector machine.

Figure D.2: Feature importance of other models trained on data for all clinical trials (continued).



(e) Multilayer perceptron.

Figure D.2: Feature importance of other models trained on data for all clinical trials (continued).

Table D.4: Feature ablation by sponsor-history subgroup.

Model	N	Full AUC (%)	AUC Loss from Removing Block (Full − Ablated)		
			Sponsor	BERT	Both
Panel A: All Trials					
RF	67,089	60.84	+2.95	+0.51	+3.66
LR	67,089	58.10	+2.04	+0.20	+3.61
XGB	67,089	59.24	+2.42	+1.26	+3.35
GBT	67,089	58.83	+1.49	+0.27	+2.53
MLP	67,089	55.25	+2.05	-1.13	+4.21
SVM	67,089	57.92	+2.37	+0.15	+2.70
KNN	67,089	53.95	+0.83	+0.03	+0.95
Panel B: Sponsors with Prior Record					
RF	50,894	62.04	+2.37	+0.39	+3.16
LR	50,894	58.84	+1.43	+0.26	+3.08
XGB	50,894	59.80	+1.46	+1.12	+2.34
GBT	50,894	59.62	+1.02	+0.26	+2.23
MLP	50,894	56.20	+1.73	-0.47	+4.24
SVM	50,894	59.27	+2.02	+0.15	+2.41
KNN	50,894	55.34	+0.80	+0.03	+0.81
Panel C: New Sponsors					
RF	16,195	51.52	-0.78	+0.45	-0.62
LR	16,195	52.32	-0.50	+0.07	+0.48
XGB	16,195	53.01	+0.27	+1.79	+1.11
GBT	16,195	53.07	-0.38	+0.16	+0.07
MLP	16,195	49.87	+1.57	-1.50	+2.16
SVM	16,195	50.43	-0.38	+0.19	-0.28
KNN	16,195	47.74	-1.02	+0.03	-0.63

Notes: Entries are weighted averages across evaluation years. N is the pooled number of realized-outcome test evaluations (trial–years) across the evaluation years—repeated, non-independent evaluations of the underlying trials; the two sponsor-history panels partition the All-Trials evaluations. Full AUC is in percent; each remaining column is the AUC loss from removing the indicated feature block, computed as full-model AUC minus the ablated-model AUC (in percentage points). Sponsor, BERT, and Both denote removing the sponsor-history block, the BERT text score, and both, respectively; a positive value means the block improved discrimination (i.e., removing it lowered AUC).

default historical PoS (the Beta$(1, 1)$ prior mean of 0.5), so the feature carries no information for them in the first place. For genuinely new entrants—spin-outs, first-time sponsors, or licensors without a registry history—the model therefore adds little beyond chance, precisely the cases where an external benchmark would be most useful.

D.3.2 Predictive Performance by Novelty Stratum

As a further robustness check, Table D.5 reports the models' out-of-sample AUC within the same novelty strata used for the PoS analysis (Appendix C.5.2), overall and by therapeutic area. Discrimination varies across strata—pooled AUC is higher for globally launched than for not-yet-launched programs (68.5% versus 57.4% for the random forest), with larger gaps in some therapeutic areas—but the models retain discriminating power in most strata, so predictive performance is not confined to any single novelty class.

Table D.5: Model AUC (%) by novelty proxy and therapeutic area.

