Bringing Back the Frontier: Biopharma Returnees and Knowledge Flow

Panle Jia Barwick Chenyan Gong Hongyuan Xia Tianli Xia*

July 4, 2026

DRAFT.
PLEASE DO NOT POST OR CIRCULATE.

Abstract

We study how return migration shapes innovation in China's biopharmaceutical sector. We link clinical trials to firms and founders' employment histories, identifying entrepreneurs who trained or worked outside China before returning to found biopharma firms. Returnee-founded firms conduct 17.3% more trials and disproportionately undertake 39.0% more innovative trials in frontier areas, especially oncology and gene therapy. Their trials involve more diseases, patients, and sites. The returnee premium persists after founders return and follows the U.S. innovation frontier, consistent with a knowledge-flow channel rather than a one-time transfer of expertise.

Keywords: Immigration, Innovation, Pharmaceutical, Entrepreneurship

*Barwick: Department of Economics, UW-Madison, NBER, and CEPR, pbarwick@wisc.edu. Gong: Department of Economics, UW-Madison, chenyan.gong@wisc.edu. H. Xia: Department of Economics, Cornell University, hx276@cornell.edu. T. Xia: Simon Business School, University of Rochester, tianli.xia@rochester.edu.

1 Introduction

Talent migrates across borders, and the United States has long been a hub for global talent (Amanzadeh et al. 2024; Prato 2025). Some foreign-born talent comes to the United States and stays, contributing to U.S. innovation and economic growth (Azoulay et al. 2022; Bernstein et al. 2022; Borjas and Doran 2012; Glennon 2024; Hunt and Gauthier-Loiselle 2010; Kerr and Lincoln 2010; Moser et al. 2014). A well-known example is the influx of German and Austrian Jewish scientists and inventors who fled Nazi Germany and later made important contributions to U.S. science and technology. Others, however, return to their countries of origin after receiving education, training, or work experience in the United States (Bahar et al. 2024; Batista et al. 2025; Filatotchev et al. 2011). For example, roughly a quarter of U.S.-trained scientists eventually leave the United States (Shvadron et al. 2025). This trend appears to have accelerated in recent years. While much effort has been devoted to understanding how immigrants boost innovation and economic growth in receiving countries, much less is known about those who return to their countries of origin. Why do they return? How do they contribute to innovation and economic growth at home? And what are the welfare implications for both the originating country and the receiving country?

In this paper, we focus on biopharma returnees from China, in particular those who received education or work experience in the United States and later returned to China to found biopharma firms. We use newly assembled founder–firm–trial data to examine whether returnee-founded firms differ from local-founded firms in the scale, novelty, and technological direction of clinical innovation in China.

This setting is meaningful for several reasons. First, China represents one of the largest and most economically important cases of high-skilled return migration. Chinese students remain one of the largest groups of international students in the United States, and the scale of return migration to China has increased substantially in recent years (Mallapaty 2025; Shi et al. 2023). Quantifying the contribution of these returnees is therefore important for understanding the development of China's innovation ecosystem. Second, biopharma is one

of the most active and globally connected innovation ecosystems, where scientific knowledge and clinical development are closely intertwined. China's biopharma sector is also rapidly becoming a central part of this global ecosystem (Barwick et al. 2026). Third, instead of focusing on inventors or academic scientists, as much of the prior literature has done, we focus on entrepreneurs. Entrepreneurs are a distinct group because they not only produce knowledge but also connect scientific ideas to product development and commercialization. This makes them a particularly useful lens for studying how return migration translates foreign training and experience into domestic innovation.

To understand how returnee entrepreneurs contribute to domestic innovation, we assemble comprehensive data on biopharma firms conducting clinical trials in China. The clinical trial information comes from Citeline TrialTrove, which covers clinical trials conducted in China and provides a comprehensive picture of the country's innovation and commercialization landscape. We then link each firm to its founders and their backgrounds using employee resume data from Revelio, company profile information from PitchBook, and manually collected information from public sources. This data construction allows us to identify whether biopharma firms were founded by returnees, where these founders were trained or worked before returning, and how their firms subsequently participated in China's clinical development ecosystem. Our matching algorithm identifies the founders and firms associated with more than 79% of clinical trials conducted by Chinese firms, giving us a broad and systematic view of how returnee entrepreneurs contribute to domestic biopharma innovation.

We find that returnee-founded firms are younger but employ more workers than local-founded firms. In our sample of 1,608 biopharma firms, 940 firms were founded by 1,782 returnee entrepreneurs. A simple calculation shows that returnee-founded firms account for more than 70% of clinical trials conducted in China between 2010 and 2024. In other words, a relatively small group of returnee entrepreneurs is associated with a disproportionately large share of China's biopharma clinical development activity.

Taking the analysis to the firm level, we show that returnee-founded firms conduct 17.3%

more clinical trials than local-founded firms. The difference is even larger for innovative clinical trials: returnee-founded firms conduct 6.8% to 39.0% more innovative trials, depending on the innovative measure. These differences are particularly pronounced among smaller firms and among firms whose founders have prior industry experience. The returnee premium is concentrated in frontier therapeutic areas, especially oncology and gene therapy. Moreover, trials conducted by returnee-founded firms cover a wider range of diseases and involve more patients and trial sites, suggesting more complex and ambitious trial designs.

While the goal of this paper is not to fully decompose all sources of the returnee advantage, we provide evidence on two mechanisms that may explain why returnee-founded firms conduct more and more innovative trials. The first is a static knowledge-stock mechanism: returnee entrepreneurs bring back accumulated knowledge, training, and experience from abroad, which may gradually depreciate as the frontier moves forward. The second is a dynamic knowledge-flow mechanism: returnee entrepreneurs remain exposed to the U.S. innovation frontier and continue to channel new scientific and clinical opportunities into China. Our evidence is more consistent with the latter view. The returnee premium persists long after founders return to China and tends to follow the recent U.S. innovation frontier. This suggests that returnee entrepreneurs do not simply bring back a fixed stock of foreign knowledge; they help sustain an ongoing flow of frontier knowledge into China's biopharma innovation ecosystem.

We are currently extending the analysis by constructing instrumental variables and conducting counterfactual welfare analysis. The IV strategy addresses the endogeneity concern: returnee entrepreneurs may differ systematically from local founders in ability, motivation, and networks, and they may return to sectors or regions that are already experiencing faster innovation growth. To isolate exogenous variation in return migration, we exploit two instruments that capture push and pull forces. The push-side instrument uses variation in U.S. labor-market tightness, which affects the opportunity cost of staying in the United States. The pull-side instrument uses variation in Chinese talent-attraction policies, which affects the

incentive to return. These two instruments also speak to two competing policy approaches: whether return migration is shaped more by conditions in the receiving country that push talent out, or by origin-country policies that pull talent back. Finally, we use the estimated returnee premium in clinical-trial quantity and novelty to conduct counterfactual analysis, comparing the observed innovation ecosystem with two benchmark scenarios: one in which no U.S.-trained talent returns to China and another in which all U.S.-trained talent returns.

This paper contributes to the literature on immigration and innovation. Existing work shows that high-skilled immigrants contribute substantially to innovation in destination countries, both directly through their own inventive output and indirectly through spillovers to local inventors (Azoulay et al. 2022; Bernstein et al. 2022; Borjas and Doran 2012; Glennon 2024; Hunt and Gauthier-Loiselle 2010; Kerr and Lincoln 2010; Moser et al. 2014). This literature has also debated whether immigrant scientists and inventors complement or crowd out native innovators. Related work further shows that migrant inventors can diffuse knowledge across borders (Bahar et al. 2024). We extend this literature in two ways. First, we shift the focus from inventors to entrepreneurs, who not only produce knowledge but also organize firms and translate scientific ideas into product development and commercialization. Second, we distinguish between a static knowledge-stock channel and a dynamic knowledge-flow channel, showing how returnee entrepreneurs can serve as a persistent bridge between the U.S. innovation frontier and China's biopharma ecosystem.

Second, this paper contributes to the literature on global competition for high-skilled talent. A central tension in this literature is that origin countries worry about brain drain, while receiving countries benefit from immigrant talent but may also worry about crowding out local workers and losing talent to return migration (Agrawal et al. 2011; Borjas and Doran 2012; Docquier and Rapoport 2012; Gibson and McKenzie 2014). These concerns have become increasingly important in policy debates over how to attract, retain, or bring back globally trained scientists and entrepreneurs (Prato 2025). We contribute by studying return migration as the outcome of competing policy forces. U.S. labor-market conditions

shape the incentives to stay, while Chinese talent-attraction policies shape the incentives to return. This push–pull framework allows us to examine which forces matter more for return migration and how the allocation of U.S.-trained talent affects innovation across countries. We further use counterfactual analysis to quantify the aggregate implications of alternative talent-allocation regimes, including scenarios in which no U.S.-trained talent returns to China or all U.S.-trained talent returns.

