Learning About Roads Not Taken

Daniel Schliesmann
McCombs School of Business
dschlies@mccombs.utexas.edu

Matteo Tranchero
The Wharton School
mtranc@wharton.upenn.edu

ABSTRACT

Learning from failure is central to innovation. When experimentation is costly, firms might generalize from failed projects to evaluate related but untested opportunities. We argue that doing so can misdirect search precisely when feedback from one project is informative about nearby alternatives. Because firms tend to test promising leads, failures occur disproportionately in regions of the innovation landscape where nearby alternatives are also likely to be valuable. Generalizing from failures, therefore, mostly screens out valuable alternatives. We test this mechanism in pharmaceutical R&D using data on patenting, clinical trial failures, and biological relatedness among drug targets. Following a failure, firms reduce investment in both the focal target and related targets, especially high-potential ones, which worsens the allocation of innovative effort. This distortion is strongest in smooth landscapes, where nearby alternatives are more informative about one another, making generalization both more effective and more costly. Taken together, our findings identify how a seemingly efficient learning strategy can redirect search away from promising alternatives.

Keywords: organizational learning, search, generalization, pharmaceutical innovation, difference-in-differences regression

1. INTRODUCTION

In the pursuit of treatments for Alzheimer's disease, multiple firms invested in drugs targeting the β-secretase 1 (BACE1) protein. In February 2017, however, Merck reported the early termination of a large-scale clinical trial on this target due to disappointing results. This failure undermined confidence in BACE1 and prompted other firms pursuing the same approach to reassess their clinical investments (Krieger, 2021). Yet industry experts interpreted the failure as potentially informative beyond the focal target (Carroll, 2017; Garde, 2016). This broader inference is plausible because BACE1 is embedded in a neighborhood of functionally related proteins. The clinical failure could therefore shape beliefs not only about the tested target but also about related targets that had not been directly evaluated. BACE2 provides a natural example, as its name suggests, given its close biological relationship to BACE1.

Consistent with this example, an emerging body of work has begun to investigate the process by which organizations generalize — defined as the ability to learn from "prior experience that is distinct from, but related to, one's current circumstance" (Choi and Levinthal, 2023: 1073). Underpinning the efficacy of this mechanism is a simple logic: when related alternatives tend to generate similar outcomes, feedback about one alternative provides information about others (Martignoni et al., 2016). Generalization is thus especially valuable when experimentation is costly, and organizations face more opportunities than they can evaluate directly. In such settings, firms can use limited feedback to discipline search across a broader opportunity space (Schliesmann, 2025). Applied to failure, this logic implies that negative feedback about one alternative can help avoid investments in alternatives that are likely to share the same limitations, thereby improving the allocation of innovation effort (Choi and Levinthal, 2023).

We argue, however, that this learning strategy can undermine innovation search by steering firms away from valuable opportunities. The reason is that neither the distribution of observed failures nor the structure of the opportunity space in which firms operate is random. Firms are more

likely to test promising alternatives. At the same time, the opportunity space is locally correlated, meaning that the value of one alternative is often informative about the quality of its neighbors (Levinthal, 1997; Fleming, 2001). These features imply that when firms select promising alternatives for experimentation, they also implicitly select neighborhoods that are likely to contain other promising alternatives. Therefore, failed experiments disproportionately occur in regions of the landscape that, in expectation, remain more attractive than the average neighborhood. When firms generalize from failures, they extrapolate negative feedback to nearby alternatives that are themselves likely to be valuable. The counterintuitive implication is that learning from failure may be distortionary precisely when local correlation makes generalization appear most justified.

Testing this argument requires observing both credible negative feedback and the set of related opportunities to which that feedback might apply. These conditions are rarely met — most innovation settings involve selectively disclosed experiments, difficult-to-interpret outcomes, and unclear boundaries of the relevant opportunity space. Pharmaceutical R&D offers a rare exception (Kao, 2025b). Clinical trials are large, costly, and tightly regulated, making failures salient events that are publicly reported and widely observed by both sponsoring firms and their rivals (Krieger, 2021). At the same time, because drug development begins with selecting drug targets from millions of possible protein-disease combinations, the setting provides a structured landscape in which focal and related opportunities can be identified (Tranchero, 2026).

We exploit these features by combining multiple data sources. First, records from ClinicalTrials.gov identify clinical trial failures at the level of specific protein-disease pairs. Second, USPTO patent applications allow us to trace how firms reallocate innovative effort across the same opportunity space after these failures. Third, we use Open Targets scores to proxy for the therapeutic potential of each protein-disease pair. These scores summarize the extent to which genetic evidence supports a target's relevance for a given disease, providing a measure of how promising each opportunity is, irrespective of the focal trial outcome (Buniello et al., 2025). Finally, we use protein-protein functional interactions from STRING to construct granular measures of biological

distance between drug targets (Szklarczyk et al., 2025). Together, these data allow us to identify which alternatives are proximate enough to plausibly receive negative feedback from a failed target, and to examine how such generalization shapes the allocation of innovative effort.

Our results document "spatial spillovers" consistent with learning via generalization. Following a clinical trial failure, firms reduce patenting on both the focal protein-disease pair and on biologically proximate pairs, with the magnitude of this effect attenuating as distance from the failed target increases. The effect is economically meaningful, reducing investment in neighboring opportunities by roughly 17.4% over the sample mean. Importantly, this response is not concentrated on less promising alternatives. Conditional on the same biological distance from the failed target, firms reduce investment by a greater amount for high-potential pairs than for low-potential ones. This asymmetric response has aggregate consequences. Generalization does screen out some low-potential opportunities, but it also screens out many high-potential ones. As a result, we document that the overall allocation of R&D investment worsens due to generalization.

Additional analyses clarify the mechanisms behind this finding. Firms concentrate clinical experimentation on targets with higher ex ante therapeutic potential, as measured by their Open Targets scores. We also find that high-potential targets are spatially clustered. Taken together, in selecting promising targets for clinical experimentation, firms also select neighborhoods with a much higher frequency of valuable alternatives. This implies that firms should be more conservative when generalizing from observed failures, because the negative signal from a failed trial is applied to nearby opportunities that remain positively selected. Consistent with this logic, the misallocation effect is strongest in smoother landscapes, where local correlation makes the surrounding neighborhood more likely to contain valuable targets, and attenuates as landscapes become more rugged. Together, these findings show how economizing on scarce experience can systematically redirect innovative effort away from valuable opportunities.

Finally, recognizing that the same observed failure may not carry the same meaning across organizations, we supplement these tests with a post hoc analysis of which firms rely more on gener-

alization. Specifically, we use firms' prior scientific publications to measure target-specific expertise and examine how it moderates responses to failed trials. We find that firms without prior publications on the relevant protein generalize more broadly from failure. In doing so, they more aggressively screen out related opportunities, reducing investment in some low-potential neighbors while also increasing the likelihood that promising ones are prematurely abandoned. Firms with target-specific knowledge, by contrast, respond more cautiously (Levinthal and Schliesmann, 2025). Consistent with research on absorptive capacity (Cohen and Levinthal, 1990; Tranchero, 2026), prior knowledge thus appears to discipline generalization, improving firms' ability to interpret what a failure does, and does not, imply for related opportunities.

Taken together, this paper makes several contributions. We provide the first large-sample evidence on how organizations enrich scarce experience through generalization (Choi and Levinthal, 2023; Schliesmann, 2025). In doing so, we clarify the conditions in which generalization can serve as a substitute for direct experience and when it becomes distortionary. We also add to work on learning from failure (Madsen and Desai, 2010; Maslach, 2016; Lee and Park, 2024) and document a novel behavioral bias that emerges from extrapolating negative feedback. Furthermore, we contribute to research on R&D and innovation by offering a cognitive microfoundation for the diffusion of learning across innovation activities. Prior work has shown that firms respond to rivals' failed clinical trials (Krieger, 2021) and learn from experience across related activities (Eggers, 2012a; Zollo and Reuer, 2010). We extend this work by showing that the magnitude and direction of learning spillovers depend on which opportunities firms treat as related to a failed experiment (Kang, 2025).

Methodologically, we show how core constructs from research on innovation and learning can be operationalized in an empirical setting. Fitness landscapes are foundational to theories of organizational search (Levinthal, 1997), but studies that measure the structure of real innovation landscapes remain rare (Fleming and Sorenson, 2001). Our paper provides a blueprint for constructing such landscapes by exploiting various features of the empirical context. We define al-

ternatives as protein-disease pairs, map their biological proximity, and measure their therapeutic potential using genetic evidence. We then measure landscape ruggedness by examining whether nearby alternatives have similar potential, using the local relationship between proximity and genetic potential in each neighborhood (Jones and Forrest, 1995). This setup turns the opportunity landscape from a metaphor into an empirical object, allowing us to better understand how the structure of the search space shapes what organizations learn from failure.

2. THEORETICAL FRAMEWORK

We develop our theoretical framework in three steps. First, we situate the argument within research on learning from failure in innovation. Second, we use the imagery of fitness landscapes to explain how firms generalize from failures to related alternatives. Third, we examine when this process improves or worsens the allocation of innovative effort. In Appendix A, we formalize the intuition underlying the theoretical framework and provide additional analysis to empirically corroborate each step of the argument.

2.1. Learning from Failures in Innovation

Failure is pervasive in entrepreneurship and innovation: only a small fraction of new product ideas are ultimately commercialized (Stevens and Burley, 1997), and most new ventures do not survive (Levinthal, 1991; Knott and Posen, 2005; Chen et al., 2018). In such settings, progress depends less on the steady accumulation of successes than on how subsequent efforts build on repeated failed attempts (Yin et al., 2019). Reflecting this logic, the organizational learning literature has long emphasized the dual nature of failure. Although privately costly, failure can be productive because it signals a need to change (Posen et al., 2018). As a result, organizations often respond to failure by engaging in exploration (Greve, 2003; Maslach, 2016) and by increasing

risk-taking (Audia and Greve, 2006; Shapira, 1995). In innovative and creative settings, these responses can be especially valuable because failure not only reveals what does not work, but also expands the diversity of approaches organizations consider (March, 1991; Nelson, 1990).

Building on these insights, a large body of work has examined how organizations learn from failure (Desai et al., 2020). A central finding is that failure can improve subsequent performance, both for the focal firm and for the broader industry (Madsen and Desai, 2010, 2018; Knott and Posen, 2005). However, whether failure engenders subsequent learning is highly contingent and depends on the organization's structure (Gaba and Joseph, 2013; Desai, 2015), on the prior experience of the firm and its members (Madsen, 2009; Muehlfeld et al., 2012; Eggers, 2012a; Lee and Park, 2024), and on how noisy or ambiguous outcomes are interpreted (Levinthal and Rerup, 2021; Maslach et al., 2018). Firms also learn vicariously from similar or successful peers (Madsen, 2009; Kim and Miner, 2007; Krieger, 2021).

However, even when organizations learn from both direct and vicarious failure, relevant experience is often limited relative to the size and complexity of the search space (Maslach et al., 2018; Rerup and Zbaracki, 2021). As March et al. (1991: 1) note, "the paucity of historical events conspires against effective learning." This problem is especially acute in innovation, where experimentation is costly and at least partly irreversible. Pharmaceutical drug development illustrates the challenge well. In choosing which drug target to pursue, firms face more than 19,000 proteins and millions of possible protein-disease combinations (Tranchero, 2026). Yet the experiments needed to evaluate these opportunities are expensive and constrained by patient safety concerns, with each trial taking several years to complete (Kao, 2025b; Martin et al., 2017). As a result, innovation efforts remain concentrated on a small subset of targets (Hoelzemann et al., 2024), forcing firms to search and adapt on the basis of a thin experience base.

When experience is scarce, firms necessarily adopt learning strategies to compensate for their inability to evaluate each alternative in isolation. A key approach is generalization: using feedback from one alternative to revise beliefs about related but untested ones (Martignoni et al., 2016;

Choi and Levinthal, 2023; Schliesmann, 2025; Sen et al., 2026).[1] This strategy is especially important in innovation settings because a single experiment often carries information that can be extrapolated beyond the focal project. For instance, a failed clinical trial may cast doubt not only on a drug target, but also on a mechanism of action common to other targets or a broader therapeutic strategy (Krieger, 2021). In this sense, generalization allows firms to decide which alternatives deserve further investment without testing each one directly.

Despite its importance, relatively little is known about how organizations generalize from sparse failures to related alternatives. Existing work on generalization is largely theoretical, while empirical studies of learning from failure typically examine responses to the focal action. For example, Lee and Park (2024) examine how heart surgeons learn from failed procedures, but not whether those failures affect related procedures. This gap is consequential because organizations often learn from "samples of one or fewer" (March et al., 1991). Generalizing too little may leave firms investing in related, low-value alternatives; generalizing too much may lead them to abandon promising opportunities that superficially resemble past failures. The key question, therefore, is how inferences from failure propagate across the opportunity space, and with what consequences.

2.2. Generalizing on Fitness Landscapes

At the core of our theorizing is the idea that organizations search on a fitness landscape. In this imagery, each position corresponds to a distinct alternative, and each alternative has a fitness value reflecting its potential to solve the focal problem (Levinthal, 1997; Fleming and Sorenson, 2001; Baumann et al., 2019). Innovation search is therefore conceptualized as a series of experiments that test various positions to ascertain their true value (Chen et al., 2024). Locations on the landscape reflect latent sources of similarity (Ethiraj and Levinthal, 2004), such that proximate

[1] In this regard, it is important to distinguish between analogical reasoning and generalization. Analogical reasoning involves a one-to-one mapping between a novel experience and a familiar prior domain (Gavetti et al., 2005; Miller and Lin, 2015; Carroll and Sørensen, 2024). Generalization, by contrast, refers to the broader process by which knowledge from a prior domain is extended to new ones. Thus, analogical reasoning represents one specific way through which strategic actors generalize (Choi and Levinthal, 2023; Schliesmann, 2025; Sen et al., 2026).

alternatives tend to share more of their underlying attributes. This notion of similarity has long been recognized as a basis for generalization, dating back at least to Woodworth and Thorndike (1901), whose "identical elements" theory proposed that learning transfer depends on the number of shared features. Applied to organizational search, the extent to which experience is extrapolated across alternatives should depend on their proximity.

To make this more concrete, consider again Merck's efforts to develop a treatment for Alzheimer's disease. In this setting, the landscape is the space of possible drug targets, such as proteins that, when modulated by a drug, may elicit a therapeutic response. BACE1 occupies one position in this landscape, and its fitness corresponds to its therapeutic potential for treating Alzheimer's. Evaluating this position required costly and time-consuming experimentation. In Merck's case, the clinical program cost hundreds of millions, eventually resulting in a $226 million write-off directly tied to its failure (Merck & Co., 2017). The trial failure in 2017 publicly revealed that BACE1 was less promising than initially believed, leading multiple firms to update their beliefs about the focal target and to reduce subsequent investment in the approach (Begley, 2018; Krieger, 2021).

The same example illustrates why failure may also have implications beyond the focal alternative. Consistent with the theoretical discussion above, drug targets are not randomly distributed across the landscape. Those that participate in similar biological processes or interact with similar proteins are located closer together, forming neighborhoods of related alternatives (Kim, 2026; Srivastava et al., 2026). In the case of Merck's failure, since BACE1 is close to targets such as BACE2 and SORL1, negative evidence about BACE1 may inform beliefs about them. More generally, when direct experimentation is costly and feasible tests cover only a small part of the opportunity space, firms can use feedback from one experiment to update their beliefs about untested options. Generalization thus creates a "learning spillover" from the focal experiment to its surrounding neighborhood. The strength of this spillover should decline with distance, because closer alternatives share more underlying attributes with the failed target.

However, this spillover need not be uniform even among alternatives at the same distance from the focal failure. Within a given neighborhood, firms hold heterogeneous beliefs about the promise of different opportunities, and these priors shape how strongly they update their beliefs. The intuition follows Harrison and March (1984), who show that failure generates a negative post-decision surprise equal to the gap between prior expectations and realized performance. For alternatives viewed as more promising ex ante, the same focal failure represents a larger negative surprise and should therefore induce a stronger downward revision in beliefs. When firms generalize from a focal failure to related alternatives, these asymmetric revisions extend to the surrounding neighborhood. As a result, even among alternatives equally related to the failed target, investment should decline more sharply for high-potential alternatives than for lower-potential ones.

Hypothesis 1: *Following a failure, firms will reduce investment more strongly in high-potential opportunities than in low-potential opportunities at the same level of relatedness.*

2.3. The Effect of Generalization on the Allocation of Innovation Efforts

While the preceding hypothesis characterizes how failure affects individual alternatives at a given distance, the broader question is how generalization from failure shapes the allocation of innovation effort. A natural intuition is that generalization should improve such efforts. In most landscapes, similar alternatives tend to be associated with outcomes that are similar, though not identical (Ethiraj and Levinthal, 2004; Levinthal, 1997). A failed focal alternative should therefore provide a negative signal about related alternatives as well. Reducing investment in these alternatives may help organizations avoid continued search in unpromising regions of the landscape, lowering false positives while generating at most a limited increase in false negatives (Gavetti and Levinthal, 2000; Choi and Levinthal, 2023). In this sense, generalization appears to be an efficient mechanism for guiding search under scarce experience (Schliesmann, 2025).

However, this intuition assumes that failures are drawn from average regions of the search space.

In most innovation settings, this is unlikely to be the case. Organizations do not sample alternatives at random — they concentrate experimentation on candidates that appear promising ex ante, a feature of search known as endogenous sampling (Denrell and March, 2001). Observed failures therefore arise from targets that were attractive enough to justify costly experimentation in the first place. In locally correlated landscapes, this selection process has implications beyond the focal alternative. When firms choose a promising target for experimentation, they also tend to choose a neighborhood whose nearby alternatives are more valuable than the average ones in the broader landscape. As a result, generalizing from these failures applies negative feedback to a positively selected set of alternatives.