Dimension	Stratum	N	RF	LR	XGB	GBT	MLP	SVM	KNN
Panel A: Cardiovascular									
Global Launch Status	Launched Globally (Follow-On)	1,248	68.99	72.83	70.87	67.67	49.64	67.98	46.26
Global Launch Status	Not Yet Launched (Novelty Proxy)	3,156	52.01	54.83	52.61	53.43	51.94	54.05	49.74
Domestic/Imported	Domestic	3,471	55.69	57.60	56.10	56.69	53.84	56.17	50.66
Domestic/Imported	Imported	915	55.35	54.26	55.73	54.42	48.74	53.80	44.32
Biologic/Chemical	Biologic	946	69.49	65.92	66.16	64.87	62.97	64.45	55.83
Biologic/Chemical	Chemical	3,458	52.59	56.62	54.17	54.46	50.20	54.20	48.28
Panel B: Dermatology									
Global Launch Status	Launched Globally (Follow-On)	990	79.80	74.15	75.27	74.60	63.54	73.80	55.77
Global Launch Status	Not Yet Launched (Novelty Proxy)	2,646	58.32	56.58	56.88	58.14	53.24	54.45	50.84
Domestic/Imported	Domestic	2,936	62.38	60.90	60.93	62.71	55.73	59.96	51.44
Domestic/Imported	Imported	716	66.83	62.91	61.46	60.58	58.97	59.32	57.16
Biologic/Chemical	Biologic	1,491	61.29	62.21	58.10	59.89	56.48	61.86	51.58
Biologic/Chemical	Chemical	2,161	63.47	58.89	62.00	63.01	55.47	56.90	51.13
Panel C: Endocrinology									
Global Launch Status	Launched Globally (Follow-On)	1,829	60.11	57.10	57.30	58.25	55.69	55.63	49.38
Global Launch Status	Not Yet Launched (Novelty Proxy)	5,278	60.90	59.61	60.04	61.49	51.57	56.04	54.05
Domestic/Imported	Domestic	5,890	62.40	61.49	60.79	61.73	53.04	57.82	53.19
Domestic/Imported	Imported	1,213	62.04	62.57	62.23	61.62	52.19	59.64	52.80
Biologic/Chemical	Biologic	2,203	53.25	51.94	51.35	53.53	52.41	53.86	48.02
Biologic/Chemical	Chemical	4,918	65.81	65.11	65.31	65.28	52.44	59.56	55.72
Panel D: Gastroenterology									
Global Launch Status	Launched Globally (Follow-On)	1,263	60.01	51.46	57.04	54.68	62.34	56.80	58.86
Global Launch Status	Not Yet Launched (Novelty Proxy)	3,001	53.35	53.20	51.80	52.63	46.96	51.23	52.87
Domestic/Imported	Domestic	3,408	58.28	55.83	55.75	56.05	52.38	55.27	55.03
Domestic/Imported	Imported	857	54.77	52.14	52.11	51.85	51.97	51.93	55.40
Biologic/Chemical	Biologic	1,023	55.60	54.34	49.92	51.38	50.86	56.64	57.10
Biologic/Chemical	Chemical	3,258	58.77	56.10	57.13	56.68	53.04	55.49	54.81
Panel E: Hematology									
Global Launch Status	Launched Globally (Follow-On)	696	67.29	68.58	63.26	66.21	54.15	63.95	49.72
Global Launch Status	Not Yet Launched (Novelty Proxy)	1,174	55.47	55.16	53.19	54.12	55.67	53.50	49.96
Domestic/Imported	Domestic	1,433	62.34	57.71	59.34	59.78	58.03	57.20	50.54
Domestic/Imported	Imported	436	50.79	56.01	45.46	47.65	51.12	55.83	53.58
Biologic/Chemical	Biologic	973	56.22	56.98	49.24	52.15	55.35	57.05	54.56
Biologic/Chemical	Chemical	910	65.60	64.48	64.42	64.09	56.19	59.94	51.57
Panel F: Immunology									
Global Launch Status	Launched Globally (Follow-On)	2,299	63.93	60.93	64.65	64.41	58.35	64.58	52.97
Global Launch Status	Not Yet Launched (Novelty Proxy)	5,650	55.95	53.98	55.15	55.36	52.29	53.45	51.38

Continued on next page

Table D.5: Model AUC (%) by novelty proxy and therapeutic area (continued).