Third, this paper contributes to the literature on founder human capital and science-based entrepreneurship. Prior work shows that founders' backgrounds, experience, and social networks shape new ventures' strategies and performance (Choi et al. 2025; Gompers et al. 2010; Zucker et al. 1994). We extend this literature by studying returnee founders in a science-based industry and showing that foreign education and experience affect not only whether firms innovate, but also the direction, novelty, and scale of their innovation. This contribution is especially relevant in biopharma, where entrepreneurs play a central role in translating scientific knowledge into clinical development.

2 Data and Research Design

2.1 Data

To study how returnee entrepreneurs shape pharmaceutical innovation in China, we assemble a new founder–firm–trial dataset that links each clinical trial conducted by a private biopharma firm in China to the firm's founders and to the founders' education and career histories. We combine three data sources: clinical-trial records from Citeline TrialTrove, employee résumés from Revelio Labs, and company and founder profiles from PitchBook.

Citeline TrialTrove Database. Our clinical-trial data come from Citeline TrialTrove, a commercial database that compiles trial records from national registries, including ClinicalTrials.gov in the U.S. and Chinese Clinical Trial Registry (ChiCTR), and supplements

them with sponsor announcements and other public filings. Each record describes a single trial: the drug under investigation and its biological mechanism, the sponsoring organization (classified as a private company or a public institution and geocoded to its parent firm's headquarters), the therapeutic area and indication, the development phase, the dates on which the trial began and concluded, and indicators of trial scale such as planned enrollment and the number of clinical sites. We link these records to our other sources through the sponsoring firm.

Revelio Labs. Revelio Labs assembles individual résumés from a range of online sources, comparable in content to professional networking platforms such as LinkedIn, with broad coverage of white-collar employment. The data over-represent management, business, and technology occupations and under-represent fields such as agriculture and manufacturing (Tambe 2026). We use Revelio to identify each firm's founders and to recover their education and employment histories. Specifically, we define founders as the earliest employees observed in the firm, which allows us to construct a consistent founder-level panel across a broad set of firms.

PitchBook. PitchBook is among the most comprehensive databases of startups and venture-capital activity, with especially strong coverage in the post-2000 period (Marx et al. 2026; Retterath and Braun 2020). It records founders, funding rounds, and firm characteristics, and provides company- and member-level profile links that we use to connect firms to the Revelio résumé data. We use PitchBook as a complementary source of founder education for firms with limited Revelio coverage, and as the source for two firm attributes: the firm's founding year and whether the firm appears in PitchBook's venture-funded universe. Because PitchBook directly identifies a firm's actual founders for most companies, we also use it to validate our tenure-based founder rule, confirming that our returnee-firm classification is not an artifact of how we select founders from the Revelio panel.

Data Linkage. We link each clinical trial to a founder in four steps: each trial is associated with its sponsoring organization in TrialTrove; each private, China-based sponsor is matched to a firm in the Revelio and PitchBook data; each firm is associated with its founders; and each founder is linked to the country of their educational institutions. Sponsor-to-firm matching combines name normalization, fuzzy string matching, and large-language-model verification, together with manual review. Coverage is high on the dimensions that matter for our analysis. Of the 31,588 trials sponsored by 3,819 distinct Chinese firms, we match 1,750 sponsors (45.8%) to a firm in the Revelio or PitchBook data; because matched sponsors are larger, more active firms, these account for 79.5% of all trials conducted by Chinese firms. Of these matched sponsors, 1,608 (91.9%) have at least one founder or early employee with observed education, forming our regression sample.

2.2 Variable Construction

Table 1 documents the firm-level summary statistics for key variables in our sample.

Returnee founders and returnee-founded firms. We define a firm's *founders* as the five employees with the earliest start dates on or after the firm's founding year for whom we observe at least one education record with an identifiable institution country.[1] We classify each institution as Chinese, U.S., or other, where "Chinese" includes mainland China, Hong Kong, and Macau. A founder is a *returnee* if they hold at least one degree from a non-Chinese institution, and a firm is *returnee-founded* if at least one of its founders is a returnee and *local-founded* otherwise. The treatment indicator ReturneeFounded_i equals one for returnee-founded firms, which make up 58.5% of our sample. We adopt the at-least-one-returnee threshold because a single internationally trained founder can shape a firm's scientific direction. We also test robustness using alternative definitions, including the ear-

[1] Using the earliest employees rather than current titles captures the individuals present at firm formation and avoids relying on seniority labels, which are missing for many résumé records. When a firm has no matched Revelio founders, we use its PitchBook-listed founders instead.

liest three employees and the share of returnee founders. We additionally record narrower dimensions of foreign exposure: a foreign doctorate, capturing advanced foreign education (25% of firms), and foreign work experience in industry (53%) or academia (11%). Each is nested within the returnee definition and used in the heterogeneity analysis below.

Trial novelty. We mainly characterize a firm's clinical-trial activity along two dimensions: the number of trials a firm sponsors and the number of novel trials conducted by a firm.

We quantify trial novelty using three complementary measures, each designed to separate frontier R&D from incremental, "me-too" development. First, following Dranove et al. (2022), we construct a mechanism-of-action novelty index that captures how often a drug's molecular target has been pursued in earlier trials.[2] We classify a trial as high-novelty if it lies in the top half of its start-year cohort.

Second, we apply a fine-tuned large-language-model classifier to the textual registry record, which labels each trial's tested drug as a new molecular entity, a novel combination, a new indication for an existing molecule, or a generic; we treat the first three as non-generic, innovative trials.

Third, we classify a trial as *first-in-China* if it is the first China-firm trial to test a given mechanism of action, meaning that no earlier China-firm trial at the same or a more advanced phase had tested the same mechanism. For each measure we count, at the firm level, the number of trials meeting the definition. On average a firm runs 7.0 high-novelty and 12.8 non-generic trials, while first-in-China trials are considerably rarer (a mean of 0.56 per firm).

[2]Following Dranove et al. (2022), the novelty of trial i with mechanism $m(i)$ at phase $j(i)$ is

$$\text{Novelty}_i = \frac{1}{1 + \#\{\text{prior trials at phase } j(i) \text{ using mechanism } m(i)\}}.$$

This separates drugs whose mechanisms have rarely been explored from those relying on well-established pathways. We calculate the novelty median within each year, pooling observations across all countries to benchmark novelty against the global clinical development frontier rather than country-specific baselines.

Trial scale. We also characterize the scale of firms' trials. Scale is measured by the average target enrollment (patients), the average number of identified clinical sites, and the average trial duration. Trial activity is highly concentrated: firms sponsor 19 trials on average but only 6 at the median, while the most active firm runs more than 1,500. The underlying trials average 469 days in duration, a target enrollment of 245 patients, and fewer than 6 identified clinical sites.

Firm controls. We control for the firm's employee headcount (from Revelio, with PitchBook as a fallback), its founding year, an indicator for whether it appears in PitchBook's venture-funded universe, and its headquarters province. Employment is heavily skewed toward a few large firms. The median firm has about 20 employees, the mean exceeds 110, and the largest reports more than 12,000. The median founding year is 2010, consistent with the recent expansion of China's biopharmaceutical sector, and about 35% of firms appear in PitchBook's venture-funded universe. Figure 1 shows how the two groups differ in age and size. Returnee-founded firms are younger, with most founded after 2010 and a large share after 2015, while local-founded firms date mainly from the early-to-mid 2000s. They are also larger, and the gap is widest at the extremes, where the largest firms in the sample are almost all returnee-founded and the smallest are mostly local.

2.3 Research Design

Our baseline analysis is a firm-level cross-sectional comparison of returnee-founded and local-founded firms. For firm i, we estimate

$$Y_i = \beta\, \text{ReturneeFounded}_i + \gamma' X_i + \delta_{p(i)} + \theta_{f(i)} + \varepsilon_i, \tag{1}$$

where Y_i is a measure of firm i's clinical-trial activity—the number of trials, one of the three novel-trial counts, or a measure of trial scale; ReturneeFounded_i indicates a returnee-founded firm; X_i is a vector of firm controls, including employee headcount and the venture-funded

indicator; $\delta_{p(i)}$ and $\theta_{f(i)}$ are headquarters-province and founding-year fixed effects; and ε_i is an error term. The coefficient of interest, β, is the conditional difference between returnee- and local-founded firms. Standard errors are heteroskedasticity-robust.

Because the trial-count outcomes are non-negative and highly right-skewed, we estimate equation (1) two ways: by ordinary least squares on $\ln(Y_i + 1)$, and by Poisson pseudo-maximum-likelihood (PPML) on the untransformed count. We report the OLS estimates below and the results are robust in PPML specification. The province and founding-year fixed effects absorb regional and cohort differences in the clinical-development environment, so β compares returnee- and local-founded firms of the same age operating in the same province.