This logic changes the aggregate implication of generalization. The same selection process that makes a failure informative also makes it dangerous to extrapolate due to endogenous sampling. Generalizing from failure, then, can create a systematic distortion. Firms may reduce investment in some alternatives that are truly unpromising, thereby lowering the number of false positives pursued. But they also withdraw from valuable related opportunities that have not been tested directly, thereby increasing the number of false negatives. Because observed failures are concentrated in promising neighborhoods, this second effect will likely dominate the first. As a result, generalization from failure could worsen the allocation of innovative effort by redirecting investment away from valuable opportunities.

Hypothesis 2: *Generalization from failure will worsen the allocation of innovative effort by increasing false negatives among related opportunities.*

Furthermore, the extent of this distortion should vary with the structure of the innovation landscape. When local correlation is high, the value of a focal alternative is more closely tied to the value of nearby alternatives. This tight link amplifies the consequences of endogenous sampling. A firm that tests a high-value alternative has implicitly selected not only a focal position but also a neighborhood that, in expectation, is above average. A failed outcome provides negative infor-

mation about that neighborhood, but its meaning is not equivalent to observing failure in a randomly chosen part of the landscape. Unless nearby alternatives are perfectly correlated with the focal one, the failure lowers expected value without necessarily eliminating the positive selection effect created by the initial decision to experiment there.

This creates a counterintuitive boundary condition. The landscapes in which generalization appears most justified are also the landscapes in which generalizing from failure is most likely to create false negatives in search. In smoother landscapes, related alternatives are more informative about one another, making it reasonable to extrapolate from a focal failure. Yet, because firms selected the focal alternative from an above-average neighborhood, the negative signal also reaches nearby opportunities that may remain valuable relative to the broader landscape. In more rugged landscapes, by contrast, the value of the focal alternative is less tightly linked to the value of its neighborhood. Failures are therefore less likely to be concentrated in regions where high-potential alternatives are clustered, weakening the tendency for generalization to suppress investment in valuable related opportunities.

Hypothesis 3: *The increase in false negatives from generalizing failure will be weaker in more rugged, that is, less locally correlated, innovation landscapes.*

3. EMPIRICAL SETTING

We test our hypotheses with an empirical study of how pharmaceutical firms learn from clinical trial failures. Pharmaceutical R&D is well-suited to our theoretical framework because it lays bare the mismatch between the scale of the opportunity space and the limited experience base available to firms. One of the earliest and most consequential decisions in drug development is selecting a drug target, typically a protein whose modulation can produce a therapeutic effect (Nelson et al., 2015; Razuvayevskaya et al., 2024). That choice is difficult because firms must

search across more than 19,000 protein-coding genes and millions of possible protein-disease combinations, most of which offer little therapeutic value (Kang, 2025; Tranchero, 2026).

Direct experience is scarce because the experiments needed to evaluate any given target are slow, costly, and tightly constrained. Once a target is selected, compounds proceed through preclinical testing in animal models and then through three sequential phases of human clinical trials that increase in scale and regulatory scrutiny. Phase I assesses safety in a small group of subjects, phase II evaluates efficacy in a larger sample, and phase III confirms both in a broader population over a longer time horizon. Median costs reach $8.6 million for phase II and $21.4 million for phase III trials (Kao, 2025a). Yet despite this structured pipeline and extensive prior scientific research, attrition remains high: only one in ten drugs that enter clinical trials ultimately receives approval from the U.S. Food and Drug Administration (Hay et al., 2014). This is because the therapeutic potential of a target often becomes clear only after firms commit to large human trials.

Against this backdrop, the transparency and regulatory structure of clinical trials create a rich setting for studying organizational learning. Prior work shows that firms in this industry learn not only from their own efforts (Khanna et al., 2016; Maslach, 2016), but also from rivals' actions and outcomes (Baum and Dahlin, 2007; Krieger, 2021). Disclosure requirements, including the registration of trials on ClinicalTrials.gov, make competitors' progress and trial outcomes visible to the broader industry (Kao, 2025b). Pharmaceutical firms can therefore observe not only which targets rivals advance into human testing, but also whether those targets survive clinical evaluation. For instance, when Merck terminated its BACE1 trial in 2017 after lackluster results,[2] the failure became a salient public signal — as reflected by its shares trading 2.45% lower on the day of the announcement (Carroll, 2017).

Yet the relevant question for our purposes is whether this learning remains confined to the failed target. In the case of Merck's discontinued trial, some industry observers argued that the implications extended beyond the target directly tested (Garde, 2016). Rather than treating the episode as

[2]The details of Merck's EPOCH trial are available at: `https://clinicaltrials.gov/study/NCT01739348`.

an isolated setback, they interpreted it as a broader signal that weakened confidence in functionally related but untested alternatives hinging on the same therapeutic hypothesis (Begley, 2018). BACE2 provides a natural example. As its name suggests, it is closely related to BACE1 and participates in similar biological processes (Yeap et al., 2023). This biological proximity creates a basis for generalization: evidence against BACE1 may also reduce confidence in BACE2 and other nearby targets.

While such generalization may seem sensible, our theorizing suggests a more nuanced organizational reality. Firms select protein-disease pairs they view as especially promising, and proteins involved in similar biological processes are often locally clustered. As a result, targets that fail in clinical trials are likely to be embedded in neighborhoods that contain other high-potential alternatives. Generalizing from failure may therefore extend negative feedback over a positively selected set of untested targets. Put simply, generalizing from failures may lead to a disproportionate increase in false negatives in drug discovery. This logic challenges a simplistic view that learning from failure always improves search and the aggregate allocation of innovation effort. In what follows, we examine whether the underlying mechanisms of endogenous sampling and local correlation hold empirically.

4. DATA AND MEASUREMENT

4.1. Data Sources

In our setting, protein–disease pairs are the relevant alternatives that firms search over. Each protein represents a potential drug target for a given disease, and together they define the landscape in which firms learn from clinical trials and allocate their R&D investments. Relatedness across alternatives is captured through protein–protein interactions, which provide a biologically grounded measure of genetic distance. The latent potential of each alternative is proxied by Open

Targets scores, a measure of the genetic evidence supporting each protein–disease pair. We summarize each data source below and provide additional details in Appendix B.

4.1.1. Clinical Trials. We use data from ClinicalTrials.gov, an online registry maintained by the National Institutes of Health (NIH) that records ongoing and completed clinical trials. Following prior work, we focus on phase II and phase III trials, which represent substantial investments in drug development (Krieger, 2021; Martin et al., 2017). We exclude phase I trials because they are reported less consistently and often lack complete metadata (Kang, 2025; Kao, 2025b). Prior work also suggests that they are less informative, as early-stage failures tend to yield weaker signals (Eggers, 2012b; Khanna et al., 2016). Accordingly, we construct a dataset of all phase II and III trials completed between 2001 and 2019, including information on the drug target(s) and disease studied in each case. We end the sample in 2019 to avoid contamination from the COVID-19 pandemic and truncation due to reporting delays. We map protein names to Gene IDs from the National Center for Biotechnology Information (NCBI) and diseases to Medical Subject Headings (MeSH) terms, which allows us to trace trial outcomes at the level of specific protein-disease pairs. To identify failures, we use the reported trial status and classify a trial as failed if it was terminated before completion. Although early terminations can occur for multiple reasons, prior work shows that lack of therapeutic efficacy and the emergence of target-related side effects are the main predictors (Razuvayevskaya et al., 2024).

4.1.2. Firm Patent Applications. Because R&D spending is typically observed only at the firm level, it is difficult to determine which specific protein-disease pairs firms choose to pursue. Instead, we use patent applications to trace where firms direct their innovation efforts (Eggers and Kaplan, 2009). This measure is particularly well suited to pharmaceutical R&D, where firms patent early in the research process, before it is known whether a candidate will succeed clinically (Cohen et al., 2000). Patent applications, therefore, capture upstream commitment to a target rather than successful innovation outcomes. Through a partnership with the European Bioin-

formatics Institute, we use proprietary data compiled by SciBite's TERMite software, which extracts biological entities from the full text of USPTO patent applications filed between 2001 and 2019 (Tranchero, 2026). TERMite reliably distinguishes true biological entities from incidental mentions and maps proteins and diseases to the same standardized identifiers used in our clinical trial data. This allows us to place each patent application in the same protein-disease landscape in which firms receive feedback from clinical trials and to observe how negative signals reshape subsequent investment choices.

4.1.3. Genetic Distance. To capture generalization across related targets, we measure genetic distance using protein-protein interactions from the STRING database (Szklarczyk et al., 2025). STRING aggregates evidence on protein-protein associations in human biological processes and assigns each pair a combined confidence score based on the frequency and strength of those interactions. This yields a biologically grounded measure of functional proximity and, in our setting, a way to assess how informative feedback about one target is likely to be for another. Conversations with chemistry researchers reinforced this interpretation. They described STRING as a common tool for evaluating whether proteins are connected and identifying which pathways are likely to be affected when a drug target is perturbed. For example, BACE1 and BACE2 share key biological functions, which is reflected in their high interaction score (Appendix Figure B1). A key advantage of STRING is that it captures functional relationships among proteins without being tied to any single disease domain. This feature is important in drug development because firms often deploy the same target across multiple therapeutic areas. It also distinguishes our approach from disease-specific measures of relatedness based on genetic overlap, such as those that infer similarity across cancers from shared mutations (Kang, 2025).

4.1.4. Genetic Potential. To measure the promise of each protein-disease pair independent of clinical outcomes, we use the Open Targets score developed by the Open Targets Platform (Buniello et al., 2025). Open Targets is a public-private initiative that aggregates and weights ev-

idence on protein-disease pairs to support clinical prioritization. Each score reflects the strength of the genetic evidence linking a protein to a disease, adjusted for the quality and reliability of the sources. Recent work shows that Open Targets scores predict clinical success (Razuvayevskaya et al., 2024) and are positively associated with the technological and economic value of patents targeting the corresponding protein-disease pair (Tranchero, 2026). We merge these scores with our data using Gene IDs and MeSH terms. The resulting measure provides a benchmark for the genetic promise of each protein-disease pair, independent of firms' R&D choices. This allows us to assess both false positives (when firms pursue low-potential alternatives) and false negatives (when they ignore more promising ones) in the context of drug target selection.

4.2. Constructing Empirical Fitness Landscapes

The organizational learning literature has long described search as unfolding on landscapes in which alternatives differ in both their relatedness and their underlying potential (Levinthal, 1997). Pharmaceutical R&D offers an opportunity to observe such a landscape empirically. In our setting, the alternatives are protein–disease pairs, each representing a potential direction for drug discovery. Their location in the landscape is determined by biological relatedness across targets, which we capture using STRING protein-protein interaction scores. Pairs involving more closely related targets occupy nearby positions, whereas pairs involving more dissimilar targets lie farther apart. The potential of each protein–disease pair is proxied by its Open Targets score, which aggregates genetic evidence on the drug target. This corresponds to the "fitness" values in traditional NK models, with the interaction between distance and promise determining the degree of spatial autocorrelation, or landscape ruggedness. The result is an empirical landscape in which firms search across alternatives that vary in both proximity and promise.

Clinical trials provide the events through which firms learn in this landscape (Krieger, 2021). Each trial yields a discrete outcome for a focal protein–disease pair, revealing whether that target is more or less promising than previously believed. Patent applications then allow us to observe

how firms adjust subsequent innovation effort in response. By relating changes in patenting activity to the timing of trial discontinuations, we can estimate the direct effect of failure on the focal target. Panel (a) of Figure 1 illustrates this logic using Merck's failed BACE1 trial, showing in stylized form how a negative clinical outcome can reduce investment in the focal opportunity.

[INSERT FIGURE 1 ABOUT HERE]

To assess whether failure affects only the tested target or also nearby opportunities, we define a local neighborhood around each protein–disease pair. For each pair (p, d) that receives a trial, we collect all other pairs (p', d) whose proteins are sufficiently close to p in the STRING network. In the baseline specification, neighborhood boundaries are defined using the sample median of the STRING score.[3] This construction also yields a natural control group: treated neighborhoods are centered on pairs that have ever received a phase II or III clinical trial, whereas control neighborhoods are centered on never-trialed pairs that lie outside the neighborhood of any treated target. Panel (b) of Figure 1 shows part of the neighborhood around BACE1, with targets positioned by their STRING distance from BACE1 and colored by their Open Targets scores. Scaled to the full sample, this structure allows us to test whether a focal clinical failure remains confined to the tested target or spills over to nearby alternatives.

To capture variation in landscape structure, we compute a neighborhood-level ruggedness measure. For each focal protein–disease pair, we examine how the genetic potential of surrounding alternatives varies with biological distance from the focal target. The neighborhood is smoother when nearby alternatives tend to have similar Open Targets scores, and more distant ones differ; by contrast, a neighborhood is more rugged when biological distance is only weakly related to Open Targets scores. We operationalize this intuition with a local fitness-distance correlation measure following Jones and Forrest (1995) (details in Appendix B). Specifically, for each focal pair, we calculate the correlation between the Open Targets scores of neighboring alternatives

[3] Our results are robust to alternative proximity thresholds, such as the first or third quartiles of the distance distribution.

and their STRING distance from the focal target. This yields a measure of spatial correlation centered on each focal target. We then use it to test whether spillovers from failure are stronger in smoother neighborhoods, where proximity is more informative, than in more rugged ones.

4.3. Descriptive Statistics

Table 1 reports summary statistics for the empirical landscape and the panel structures used in the analysis. Panel A presents cross-sectional descriptive statistics at the protein-disease pair level. The data include 7,788,369 protein-disease pairs, constructed from 16,136 human proteins and 483 diseases represented in the patent data. We have information on 8,628 unique phase II and III clinical trials, of which 1,646 were terminated prior to completion and are therefore classified as failures (Appendix Figure C1). These trials provide direct information on 7,007 protein-disease pairs and, as a function of their functional proximity, indirect information on an additional 2,283,284 related pairs.

Panel B presents descriptive statistics for the panel of protein-disease pairs observed annually from 2001 to 2019. On average, a protein-disease pair receives 0.084 patent applications per year, although the distribution is highly skewed, with some targets receiving more than 1,400 applications in a single year. Patenting is substantially more intense for higher-potential pairs, defined as those with a positive Open Targets score. Only 0.05% of observations fall in the post-period following a direct clinical trial, whereas 15.9% fall in the post-period following a trial in a genetically related pair. This contrast highlights a central feature of the setting: firms rarely observe direct evidence for a given pair, but often face signals from nearby trials.

Panel C summarizes the neighborhood-year panel used to study the aggregate implications of generalization. Average patenting in a focal neighborhood is 46.1 applications per year. Although this activity is split almost evenly between high-promise and low-promise pairs, the underlying neighborhoods contain far more low-promise alternatives. On average, a neighborhood contains

302 protein-disease pairs, of which 23.8 are high-potential, and 278.2 are low-potential. Together, these descriptive statistics underscore the core empirical features of the setting: a vast and sparse opportunity space, constrained direct experimentation, and substantial scope for generalizing across nearby alternatives.

[INSERT TABLE 1 ABOUT HERE]

5. RESULTS

5.1. Research Design

Empirically studying how firms learn from failure poses two main challenges. The first is measurement. To identify learning, researchers must observe both the actions firms take and the feedback those actions generate. Without clear information on both, it is difficult to link subsequent behavior to the underlying learning process. The second challenge is causal inference. Firms are more likely to invest in areas where they already possess expertise, such as genes they have previously studied successfully, which can bias estimates upward. In an ideal setting, firms would receive exogenous information about the therapeutic potential of specific protein-disease pairs. The causal effect of that information could then be identified from changes in patenting activity on treated pairs relative to otherwise similar pairs that remain unaffected.

We approximate this ideal experiment by exploiting the staggered timing of clinical trial discontinuations. These events are publicly disclosed and generate shared information observed by all pharmaceutical firms, regardless of their internal data or capabilities. Following Krieger (2021), we focus on firms' responses to the failures of other companies, which helps mitigate concerns that the feedback firms receive is mechanically tied to their own prior research choices. Because firms do not initiate trials expecting them to fail, these discontinuations are plausibly unexpected

for the sponsoring firm and even more so for outside observers. This setup provides an opportunity to identify how organizations adjust their search in response to publicly disclosed negative outcomes.

An important advantage of our approach is that it avoids a well-known pitfall in the empirical study of organizational learning. Much of the prior literature relies on cumulative counts of past failures to explain subsequent performance. Yet cumulative measures can generate significant estimates even in the absence of genuine learning effects (Bennett and Snyder, 2017). The problem is that correlations between past failures and future outcomes may arise mechanically from shared time trends rather than from behavioral changes. By contrast, the discontinuation of a clinical trial is a discrete and externally visible event. This allows us to distinguish genuine responses to negative feedback from statistical artifacts. More broadly, it provides an empirical design that allows failures to be directly linked to subsequent search behavior.

5.2. Validation of the Research Design

We first validate whether clinical trial failures act as unanticipated learning shocks by examining their impact on subsequent innovation for the same protein-disease pair. To do so, we estimate their direct effect with a difference-in-differences specification at the protein-disease level:

$$\begin{aligned} Y_{i,j,t} = \alpha + \beta_1 \left(Post_t \times ClinicalTrial_{i,j}\right) + \beta_2 \left(Post_t \times ClinicalTrial_{i,j} \times Failure_{i,j}\right) \\ + \gamma PD_{i,j} + \delta Protein_i + \omega Disease_j + \sigma Year_t + \epsilon_{i,j,t}, \end{aligned} \tag{1}$$

where $Y_{i,j,t}$ denotes the number of patent applications filed in year t for inventions targeting protein i and disease j. The term $Post_t \times ClinicalTrial_{i,j}$ captures the change in patenting after the conclusion of a clinical trial on protein-disease pair $\langle i, j \rangle$, while the interaction $Post_t \times ClinicalTrial_{i,j} \times Failure_{i,j}$ captures the additional change following a discontinued trial. Accordingly, β_1 measures the average post-trial change after completed trials, while β_2 measures the causal effect of unanticipated failures relative to that baseline. Fixed effects at the protein-disease,

protein, disease, and year levels absorb differences in baseline research intensity and common trends across technologies and therapeutic areas. Standard errors are clustered by protein and disease.