Dimension	Stratum	N	RF	LR	XGB	GBT	MLP	SVM	KNN
Domestic/Imported	Domestic	5,913	61.15	57.96	59.67	58.95	55.89	58.72	50.98
Domestic/Imported	Imported	2,036	61.49	60.35	59.29	60.32	54.91	57.86	56.88
Biologic/Chemical	Biologic	3,370	59.32	58.87	57.12	56.38	56.92	60.22	54.13
Biologic/Chemical	Chemical	4,579	62.09	58.57	61.07	61.23	55.32	57.08	51.17
Panel G: Infectious									
Global Launch Status	Launched Globally (Follow-On)	1,867	70.42	61.75	66.74	65.43	57.01	65.89	56.15
Global Launch Status	Not Yet Launched (Novelty Proxy)	4,437	54.50	52.33	54.36	55.25	48.30	53.34	54.31
Domestic/Imported	Domestic	5,481	59.23	54.64	57.50	57.44	51.78	56.46	56.03
Domestic/Imported	Imported	807	63.84	59.55	60.70	61.13	52.01	65.77	57.03
Biologic/Chemical	Biologic	2,664	55.33	51.92	52.63	53.97	49.39	55.40	55.48
Biologic/Chemical	Chemical	3,640	61.92	57.55	61.48	60.96	51.75	58.07	54.62
Panel H: Musculoskeletal									
Global Launch Status	Launched Globally (Follow-On)	633	73.56	71.72	71.57	71.18	64.06	72.18	63.14
Global Launch Status	Not Yet Launched (Novelty Proxy)	1,764	51.22	52.76	54.08	54.11	50.36	52.09	48.89
Domestic/Imported	Domestic	1,861	59.61	62.17	61.49	59.89	54.02	59.68	51.98
Domestic/Imported	Imported	530	60.88	58.01	57.49	60.37	56.21	59.47	53.97
Biologic/Chemical	Biologic	958	56.68	55.59	54.60	55.56	54.80	57.84	52.69
Biologic/Chemical	Chemical	1,450	59.95	62.16	62.91	62.64	53.81	58.41	51.58
Panel I: Neurology									
Global Launch Status	Launched Globally (Follow-On)	1,084	62.28	63.41	59.30	59.45	49.47	60.58	50.44
Global Launch Status	Not Yet Launched (Novelty Proxy)	3,152	58.18	56.60	57.15	57.12	50.78	54.83	48.47
Domestic/Imported	Domestic	3,272	59.05	57.41	57.97	57.80	52.33	56.69	48.85
Domestic/Imported	Imported	964	63.00	64.58	59.44	60.14	47.71	59.34	49.73
Biologic/Chemical	Biologic	761	66.86	68.58	60.01	59.22	57.28	66.58	51.89
Biologic/Chemical	Chemical	3,475	58.99	58.10	58.80	59.11	49.42	55.64	49.83
Panel J: Oncology									
Global Launch Status	Launched Globally (Follow-On)	9,900	61.31	57.43	61.28	56.43	51.22	58.16	62.24
Global Launch Status	Not Yet Launched (Novelty Proxy)	17,161	59.23	56.25	60.09	58.75	54.51	56.07	55.29
Domestic/Imported	Domestic	21,570	61.63	56.56	60.61	59.19	56.42	56.86	54.73
Domestic/Imported	Imported	5,491	58.19	57.16	56.41	54.94	58.84	56.75	52.73
Biologic/Chemical	Biologic	12,844	58.75	56.93	56.76	55.22	59.31	56.49	53.12
Biologic/Chemical	Chemical	14,217	61.15	56.99	60.83	60.16	55.71	57.02	53.94
Target	Other Targets	21,454	59.43	56.36	59.41	58.40	56.18	56.60	53.64
Target	PD-1/PD-L1	5,607	60.51	56.51	55.09	53.90	60.11	55.03	52.01
Panel K: Ophthalmology									
Global Launch Status	Launched Globally (Follow-On)	397	64.11	60.03	55.65	59.79	62.26	61.14	52.96
Global Launch Status	Not Yet Launched (Novelty Proxy)	1,287	58.10	62.42	55.66	56.49	53.56	56.63	51.49
Domestic/Imported	Domestic	1,305	56.92	60.07	53.65	55.09	54.62	54.44	52.63
Domestic/Imported	Imported	397	66.33	68.02	57.57	60.01	60.70	69.76	51.42
Biologic/Chemical	Biologic	951	60.70	63.81	52.66	55.40	59.74	60.82	53.61
Biologic/Chemical	Chemical	743	57.11	60.68	56.85	57.02	51.65	55.46	52.21
Panel L: Other									
Global Launch Status	Launched Globally (Follow-On)	2,258	65.96	62.21	62.61	59.92	60.86	63.06	58.97
Global Launch Status	Not Yet Launched (Novelty Proxy)	4,482	57.25	54.60	56.79	57.56	53.64	54.45	49.88
Domestic/Imported	Domestic	5,881	61.35	58.21	60.14	59.66	56.16	59.51	52.44
Domestic/Imported	Imported	840	60.85	61.41	57.65	58.93	55.61	58.86	54.34
Biologic/Chemical	Biologic	1,428	65.62	65.07	60.56	60.98	59.41	64.77	53.96
Biologic/Chemical	Chemical	5,312	59.92	55.87	59.36	59.02	55.15	56.65	52.19
Panel M: Otolaryngology									
Global Launch Status	Launched Globally (Follow-On)	197	74.53	68.12	68.53	66.97	55.06	68.81	69.90
Global Launch Status	Not Yet Launched (Novelty Proxy)	409	58.36	59.13	58.67	56.12	50.49	56.12	55.31
Domestic/Imported	Domestic	551	64.91	62.85	63.27	63.02	54.45	62.14	60.19
Biologic/Chemical	Biologic	337	60.45	57.15	56.62	52.45	46.01	54.64	54.88
Biologic/Chemical	Chemical	283	57.20	63.12	58.55	57.01	54.60	54.96	53.87
Panel N: Psychiatry									
Global Launch Status	Launched Globally (Follow-On)	431	67.61	74.50	69.76	70.00	53.02	68.06	48.75
Global Launch Status	Not Yet Launched (Novelty Proxy)	1,828	61.96	58.91	58.79	58.93	51.39	56.50	45.88