Disease breadth and within-field novelty. The firm-level premium in equation (1) combines two distinct margins: the *breadth* of disease fields in which a firm is active, and the *novelty rate* of its trials within a field. To separate the extensive and intensive margins, we estimate a two-part model. The first part models whether a firm conducts any trial in a disease (logit), while the second models the share of trials that are novel conditional on entry using a fractional logit.

The first part is estimated over all firm–disease cells:

$$\Pr(\text{Enter}_{id} = 1) = \Lambda\Big(\beta\,\text{ReturneeFounded}_i + \gamma' X_i + \delta_{p(i)} + \theta_{f(i)}\Big), \qquad \Lambda(z) = \frac{e^z}{1+e^z}, \quad (2)$$

where Enter_{id} equals one if firm i conducts at least one trial in disease field d. The coefficient β captures whether returnee-founded firms are active across a wider set of disease fields.

Conditional on entry, the second part models how novel a firm's trials in a field are. Treating the firm's n_{id}^{novel} novel trials as successes out of its n_{id}^{trials} trials in field d, we estimate

a grouped binomial logit for the novelty share π_{id}:

$$n_{id}^{\text{novel}} \sim \text{Binomial}\Big(n_{id}^{\text{trials}},\, \pi_{id}\Big)\,, \qquad \text{logit}(\pi_{id}) = \alpha\, \text{ReturneeFounded}_i + \eta' X_i + \delta_{p(i)} + \theta_{f(i)} + \lambda_d, \tag{3}$$

where λ_d is a disease-field fixed effect, so α is identified from within-field differences in novelty rates between returnee- and local-founded firms. The logit coefficients are on the log-odds scale; alongside each, we report the implied percentage-point change in the entry or novelty probability as ReturneeFounded_i moves from zero to one, evaluated at the sample mean of the outcome.

Instrument variables. The returnee coefficient in each of these specifications identifies a causal effect only under the assumption that, conditional on the controls and fixed effects, returnee status is uncorrelated with unobserved determinants of clinical innovation. This assumption is strong, since returnee entrepreneurs may differ from local founders in ability, motivation, and networks, and may sort into faster-growing market segments. The OLS estimates therefore measure conditional differences; to recover the causal effect of returnee founding, we instrument returnee status with variation driven by the *supply* of returnable talent in a firm's disease field and founding cohort, rather than by the firm's own quality. Because this variation lives in the interaction of field with founding cohort, it is identified within the province, founding-year, and disease fixed effects our specifications absorb.

The instruments exploit two sources of variation. *Push* instruments capture slack in the foreign labor market. When the market abroad cannot absorb the doctorates trained in a given field around a given year, more of those graduates return to China and start or staff firms, so a field and cohort with weaker foreign absorption offers a larger pool of returnable scientists. *Pull* instruments capture Chinese policy demand. The country's talent-recruitment programs, led by the Thousand Talents Program launched in 2008 and its staggered provincial analogues, raised return rates for overseas-Chinese scientists, and their differential bite takes a shift-share form that combines a field's pre-policy diaspora depth

with the policy intensity facing a firm's founding cohort and province. Both instruments shift the local supply of returnable founders while remaining plausibly unrelated to a firm's own determinants of clinical innovation, the exclusion restriction our estimates require.

3 Results

3.1 Main Results

More trials, and more novel trials. Table 2 reports our baseline firm-level regressions (equation (1)). Returnee-founded firms run more trials than otherwise-similar local firms: the coefficient on Returnee_i in column 1 is 0.16, implying about 17.3% more total trials ($e^{0.16} - 1$), significant at the 1% level. The premium widens for novel trials, roughly 21.5% more Dranove-novel ($e^{0.195} - 1$) and 39.0% more GPT-novel trials ($e^{0.329} - 1$, columns 2 and 3), and remains positive but smaller and less precise for the rare first-in-China outcome, at about 6.8% ($e^{0.066} - 1$, column 4, significant at 5%). The ordering is itself informative: the premium is largest for the measures that most directly capture genuine novelty and smallest for total volume, a first indication that returnee firms' portfolios are not merely larger but tilted toward the frontier. The effect is not an artifact of venture funding. The venture-backed indicator enters strongly on its own (columns 1–3), yet the returnee premium survives its inclusion, and is stable in sign and significance across all four constructions. The Poisson estimates are larger still, implying about 31% more total and 61% more GPT-novel trials, and are significant on every margin.

Breadth and within-field novelty. This firm-level premium blends two margins, which the two-part model in Table 3 separates. On the extensive margin, returnee firms are more likely to enter a disease field at all: switching on the returnee indicator raises a firm's entry probability by 0.42 percentage points on a 2.8% baseline, a 15% relative increase. On the intensive margin we examine the *share* of a firm's trials in a field that are novel; here too

returnee firms do more. The novel share rises by 3.44 percentage points under the Dranove measure (a 9.8% relative increase on a 35.1% baseline) and by 8.38 points under the GPT measure (a 13.6% relative increase on a 61.7% baseline), with a smaller, marginally significant effect of about 0.7 points on the first-in-China share.[3] Because the second stage is a rate rather than a count, this within-field effect shows that returnee firms' novelty advantage is not merely a by-product of their running more trials—trial for trial, their work is more likely to reach the frontier. The two margins are comparable in size, with entry the largest in relative terms, so returnee firms' edge comes as much from spreading across more disease areas as from doing more novel science within each.

The aggregate contribution of returnees. These firm-level premia accumulate into a sizable aggregate presence, which Table 4 quantifies. Returnee-led firms already dominate Chinese clinical development: they run 71.4% of the trials in our sample and an even larger share of the novel ones—76.3% of Dranove-novel, 78.1% of GPT-novel, and 74.9% of first-in-China trials. This raw share reflects both how prevalent returnee firms are and their per-firm premium; to isolate the latter, we ask how many of these trials the returnee premium itself accounts for. Applying the baseline coefficients, had returnee-led firms instead behaved like otherwise-similar local firms, China would have run about 2,600 fewer trials (11% of the total), 1,200 fewer Dranove-novel trials (14% of the total), and 3,500 fewer GPT-novel trials (22% of the total). The returnee channel thus accounts for roughly 22% of all GPT-novel trials but only 11% of total volume.[4]

Larger, more resource-intensive trials. Beyond the volume and novelty margins documented above, returnee firms also run larger and more operationally demanding trials.

[3]The reported coefficients are log-odds; a coefficient $\hat{\alpha}$ is converted to a probability change at the baseline mean $\bar{\pi}$ in three steps. Take baseline odds $\text{odds}_0 = \bar{\pi}/(1-\bar{\pi})$; scale by the odds ratio, $\text{odds}_1 = \text{odds}_0\, e^{\hat{\alpha}}$; and convert back, $\pi_1 = \text{odds}_1/(1+\text{odds}_1)$. The implied change is $\pi_1 - \bar{\pi}$. For the GPT column ($\hat{\alpha} = 0.374$, $\bar{\pi} = 0.617$): $\text{odds}_0 = 1.610$, $\text{odds}_1 = 1.610 \times e^{0.374} = 2.341$, $\pi_1 = 0.701$, so $\pi_1 - \bar{\pi} \approx 0.084$, i.e. +8.38 as reported.

[4]These are back-of-envelope magnitudes: they apply the firm-level proportional premia from Table 2 to aggregate trial counts, hold the number and composition of firms fixed.

Table 5 reports firm-level regressions of three measures of trial scale. Relative to otherwise-similar local firms, returnee firms' trials last 24.7% longer, target 11.3% larger enrollments, and use 24.5% more clinical sites. Longer, larger, multi-site trials are typical of more ambitious and later-stage development, consistent with returnee firms undertaking more demanding clinical programs.

3.2 Heterogeneity

Heterogeneity across therapeutic areas. The returnee premium documented above varies substantially across therapeutic areas. Figure 2 re-estimates the firm-level regressions in Table 2 separately for trials in ten therapeutic areas and plots the returnee coefficient for each trial outcome. The effects are mainly concentrated in oncology and gene therapy.

In oncology, returnee firms conduct significantly more total trials and novel trials (using three novelty measurements). Gene therapy shows a similar pattern for total trials and both novelty measures, though not for first-in-China. These two areas are also where China's recent clinical expansion has been most closely linked to global scientific breakthroughs (Barwick et al. 2026), suggesting that returnee founders' exposure to the international research frontier matters most in research-intensive fields. In contrast, the returnee premium is largely absent in most other areas, including infectious disease, cardiovascular, autoimmune, and ophthalmology, where the estimates are small and statistically insignificant. Overall, the returnee advantage does not reflect a broad-based expansion of clinical activity, but is concentrated in a few frontier therapeutic areas.[5] We examine the forces behind this premium in Section 4.