[INSERT TABLE 2 ABOUT HERE]

Columns 1 and 2 of Table 2 report the baseline results. Although the successful completion of a clinical trial is generally associated with more patent applications, there is a large and significant decline when the trial is terminated early. Appendix Figure C2 examines the stability of this estimate across alternative fixed-effect structures. The largest change in magnitude occurs when protein-disease pair fixed effects are introduced, suggesting that part of the raw association reflects persistent cross-sectional differences in research intensity across pairs. Once those differences are absorbed, the remaining effect is identified from within-pair changes following failure. Additional robustness checks show that the results remain unchanged when phase III trials, which are typically closer to regulatory approval, are excluded from the sample (Appendix Table C1).[4]

The central identifying assumption of our difference-in-differences design is that, absent a trial discontinuation, patenting trends for protein-disease pairs targeted by discontinued trials would have evolved in parallel to those targeted by completed trials. Figure 2 assesses this assumption using an event-study version of Equation 1. The estimates show flat and statistically indistinguishable pre-trends, followed by a persistent decline in patenting after the public disclosure of a failure on ClinicalTrials.gov. The post-treatment coefficients stabilize at a level consistent with the average treatment effect reported in Table 2. Taken together, these results support the interpretation of clinical trial failures as salient and well-identified negative signals, consistent with prior work on organizational responses to failure (Greve, 2003; Krieger, 2021).

[INSERT FIGURE 2 ABOUT HERE]

[4]Because patenting is skewed, we also estimate the main specifications using a logarithmic transformation of the dependent variable. The results are substantively unchanged, as shown in Appendix Table C2.

5.3. Estimating Learning Spillovers from Failure

We next examine whether clinical trial failures generate spillover effects on related protein-disease pairs, rather than affecting only the targets directly tested. Specifically, for a focal clinical trial on protein-disease pair $\langle i, j \rangle$, we estimate the following difference-in-differences specification for functionally related pairs $\langle p, j \rangle$:

$$\begin{aligned} Y_{p,j,t} =& \alpha + \beta_1 \left(Post_t \times ClinicalTrial_{p,j}^{D(i,j)} \right) + \beta_2 \left(Post_t \times ClinicalTrial_{p,j}^{D(i,j)} \times Failure_{i,j} \right) \\ &+ \gamma PD_{p,j} + \delta Protein_p + \omega Disease_j + \sigma Year_t + \epsilon_{p,j,t}, \end{aligned} \tag{2}$$

where $Y_{p,j,t}$ denotes the number of patent applications filed in year t for inventions targeting protein p and disease j, excluding the focal pair that received the clinical trial. The term $ClinicalTrial_{p,j}^{D(i,j)}$ equals one if pair $\langle p, j \rangle$ lies in distance quartile D from the focal pair $\langle i, j \rangle$. The interaction $Post_t \times ClinicalTrial_{p,j}^{D(i,j)}$ captures the spillover associated with the conclusion of a focal clinical trial for related pairs in quartile D. The triple interaction $Post_t \times ClinicalTrial_{p,j}^{D(i,j)} \times Failure_{i,j}$ captures the additional spillover following a discontinued trial.[5] The specification includes fixed effects at the protein-disease, protein, disease, and year levels. Standard errors are clustered by protein and disease. The triple-difference coefficient β_2 identifies whether patenting in related pairs falls following a focal failure, relative to pairs that are not related to the failed trial.

Our results show that clinical failures reverberate beyond the tested targets. Columns 3 and 4 of Table 2 provide clear evidence of negative spillovers, with patenting declining by 17.4% relative to the sample mean.[6] Panel (a) of Figure 3 sharpens this result by dividing protein-disease pairs

[5] We define spillover treatment using the first related clinical trial observed for each protein-disease pair. Specifically, for each pair $\langle p, j \rangle$, we identify the first phase II or III trial in the sample that targets a genetically related protein for the same disease. The post indicator turns on after the conclusion of that trial and remains on thereafter, and the failure indicator is defined by that trial's outcome. Later related trials do not change treatment timing or failure status in the baseline specification. This coding choice assigns each pair to a single spillover event, thereby preserving a standard staggered-treatment interpretation.

[6] Appendix Table C3 shows that these results are robust to alternative definitions of the estimation sample.

into quartiles based on their genetic distance from the target of the discontinued trial. The pattern closely matches the theoretical intuition. Firms reduce innovation activity more strongly for pairs that are functionally close to failed targets, and the effect declines steadily with distance. Spillovers remain negative throughout and become statistically insignificant only in the most distant quartile. As a falsification test, we also examine whether spillovers are stronger when the failure signal is more precise. We use enrollment size as a proxy for signal strength, since larger trials provide more informative clinical evidence (Catillon, 2019). Appendix Table C4 shows that failures of higher-enrollment trials generate larger spillovers onto related targets. This pattern is consistent with the idea that firms generalize more when the signal is clearer.

[INSERT FIGURE 3 ABOUT HERE]

However, these average spillovers conceal the key asymmetry predicted by Hypothesis 1. Panel (b) of Figure 3 separates protein-disease pairs into high- and low-genetic-potential categories using the Open Targets score, with supporting regressions reported in Appendix Table C5. Among pairs equally related to the discontinued trial, there is a larger decline in patenting among high-potential pairs than among low-potential ones. Consistent with our theory, the difference is most pronounced among the closest neighbors. Related failures appear to generate larger downward revisions for alternatives that firms had stronger reasons to view ex ante favorably, echoing Harrison and March (1984)'s argument that failure creates greater post-decision surprise when prior expectations are higher. These results support Hypothesis 1 and show that generalization from failure is strongest for the alternatives that firms had the greatest reason to pursue.

5.4. Does Generalizing Failures Hurt Search?

We now move the level of analysis from individual protein-disease pairs to local neighborhoods. Appendix B illustrates this logic in the BACE1-Alzheimer's case. Consistent with our theoretical framework, Merck's trial failure was followed by reduced patenting on the most promising

neighboring drug targets, but not on the least promising ones. We now test whether this pattern holds systematically by estimating the following difference-in-differences specification on the neighborhood-year panel:

$$Y_{n,t} = \alpha + \beta_1 \left(Post_t \times ClinicalTrial_n\right) + \beta_2 \left(Post_t \times ClinicalTrial_n \times Failure_n\right) + \mu_N + \delta_P + \omega_D + \sigma_t + \varepsilon_{n,t}, \tag{3}$$

where $Y_{n,t}$ denotes patenting in neighborhood n in year t, excluding the focal protein–disease pair at the center of the neighborhood. Depending on the specification, the outcome is total patenting in the neighborhood or patenting separately in high-potential and low-potential neighboring pairs. As in the previous specifications, the interaction $Post_t \times ClinicalTrial_n \times Failure_n$ captures the additional change following a terminated trial. Neighborhood, protein, disease, and year fixed effects absorb time-invariant heterogeneity and common shocks, and standard errors are clustered at the neighborhood level. The coefficient of interest is β_2, which identifies how a failure changes subsequent patenting in the surrounding neighborhood relative to a successfully completed trial.

Returning to our theoretical framework, Hypothesis 2 predicts that generalization from failure can worsen the allocation of innovative effort by increasing false negatives among related opportunities. Table 3 provides direct support for this prediction. At the neighborhood level, the early termination of a focal clinical trial leads to a substantial decline in subsequent patenting. This decline is not limited to low-potential alternatives that firms would ideally abandon. Instead, it is larger for high-potential neighboring pairs than for low-potential ones. Column 2 shows that patenting in high-potential neighbors is reduced by 59.5 applications after a discontinued trial, compared with 26.3 applications for low-potential neighbors in Column 3. This difference is especially consequential because only about 8% of neighboring pairs are, on average, high potential. The reduction in innovative effort is therefore concentrated on a relatively small set of valuable alternatives. Consistent with Hypothesis 2, generalization from failure withdraws innovation effort from precisely the most promising parts of the landscape.

[INSERT TABLE 3 ABOUT HERE]

Figure 4 clarifies why generalization from failure can worsen the allocation of innovation effort. As theorized in Section 2 and Appendix A, the key mechanism is the interaction between endogenous sampling and local correlation in target quality. Panel (a) shows that firms do not test protein-disease pairs at random. Clinical trials are disproportionately concentrated among pairs with higher Open Targets scores, indicating that experimentation is directed toward opportunities that appear more promising ex ante. Panel (b) shows that these promising focal pairs also sit in more promising neighborhoods. Because genetic potential is locally correlated, neighborhoods around tested targets, including failed ones, contain more high-potential neighbors than neighborhoods around pairs that never receive a trial. This creates a central paradox. Because nearby alternatives are informative about one another, it is sensible to generalize from failure. Yet because firms test the opportunities they view as most promising, failures occur disproportionately where valuable neighbors are concentrated. Generalizing from those failures, therefore, mostly screens out valuable drug targets.

[INSERT FIGURE 4 ABOUT HERE]

Finally, we examine whether the effect identified in Hypothesis 2 varies across landscape structures. To do so, we split neighborhoods using the local fitness-distance correlation measure, which captures how tightly the promise of surrounding alternatives covaries with their distance from the focal target (Jones and Forrest, 1995). Neighborhoods are smoother when distance is more informative about underlying potential, and more rugged when it is less so. Figure 5 shows that in smoother neighborhoods, a focal failure leads to a large and significant decline in patenting on high-potential neighboring pairs. In more rugged neighborhoods, by contrast, this effect largely disappears. This pattern strongly supports Hypothesis 3. When nearby alternatives are more tightly linked in their underlying promise, generalizing from a focal failure becomes both

more compelling and more costly, because valuable opportunities lie close to the positively selected target. When that local correlation weakens, failure is less informative about nearby alternatives, and the distortion correspondingly attenuates.

[INSERT FIGURE 5 ABOUT HERE]

Taken together, the evidence closely aligns with the predictions of our theoretical framework. Firms learn directly from failure and generalize that feedback to biologically related opportunities in a distance-dependent way. Although this process may help screen out weak nearby alternatives, it also deters firms from pursuing high-potential ones, especially when failures occur in smooth, promising regions of the landscape. More broadly, the results highlight both the value and the cost of generalization, showing how learning from sparse experiments can systematically redirect search away from valuable opportunities.

6. WHICH FIRMS GENERALIZE MORE?

In light of these results, an important question is whether all firms rely on generalization to the same extent. The answer is not obvious ex ante. Richer knowledge bases may help firms recognize meaningful similarities across opportunities and generalize more productively (Gavetti et al., 2005; Miller and Lin, 2015). For example, it was only because of his prior experience that Charles Merrill was able to discern similarities between retailing and financial services that others missed (Gavetti and Menon, 2016). Yet deeper domain knowledge may also reduce the need to generalize from any single failure. Firms with more direct experience can better contextualize new observations (Tranchero, 2026), making them less reliant on relatedness alone as a basis for inference. Less experienced firms, by contrast, may use generalization to enrich their scarce knowledge, even when doing so exposes them to the distortions documented above.

We explore this question by linking firms' reactions to clinical failures with their prior scientific

experience with drug targets. We use publication data from PubTator Central, which identifies protein mentions in PubMed-indexed articles through computer annotation (Wei et al., 2024), and match these publications to firms using author affiliations. Additional details on the data are provided in Appendix B. This allows us to observe whether a firm had studied a given protein before filing a patent that mentions it. For each firm-patent observation, we classify a firm as experienced if it had previously published on the protein targeted by the patent application. We then re-estimate the spillover specification in Equation 2 separately for experienced and inexperienced firms, asking whether target-specific knowledge changes the tendency to generalize negative feedback to biologically related alternatives.

Appendix Figure C3 presents the first set of results, with the corresponding regressions reported in Appendix Table C6. The evidence shows that firms differ substantially in how broadly they generalize from failure. Firms without prior scientific publications on the relevant target exhibit stronger reactions from failed trials, with effects that attenuate as biological distance from the target increases. By contrast, firms with target-specific knowledge respond less strongly, suggesting that they rely less on generalization. Figure 6 illustrates the consequences of these differences in learning behavior. Inexperienced firms reduce patenting on low-potential neighboring pairs after a failure, consistent with generalization helping them screen out some false positives. However, this gain is offset by a much sharper decline in patenting on high-potential neighboring pairs. Taken together, these results suggest that generalization operates as an imperfect substitute for knowledge capabilities.

[INSERT FIGURE 6 ABOUT HERE]

Prior knowledge shapes not only what organizations know, but also how they interpret failure. Firms with deeper target-specific knowledge appear better able to contextualize a focal failure, thus being less likely to withdraw from promising neighbors. This pattern is consistent with research on absorptive capacity, which emphasizes that prior knowledge shapes how organizations

interpret and learn from new information (Cohen and Levinthal, 1990). More broadly, experience operates as both a pipe and a prism (Levinthal and Wu, 2025): it expands the information firms can access and changes how they evaluate new evidence. In this setting, specialized knowledge fosters a degree of healthy skepticism, helping firms persist with promising opportunities despite negative feedback (Levinthal and Schliesmann, 2025). By contrast, firms with less direct experience rely more heavily on generalization, bearing the costs of markedly higher false negatives.

7. DISCUSSION AND CONCLUSION

In many innovation settings, organizations operate in environments where the number of opportunities greatly exceeds the number that can be tested directly. In such contexts, organizations may compensate by generalizing — extrapolating experience from one domain to related but untested alternatives. Through this strategy, organizations can augment meager experience, form beliefs about the potential of adjacent possibilities, and conceivably improve the allocation of innovation efforts (Choi and Levinthal, 2023; Schliesmann, 2025). However, we identify an important boundary condition to the efficacy of this mode of learning. When firms sample opportunities endogenously, focusing experimentation on alternatives they believe are most promising, and there is spatial correlation in the fitness landscape, generalizing from failure can backfire. Although such learning reduces false positives, it also leads to a disproportionate increase in false negatives because failures are predominantly observed in promising neighborhoods.

Prior work has recognized endogenous sampling as central to innovation and organizational learning (Denrell and March, 2001). Because organizations learn from the alternatives they choose to pursue, sampling decisions shape both what they discover and what they fail to learn. This mechanism provides a dynamic foundation for the exploration-exploitation tradeoff (March, 1991), but it also creates systematic biases in learning. In particular, organizations can fall victim to the hot-stove effect: after receiving negative feedback, they may abandon promising alternatives and

foreclose the additional experimentation needed to ascertain their true value (Denrell and March, 2001). Our findings expand the scope of this mechanism. Failure with a focal alternative does not merely foreclose future opportunities to learn about that alternative but may also suppress investment in similar, high-potential targets. Put simply, our theorizing and empirical results suggest that the hot-stove effect in innovation search may be more pernicious than previously demonstrated.

Furthermore, our findings point to an important distinction between the accuracy and representativeness of performance feedback. Research on organizational search has long emphasized the challenges of evaluating noisy and potentially misleading feedback (Knudsen and Levinthal, 2007), especially when learning from rare events (Maslach et al., 2018). Yet the problem is not limited to whether information correctly characterizes the alternative tested. Even accurate information can distort search if it draws attention toward a narrow part of the opportunity space (Hoelzemann et al., 2024). Our results show that generalization can amplify this problem by extending unrepresentative observations across related opportunities. A failure may accurately reveal the limited promise of the focal alternative, yet provide a poor basis for evaluating its neighbors. Conversely, when observed outcomes are representative of the broader neighborhood, generalization can improve search efficiency even if the focal signal is noisy.

Taken together, our findings reveal a fundamental tension in organizational learning from failure. Spatial correlation is what makes generalization valuable: when nearby alternatives share underlying attributes, feedback from one experiment can help organizations draw inferences about opportunities they have not tested. But the same structure becomes costly once feedback is generated through endogenous sampling. By selecting targets that they see as promising, firms tend to experiment in positively selected regions, so even failed trials occur in neighborhoods that remain above average. The implication is that organizations must evaluate not only whether feedback is informative, but also whether the observation from which they learn is representative of the alternatives to which it is applied. In innovation settings, the risk is that firms learn too broadly from

failures that are locally accurate but unrepresentative, turning a useful learning heuristic into a source of systematic underexploration.

REFERENCES

Audia, P. G. and H. R. Greve (2006): "Less Likely to Fail: Low Performance, Firm Size, and Factory Expansion in the Shipbuilding Industry," *Management Science*, 52, 83–94.

Baum, J. A. and K. B. Dahlin (2007): "Aspiration performance and railroads' patterns of learning from train wrecks and crashes," *Organization Science*, 18, 368–385.

Baumann, O., J. Schmidt, and N. Stieglitz (2019): "Effective search in rugged performance landscapes: A review and outlook," *Journal of Management*, 45, 285–318.

Begley, S. (2018): "What can we learn from the latest Alzheimer's drug failure?" *STAT News*.

Bennett, V. M. and J. Snyder (2017): "The empirics of learning from failure," *Strategy Science*, 2, 1–12.

Buniello, A., D. Suveges, C. Cruz-Castillo, M. B. Llinares, et al. (2025): "Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery," *Nucleic Acids Research*, 53, D1467–D1475.

Bush, R. R. and F. Mosteller (1955): *Stochastic models for learning*, John Wiley & Sons, Inc.

Carroll, G. R. and J. B. Sørensen (2024): "Strategy theory using analogy: Rationale, tools and examples," *Strategy Science*, 9, 483–498.

Carroll, J. (2017): "Another Alzheimer's drug flops in pivotal clinical trial," *Science*, Endpoints News.

Catillon, M. (2019): "Trends and predictors of biomedical research quality, 1990–2015: a meta-research study," *BMJ open*, 9, e030342.