Continued on next page

Table D.5: Model AUC (%) by novelty proxy and therapeutic area (continued).

Dimension	Stratum	N	RF	LR	XGB	GBT	MLP	SVM	KNN
Domestic/Imported	Domestic	1,876	62.62	60.46	60.03	60.04	52.54	56.90	46.64
Domestic/Imported	Imported	401	58.14	60.65	58.35	55.95	45.92	60.34	48.68
Biologic/Chemical	Chemical	2,154	61.01	59.80	59.03	58.94	50.45	56.82	45.99
Panel O: Rare									
Global Launch Status	Launched Globally (Follow-On)	1,252	64.46	64.84	62.94	62.72	60.53	66.77	57.66
Global Launch Status	Not Yet Launched (Novelty Proxy)	3,202	57.89	60.51	58.65	60.08	52.08	57.29	48.39
Domestic/Imported	Domestic	3,147	58.31	58.00	56.92	58.93	52.53	58.36	50.45
Domestic/Imported	Imported	1,307	63.21	67.49	60.56	61.49	57.93	61.95	54.26
Biologic/Chemical	Biologic	2,158	59.23	62.19	55.42	57.57	55.33	60.53	52.43
Biologic/Chemical	Chemical	2,296	60.62	61.40	60.80	61.54	52.73	59.42	51.38
Panel P: Reproductive									
Global Launch Status	Launched Globally (Follow-On)	327	49.72	41.74	50.77	48.47	51.80	44.71	40.02
Global Launch Status	Not Yet Launched (Novelty Proxy)	725	62.11	58.18	59.43	61.72	48.50	54.38	50.21
Domestic/Imported	Domestic	908	60.54	56.99	58.39	59.91	49.30	55.54	47.85
Domestic/Imported	Imported	141	34.72	36.85	39.97	35.41	46.85	37.86	39.92
Biologic/Chemical	Biologic	184	44.85	48.68	43.56	47.72	45.75	50.13	44.00
Biologic/Chemical	Chemical	849	60.47	54.06	58.86	59.48	50.15	52.03	47.48
Panel Q: Respiratory									
Global Launch Status	Launched Globally (Follow-On)	993	69.73	63.29	70.91	69.47	50.14	67.12	58.44
Global Launch Status	Not Yet Launched (Novelty Proxy)	2,831	59.70	59.17	60.26	58.95	52.72	58.02	51.50
Domestic/Imported	Domestic	2,966	59.79	57.42	60.25	59.51	51.59	57.89	51.52
Domestic/Imported	Imported	858	63.21	60.67	60.20	58.96	51.83	61.36	56.27
Biologic/Chemical	Biologic	1,360	61.54	59.69	61.71	59.51	52.01	59.67	51.20
Biologic/Chemical	Chemical	2,478	60.41	57.92	60.28	60.03	52.27	58.91	52.72
Panel R: Urology									
Global Launch Status	Launched Globally (Follow-On)	677	77.91	76.19	73.72	72.95	52.41	72.75	48.62
Global Launch Status	Not Yet Launched (Novelty Proxy)	1,449	59.58	61.19	61.11	62.13	50.65	56.63	48.45
Domestic/Imported	Domestic	1,530	63.24	62.05	63.92	64.63	51.86	59.91	48.52
Domestic/Imported	Imported	596	66.79	67.29	61.82	60.85	56.71	63.08	47.72
Biologic/Chemical	Biologic	484	53.41	56.69	50.93	50.15	48.82	53.54	48.01
Biologic/Chemical	Chemical	1,641	67.65	67.54	67.26	67.90	55.18	64.79	49.66
Panel S: All Therapeutic Areas (Pooled)									
Global Launch Status	Launched Globally (Follow-On)	20,973	68.45	64.30	65.92	63.31	59.34	65.82	59.33
Global Launch Status	Not Yet Launched (Novelty Proxy)	46,116	57.41	55.24	57.13	57.27	51.92	54.41	52.53
Domestic/Imported	Domestic	53,967	60.91	57.23	59.52	59.15	54.70	57.33	53.85
Domestic/Imported	Imported	13,122	60.53	59.86	57.99	57.48	56.60	58.90	53.97
Biologic/Chemical	Biologic	25,328	60.12	57.76	57.17	56.23	57.97	58.93	55.03
Biologic/Chemical	Chemical	41,761	61.15	58.44	60.45	60.41	53.75	57.30	53.26