Foreign work experience: industry and academia. The baseline returnee indicator is defined by *education*, capturing whether a founder studied outside China. Panel A of Table 6

[5]We use *frontier* to denote the areas at the leading edge of recent clinical expansion. Among U.S. trials over our sample period, oncology is the single largest therapeutic area which accounts for about 30% of all U.S. trials, and records the largest absolute increase in activity, while gene therapy is the fastest-growing area, with the number of U.S. gene-therapy trials roughly quadrupling between the early 2010s and the early 2020s. These are also the areas where China's recent clinical development has been most closely tied to global scientific breakthroughs (Barwick et al. 2026).

takes a different angle, defining foreign exposure instead through founders' *work experience* and splitting it into two channels: foreign *industry* experience and foreign *academic* experience (= 1 if the founder has postgraduate academic positions in a research institution). Industry experience seems more important than academic experience— industry experience raises total trials by about 21%, Dranove-novel trials by 28%, GPT-novel trials by 47%, and first-in-China trials by 8%, all bigger than effects from the academic experience. In addition, foreign industry experience has positive and significant effects on every outcome, while foreign academic experience works through the novelty channel. It leaves the *number* of trials a firm runs unchanged, yet it significantly raises their *novelty*, with about 34% more GPT-novel trials and a positive, but weaker, effect on Dranove-novel. Industry experience thus contributes to both the scale and the novelty of clinical development, whereas academic experience contributes specifically to novelty, the dimension on which frontier scientific training is most likely to bear.

Concentrated in small firms. The premium also depends sharply on firm size. Panel B interacts the returnee indicator with an indicator for firms at or above the median headcount. In small firms, the returnee premium is large, raising total trials by about 31% and GPT-novel trials by 51%. Its interaction with size is negative and significant on all three count outcomes, so the advantage shrinks steeply, and for large firms the implied returnee effect on total trials is essentially zero. The premium is therefore concentrated in small, founder-centric firms, where a single internationally trained founder can shape the entire research agenda; in larger firms, with deeper organizations and established routines, the founder's own background matters far less. This pattern is consistent with the view that a founding team's human capital is embodied in the organization at its formation and matters most in young, small ventures, where the team is the organization (Choi et al. 2025; Ling et al. 2007).

4 Mechanism

The observed returnee premium is consistent with two distinct mechanisms. Under a static knowledge-stock mechanism, founders bring back a fixed stock of foreign knowledge, training, and experience when they return to China. This knowledge provides an initial advantage but gradually depreciates as the global technological frontier advances. Under a dynamic knowledge-flow mechanism, returnee founders remain connected to the U.S. innovation frontier and continue to transmit new scientific and clinical opportunities into China long after they return. Although both mechanisms can generate higher innovation among returnee-founded firms, they imply different predictions about the persistence and content of the returnee premium.

We distinguish between these mechanisms using three complementary tests. First, we examine how the returnee premium evolves over a firm's life cycle. A static knowledge-stock mechanism predicts that the premium should be strongest immediately after founding and decline as the initial knowledge advantage depreciates, whereas a dynamic knowledge-flow mechanism predicts a more persistent advantage. Second, we examine whether the premium depends on the time elapsed between a founder's return to China and the firm's founding. If foreign knowledge depreciates after the founder leaves the frontier, firms founded soon after return should enjoy the largest advantage, which should diminish as the gap widens. Third, we ask whether returnee-founded firms continue to draw disproportionately from the U.S. innovation frontier. If the premium reflects an ongoing flow of knowledge rather than a one-time transfer, their research should remain more closely aligned with emerging U.S. scientific and clinical opportunities than that of otherwise similar domestic firms.

The premium builds with firm age. To see when the returnee gap appears, we follow firms over their life cycle on a firm-by-year flow panel whose outcome is the number of trials

a firm starts each year. For each outcome we estimate

$$\ln(y_{i,t}+1) = \sum_{\tau=0}^{12} b_\tau \left(\text{ReturneeFounded}_i \times \mathbf{1}[a_{i,t}=\tau]\right) + \alpha_{a_{i,t}} + \gamma_t + \delta_{p(i)} + X'_{i,t}\beta + \varepsilon_{i,t}, \quad (4)$$

where $a_{i,t}$ is firm i's age in year t, α_a is a full set of firm-age fixed effects, γ_t and δ_p are calendar-year and HQ-province fixed effects, and $X_{i,t}$ includes employees and a VC-backing indicator, with standard errors clustered at the firm level. The age fixed effects absorb the common life-cycle profile of trial activity, so each b_τ measures the gap in annual trial starts between returnee and non-returnee firms of the same age, in the same year and province and with the same size and funding.

The coefficients are shown in Figure 3. The returnee gap widens steadily with firm age. Total trials, Dranove-novel, and GPT-novel all begin slightly negative in the first year or two, turn significantly positive around year three, and rise to roughly 5 to 8 percent more per year by ages six through twelve, with GPT-novel the largest throughout. First-in-China is the exception: it hovers near zero and is mostly insignificant across the age profile. The premium is thus something firms grow into rather than start with.

The premium does not fade with time since return. A natural concern is that the premium reflects fresh knowledge a founder brings straight from abroad and that depreciates once they are back in China. If so, the advantage should be largest for founders who start a firm soon after returning and should shrink the longer they have been back. We test this on the firm cross section by interacting the returnee indicator with bins of the gap between a founder's return to China and the firm's founding,

$$\ln(y_i+1) = \sum_k b_k \left(\text{ReturneeFounded}_i \times \mathbf{1}[\text{gap}_i \in k]\right) + \lambda_{f(i)} + \delta_{p(i)} + X'_i\beta + \varepsilon_i, \quad (5)$$

where gap_i is measured in years and grouped into the seven cohorts shown in Figure 4, running from zero to eleven or more, λ_f and δ_p are founding-year and HQ-province fixed

effects, and X_i again holds employees and VC-backing. Non-returnee firms enter every bin at zero and form the common baseline, so b_k is the trial-output gap between returnee firms in cohort k and otherwise comparable non-returnee firms. A depreciating-knowledge account predicts that b_k declines as the gap widens.

Figure 4 shows no such decline. The premium is positive across the full range of return-to-founding gaps and, for GPT-novel trials, significant in nearly every cohort, including the longest gaps of nine to ten years and eleven or more. Total trials follow the same broad pattern, significant at the shorter gaps and again at nine to ten years, though the intermediate cohorts are imprecise. The individual bins rest on few firms and are correspondingly noisy, but the profile as a whole is flat rather than downward-sloping. The returnee advantage does not appear to depend on how recently the founder returned, which is suggestive of a durable founder contribution rather than a stock of foreign knowledge that quickly depreciates.

Together, the two figures point toward the founder's durable human capital as the source of the premium. It is not a head start handed to the firm at birth, nor does it fade with the time since the founder returned home; instead, it emerges as the firm matures and draws on capabilities the founder retains. What those capabilities consist of, and whether they surface in the science returnee firms build on, is the question we turn to next.

Knowledge flow from the US frontier. The event studies place the premium in the founder's durable human capital, but they do not say what that capital consists of. A natural candidate is knowledge of the US research frontier. If returnee founders carry frontier science home, returnee firms should lean toward diseases where the US frontier is most active, and they should do so in their novel work rather than in routine trials. We test this on the firm-by-disease panel by interacting the returnee indicator with the size of the US frontier in each disease. For firm i and disease d we estimate

$$\ln(1 + y_{id}) = \beta \left(\text{Returnee}_i \times \ln(1 + F_d) \right) + \alpha \ln(1 + F_d) + \phi_i + \lambda_d + \varepsilon_{id}, \tag{6}$$

where y_{id} counts firm i's trials of a given type in disease d, F_d is the number of total US trials in disease d over a window after the founder returns to China, and ϕ_i and λ_d are firm and disease fixed effects. Because returnee status does not vary within a firm, it is absorbed by ϕ_i, and the coefficient of interest is the interaction β, which asks whether a returnee firm's activity in a disease responds more strongly to the US frontier there than a domestic firm's does.

Table 7 provides evidence for the dynamic knowledge-flow mechanism. Local firms do not systematically increase innovation in disease areas where the U.S. frontier advances. In contrast, returnee-founded firms significantly increase their novel clinical development in those same diseases, while showing no comparable increase in overall trial activity. This pattern suggests that returnees do not simply expand research effort; they selectively bring frontier scientific opportunities into China. The effects are economically meaningful. A doubling of U.S. frontier activity in a disease increases GPT-novel trials by returnee firms by 0.21–0.25 percentage points relative to local firms, equivalent to roughly 10–12% of the baseline mean. The corresponding effect is 0.12–0.15 percentage points (9–12% of the mean) for Dranove novelty, while the effect on first-in-China trials is positive but smaller, at 0.01–0.02 percentage points (10–15% of the mean). The result is stable across horizons. Measuring the frontier over the first five years after return and over years five to ten gives coefficients of similar size, slightly larger at the longer horizon, indicating the frontier-tracking is not a transient response to conditions at founding.