Chen, J. S., D. C. Croson, D. W. Elfenbein, and H. E. Posen (2018): "The impact of learning and overconfidence on entrepreneurial entry and exit," *Organization Science*, 29, 989–1009.

Chen, J. S., D. W. Elfenbein, H. E. Posen, and M. Z. Wang (2024): "Programs of experimentation and pivoting for (overconfident) entrepreneurs," *Academy of Management Review*, 49, 80–106.

Choi, J. and D. Levinthal (2023): "Wisdom in the wild: Generalization and adaptive dynamics," *Organization Science*, 34, 1073–1089.

Cohen, W. M. and D. A. Levinthal (1990): "Absorptive capacity: A new perspective on learning and innovation," *Administrative science quarterly*, 35, 128–152.

Cohen, W. M., R. Nelson, and J. P. Walsh (2000): "Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)," *NBER Working Paper w7552*.

Denrell, J. and J. G. March (2001): "Adaptation as information restriction: The hot stove effect," *Organization Science*, 12, 523–538.

Desai, V. (2015): "Learning through the distribution of failures within an organization: Evidence from heart bypass surgery performance," *Academy of Management Journal*, 58, 1032–1050.

Desai, V. M., D. Maslach, and P. Madsen (2020): "Organizational learning from failure," *The Oxford handbook of group and organizational learning*, 109.

Eggers, J. P. (2012a): "All experience is not created equal: Learning, adapting, and focusing in product portfolio management," *Strategic Management Journal*, 33, 315–335.

——— (2012b): "Falling flat: Failed technologies and investment under uncertainty," *Administrative Science Quarterly*, 57, 47–80.

Eggers, J. P. and S. Kaplan (2009): "Cognition and renewal: Comparing CEO and organizational effects on incumbent adaptation to technical change," *Organization Science*, 20, 461–477.

Ethiraj, S. K. and D. Levinthal (2004): "Modularity and innovation in complex systems," *Management Science*, 50, 159–173.

Fleming, L. (2001): "Recombinant uncertainty in technological search," *Management science*, 47, 117–132.

Fleming, L. and O. Sorenson (2001): "Technology as a complex adaptive system: Evidence from patent data," *Research Policy*, 30, 1019–1039.

Gaba, V. and J. Joseph (2013): "Corporate structure and performance feedback: Aspirations and adaptation in M-form firms," *Organization Science*, 24, 1102–1119.

Garde, D. (2016): "A big Alzheimer's drug trial now wrapping up could offer real hope—Or crush it," *STAT News*.

Gavetti, G. and D. Levinthal (2000): "Looking forward and looking backward: Cognitive and experiential search," *Administrative Science Quarterly*, 45, 113–137.

Gavetti, G., D. A. Levinthal, and J. W. Rivkin (2005): "Strategy making in novel and complex worlds: The power of analogy," *Strategic Management Journal*, 26, 691–712.

Gavetti, G. and A. Menon (2016): "Evolution cum agency: Toward a model of strategic foresight," *Strategy Science*, 1, 207–233.

Greve, H. R. (2003): *Organizational learning from performance feedback: A behavioral perspective on innovation and change*, Cambridge University Press.

Harrison, J. R. and J. G. March (1984): "Decision making and postdecision surprises," *Administrative Science Quarterly*, 26–42.

Hay, M., D. W. Thomas, J. L. Craighead, C. Economides, and J. Rosenthal (2014): "Clinical development success rates for investigational drugs," *Nature Biotechnology*, 32, 40–51.

Hoelzemann, J., G. Manso, A. Nagaraj, and M. Tranchero (2024): "The streetlight effect in data-driven exploration," *NBER wp 32401*.

Jones, T. and S. Forrest (1995): "Fitness distance correlation as a measure of problem difficulty for genetic algorithms." in *ICGA*, vol. 95, 184–192.

Kang, S. (2025): "From outward to inward: Reframing search with new mapping criteria," in *Academy of Management Proceedings*, Academy of Management Valhalla, NY 10595, vol. 2025, 10129.

Kao, J. (2025a): "Charted territory: Mapping the cancer genome and R&D decisions in the pharmaceutical industry," *UCLA Anderson*.

——— (2025b): "Information disclosure and competitive dynamics: Evidence from the pharmaceutical industry," *Management Science*, 71, 5948–5970.

Khanna, R., I. Guler, and A. Nerkar (2016): "Fail often, fail big, and fail fast? Learning from small failures and R&D performance in the pharmaceutical industry," *Academy of Management Journal*, 59, 436–459.

Kim, J.-Y. and A. S. Miner (2007): "Vicarious learning from the failures and near-failures of others: Evidence from the US commercial banking industry," *Academy of Management Journal*, 50, 687–714.

Kim, S. (2026): "Navigating the rugged data landscape: The impact of data-extrapolation technologies on knowledge production," *Columbia Business School*.

Knott, A. M. and H. E. Posen (2005): "Is failure good?" *Strategic Management Journal*, 26, 617–641.

Knudsen, T. and D. A. Levinthal (2007): "Two Faces of Search: Alternative Generation and Alternative Evaluation," *Organization Science*, 18, 39–54.

Krieger, J. L. (2021): "Trials and terminations: Learning from competitors' R&D failures," *Management Science*, 67, 5525–5548.

Lee, S. and J. Park (2024): "Giving up learning from failures? An examination of learning from one's own failures in the context of heart surgeons," *Strategic Management Journal*, 45, 2063–2094.

Levinthal, D. A. (1991): "Random Walks and Organizational Mortality," *Administrative Science Quarterly*, 36, 397–420.

——— (1997): "Adaptation on rugged landscapes," *Management Science*, 43, 934–950.

Levinthal, D. A. and C. Rerup (2021): "The plural of goal: Learning in a world of ambiguity," *Organization science*, 32, 527–543.

Levinthal, D. A. and D. Schliesmann (2025): "Cautious exploitation: Learning and search in problems of evaluation and discovery," *Organization Science*, 36, 903–917.

Levinthal, D. A. and B. Wu (2025): "Resource redeployment and the pursuit of the new best use: Economic logic and organizational challenges," *Strategy Science*, 10, 32–47.

Madsen, P. M. (2009): "These Lives Will Not Be Lost in Vain: Organizational Learning from Disaster in U.S. Coal Mining," *Organization Science*, 20, 861–875.

Madsen, P. M. and V. Desai (2010): "Failing to Learn? The Effects of Failure and Success on Organizational Learning in the Global Orbital Launch Vehicle Industry," *The Academy of Management Journal*, 53, 451–476.

——— (2018): "No firm is an island: The role of population-level actors in organizational learning from failure," *Organization Science*, 29, 739–753.

March, J. G. (1991): "Exploration and exploitation in organizational learning," *Organization science*, 2, 71–87.

March, J. G., L. S. Sproull, and M. Tamuz (1991): "Learning from samples of one or fewer," *Organization Science*, 2, 1–13.

Martignoni, D., A. Menon, and N. Siggelkow (2016): "Consequences of misspecified mental models: Contrasting effects and the role of cognitive fit," *Strategic Management Journal*, 37, 2545–2568.

Martin, L., M. Hutchens, C. Hawkins, and A. Radnov (2017): "How much do clinical trials cost?" *Nature Reviews Drug Discovery*, 16, 381–382.

Maslach, D. (2016): "Change and persistence with failed technological innovation," *Strategic Management Journal*, 37, 714–723.

Maslach, D., O. Branzei, C. Rerup, and M. J. Zbaracki (2018): "Noise as signal in learning from rare events," *Organization Science*, 29, 225–246.

Merck & Co. (2017): "Annual Report: 2017," *Form 10-K*.

Miller, K. D. and S.-J. Lin (2015): "Analogical reasoning for diagnosing strategic issues in dynamic and complex environments," *Strategic Management Journal*, 36, 2000–2020.

Muehlfeld, K., P. Rao Sahib, and A. Van Witteloostuijn (2012): "A contextual theory of organizational learning from failures and successes: A study of acquisition completion in the global newspaper industry, 1981–2008," *Strategic management journal*, 33, 938–964.

Nelson, M. R., H. Tipney, J. L. Painter, et al. (2015): "The support of human genetic evidence for approved drug indications," *Nature Genetics*, 47, 856–860.

Nelson, R. R. (1990): "Capitalism as an engine of progress," *Research Policy*, 19, 193–214.

Posen, H. E., T. Keil, S. Kim, and F. D. Meissner (2018): "Renewing research on problemistic search—A review and research agenda," *Academy of Management Annals*, 12, 208–251.

Razuvayevskaya, O., I. Lopez, I. Dunham, and D. Ochoa (2024): "Genetic factors associated with reasons for clinical trial stoppage," *Nature Genetics*, 56, 1862–1867.

Rerup, C. and M. J. Zbaracki (2021): "The politics of learning from rare events," *Organization Science*, 32, 1391–1414.

Schliesmann, D. (2025): "The where of search," in *Academy of Management Proceedings*, Academy of Management Valhalla, NY 10595, vol. 2025, 10172.

Sen, P., M. Workiewicz, and P. Puranam (2026): "Can LLMs aid analogical reasoning for strategic decisions? A comparative study," *Strategy Science*, 11, 118–136.

Shapira, Z. (1995): *Risk Taking: A Managerial Perspective*, Russell Sage Foundation.

Srivastava, M., A. A. Louis, and N. S. Martin (2026): "Predicting the topography of fitness landscapes from the structure of genotype-phenotype maps," *Genetics*, iyag026.

Stevens, G. A. and J. Burley (1997): "3,000 Raw Ideas = 1 Commercial Success!" *Research-Technology Management*, 40, 16–27.

Szklarczyk, D., K. Nastou, M. Koutrouli, et al. (2025): "The STRING database in 2025: Protein networks with directionality of regulation," *Nucleic Acids Research*, 53, D730–D737.

Tranchero, M. (2026): "Finding diamonds in the rough: Data-driven opportunities and pharmaceutical innovation," *University of Pennsylvania*.

Wei, C.-H., A. Allot, P.-T. Lai, R. Leaman, et al. (2024): "PubTator 3.0: An AI-powered literature resource for unlocking biomedical knowledge," *Nucleic Acids Research*, 52, W540–W546.

Woodworth, R. S. and E. L. Thorndike (1901): "The influence of improvement in one mental function upon the efficiency of other functions." *Psychological review*, 8, 247.

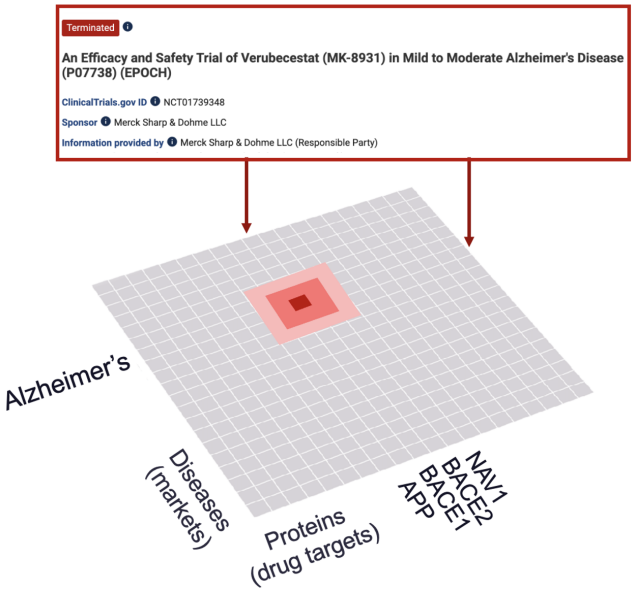
Yeap, Y. J., N. Kandiah, D. Nizetic, and K.-L. Lim (2023): "BACE2: A promising neuroprotective candidate for Alzheimer's disease," *Journal of Alzheimer's Disease*, 94, S159–S171.

Yin, Y., Y. Wang, J. A. Evans, and D. Wang (2019): "Quantifying the dynamics of failure across science, startups and security," *Nature*, 575, 190–194.

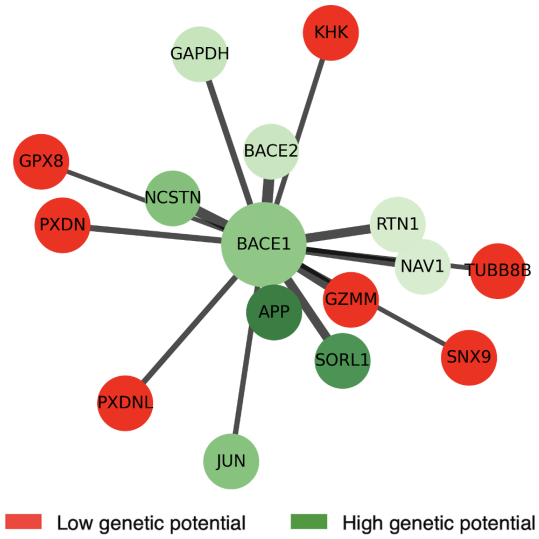
Zollo, M. and J. J. Reuer (2010): "Experience spillovers across corporate development activities," *Organization Science*, 21, 1195–1212.

Figure 1: Generalization from Clinical Trial Failure in Pharmaceutical Search.

(a) *Generalization of Clinical Trial Failures.*

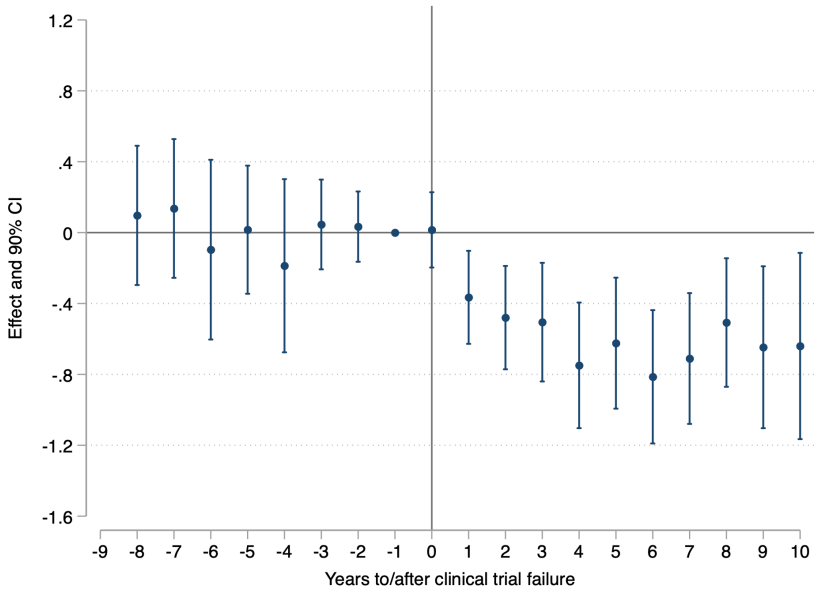


(b) *Genetic Neighborhood of BACE1.*



Note: The figure illustrates how failure in a focal clinical trial can reshape beliefs about related opportunities in the surrounding search space. Panel (a) provides a stylized representation using Merck's EPOCH trial on BACE1 inhibitors for Alzheimer's disease. The red shading indicates the strength of the negative signal transmitted to nearby opportunities as a function of their distance from BACE1, with darker shades denoting stronger spillovers. Panel (b) shows the empirical local neighborhood around BACE1. Nodes are positioned according to their proximity to BACE1 in the STRING protein-protein interaction network. Node color indicates each target's genetic potential for Alzheimer's disease, as measured by the Open Targets score, with green denoting higher potential and red denoting lower potential. See text for details.

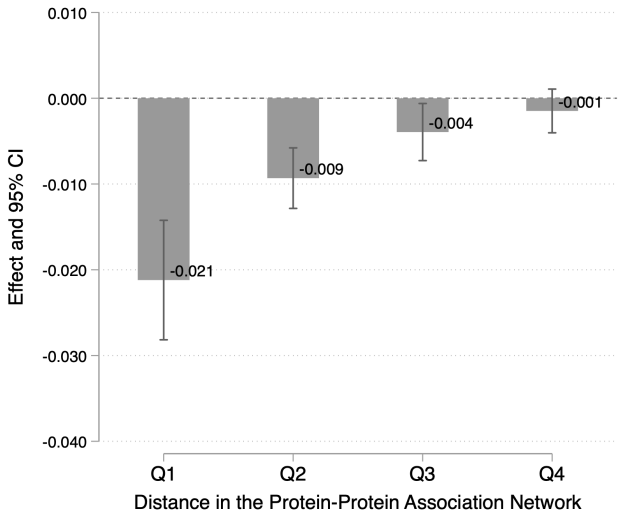
Figure 2: Patents on Focal Protein-Disease Pair Following a Clinical Failure.



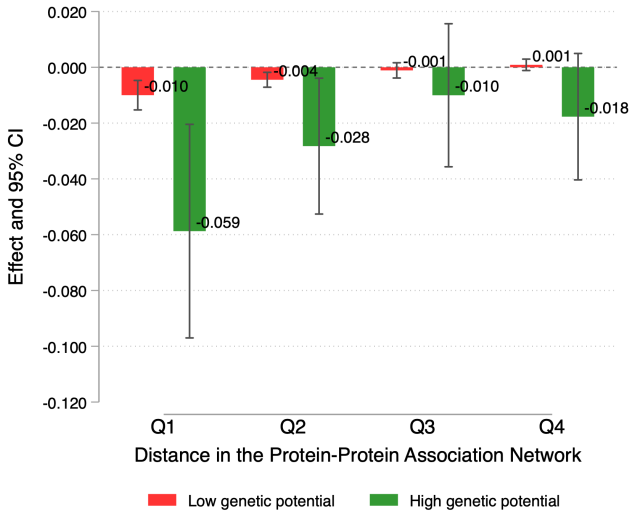
Note: The figure shows the event study coefficients estimated from the following panel OLS specification: $Y_{i,j,t} = \alpha + \sum_z \beta_{1,z} ClinicalTrial_{i,j} \times 1(z) + + \sum_z \beta_{2,z} ClinicalTrial_{i,j} \times Failure_{i,j} \times 1(z) + \gamma PD_{i,j} + \delta Protein_i + \omega Disease_j + \sigma Year_t + \epsilon_{i,j,t}$ The dependent variable is the number of USPTO patent applications for innovations focusing on a specific protein-disease combination $\langle i, j \rangle$ in a given year t. The chart plots values of the triple-interaction $\beta_{2,z}$ for different lags z before and after the failure of the first phase II or phase III clinical trial targeting the protein-disease pair. Regressions include protein, disease, year, and protein-disease combination fixed effects. Standard errors are clustered at the protein-disease level. See text for details.