Notes: Entries are realized AUC values averaged across evaluation years using the number of realized future outcomes as weights. N is the pooled number of realized-outcome test evaluations (trial–years) in the stratum across the evaluation years. Novelty strata are registry-derived proxies and should be interpreted descriptively.

E English–Chinese Variable Translations

Because the original DXY database is documented in Chinese, in this appendix, we provide a translation reference. Tables E.1–E.5 list the Chinese variable names and category labels together with the English terminology used throughout the main text.

Table E.1: Chinese names related to the registration platform and the database and their English translations used in our main article.

English Name	Chinese Name
Drug Clinical Trial Registration and Information Disclosure Platform	药物临床试验登记与信息公示平台
Center for Drug Evaluation (CDE)	药品审评中心
China Food and Drug Administration (CFDA)	国家食品药品监督管理局
National Medical Products Administration (NMPA)	国家药品监督管理局
DXY Insight	丁香园 Insight
Project Dataset	国内项目进度
Trial Dataset	CDE 临床试验
Opinions on Deepening the Reform of the Review & Approval System to Encourage Innovation in Drugs and Medical Devices	关于深化审评审批制度改革鼓励药品医疗器械创新的意见

Table E.2: Chinese names of variables in the Project Dataset and their English translations used in our main article.

English Name	Chinese Name
Sponsor	企业名称
Biologic/Chemical	药物类型
Domestic/Imported	国产/进口
Global Launch Status	成分全球上市情况
Current Project Status	当前项目进度（官方公示）

Table E.3: Chinese names of variables in the Trial Dataset and their English translations used in our main article.