5 Discussion and Conclusion

This paper studies how returnee entrepreneurs shape domestic innovation in a science-based industry. Using newly assembled founder-firm-trial data from China's biopharma sector, we show that returnee-founded firms account for a disproportionately large share of clinical development activity and conduct more, more novel, and broader clinical trials than local-

founded firms. These differences are especially pronounced among smaller firms and among firms whose founders have prior industry experience, suggesting that returnee founders matter not only through firm scale, but also through the knowledge and capabilities they bring into entrepreneurial organizations.

The evidence also speaks to how return migration transmits knowledge across borders. Returnee entrepreneurs do not appear to simply bring back a fixed stock of foreign knowledge that depreciates over time. Instead, their advantage persists after return and tends to follow the recent U.S. innovation frontier. This pattern suggests that returnee entrepreneurs can serve as a continuing bridge between global frontier knowledge and domestic innovation, translating exposure to foreign science, industry experience, and technological opportunities into clinical development in China.

Our ongoing work extends this analysis in two directions. First, we are developing push- and pull-side instruments to better isolate the causal effect of return migration and to distinguish whether returnee entrepreneurship is shaped more by conditions in the United States that affect the incentive to stay or by Chinese policies that affect the incentive to return. Second, we use these estimates to conduct counterfactual welfare analysis, comparing China's observed biopharma innovation ecosystem with scenarios in which no U.S.-trained talent returns or all U.S.-trained talent returns. These exercises are especially relevant for current policy debates over global talent competition: policies that attract, retain, or restrict high-skilled talent do not only affect individual careers, but can also reshape where frontier knowledge is commercialized and which countries benefit from scientific training.

References

Agrawal, Ajay, Devesh Kapur, John McHale, and Alexander Oettl. 2011. "Brain drain or brain bank? The impact of skilled emigration on poor-country innovation." *Journal of Urban Economics* 69 (1): 43–55.

Amanzadeh, Naser, Amir Kermani, and Timothy McQuade. 2024. *Return migration and human capital flows.* Technical report. National Bureau of Economic Research.

Azoulay, Pierre, Benjamin F Jones, J Daniel Kim, and Javier Miranda. 2022. "Immigration and entrepreneurship in the United States." *American Economic Review: Insights* 4 (1): 71–88.

Bahar, Dany, Prithwiraj Choudhury, Ernest Miguelez, and Sara Signorelli. 2024. "Global mobile inventors." *Journal of Development Economics* 171:103357.

Barwick, Panle Jia, Hongyuan Xia, and Tianli Xia. 2026. *From Free Rider to Innovator: The Rise of China's Drug Development.* Technical report. National Bureau of Economic Research.

Batista, Catia, Daniel Han, Johannes Haushofer, Gaurav Khanna, David McKenzie, Ahmed Mushfiq Mobarak, Caroline Theoharides, and Dean Yang. 2025. "Brain drain or brain gain? Effects of high-skilled international emigration on origin countries." *Science* 388 (6749): eadr8861.

Bernstein, Shai, Rebecca Diamond, Abhisit Jiranaphawiboon, Timothy McQuade, and Beatriz Pousada. 2022. *The contribution of high-skilled immigrants to innovation in the United States.* Technical report. National Bureau of Economic Research.

Borjas, George J, and Kirk B Doran. 2012. "The collapse of the Soviet Union and the productivity of American mathematicians." *The Quarterly Journal of Economics* 127 (3): 1143–1203.

Choi, Joonkyu, Nathan Goldschlag, John Haltiwanger, and J Daniel Kim. 2025. "Early joiners and startup performance." *Review of Economics and Statistics* 107 (6): 1485–1500.

Docquier, Frédéric, and Hillel Rapoport. 2012. "Globalization, brain drain, and development." *Journal of economic literature* 50 (3): 681–730.

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

Filatotchev, Igor, Xiaohui Liu, Jiangyong Lu, and Mike Wright. 2011. "Knowledge spillovers through human mobility across national borders: Evidence from Zhongguancun Science Park in China." *Research Policy* 40 (3): 453–462.

Gibson, John, and David McKenzie. 2014. "Scientific mobility and knowledge networks in high emigration countries: Evidence from the Pacific." *Research policy* 43 (9): 1486–1495.

Glennon, Britta. 2024. "Skilled immigrants, firms, and the global geography of innovation." *Journal of Economic Perspectives* 38 (1): 3–26.

Gompers, Paul, Anna Kovner, Josh Lerner, and David Scharfstein. 2010. "Performance persistence in entrepreneurship." *Journal of financial economics* 96 (1): 18–32.

Hunt, Jennifer, and Marjolaine Gauthier-Loiselle. 2010. "How much does immigration boost innovation?" *American Economic Journal: Macroeconomics* 2 (2): 31–56.

Kerr, William R, and William F Lincoln. 2010. "The supply side of innovation: H-1B visa reforms and US ethnic invention." *Journal of Labor Economics* 28 (3): 473–508.

Ling, Yan, Hao Zhao, and Robert A Baron. 2007. "Influence of founder—CEOs' personal values on firm performance: Moderating effects of firm age and size." *Journal of Management* 33 (5): 673–696.

Mallapaty, Smriti. 2025. "How China's bold talent recruitment has shaped science." *Nature* 644 (8075): 18–19.

Marx, Matt, Qian Wang, and Emmanuel Yimfor. 2026. "Minimum viable signal: Venture funding, social movements, and race." *Management Science* 72 (4): 3332–3350.

Moser, Petra, Alessandra Voena, and Fabian Waldinger. 2014. "German Jewish émigrés and US invention." *American Economic Review* 104 (10): 3222–3255.

Prato, Marta. 2025. "The global race for talent: Brain drain, knowledge transfer, and growth." *The Quarterly Journal of Economics* 140 (1): 165–238.

Retterath, Andre, and Reiner Braun. 2020. "Benchmarking venture capital databases." *Available at SSRN 3706108.*

Shi, Dongbo, Weichen Liu, and Yanbo Wang. 2023. "Has China's Young Thousand Talents program been successful in recruiting and nurturing top-caliber scientists?" *Science* 379 (6627): 62–65.

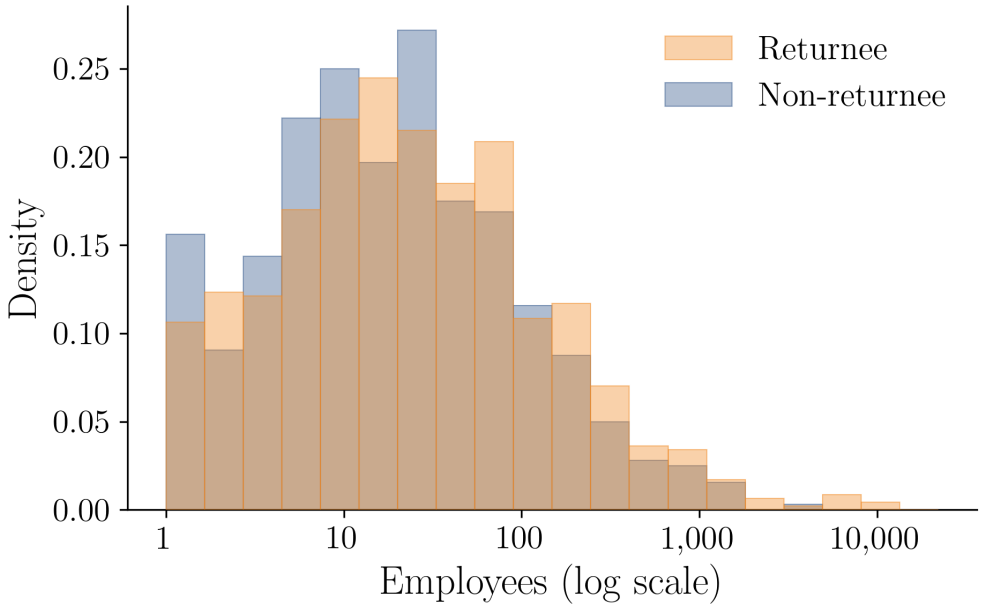
Shvadron, Dror, Hansen Zhang, Lee Fleming, and Daniel P Gross. 2025. "A Quarter of US-Trained Scientists Eventually Leave. Is the US Giving Away Its Edge?" *arXiv preprint arXiv:2512.11146.*

Tambe, Prasanna B. 2026. "Reskilling the workforce for AI: Domain expertise and algorithmic literacy." *Management Science* 72 (1): 515–537.