Figure 3: Patents on Related Protein-Disease Pairs Following a Clinical Failure.

(a) *Generalization of Clinical Trial Failures.*



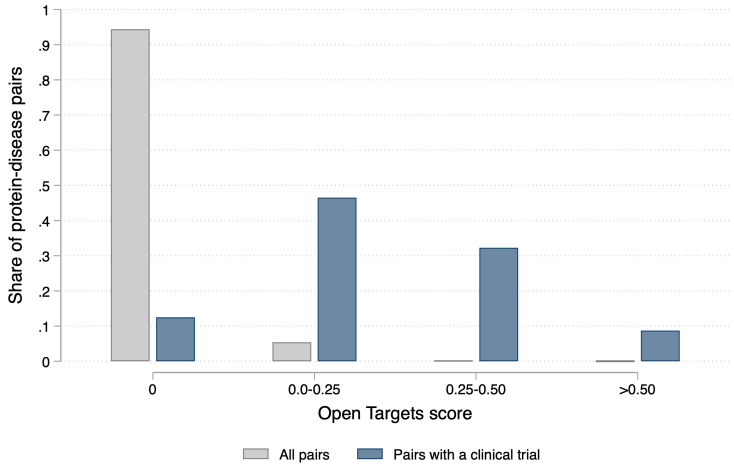
(b) *Generalization of Clinical Trial Failures by Genetic Potential.*



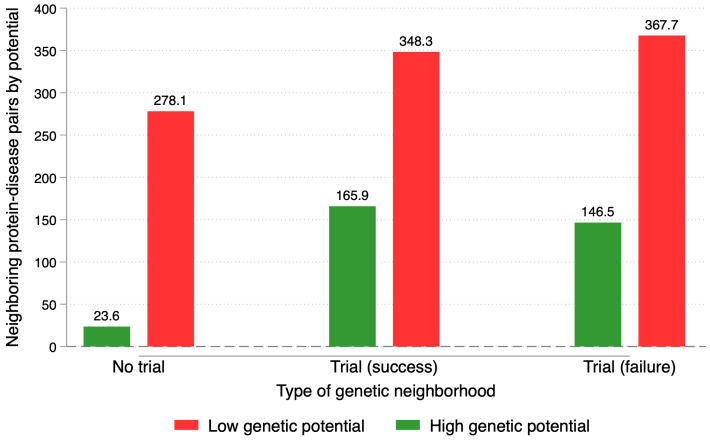
Note: The figure presents the spillover effects of clinical failures on drug targets functionally close to the failed target. Panel (a) shows the decrease in patenting for protein-disease pairs at varying quartiles of biological relatedness to the genetic targets of failed clinical trials. Panel (b) shows the decrease in patenting for protein-disease pairs at varying quartiles of biological relatedness to the genetic targets of failed clinical trials, reported separately for pairs with different genetic potential. Low-potential pairs are defined as those with an Open Targets score of zero, while high-potential pairs have a positive score. The regressions use standardized variables to enable comparison across split-sample regressions based on Equation 2 (i.e., bars represent beta coefficients). Standard errors are clustered at the protein-disease level. See text for details.

Figure 4: Descriptive Evidence on Endogenous Sampling and Spatial Correlation in Genetic Potential.

(a) *Distribution of targets by Open Targets score.*

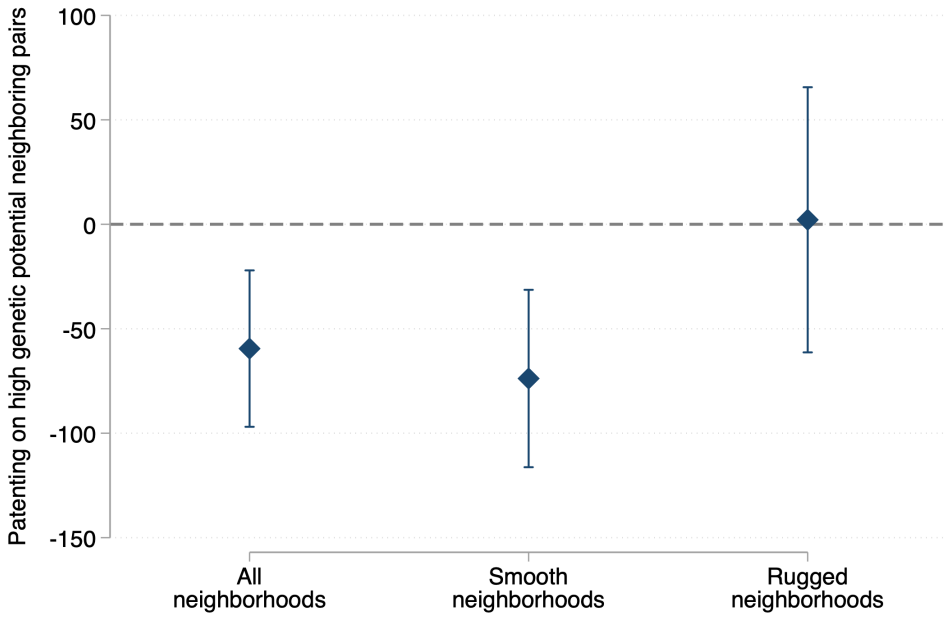


(b) *Average composition of genetic neighborhoods by type.*



Note: The figure provides descriptive evidence on two features of the empirical landscape that clarify the mechanism behind our results. Panel (a) compares the distribution of protein-disease pairs across Open Targets score bins for all pairs in the landscape and for the subset that receives a phase II or phase III clinical trial. Pairs with clinical activity are disproportionately concentrated in higher-score bins, consistent with endogenous sampling of more promising opportunities. Panel (b) reports the average composition of local genetic neighborhoods centered on protein-disease pairs that receive no clinical trial, a completed trial, or a failed trial. For each neighborhood type, the bars show the average number of neighboring pairs with high and low genetic potential, as measured by the Open Targets score. Neighborhoods surrounding trialed pairs, including those surrounding failed trials, contain more high-potential neighbors than those centered on pairs without clinical activity. This shows that failures are disproportionately realized in relatively promising regions of the search space. See text for details.

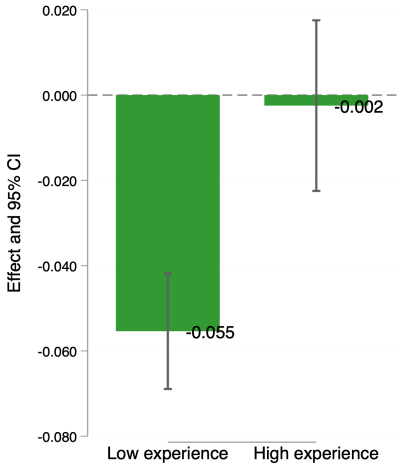
Figure 5: Heterogeneity in False Negatives by Landscape Ruggedness.



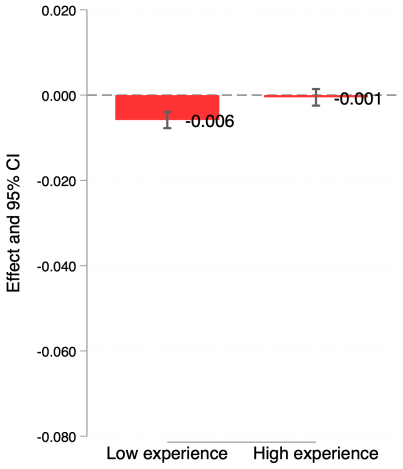
Note: The figure shows the aggregate effect of clinical trial failures on patenting in high genetic potential neighboring protein-disease pairs. The left coefficient reports the average effect across all neighborhoods, while the middle and right coefficients report the corresponding estimates for smooth and rugged neighborhoods, respectively. High-potential neighboring pairs are defined as those with a positive Open Targets score. Smooth and rugged neighborhoods are defined using a median split of the neighborhood-level fitness-distance correlation measure (Jones and Forrest, 1995). Bars indicate 95% confidence intervals. The corresponding regression coefficients are in Appendix Table C7. See text for details.

Figure 6: Firm Knowledge Base and Generalization from Clinical Trial Failures.

(a) *High-potential neighboring pairs.*



(b) *Low-potential neighboring pairs.*



Note: The figure presents heterogeneity in the extent to which firms generalize from clinical trial failures as a function of their prior knowledge. Panel (a) restricts the outcome to neighboring pairs with positive Open Targets scores, while Panel (b) restricts the outcome to neighboring pairs with Open Targets scores equal to zero. Firms with low experience are defined as those without prior publications on the relevant drug target, whereas firms with high experience are defined as those with prior publications on the relevant drug target. Bars report the triple-difference coefficient β_2 from Equation 2, estimated separately for firms with and without prior publications on the target. The regressions use standardized variables to enable comparison across split-sample regressions, so bars represent beta coefficients. Standard errors are clustered at the protein-disease level. See text for details.

Table 1: Descriptive statistics.

	Panel A: protein-disease pair cross-sectional descriptives					
	mean	median	st. d.	min	max	N
Total patent applications	1.597	0	19.106	0	13900	7,788,369
Ever received clinical trial	0.0009	0	0.030	0	1	7,788,369
Ever received terminated clinical trial	0.0002	0	0.0156	0	1	7,788,369
Ever spillovers from clinical trial	0.2930	0	0.455	0	1	7,788,369
Ever spillovers from terminated clinical trial	0.0617	0	0.241	0	1	7,788,369
Average Protein-Protein Distance	85.203	0	160.909	0	999	7,788,369
Average Open Targets score	0.0017	0	0.0187	0	0.908	7,788,369
	Panel B: protein-disease pair panel descriptives					
	mean	median	st. d.	min	max	N
Yearly patent applications	0.0840	0	1.310	0	1465	147,979,011
Yearly patent applications targeting high potential pairs	0.6310	0	4.287	0	1465	8,290,175
Yearly patent applications targeting low potential pairs	0.0520	0	0.842	0	1079	139,688,836
Post Clinical Trial (Direct)	0.0005	0	0.0216	0	1	147,979,011
Post Clinical Trial (Spillover)	0.1590	0	0.3660	0	1	147,979,011
Year	2010	2010	5.477	2001	2019	147,979,011
	Panel C: neighborhood-year panel descriptives					
	mean	median	st. d.	min	max	N
Yearly patent applications in neighborhood	46.149	4	192.205	0	38833	66,689,411
Yearly patent applications in high potential neighborhood pairs	23.136	1	124.211	0	36934	66,689,411
Yearly patent applications in low potential neighborhood pairs	23.012	2	101.654	0	10896	66,689,411
Post Clinical Trial	0.0007	0	0.0258	0	1	66,689,411
Year	2010	2010	5.477	2001	2019	66,689,411

Note: This table lists summary statistics at the protein-disease level for 7,788,369 pairs (Panel A), at the protein-disease-year level for a balanced panel of 147,979,011 observations (Panel B), and at the neighborhood-year level for 66,689,411 observations (Panel C). *Total patent applications*: count of USPTO patent applications for inventions targeting a given protein-disease pair; *Ever received clinical trial*: 0/1 = 1 for protein-disease pairs that have been directly targeted by a phase II or phase III clinical trial; *Ever received terminated clinical trial*: 0/1 = 1 for protein-disease pairs that have been directly targeted by a phase II or phase III clinical trial that has been terminated; *Ever spillovers from clinical trial*: 0/1 = 1 for protein-disease pairs genetically related to targets of a phase II or phase III clinical trial; *Ever spillovers from terminated clinical trial*: 0/1 = 1 for protein-disease pairs genetically related to targets of a phase II or phase III clinical trial that has been terminated; *Average Protein-Protein Distance*: average combined score of a protein-protein interaction in the STRING database; *Average Open Targets score*: average value of the Open Targets score; *Yearly patent applications*: count of yearly USPTO patent applications for inventions targeting a given protein-disease pair; *Yearly patent applications targeting high potential pairs*: count of yearly USPTO patent applications for inventions targeting a given protein-disease pair with an Open Targets score greater than zero; *Yearly patent applications targeting low potential pairs*: count of yearly USPTO patent applications for inventions targeting a given protein-disease pair with an Open Targets score equal to zero; *Post Clinical Trial (Direct)*: 0/1 = 1 in all years after conclusion of the first phase II or III clinical trial targeting a focal protein-disease pair; *Post Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase II or III clinical trial targeting a genetically related protein-disease pair; *Yearly patent applications in neighborhood*: yearly count of USPTO patent applications in the neighborhood of the focal protein-disease pair; *Yearly patent applications in high potential neighborhood pairs*: yearly count of USPTO patent applications in neighboring protein-disease pairs with Open Targets score greater than zero; *Yearly patent applications in low potential neighborhood pairs*: yearly count of USPTO patent applications in neighboring protein-disease pairs with Open Targets score equal to zero; *Post Clinical Trial*: 0/1 = 1 in all years after conclusion of the focal clinical trial for a given neighborhood; *Year*: average year of observations in the panel.

Table 2: Direct and Spillover Effects of Clinical Trial Failures on Pharmaceutical Firm Patenting.

	Firm Patents Targeting a Protein-Disease Pair			
	(1)	(2)	(3)	(4)
Post × Clinical Trial (Direct)	0.980***	1.110***		
	(0.123)	(0.000650)		
... × Failure		-0.459**		
		(0.164)		
Post × Clinical Trial (Spillover)			0.0269***	0.0302***
			(0.000630)	(0.000713)
... × Failure				-0.0146***
				(0.00147)
Protein-disease FE	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	147,979,011	147,979,011	147,845,878	147,845,878

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. The sample in columns (1) and (2) includes all protein-disease pairs, while columns (3) and (4) exclude protein-disease pairs that directly received a clinical trial. *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Direct)*: 0/1 = 1 in all years after conclusion of the first phase II or III clinical trial targeting a focal protein-disease pair; *Post × Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase II or III clinical trial targeting a genetically related protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

Table 3: Neighborhood-Level Effects of Clinical Trial Failures on Pharmaceutical Firm Patenting.

	Firm Patents in Neighboring Protein-Disease Pairs		
	All Pairs	High Genetic Potential Pairs	Low Genetic Potential Pairs
	(1)	(2)	(3)
Post × Clinical Trial	288.2***	207.9***	80.28***
	(13.47)	(11.18)	(5.623)
$\cdots \times$ Failure	-85.83***	-59.50**	-26.34**
	(22.90)	(19.10)	(9.500)
Neighborhood FE	YES	YES	YES
Gene FE	YES	YES	YES
Disease FE	YES	YES	YES
Year FE	YES	YES	YES
Observations	66,689,411	66,689,411	66,689,411

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the neighborhood-year level. Std. err. clustered at the neighborhood level. All models include neighborhood, gene, disease, and year fixed effects. Column (1) uses total patenting in the neighborhood as the dependent variable, column (2) uses patenting on high genetic potential neighboring proteins (OT > 0), and column (3) uses patenting on low genetic potential neighboring proteins (OT $= 0$). *Firm Patents in Neighboring Protein-Disease Pairs*: yearly count of USPTO patent applications in the neighborhood of the focal protein-disease pair; *Post*: 0/1 = 1 in all years after conclusion of the focal clinical trial; *Failure*: 0/1 = 1 for focal clinical trials that are terminated before their natural completion. See text for details.

Learning About Roads Not Taken
Appendix

A. ANALYTICAL INTUITION FOR THE HYPOTHESES

This appendix formalizes the theoretical framework verbally developed in Section 2. Its purpose is to make explicit the logic underlying the main predictions that we test empirically. We first derive Hypothesis 1 by showing that generalizing from failure produces larger downward revisions for alternatives with stronger prior beliefs. We then show that, under endogenous sampling and local correlation, failures are disproportionately observed in promising neighborhoods, providing the logic for Hypothesis 2. Finally, we show that this distortion weakens as landscapes become more rugged, corresponding to Hypothesis 3.

Let $i, j \in I$ index innovation opportunities, where:

μ_i : the true, underlying potential of opportunity i,
q_i : belief as to the potential of opportunity i,
d_{ij} : distance between opportunities i and j, with $d_{ij} \geq 0$,
$\kappa(d_{ij})$: relatedness kernel, with $\kappa(d_{ij}) \geq 0$ and $\kappa'(d_{ij}) < 0$.

Suppose that a failure occurs at the focal opportunity i.

A1. Derivation of Hypothesis 1

Assume that when a failure occurs at opportunity i, firms revise their beliefs about a related opportunity j according to:

$$q'_j = q_j - \phi\kappa(d_{ij})s(q_j).$$

Where:

ϕ : learning (belief updating) rate, with $\phi > 0$,
$s(q_j)$: magnitude of negative surprise, with $s(q_j) \geq 0$ and $s'(q_j) > 0$.

We assume that $s'(q_j) > 0$, so the same failure induces a larger downward revision for alternatives that were initially viewed as more promising. This assumption is consistent with standard reinforcement learning models, where prediction error is equal to performance feedback (F) minus the prior belief (q_j). When the realized feedback is failure, $F = 0$, the magnitude of the negative prediction error increases in q_j. Thus, stronger prior beliefs generate larger negative surprises and larger belief reductions for related opportunities.