English Name	Chinese Name	English Name	Chinese Name
Trial ID	登记号	Target	靶点
First Disclosure Date	首次公示日期	Modality	成分类别
Completion Date	国内试验完成日期	Dosage Form	剂型分类
Public Title	通俗标题	Line	疗法类型
Official Title	专业标题	Objectives	试验目的
Phase	标准分期	Trial Design	试验设计
Therapeutic Area	适应症疾病领域	Inclusion Criteria	纳入标准
Indication	适应症（Insight）	Exclusion Criteria	排除标准
Scope	试验范围	Target Domestic Enrollment	目标入组人数（国内）
Sponsor Name	企业名称	Target International Enrollment	目标入组人数（国际）
Parent Group	所属集团	Minimum Enrollment Age	受试者最小年龄
Number of Participating Institutions	参加机构数量	Maximum Enrollment Age	受试者最大年龄
Drug Name	药品名称	Combination Therapy	联合用药
Drug Ingredient	成分词	Control Drugs	对照药

Table E.4: Chinese names of top 15 companies with the highest number of drug development projects in the Project Dataset (Figure B.3) and their English translations used in our main article.

English Name	Chinese Name	English Name	Chinese Name
Jiangsu Hengrui	江苏恒瑞医药股份有限公司	Innovent Biologics (Suzhou)	信达生物制药（苏州）有限公司
Novartis	诺华制药	Sanofi	赛诺菲制药
Chia Tai TianQing Group	正大天晴药业集团股份有限公司	Eli Lilly	礼来制药
AstraZeneca	阿斯利康制药	Merck & Co.	默沙东制药
Qilu	齐鲁制药有限公司	Sunshine Lake Pharma	广东东阳光药业股份有限公司
Pfizer	辉瑞制药	Bayer	拜耳医药
Jiangsu Hansoh Group	江苏豪森药业集团有限公司	CSPC Zhongqi (Shijiazhuang)	石药集团中奇制药技术（石家庄）有限公司
Roche	罗氏制药		

Table E.5: Chinese names of values of categorical variables in the Trial Dataset and their English translations used in our main article.

English Name	Chinese Name	English Name	Chinese Name
Panel A: Therapeutic Area			
Immunology	免疫系统疾病	Dermatology	皮肤疾病
Endocrinology	内分泌和代谢系统疾病	Ophthalmology	眼科疾病
Otolaryngology	口耳鼻喉疾病	Neurology	神经系统疾病
Respiratory	呼吸系统疾病	Psychiatry	精神障碍疾病
Cardiovascular	心血管系统疾病	Rare	罕见病
Infectious	感染性疾病	Oncology	肿瘤
Urology	泌尿系统疾病	Hematology	血液系统疾病
Gastroenterology	消化系统疾病	Musculoskeletal	骨骼肌肉系统疾病
Reproductive	生殖系统疾病	Other	其他
Panel B: Scope			
Domestic	国内试验	International	国际多中心试验
Panel C: Line			
First	一线	Adjuvant	辅助
Second	二线	Neoadjuvant	新辅助
Third	三线	Maintenance/Consolidation	维持/巩固
Last	末线	Other	其他
Panel D: Modality			
ASO	ASO	Antibody Fusion Protein	抗体类融合蛋白
CAR-T	CAR-T	Undetermined	暂未确定类别
PROTAC	PROTAC	Therapeutic Vaccine	治疗性疫苗
siRNA	siRNA	Therapeutic Radiopharmaceutical	治疗用放射性药物
Other Protein	其他蛋白	Mixed Antibody	混合抗体
Other Diagnostic Reagent	其他诊断试剂	Oncolytic Virus	溶瘤病毒
Chemical	化药	Cyclic Peptide	环状多肽
Monospecific Antibody	单特异性抗体	Cytokine	细胞因子类
Bispecific Antibody	双特异性抗体	Diagnostic Radiopharmaceutical	诊断用放射性药物
Trispecific Antibody	三特异性抗体	Enzyme	酶
Molecular Glue Degrader	分子胶降解剂	Mesenchymal Stem Cell	间充质干细胞 MSC
Gene Therapy	基因治疗	Non-Antibody Fusion Protein	非抗体类融合蛋白
Peptide	多肽	Non-Degrading Molecular Glue	非降解型分子胶
Microbiome-Based	微生物相关	Prophylactic Vaccine	预防性疫苗
Antibody–Drug Conjugate	抗体偶联物 ADC	Other	其他
Panel E: Dosage Form			
Gel	凝胶剂	Enteric Coated Capsule	肠溶胶囊
Ointment/Cream/Plaster	软膏剂/乳膏剂/硬膏剂	Enteric Coated Tablet	肠溶片
Topical Liquid Preparation	外用液体制剂	Granule	颗粒剂
Diagnostic and Radioactive Reagent	诊断和放射性试剂	Immediate Release Tablet	速释片
Implant	植入剂	Lozenge	口含片
Aerosol/Spray/Powder Inhaler	气雾剂/喷雾剂/粉雾剂	Orally Disintegrating Film	口溶膜剂
Inhaler	吸入剂	Oral Disintegrating Tablet	口腔崩解片
Emulsion for Injection	注射用乳剂	Powder	散剂
Injection	注射剂	Regular Capsule	普通胶囊剂
Liposomes for Injection	注射用脂质体	Regular Tablet	普通片剂
Microspheres for Injection	注射用微球	Soft Capsule	软胶囊
Nasal Preparation	鼻用制剂	Sublingual Tablet	舌下片
Ophthalmic Preparation	眼用制剂	Sustained Release/Controlled Release Capsule	缓释/控释胶囊
Drop Preparation	滴剂	Sustained Release/Controlled Release Tablet	缓释/控释片
Emulsion	乳剂	Otic Preparation	耳用制剂
Oral Liquid Preparation	口服液体制剂	Patch	贴剂
Suspension	混悬剂	Suppository	栓剂