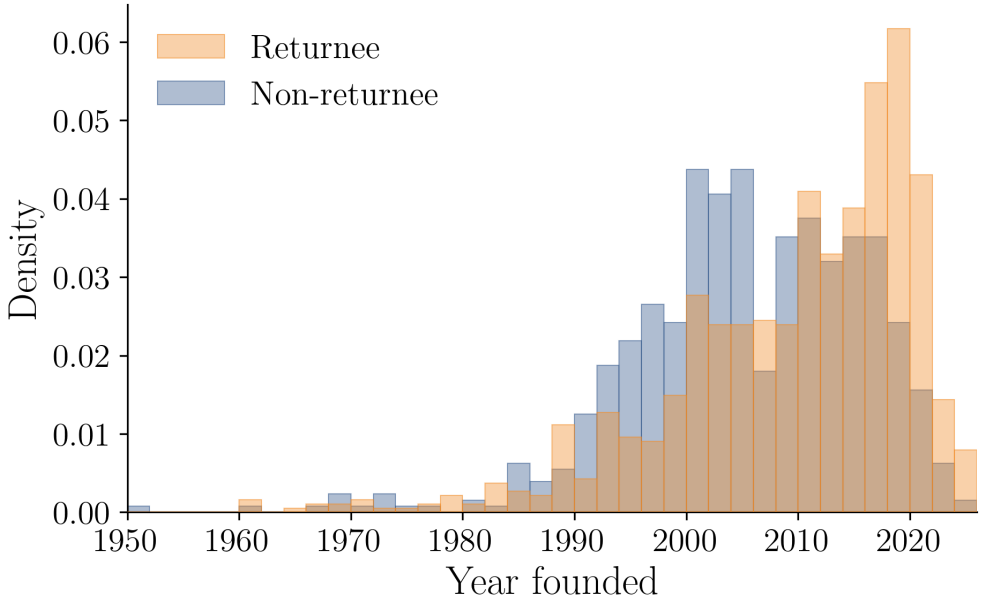
Zucker, Lynne G, Michael R Darby, and Marilynn B Brewer. 1994. *Intellectual capital and the birth of US biotechnology enterprises.*

Figures

Figure 1: Distribution of firm characteristics by returnee status



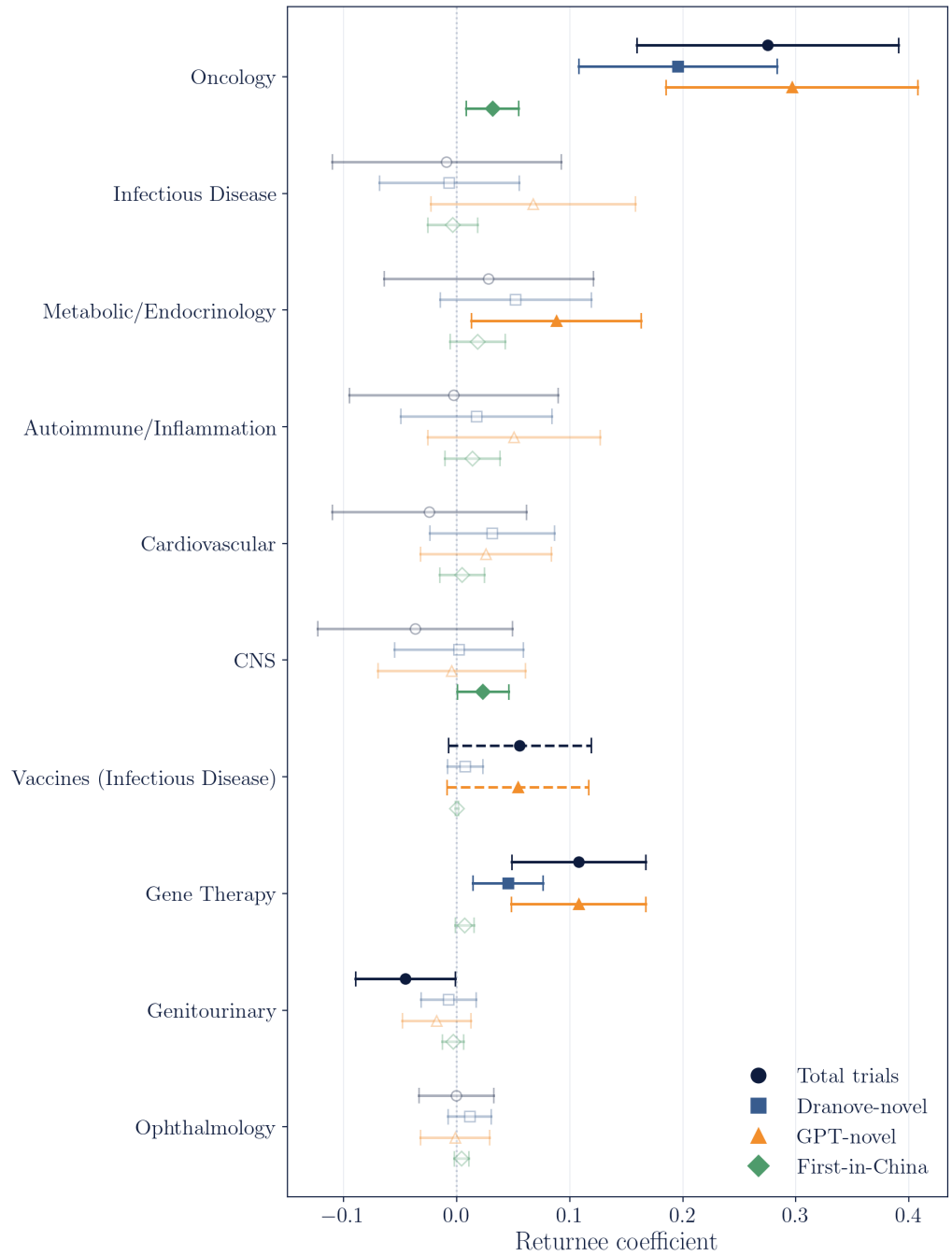
(a) Employee headcount (log scale)



(b) Year founded

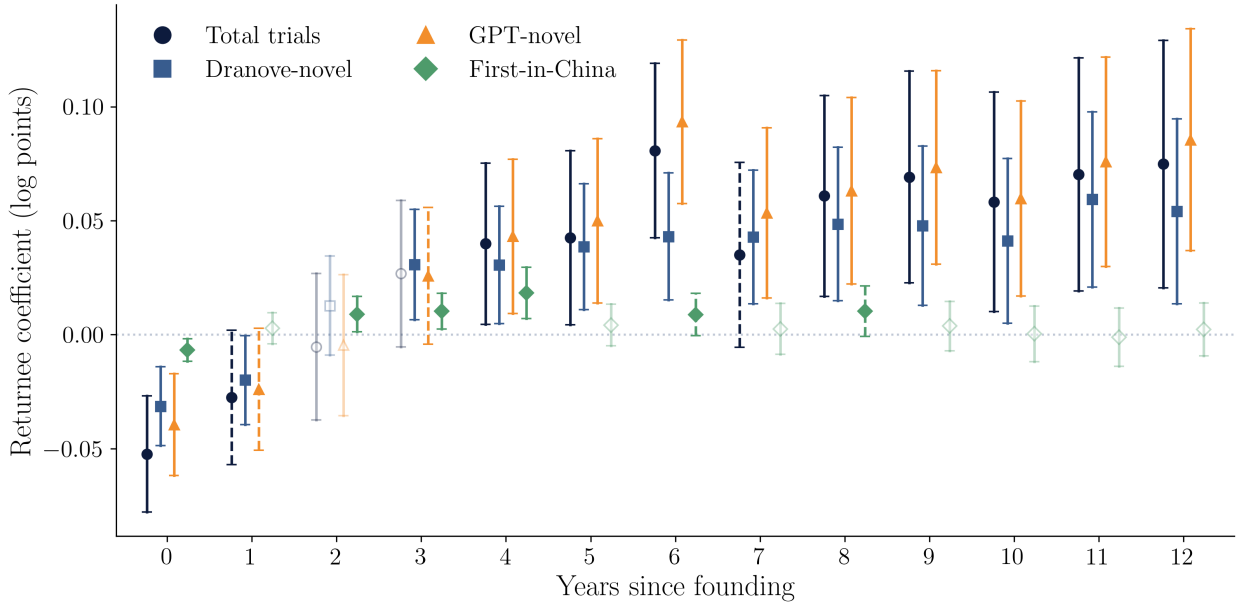
Notes: These figures compare the firm-level distributions of two key control variables across returnee and non-returnee firms in the founder-education sample. *Returnee* firms (orange) have at least one founder educated outside China; *non-returnee* firms (blue) have none. Panel (a) plots the density of employee headcount on a log scale, and panel (b) plots the density of year founded.