Hypothesis 1 (Asymmetric updating). *Holding relatedness $\kappa(d_{ij})$ constant, high-potential alternatives experience a larger decline in beliefs following a failure.*

Analytical derivation. Because the event is a failure, the belief update for unsampled alternatives is a downward revision. We define the reduction in beliefs for alternative j as Δq_j:

$$\Delta q_j = q_j - q_j' = \phi\kappa(d_{ij})s(q_j).$$

For alternatives with positive relatedness, $\kappa(d_{ij}) > 0$,

$$\frac{\partial \Delta q_j}{\partial q_j} = \phi\kappa(d_{ij})s'(q_j) > 0.$$

Thus, among alternatives that are equally related to the failed opportunity i, those with stronger prior beliefs experience larger belief reductions and, by extension, larger declines in innovation activity.

Note. The belief updating rule above can be understood as a reduced-form failure case of a standard reinforcement learning rule (Bush and Mosteller, 1955), where

$$q_j' = q_j + \phi\kappa(d_{ij})(F - q_j),$$

where $(F - q_j)$ is the signed prediction error. When outcomes are binary (success $= 1$ and failure $= 0$), then for failure, this reduces to:

$$q_j' = q_j - \phi\kappa(d_{ij})q_j.$$

A2. Derivation of Hypothesis 2

Because experimentation is costly and firms have limited resources, firms do not sample opportunities at random. Instead, they allocate experiments to alternatives that appear more promising ex ante, which we capture with the following assumption:

Assumption 1 (Endogenous Sampling). The opportunities that firms choose to test are, on average, of higher true potential than the typical opportunity in the landscape:

$$\mathbb{E}[\mu_i \mid \text{tested}] > \mathbb{E}[\mu_i].$$

We next assume that the opportunities surrounding a focal alternative are not independent of that alternative's potential. When the landscape is locally correlated, nearby opportunities tend to share underlying attributes and therefore have more similar values. As a result, a high-potential opportunity is more likely to be embedded in a high-potential neighborhood:

Assumption 2 (Local Correlation). The expected quality of the neighborhood surrounding alternative i (N_i) is increasing with its potential μ_i:

$$\mathbb{E}[N_i \mid \mu_i] \text{ is increasing in } \mu_i.$$

Hypothesis 2 (Selection into promising neighborhoods). *Failures are disproportionately realized in high-potential neighborhoods.*

Analytical derivation. Let F_i denote that alternative i was tested and failed. Because tested alternatives are, on average, of higher true potential (by Assumption 1), the distribution of trialed alternatives is skewed toward higher-μ_i positions. Since failure occurs with positive probability across the support and does not perfectly offset the selection of higher-potential opportunities into testing, this selection carries through to realized failures. This implies that:

$$\mathbb{E}[\mu_i \mid F_i] > \mathbb{E}[\mu_i].$$

By local correlation in the true underlying potential (Assumption 2),

$$\mathbb{E}[N_i \mid \mu_i] \text{ is increasing in } \mu_i.$$

Hence,

$$\mathbb{E}[N_i \mid F_i] > \mathbb{E}[N_i].$$

Thus, failures are observed in neighborhoods whose true potential is above the landscape average. When firms generalize from these failures, they reduce beliefs about nearby alternatives that are themselves disproportionately likely to be valuable. The aggregate consequence is not merely a broad reduction in related search, but a disproportionate withdrawal from high-potential related opportunities relative to low-potential ones.

Note. H2 is conceptually distinct from H1 and is not simply an "aggregation effect". H2, as shown in the preceding proof, has no reliance on $s(q_j)$. However, if H1 is true, it amplifies the effects of H2. To connect Hypothesis 1 and Hypothesis 2, assume that within neighborhoods, ex ante perceived potential is positively associated with true potential.

Per A1, let the reduction in beliefs for alternative j following failure with focal alternative i be:

$$\Delta q_j = q_j - q_j' = \phi\kappa(d_{ij})s(q_j).$$

Under potential-invariant updating, $s(q_j)$ is constant across nearby alternatives at a given level of relatedness, so the reduction in beliefs among nearby alternatives does not vary with their perceived potential. In that case, the distortion identified in H2 arises solely because failures are disproportionately realized in high-potential neighborhoods.

If H1 also holds such that $s'(q_j) > 0$, then holding relatedness fixed, alternatives with higher ex ante perceived potential experience larger reductions in beliefs following failure. Formally, for any two nearby alternatives j and j' where $d_{ij} = d_{ij'}$ and $q_j > q_{j'}$, then $\Delta q_j > \Delta q_{j'}$. By Hypothesis 2, failures disproportionately occur in neighborhoods with above-average true potential. It follows that the neighborhoods where failures are disproportionately realized also contain a disproportionately greater number of alternatives with high perceived potential. Since such alternatives receive larger downward revisions under H1, the average reduction in beliefs within those neighborhoods is greater than it would be under potential-invariant updating.

Taken together, H1 is not necessary for the distortion identified in H2, but it can amplify it by intensifying the suppression of valuable opportunities within the neighborhoods where failures occur.

A3. Derivation of Hypothesis 3

Let the ruggedness of the opportunity landscape (i.e., the strength of local correlation in the true, underlying potential of alternatives) vary across settings. In more rugged landscapes, nearby opportunities are less similar in value; in smoother landscapes, nearby opportunities are more similar.

Formally, let $r > 0$ index ruggedness such that the relationship between the true potential of a focal opportunity and the quality of its neighborhood weakens as ruggedness increases:

$$\frac{\partial \mathbb{E}[N_i \mid \mu_i]}{\partial \mu_i} \text{ is decreasing in } r,$$

where μ_i is the true, underlying potential of alternative i and N_i denotes the true quality of the surrounding neighborhood. Higher r corresponds to more rugged landscapes, whereas lower r corresponds to smoother landscapes.

Hypothesis 3 (Attenuation under ruggedness). *The concentration of failures in high-quality neighborhoods is weaker in more rugged landscapes.*

Analytical derivation. From Hypothesis 2, failures are disproportionately realized for alternatives with higher true potential. This implies that:

$$\mathbb{E}[\mu_i \mid F_i] > \mathbb{E}[\mu_i].$$

By local correlation,

$$\mathbb{E}[N_i \mid \mu_i] \text{ is increasing in } \mu_i.$$

As a result, failures occur in neighborhoods with above-average true potential:

$$\mathbb{E}[N_i \mid F_i] > \mathbb{E}[N_i].$$

The magnitude of this difference depends critically on how strongly neighborhood quality co-varies with the true potential of the focal alternative. By assumption, the sensitivity of $\mathbb{E}[N_i \mid \mu_i]$ to μ_i is decreasing in r. Therefore, as ruggedness increases, the difference between $\mathbb{E}[N_i \mid F_i]$ and $\mathbb{E}[N_i]$ declines. Conversely, as the landscape becomes smoother (more locally correlated), this difference becomes larger.

Thus, the concentration of failures in high-quality neighborhoods and the resulting distortion of innovation efforts due to generalization are attenuated in more rugged landscapes.

A4. Empirical Counterparts

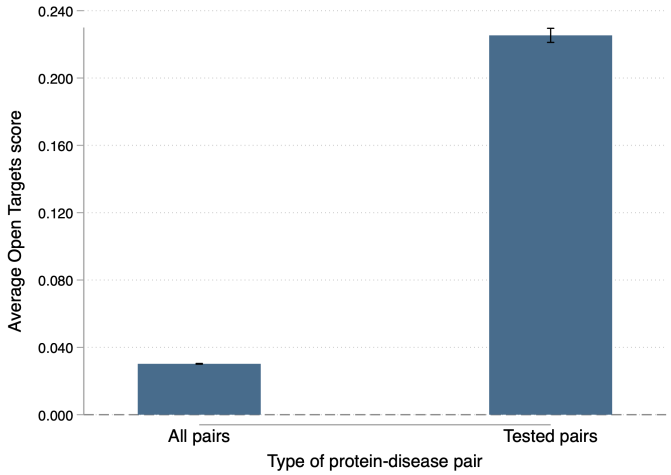
Having provided the formal intuition underpinning the hypotheses developed in the main text, we now provide empirical evidence supporting the key assumptions and theoretical predictions.

First, we provide empirical support for Assumption 1 (endogenous sampling). Specifically, if firms choose to test alternatives that they perceive to be promising, then:

$$\mathbb{E}[\mu_i \mid \text{tested}] > \mathbb{E}[\mu_i].$$

To assess this condition in our empirical context, we compare the average Open Targets score for protein-disease pairs that receive a clinical trial with that of all protein-disease pairs. If the endogenous sampling assumption holds, the average Open Targets score should be higher among trialed pairs. Figure A.1 presents this comparison. Consistent with the assumption, the mean Open Targets score conditional on receiving a trial is 0.22, whereas the average score across all protein-disease pairs is 0.03. The magnitude of this gap indicates that clinical experimentation is concentrated among protein-disease pairs that appear substantially more promising ex ante.

Figure A.1: Endogenous Sampling of Promising Protein-Disease Pairs



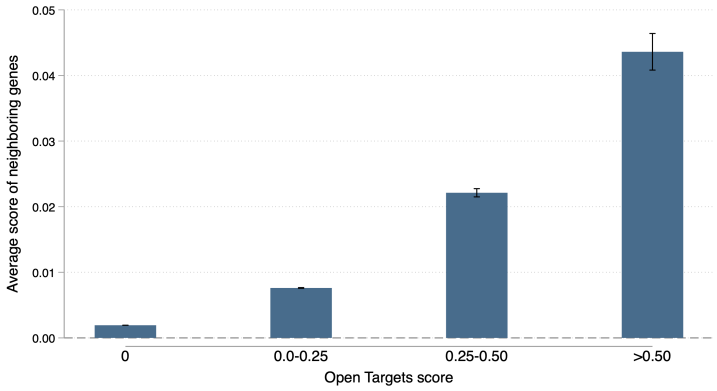
We next assess Assumption 2, which postulates local correlation in the opportunity landscape. If nearby opportunities tend to have similar underlying potential, then higher-potential focal alter-

natives should be surrounded by higher-potential neighborhoods:

$$\mathbb{E}[N_i \mid \mu_i] \text{ is increasing in } \mu_i.$$

We test the validity of this assumption in Figure A.2. The figure shows the relationship between the Open Targets score of each focal protein-disease pair and the average Open Targets score of nearby proteins, where proximity is defined by biological distance in the STRING protein-protein interaction network. The pattern is strongly monotonic. Focal targets with higher genetic support are located in neighborhoods whose neighboring proteins also have higher average Open Targets scores. For example, when the Open Targets score of the focal alternative is 0, the average score of neighboring proteins is 0.002. By contrast, when the focal score exceeds 0.5, the average score of neighboring proteins rises to 0.044. This pattern provides strong support for local correlation in the fitness landscape.

Figure A.2: Spatial Clustering of Promising Protein-Disease Pairs



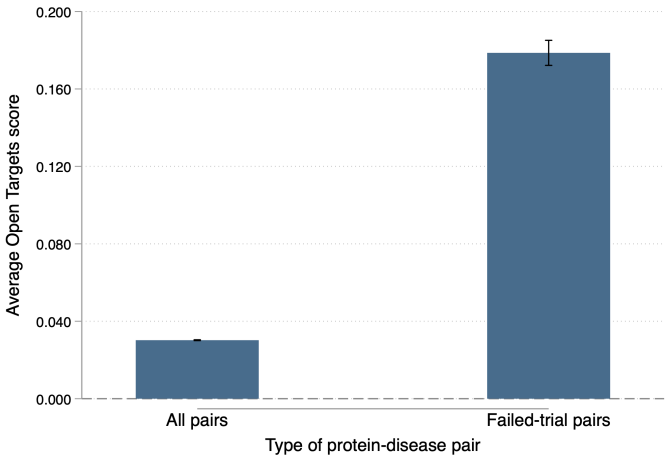
Third, Hypothesis 2 relies on the implication of Assumption 1 that clinical trial failures are drawn from a selected set of alternatives. Because firms choose which alternatives to test, failed trials should occur among alternatives whose expected value exceeds that of the full set of protein-disease pairs, including those that never enter clinical trials. Formally, this implies:

$$\mathbb{E}[\mu_i \mid F_i] > \mathbb{E}[\mu_i].$$

We assess this implication empirically by comparing the average Open Targets score of protein-disease pairs that experience a clinical trial failure with the average score across all protein-disease pairs. The results, reported in Figure A.3, strongly support this prediction. Protein-disease pairs that experience a failed clinical trial have an average Open Targets score of 0.17, only slightly below the average score of pairs that receive successful trials. By contrast, the average Open Targets score across all protein-disease pairs is much lower, equal to 0.03 on average.

Fourth, the preceding result has a neighborhood-level implication. If failed trials are drawn from higher-potential pairs, and if biological neighborhoods exhibit local correlation, then failures should also be concentrated in neighborhoods with above-average potential. Let N_i denote the average Open Targets score of the protein-disease pairs in the neighborhood around pair i.

Figure A.3: Clinical Trial Failures Are Selected from High-Potential Protein-Disease Pairs

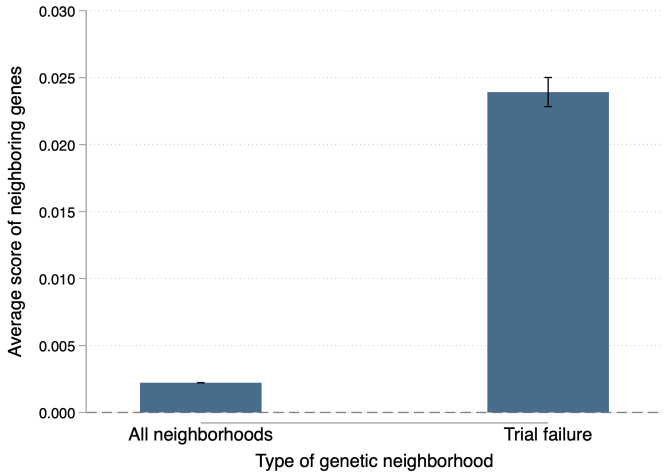


Hypothesis 2 implies:

$$\mathbb{E}[N_i \mid F_i] > \mathbb{E}[N_i].$$

We assess this implication in Figure A.4 by comparing the average potential of neighborhoods that experience a clinical trial failure with the average potential of all neighborhoods, regardless of trial status. The evidence strongly supports the mechanism. Neighborhoods around failed clinical trials have an average Open Targets score of 0.024, compared with 0.002 across all neighborhoods. Thus, failures are not only selected from above-average focal alternatives; they are also located in above-average neighborhoods. This pattern is central to Hypothesis 2: because failures occur in regions where nearby opportunities are relatively promising, generalizing from a failed trial can redirect investment away from precisely the neighborhoods in which valuable alternatives are more likely to be found.

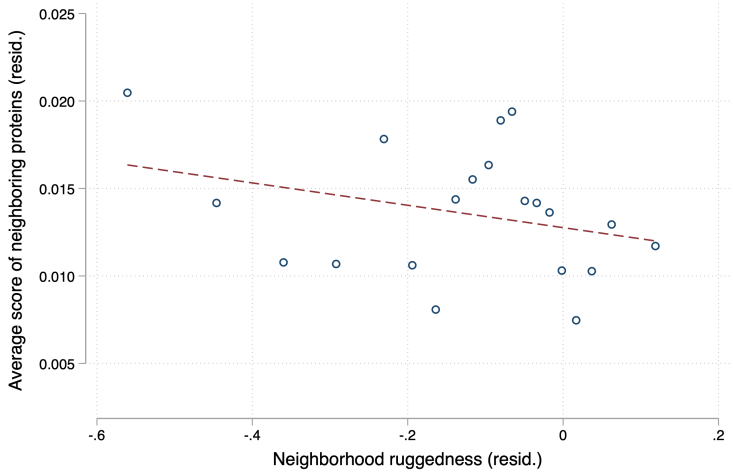
Figure A.4: Clinical Trial Failures Are Concentrated in Above-Average Neighborhoods



Finally, we examine the key empirical implication of Hypothesis 3. If ruggedness weakens local correlations in the fitness landscape, then the selection of failures into high-potential neighborhoods should diminish as ruggedness increases. Put differently, among protein-disease pairs that experience a clinical trial failure, the average potential of neighboring proteins should be lower in more rugged parts of the landscape.

We test this implication in Figure A.5. We first restrict the sample to protein-disease pairs with defined neighborhoods and non-missing measures of neighborhood potential and ruggedness. We then residualize both the average Open Targets score of neighboring proteins and the ruggedness measure with respect to protein and disease fixed effects. The figure plots the residualized neighborhood score against residualized ruggedness for neighborhoods centered on protein-disease pairs that experience a clinical trial failure. The pattern shows that failed-trial neighborhoods are less promising in more rugged regions of the landscape, consistent with the claim that ruggedness attenuates the selection of failures into high-potential neighborhoods.

Figure A.5: Failed-Trial Neighborhoods Are Less Promising in More Rugged Landscapes



B. MEASUREMENT DETAILS

Pharmaceutical Firms Search in a Protein-Disease Landscape:

Our measurement strategy builds on the idea that firms search for valuable drug targets within a vast landscape of protein–disease combinations. In our data, this landscape spans 16,136 human proteins and 483 diseases, generating more than 7.7 million possible pairs. The number of proteins is slightly below the approximately 19,000 found in the human body because we restrict attention to those mentioned at least once in a patent; however, results are robust to including proteins with no recorded R&D activity. Each pair represents a potential direction for firms' R&D, and the distances among them define the topology of the search landscape faced by firms.

Mapping clinical trial outcomes onto this landscape is the first step in our analysis. A trial that tests a specific protein–disease pair corresponds to an experiment at a single position in the landscape. The outcome of that experiment provides feedback not only about the tested pair but also about adjacent possibilities. This structure allows us to examine both direct learning on the focal protein–disease pair and spillovers to related pairs at varying biological distances. Panel (a) of Figure 1 provides a visual illustration. Empirically, we link trials to the landscape using information on the disease conditions studied in each trial, uniquely identified by MeSH IDs, and the proteins targeted by the intervention, identified by NCBI's Gene IDs.[7] Previous studies have used a similar approach to study Phase I trials (Kang, 2025); here, we focus on Phases II and III, where data coverage and reliability are substantially higher (Kao, 2025b).