English Name	Chinese Name	English Name	Chinese Name
Oral Patch	口腔贴片	Vaginal Preparation	阴道用制剂
Chewable Tablet	咀嚼片	Other	其他
Dispersible Tablet	分散片		

References

[1] Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nature Reviews Drug Discovery. 2004;3(8):711-6.

[2] Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nature Biotechnology. 2014;32(1):40-51.

[3] Wang QH, Lin J, Peng N, Wang QM, Li M, Liu GG. Clinical development success rates for innovative drugs in China. International Journal of Pharmacology. 2022;18(6):1137-50.

[4] Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics. 2019;20(2):273-86.

[5] Lo AW, Siah KW, Wong CH. Machine learning with statistical imputation for predicting drug approvals. Harvard Data Science Review. 2019;1(1).

[6] Gayvert KM, Madhukar NS, Elemento O. A data-driven approach to predicting successes and failures of clinical trials. Cell Chemical Biology. 2016;23(10):1294-301.

[7] Qi Y, Tang Q. Predicting phase 3 clinical trial results by modeling phase 2 clinical trial subject level data using deep learning. In: Machine Learning for Healthcare Conference. PMLR; 2019. p. 288-303.

[8] Fu T, Huang K, Xiao C, Glass LM, Sun J. HINT: hierarchical interaction network for clinical-trial-outcome predictions. Patterns. 2022;3(4).

[9] Zheng W, Peng D, Xu H, Li Y, Zhu H, Fu T, et al. LIFTED: multimodal mixture-of-experts for clinical trial outcome prediction. In: ICML 2024 Workshop on Foundation Models in the Wild; 2024. p. 1-13.

[10] Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database—update and key issues. New England Journal of Medicine. 2011;364(9):852-60.

[11] Friedman Y. Location of pharmaceutical innovation: 2000–2009. Nature Reviews Drug Discovery. 2010;9(11):835.

[12] Li N, Huang HY, Wu DW, Yang ZM, Wang J, Wang JS, et al. Changes in clinical trials of cancer drugs in mainland China over the decade 2009–18: a systematic review. The Lancet Oncology. 2019;20(11):e619-26.

[13] Pang T. Considerations for healthcare products' lifecycle management. In: Sietsema WK, Meacham MM, editors. Global Pharmaceutical and Biologics Regulatory Strategy. 2nd ed. Regulatory Affairs Professionals Society; 2020. p. 221-34.

[14] DiMasi JA, Feldman L, Seckler A, Wilson A. Trends in risks associated with new drug development: success rates for investigational drugs. Clinical Pharmacology & Therapeutics. 2010;87(3):272-7.

[15] Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. vol. 1; 2019. p. 4171-86.