Figure 2: The returnee premium across therapeutic areas



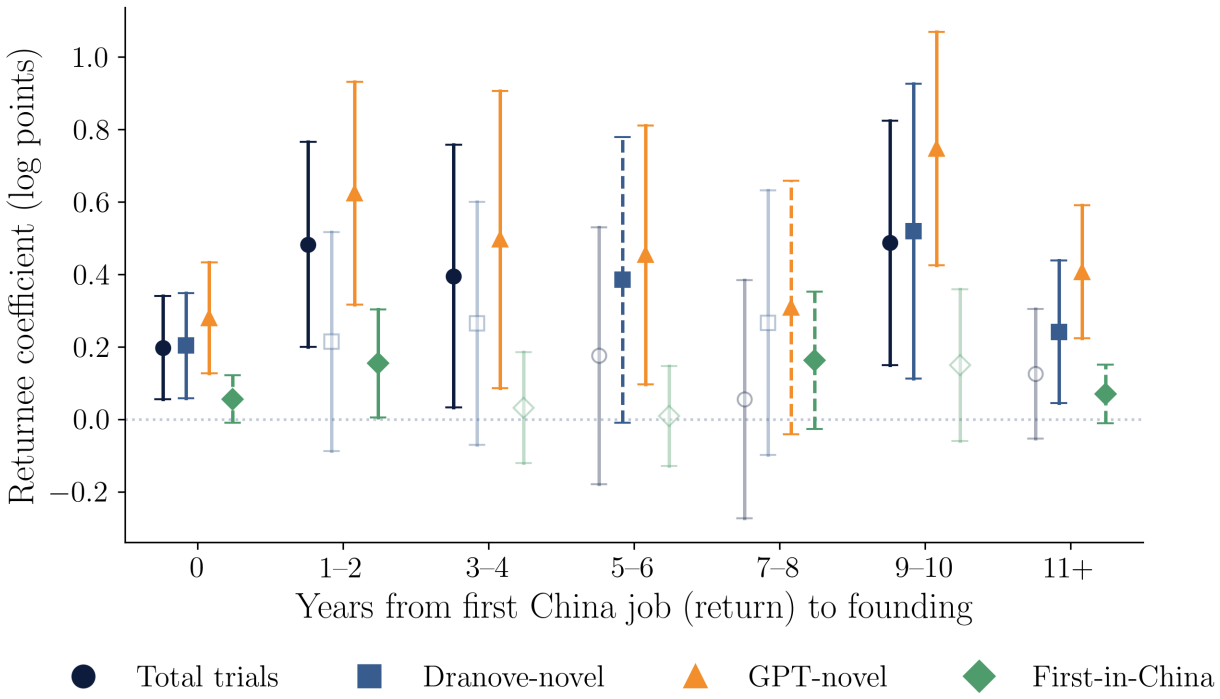
Notes: Each marker is the returnee coefficient, with its 95% confidence interval, from a firm-level $\ln(y+1)$ regression of the indicated outcome estimated separately within each therapeutic area; controls and fixed effects match Table 2. Outcomes are distinguished by both color and marker: total trials (navy circles), Dranove-novel (blue squares), GPT-novel (orange triangles), and first-in-China (green diamonds). Solid markers with full-opacity intervals are significant at the 5% level; dashed intervals are significant at the 10% level; faded, hollow markers are insignificant.

Figure 3: Returnee coefficient on trial output, by years since founding.



Notes: The figure shows how the returnee premium evolves with firm age, from founding through year twelve. Each plotted point is the returnee coefficient b_τ at firm age τ, taken from a single regression per outcome that interacts the returnee indicator with a full set of firm-age dummies and absorbs firm-age fixed effects. These age fixed effects absorb the common life-cycle profile, so b_τ measures the returnee versus non-returnee gap at age τ. The estimation panel is a group-by-calendar-year flow in which the outcome each year is the number of trials a firm starts that year. Each regression also includes calendar-year and HQ-province fixed effects and controls for employees and VC-backing, with standard errors clustered at the firm level. Whiskers are 95% confidence intervals. Solid lines indicate significance at the 5% level, dashed lines significance at the 10% level, and faded hollow markers estimates that are not statistically significant. The four outcomes are total trials, Dranove-novel, GPT-novel, and first-in-China.

Figure 4: Returnee coefficient on trial output, by years from return to founding



Notes. The figure shows how the returnee premium varies with the time between a founder's return to China and the firm's founding. Each plotted point is the returnee coefficient b_k for firms whose gap between the founder's return to China and the firm's founding falls in bin k, measured relative to non-returnee firms. A founder's return is dated by the first China-based job in the career record, and returnee founders whose return year cannot be dated are dropped. The outcomes are firm-level trial counts entered as $\ln(y+1)$, and each regression includes HQ-province and founding-year fixed effects together with controls for employees and VC-backing, with robust standard errors. Solid lines indicate significance at the 5% level, dashed lines significance at the 10% level, and faded hollow markers estimates that are not statistically significant. The four outcomes are total trials, Dranove-novel, GPT-novel, and first-in-China.

Tables

Table 1: Firm-level summary statistics

	Mean	Min	Median	Max	Std. Dev.	N
Treatments						
ReturneeFounded (non-China edu)	0.585	0.00	1.00	1.00	0.493	1,608
Foreign-PhD founder	0.252	0.00	0.00	1.00	0.435	1,608
Foreign industry experience	0.533	0.00	1.00	1.00	0.499	1,608
Foreign academia (non-postdoc)	0.105	0.00	0.00	1.00	0.307	1,608
Controls						
Employees	112.454	1.00	19.50	12,490.00	520.174	1,608
Year founded	2007	1951	2010	2025	10.917	1,608
VC-backed	0.353	0.00	0.00	1.00	0.478	1,608
Outcomes						
Trials	19.287	1.00	6.00	1,565.00	64.417	1,608
Novel trials (Dranove)	7.047	0.00	1.00	762.00	29.826	1,608
Novel trials (GPT)	12.754	0.00	3.00	1,450.00	51.285	1,608
First-in-China trials	0.559	0.00	0.00	61.00	2.306	1,608
Avg. trial duration (days)	468.607	8.00	395.00	3,486.00	401.319	1,454
Avg. target accrual (patients)	244.819	2.00	84.10	8,836.80	696.039	1,540
Avg. identified sites	5.727	1.00	3.00	102.00	7.964	1,543

Notes: This table reports firm-level summary statistics, with one observation per sponsor firm. The sample comprises firms with founder-education information (so that the returnee measure is defined); each statistic is computed over the firms in that sample for which the variable is non-missing, as reflected in the N column. *Returnee* equals one if the firm has at least one founder educated outside China, where mainland China, Hong Kong, and Macau count as China and Taiwan is treated as non-China. *Foreign-PhD founder* equals one if the firm has at least one founder with a foreign (non-China/HK/Macau) doctorate, a strict subset of the returnee dummy. *Foreign industry experience* and *Foreign academia (non-postdoc)* are founder work-experience treatments . *VC-backed* equals one if the firm has a matched PitchBook CompanyID. The novelty constructions and cost outcomes are as defined in the regression tables that follow.

Table 2: Returnee founders and firm-level clinical-trial activity

	Total trials	Dranove novel	GPT novel	First-in-China
	(1)	(2)	(3)	(4)
ReturneeFounded	0.160***	0.195***	0.329***	0.0660**
	(0.059)	(0.060)	(0.063)	(0.026)
Employees (thousands)	0.273*	0.311**	0.360*	0.138
	(0.162)	(0.154)	(0.196)	(0.088)
VC-backed	0.417***	0.400***	0.627***	0.0594**
	(0.061)	(0.062)	(0.064)	(0.028)
Constant	1.814***	0.811***	1.128***	0.145***
	(0.047)	(0.046)	(0.050)	(0.020)
HQ-province FE	Yes	Yes	Yes	Yes
Founding-year FE	Yes	Yes	Yes	Yes
Observations	1,599	1,599	1,599	1,599
R^2	0.194	0.173	0.236	0.163

Notes: Firm-level OLS regressions of clinical-trial counts on the ReturneeFounded indicator; each outcome is log-transformed as $\ln(y+1)$. *ReturneeFounded* equals one if at least one founder was educated outside China. *Employees (thousands)* is employee headcount measured in thousands. All columns include firm controls (employee headcount and a venture-funded indicator) and absorb headquarters-province and founding-year fixed effects. The estimation sample contains 1,599 of the 1,608 firms; nine singleton observations are dropped by the province and founding-year fixed effects. Heteroskedasticity-robust standard errors in parentheses. $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