We then trace firm investments using USPTO patent applications. In collaboration with the European Bioinformatics Institute, we use data compiled with SciBite's TERMite software, which extracts biological entities from full patent texts and links them to standardized identifiers (Gene IDs for proteins and MeSH terms for diseases). TERMite is a proprietary tool specifically designed for disambiguating biomedical text. These data, previously used and validated by Tranchero (2026), have been shown to be highly accurate. This approach allows us to position each patent application within the same landscape as the clinical trials. In turn, we can observe whether a firm patents the exact protein–disease pair tested in a trial, or a nearby pair in the landscape. Because project-level R&D spending data are rarely available, patent applications provide a real-time proxy for where firms allocate resources early in the innovation process.

Distance in the Protein-Disease Landscape:

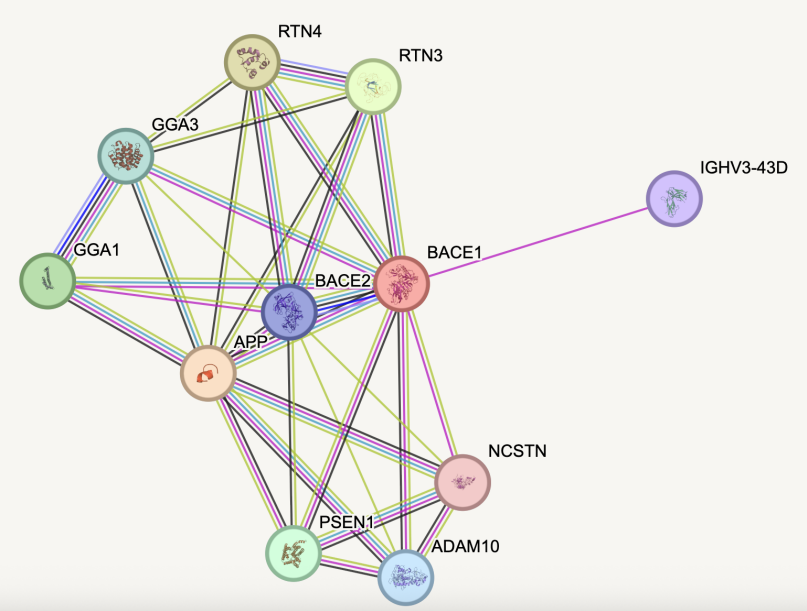
To empirically study generalization, we need a measure of distance between genetic targets. Our theory predicts that firms generalize feedback from one alternative to others in proportion to their similarity, with the effect decaying as the distance between alternatives increases. Capturing this mechanism in the data requires a biologically meaningful, continuously varying measure of protein proximity rather than relying on discrete categories or binary overlaps.

We measure this proximity using STRING (`https://string-db.org/`), a unique database that integrates known and predicted protein–protein associations from five evidence channels: genomic context predictions, high-throughput experiments, conserved co-expression, automated

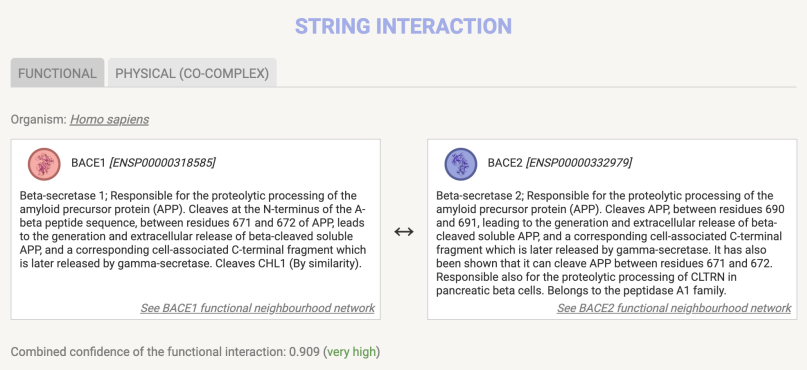
[7] We leverage the fact that proteins are coded by the homologous gene and thus use unique Gene IDs for our data matching.

Figure B1: Examples of Functional Distance between Human Proteins in the STRING Database.

(a) *Network of protein-protein interactions for the BACE1 protein*



(b) *Functional distance between the BACE1 and BACE2 proteins*



Note: Panel (a) plots the protein-protein associations network centered on the BACE1 protein according to the data in STRING. Panel (b) reports the details of the interaction between BACE1 and BACE2. See text for details.

text mining, and curated pathway databases (Szklarczyk et al., 2025). STRING combines these sources into a single confidence score for each protein-protein link, which we use as a continuous measure of functional relatedness. Higher scores denote closer proteins in the biological landscape and thus greater potential for knowledge spillovers. Our conversations with chemistry researchers confirmed that STRING is widely used in both medicinal chemistry and molecular biology. In drug discovery, researchers use it to interpret large genomic data after perturbing a target and to identify affected pathways. In experimental biology, it helps verify whether observed protein interactions align with established functional links. These applications confirm that STRING captures the kind of biological proximity along which firms are likely to generalize feedback.

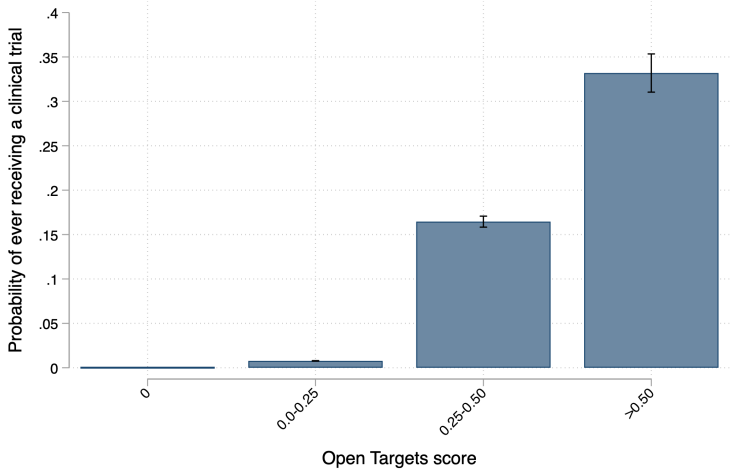
Figure B1 shows an example of how STRING captures functional proximity among proteins.

Panel (a) shows the network of associations centered on the BACE1 protein, while Panel (b) highlights its strong functional link with BACE2. As their names suggest, BACE1 and BACE2 are homologous proteins belonging to the same gene family, formed by duplication of a common ancestral gene and sharing similar biochemical functions. BACE2 shares approximately 64% amino acid sequence similarity with BACE1 and participates in the same biological processes (Yeap et al., 2023). Building on this example, we validate our measure by verifying that proteins from the same family exhibit higher STRING proximity scores, on average 14% greater than those of unrelated proteins. Importantly, STRING captures functional rather than purely structural similarity, which is appropriate for our context since proteins can be structurally similar yet play distinct biological roles, just as structurally dissimilar proteins can serve as functional substitutes.

Fitness of Protein-Disease Alternatives:

We measure the "fitness" of each protein-disease pair using the Open Targets score, an evidence-based measure of the strength of support linking a protein to a disease (Buniello et al., 2025). The Open Targets Platform is a public-private partnership that integrates publicly available evidence on target-disease associations from multiple sources, including genetic associations, animal models, and literature-mined evidence. Each source of evidence is weighted according to a scoring framework, and the resulting values are harmonized to standardized identifiers for proteins (Gene IDs) and diseases (MeSH terms), allowing us to merge Open Targets scores with the protein-disease pairs in our empirical landscape. To address the concern that the overall score may partly reflect clinical or commercial attention, we use a version of the Open Targets score that excludes evidence sources directly tied to prior drug development, including clinical trials.

Figure B2: Probability of receiving a clinical trial by OT score.



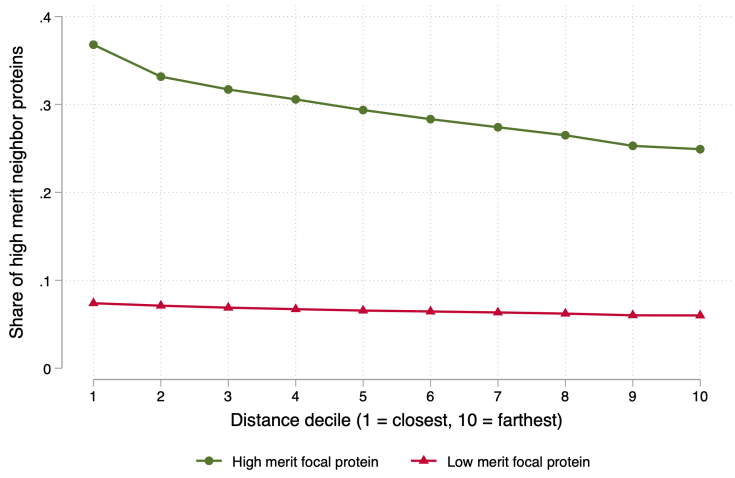
For our purposes, the Open Targets score provides a systematic proxy for the scientific promise of a protein-disease pair. We use this measure because it summarizes the publicly available evidence that a given protein is biologically relevant for a given disease and because prior work shows that genetic evidence is predictive of clinical success (Razuvayevskaya et al., 2024). We interpret the score as a proxy for the therapeutic promise of each protein-disease pair. Consis-

tent with this interpretation, Figure B2 shows that protein-disease pairs with higher Open Targets scores are substantially more likely to receive clinical trials. This pattern helps validate the score as a proxy of fitness in the search landscape.

Panel (b) of Figure 1 illustrates how our empirical setting maps into a fitness landscape, using BACE1 and its neighboring proteins for Alzheimer's disease as an example. We construct biological distance from STRING protein-protein interaction scores. Relative to BACE1, selected neighboring proteins have the following STRING scores: BACE2 = 0.909, NCSTN = 0.878, SORL1 = 0.737, NAV1 = 0.666, and FYN = 0.438. These proteins differ substantially in their Open Targets scores for Alzheimer's disease: BACE2 = 0.088, FYN = 0.126, NCSTN = 0, NAV1 = 0.004, and SORL1 = 0.661. Thus, while related proteins tend to contain valuable nearby alternatives, local neighborhoods also remain heterogeneous, underscoring the ruggedness of the pharmaceutical search landscape.

Figure B3 shows this pattern more systematically. The figure plots, by STRING distance decile, the share of neighboring proteins with high Open Targets scores, separately for high-merit and low-merit focal proteins. High-merit focal proteins have substantially more high-merit neighbors than low-merit focal proteins at every distance decile. This difference is largest among the closest neighbors and declines with distance, consistent with local correlation in the fitness landscape. At the same time, the decline is quite gradual, indicating that biological relatedness creates informative neighborhoods without eliminating substantial local heterogeneity.

Figure B3: More Promising Proteins Have More Promising Neighbors



Note: The figure plots the share of neighboring proteins classified as high merit across STRING distance deciles, separately for high-merit and low-merit focal proteins. Distance deciles are constructed from STRING protein-protein interaction scores, with 1 denoting the closest neighbors and 10 denoting the farthest neighbors. High merit is defined using the Open Targets score for the corresponding protein-disease pair.

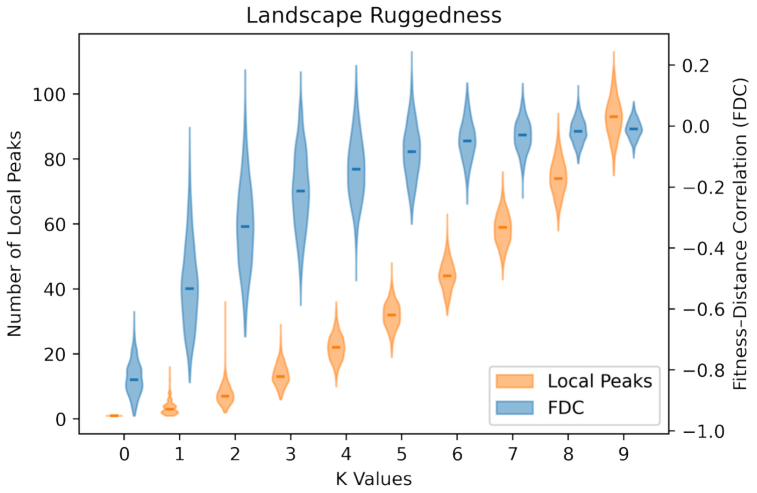
Fitness-Distance Correlation as a Proxy of Ruggedness:

In the main analysis, we construct a measure of neighborhood ruggedness (or, equivalently, the degree of spatial correlation across alternatives). Following Jones and Forrest (1995), we opera-

tionalize this construct using the fitness-distance correlation (FDC). Specifically, for each tested protein-disease pair, we calculate the correlation between the Open Targets scores of neighboring alternatives and their distance from the focal target in the STRING network. To illustrate this measure, consider the case of BACE1 and neighboring proteins such as BACE2 and SORL1. We first calculate the distance between each neighboring target and BACE1 using STRING similarity scores. For example, BACE2 and SORL1 have STRING similarity scores of 0.909 and 0.737, respectively. We convert these similarity scores into distances by subtracting each value from one, yielding distances of 0.091 and 0.263. We then compile the Open Targets scores for each neighboring protein. In Alzheimer's disease, the Open Targets scores for BACE2 and SORL1 are 0.088 and 0.661, respectively. Finally, we calculate the correlation between the distances of neighboring targets from BACE1 and their corresponding Open Targets scores. This correlation constitutes the fitness-distance correlation. Negative values indicate that nearby alternatives tend to have similar fitness and therefore suggest a relatively smooth neighborhood, whereas values closer to zero (or positive) indicate weaker spatial correlation and a more rugged local landscape.

To further assess the validity of this measure, we conducted a series of computational experiments using NK landscapes. Specifically, we generated a population of landscapes and examined how the fitness-distance correlation varied with ruggedness (K). For simplicity, we set the number of decisions (N) equal to 10 and, for each value of K between 0 (a maximally smooth, single-peak landscape) and 9 (a maximally rugged landscape), calculated the fitness-distance correlation relative to the global optimum. The results are reported in Figure B4, which plots the distributions of both the number of local peaks and the corresponding fitness-distance correlations across landscapes for each value of K.

Figure B4: Fitness Distance Correlation in an NK Landscape



As expected, increasing K leads to a monotonic increase in landscape ruggedness, as reflected by the number of local peaks. Importantly, the fitness-distance correlation exhibits a corresponding monotonic pattern. In maximally smooth landscapes (K = 0), FDC values are strongly negative, indicating that fitness declines predictably with distance from the global optimum. By contrast, in maximally rugged landscapes (K = 9), FDC values converge toward zero, indicating little systematic relationship between fitness and distance. Taken together, these results demonstrate

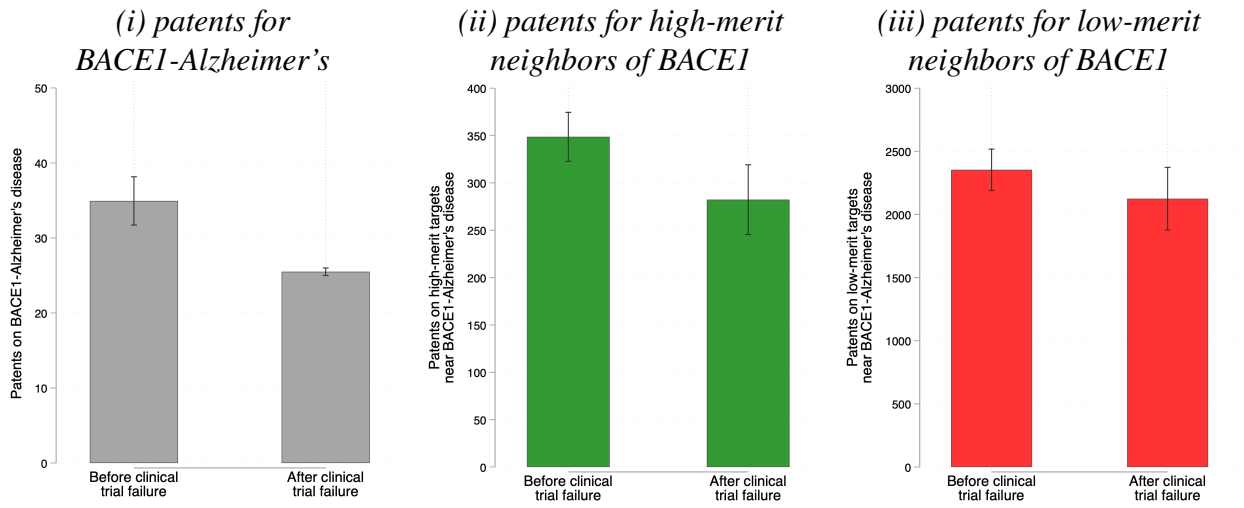
that the fitness-distance correlation provides a valid and intuitive measure of local ruggedness and spatial correlation across alternatives.

BACE1 Case Study:

The BACE1-Alzheimer's example provides a concrete illustration of how a clinical trial failure can reshape search on both the focal protein-disease pair and its surrounding neighborhood. BACE1 was a prominent Alzheimer's disease target, and Merck's drug (called verubecestat) was among the most advanced BACE inhibitor efforts. The EPOCH trial tested verubecestat in patients with mild-to-moderate Alzheimer's disease and was terminated after an external monitoring committee concluded that the study was unlikely to show a positive clinical effect. The event produced a salient negative signal about a focal position in the Alzheimer's search landscape.