Table 3: Two-part model of disease-field entry and trial novelty

	Stage 1	Stage 2: Novelty share		
	Disease entry	Dranove	GPT	First-in-China
	(1)	(2)	(3)	(4)
ReturneeFounded	0.144*	0.148**	0.374***	0.227*
	(0.075)	(0.070)	(0.113)	(0.134)
Employees (thousands)	0.181***	0.104***	0.226	0.0906**
	(0.068)	(0.023)	(0.145)	(0.045)
VC-backed	0.282***	0.0208	0.600***	0.210
	(0.074)	(0.070)	(0.145)	(0.148)
Constant	-4.133***	-0.810***	1.016**	-3.418***
	(0.162)	(0.167)	(0.480)	(0.351)
HQ-province FE	Yes	Yes	Yes	Yes
Founding-year FE	Yes	Yes	Yes	Yes
Disease FE	No	Yes	Yes	Yes
Mean of dep. var.	2.8%	35.1%	61.7%	3.0%
Implied Δ Pr	+0.42%	+3.44%	+8.38%	+0.73%
Observations	390,744	10,980	10,980	10,980

Notes: Two-part model on the firm-by-disease-field panel (1,608 firms $\times$ 243 disease fields). Column 1 is a binary logit for disease-field entry, estimated over all firm–disease cells. Columns 2–4 are grouped binomial logits for the share of a firm's trials in a field that are novel (successes out of total trials), estimated on entered cells, under the three trial-level novelty measures. *ReturneeFounded* equals one if at least one founder was educated outside China. All columns absorb headquarters-province and founding-year fixed effects; Stage 2 additionally absorbs disease-field fixed effects. *Implied* Δ Pr is the percentage-point change in the entry or novelty probability as the returnee indicator moves from zero to one, evaluated at the sample mean of the outcome. Standard errors clustered at the firm level in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4: Returnee firms' share of China's clinical trials and trial novelty

	Total trials (1)	Novel Dranove (2)	Novel GPT (3)	First-in-China (4)
Observed counts				
Total	24,259	8,789	16,059	734
Returnee-led	17,314	6,710	12,550	550
Pure-China-led	6,945	2,079	3,509	184
Returnee share	71.4%	76.3%	78.1%	74.9%
Implied returnee contribution				
Implied premium ($e^{\beta} - 1$)	+17%	+22%	+39%	+7%
Attributable to returnees	2,560	1,189	3,518	35
Share of total	10.6%	13.5%	21.9%	4.8%

Notes: Trial-level counts over the founder-covered China-firm trial sample. A trial is defined as *returnee-led* if at least one of its sponsors has a founder with a non-China degree. For each firm, founders are defined as the earliest five joiners. Columns 2–4 restrict the sample to trials classified as novel under the Dranove measure, the GPT non-generic measure, and the first-in-China measure, respectively. The lower panel reports the returnee contribution implied by the firm-level baseline estimates: the proportional premium from Table 2 ($e^{\beta} - 1$); the count attributable to the returnee channel, computed as the returnee-led count $\times (1 - e^{-\beta})$—the trials that would be lost if returnee-led firms had instead behaved like otherwise-similar local firms—and that count as a share of the total. The attributable figures apply the firm-level proportional premium to trial counts as a back-of-envelope. The Dranove column is computed on the 19,531 trials with a constructible novelty score.

Table 5: Trial scale: duration, target enrollment, and clinical sites

	Duration (days)	Target accrual	Identified sites
	(1)	(2)	(3)
ReturneeFounded	0.221***	0.107*	0.219***
	(0.058)	(0.058)	(0.057)
Employees (Thousands)	0.124**	0.0125	0.176***
	(0.061)	(0.047)	(0.059)
VC-backed	0.336***	0.0697	0.272***
	(0.058)	(0.065)	(0.061)
Constant	5.449***	4.501***	0.905***
	(0.049)	(0.049)	(0.046)
HQ-province FE	Yes	Yes	Yes
Founding-year FE	Yes	Yes	Yes
Observations	1,446	1,531	1,534
R^2	0.221	0.122	0.134

Notes: Firm-level regressions of trial scale on the returnee indicator. Each outcome is a firm's mean across its trials—average trial duration in days, average target enrollment, and average number of identified clinical sites—log-transformed ($\ln(y)$). *ReturneeFounded* equals one if at least one founder was educated outside China. All columns include firm controls (employee headcount and a venture-funded indicator) and absorb headquarters-province and founding-year fixed effects. Heteroskedasticity-robust standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 6: Heterogeneity: foreign-experience channels and firm size

	Total trials	Novel Dranove	Novel GPT	First-in-China
	(1)	(2)	(3)	(4)
Panel A. Foreign-experience channels				
Foreign industry exp	0.193***	0.246***	0.385***	0.0785***
	(0.059)	(0.060)	(0.062)	(0.027)
Foreign academia exp	0.124	0.176*	0.291***	-0.0356
	(0.094)	(0.099)	(0.097)	(0.046)
Panel B. ReturneeFounded × firm size				
ReturneeFounded	0.268***	0.267***	0.412***	0.0735***
	(0.069)	(0.070)	(0.072)	(0.027)
Large firm (≥ median)	0.684***	0.595***	0.723***	0.124***
	(0.092)	(0.091)	(0.096)	(0.040)
Returnee × Large firm	-0.360***	-0.273**	-0.324***	-0.0436
	(0.113)	(0.113)	(0.117)	(0.050)
Observations	1,599	1,599	1,599	1,599
Firm controls	Yes	Yes	Yes	Yes
HQ-province FE	Yes	Yes	Yes	Yes
Founding-year FE	Yes	Yes	Yes	Yes

Notes. Firm-level regressions of each outcome on the returnee indicator, including the firm controls (employee headcount and a venture-funded indicator) and headquarters-province and founding-year fixed effects, as in Table 2. *Panel A* decomposes founders' foreign exposure into two work-experience channels entered jointly: *Foreign industry exp* and *Foreign academia exp* indicate a founder or tracked employee with non-China industry and academic (non-postdoctoral) experience, respectively. *Panel B* interacts the ReturneeFounded indicator with a *Large firm* indicator (employee headcount at or above the sample median); the ReturneeFounded main effect is the effect for small firms and *ReturneeFounded × Large firm* is the additional effect for large firms. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Knowledge flow from the US frontier

	All trials		Dranove-novel		GPT-novel		First-in-China	
	1–5	5–10	1–5	5–10	1–5	5–10	1–5	5–10
ReturneeFounded × ln(1+Frontier)	0.0009	0.0011	0.0017*	0.0022**	0.0030**	0.0036***	0.0002**	0.0003**
	(0.0017)	(0.0019)	(0.0009)	(0.0010)	(0.0012)	(0.0014)	(0.0001)	(0.0001)
ln(1+Frontier)	-0.0048**	0.0024	-0.0038***	-0.0005	-0.0050***	-0.0012	-0.0006***	-0.0000
	(0.0022)	(0.0019)	(0.0013)	(0.0009)	(0.0015)	(0.0012)	(0.0002)	(0.0002)
Constant	0.0400***	0.0240***	0.0190***	0.0112***	0.0272***	0.0182***	0.0024***	0.0010***
	(0.0043)	(0.0038)	(0.0024)	(0.0019)	(0.0029)	(0.0026)	(0.0005)	(0.0003)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Disease FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dep. var.	0.0308	0.0308	0.0129	0.0129	0.0202	0.0202	0.0014	0.0014
Observations	390,744	390,744	390,744	390,744	390,744	390,744	390,744	390,744
R^2	0.141	0.141	0.105	0.105	0.117	0.117	0.024	0.024

Notes: This table tests whether a firm's trial activity in a disease tracks the US frontier in that disease more closely when the firm is returnee-founded. The unit of observation is a firm-by-disease cell over the full firm-by-disease grid (390,744 cells), and the dependent variable is $\ln(1+y_{id})$, where y_{id} counts the firm's trials of the indicated type in disease d; the four column groups use all trials, Dranove-novel, GPT-novel, and first-in-China counts. The key regressor is the interaction $\text{ReturneeFounded}_i \times \ln(1+\text{Frontier}_d)$, where Frontier_d is the number of total US trials in disease d, measured one to five years (first column of each group) or five to ten years (second column) after the founder's return to China; the founding year is used when the return year cannot be dated. A positive interaction indicates that returnee firms' disease-level activity tracks the US frontier more closely than local firms'. All columns include firm and disease fixed effects, and standard errors are two-way clustered by firm and disease. $^{*}\,p<0.10$, $^{**}\,p<0.05$, $^{***}\,p<0.01$.