The informativeness of this signal, however, was not confined to BACE1 itself. As discussed in the main text, BACE1 is embedded in a neighborhood of functionally related proteins (Figure 1). Some of these neighbors are close to BACE1 and have meaningful Open Targets scores for Alzheimer's disease; others are biologically proximate but have lower measured therapeutic potential. This combination of functional relatedness and heterogeneous merit makes the case useful for illustrating the central tension in our argument. A failed trial on BACE1 may reasonably reduce confidence in nearby targets, but it may also extend negative feedback to neighboring opportunities whose underlying promise remains relatively high.

Figure B5: BACE1 example: focal and neighboring patenting around trial failure.



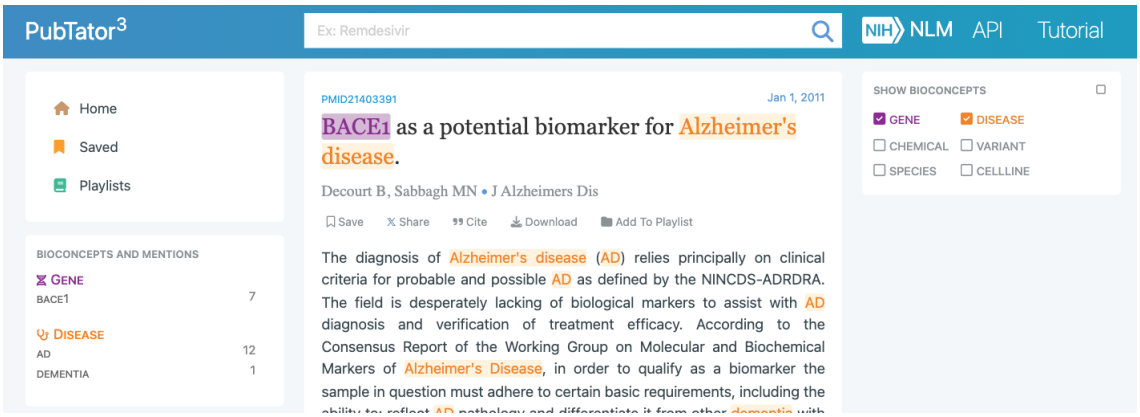
Note: This figure provides a simple pre-post comparison of patenting around the BACE1 clinical trial failure. Panel (i) reports the count of patent applications targeting the BACE1-Alzheimer's pair. Panels (ii) and (iii) report patent applications for all neighboring protein-disease pairs, separately for high-merit and low-merit neighbors of BACE1. Note that the total volume of patenting differs because in the neighborhood of BACE1, there are many more low-potential targets. The pre-period averages patenting over the years before the early-2017 failure, while the post-period averages patenting over the years after the failure. High-merit neighbors are defined using the Open Targets score. Error bars report 95% confidence intervals. See text for details.

Figure B5 provides descriptive evidence consistent with this logic. Panel (i) shows a clear decline in patenting on the focal BACE1-Alzheimer's pair after the trial failure, consistent with Krieger (2021). Panels (ii) and (iii) examine patenting in BACE1's neighborhood, distinguishing between high-merit and low-merit neighbors. Patenting on high-merit neighbors declines significantly after the failure, whereas patenting on low-merit neighbors remains statistically unchanged. Although this figure is only a descriptive case study, the pattern illustrates the mechanism developed in the paper. Negative feedback from a focal clinical trial can travel through the biological landscape. However, when the surrounding neighborhood contains promising targets, this form of generalization can turn a locally informative failure into withdrawal from valuable nearby opportunities.

Firm Heterogeneity:

We use firm-level publication histories to measure heterogeneity in how organizations generalize from failure. The objective is to distinguish firms with direct, target-specific knowledge from those that rely more heavily on inference across related targets. We operationalize this knowledge base at the firm-protein level, measuring whether a firm has prior scientific experience with the protein targeted by a patent application. Publication data come from the NIH's PubTator3 (Wei et al., 2024). Each publication is tagged with the same disease and protein identifiers mentioned above, as illustrated in Figure B6. We link these publications to firms using affiliation data from Dimensions, allowing us to capture each firm's accumulated expertise with specific drug targets prior to patenting. Combining this information with our clinical and patent data allows us to examine how a firm's knowledge base shapes its response to trial outcomes. Any measurement error in this linkage would make it harder to detect systematic heterogeneity, making our estimates conservative.

Figure B6: PubTator3 Data Extract the Proteins and Diseases Studied in Each Published Paper.



Note: The figure shows how PubTator3 annotated scientific articles by extracting the proteins and diseases mentioned in their title and abstract. See text for details.

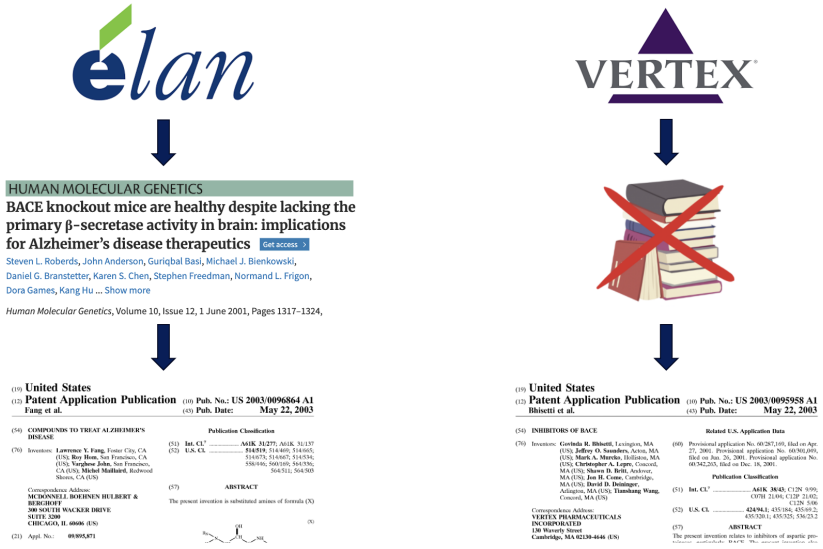
More specifically, we capture prior experience for firm f, protein p, and year t by defining the

following indicator:

$$\text{PriorPub}_{f,p,t} = \mathbb{I}\{\exists \text{ a PubTator3 publication by } f \text{ that studies } p \text{ with publication year } \leq t-1\}$$

This equals 1 if f has at least one prior publication on p before year t and 0 otherwise. In split-sample analyses, "higher experience" firms have $\text{PriorPub}_{f,p,t} = 1$ for the protein featured in their patents; "lower experience" firms have $\text{PriorPub}_{f,p,t} = 0$. Figure B7 shows an example from our data on two patents published on the same date, May 22, 2003. Élan Corporation has a BACE1-related article published from 2001, so $\text{PriorPub}_{\text{Élan, BACE1, 2003}} = 1$. Instead, Vertex Pharmaceuticals' first BACE1-related publication appeared in 2019, so $\text{PriorPub}_{\text{Vertex, BACE1, 2003}} = 0$. Both firms patented on BACE1 in 2003, but they differ in target-specific prior experience according to our PubTator3-based indicator.

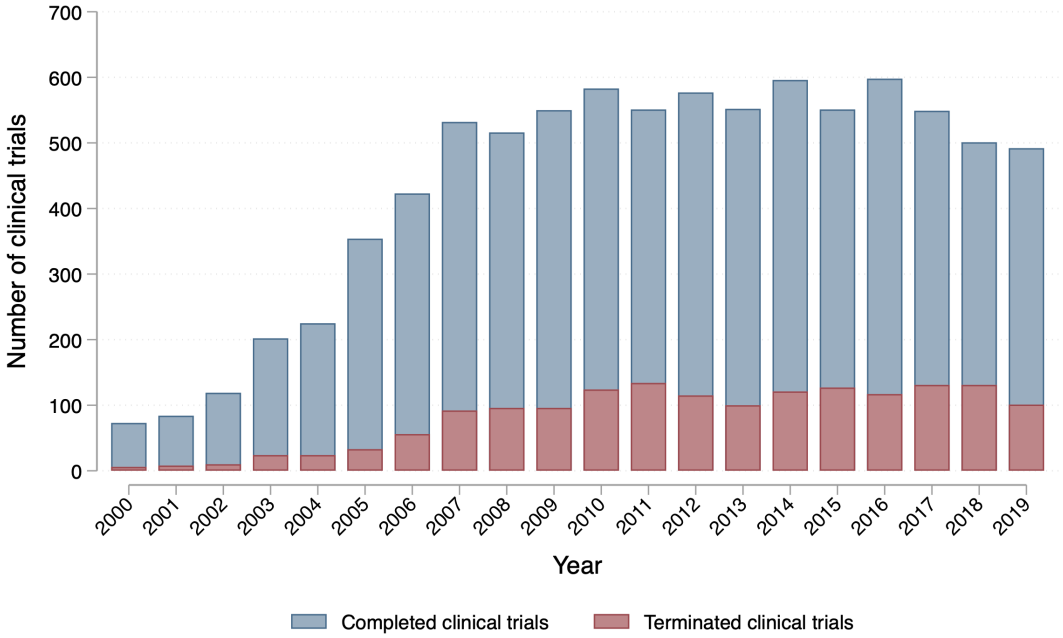
Figure B7: Example of Patenting Firms with Differing Levels of Expertise on BACE1.



Note: The figure shows the two patents published the same day, May 22, 2003, both leveraging BACE1 as a drug target for Alzheimer's. Élan Corporation has prior publications on the protein, whereas Vertex Pharmaceuticals does not, thus implying differing levels of experience with the protein. See text for details.

C. ADDITIONAL TABLES AND FIGURES

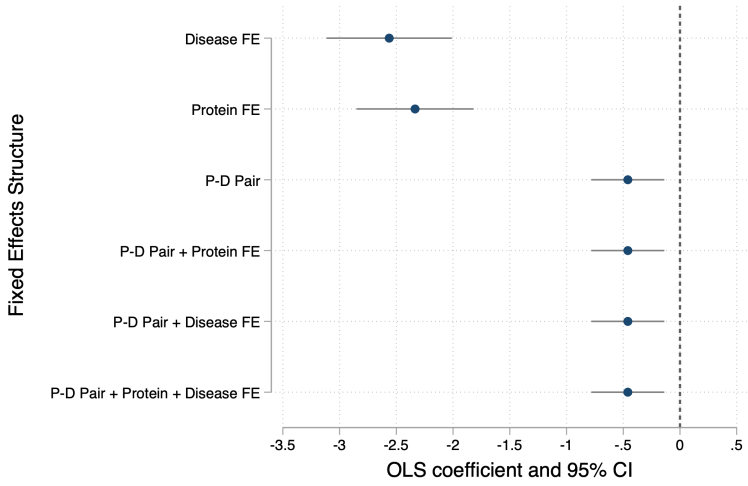
Figure C1: Distribution of Trial Failures Over Time



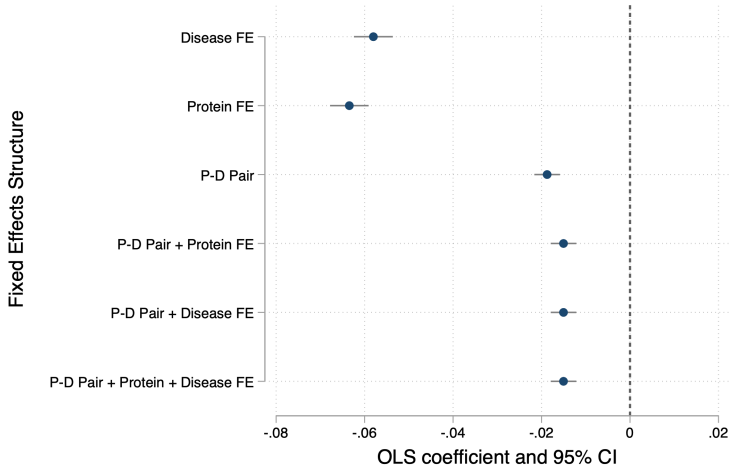
Note: The figure plots the number of phase II and phase III clinical trials in our sample. Blue bars report the number of completed clinical trials, while maroon bars report the subset of trials that were terminated, suspended, or withdrawn before completion. See text for details.

Figure C2: Direct and Spillover Effects from Clinical Trial Failures (Robustness to Alternative Fixed Effect Structures).

(a) *Direct Effects of Clinical Trial Failures.*

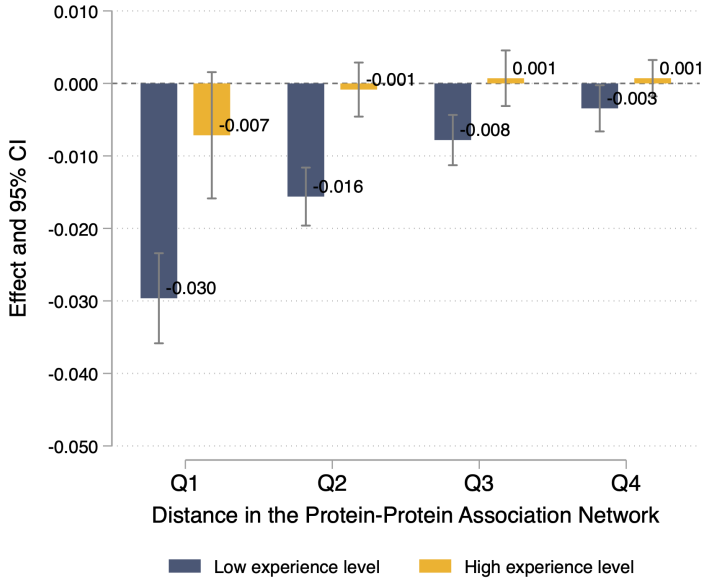


(b) *Spillover Effects of Clinical Trial Failures.*



Note: The figure presents the robustness of our main results reported in Table 2 to alternative structures of fixed effects. Each coefficient presents the OLS estimate of the interaction term with increasingly stringent specifications of gene, disease, and gene-disease pair fixed effects. Panel (a) shows the decrease in patenting for protein-disease pairs subject to failed clinical trials, depending on the fixed effect structure. The coefficients are estimated using variations of Equation 1, with the last coefficient corresponding to Column (2) of Table 2. Panel (b) shows the decrease in patenting for protein-disease pairs biologically related to the genetic targets of failed clinical trials, depending on the fixed effect structure. The coefficients are estimated using variations of Equation 2, with the last coefficient corresponding to Column (4) of Table 2. See text for details.

Figure C3: Firm Heterogeneity in Patenting Behavior Following a Clinical Trial Failure.



Note: The figure presents heterogeneity in the extent to which firms generalize from clinical failures depending on their level of experience. Specifically, the figure shows the decrease in patenting for protein-disease pairs at varying levels of biological relatedness to the genetic targets of failed clinical trials, reported separately for firms with differing levels of genetic experience. Firms with low experience are defined as those without prior publications on the protein. The bars represent standardized beta coefficients to enable comparison across split-sample regressions based on Equation 2. See text for details.

Table C1: Direct Effects of Clinical Trial Failures on Pharmaceutical Firm Patenting by Clinical Trial Stage.

	Firm Patents Targeting a Protein-Disease Pair			
	(1)	(2)	(3)	(4)
Post × Clinical Trial (Phase II and III)	0.980***	1.110***		
	(0.123)	(0.000650)		
... × Failure		-0.459**		
		(0.164)		
Post × Clinical Trial (Only Phase II)			1.202***	1.387***
			(0.105)	(0.138)
... × Failure				-0.646***
				(0.188)
Protein-disease FE	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	147,979,011	147,979,011	147,945,970	147,945,970

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. The sample in columns (1) and (2) includes all protein-disease pairs, while columns (3) and (4) exclude protein-disease pairs that received a Phase III clinical trial. *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Phase II and III)*: 0/1 = 1 in all years after conclusion of the first phase II or 3 clinical trial targeting a focal protein-disease pair; *Post × Clinical Trial (Only Phase II)*: 0/1 = 1 in all years after conclusion of the first phase II clinical trial targeting a focal protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

Table C2: Direct and Spillover Effects of Clinical Trial Failures on Pharmaceutical Firm Patenting (Logarithmic Transformation).

	Log(Firm Patents Targeting a Protein-Disease Pair + 1)			
	(1)	(2)	(3)	(4)
Post × Clinical Trial (Direct)	0.1656***	0.1879***		
	(0.00590)	(0.00723)		
... × Failure		-0.0788***		
		(0.0122)		
Post × Clinical Trial (Spillover)			-0.00284***	-0.00186***
			(0.000147)	(0.000158)
... × Failure				-0.00430***
				(0.000243)
Protein-disease FE	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	147,979,011	147,979,011	147,845,878	147,845,878

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. The sample in columns (1) and (2) includes all protein-disease pairs, while columns (3) and (4) exclude protein-disease pairs that directly received a clinical trial. *Log(Firm Patents Targeting a Protein-Disease Pair +1)*: logarithmic transformation of the yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Direct)*: 0/1 = 1 in all years after conclusion of the first phase II or III clinical trial targeting a focal protein-disease pair; *Post × Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase II or III clinical trial targeting a genetically related protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

Table C3: Spillover Effects of Clinical Trial Failures on Pharmaceutical Firm Patenting (Alternative Samples).

	Firm Patents Targeting a Protein-Disease Pair			
	(1)	(2)	(3)	(4)
Post × Clinical Trial (Spillover)	0.0269***	0.0302***	-0.00482***	-0.00220*
	(0.000630)	(0.000713)	(0.00105)	(0.00110)
... × Failure		-0.0146***		-0.0116***
		(0.00147)		(0.00147)
Protein-disease FE	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	147,979,011	147,979,011	43,281,430	43,281,430

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. The sample in columns (1) and (2) includes all protein-disease pairs that did not receive a clinical trial, while columns (3) and (4) include only protein-disease pairs genetically related to pairs receiving a clinical trial. *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase II or III clinical trial targeting a genetically related protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

Table C4: Spillover Effects of Clinical Trial Failures on Pharmaceutical Firm Patenting (Enrollment Size of Clinical Trial).

	Firm Patents Targeting a Protein-Disease Pair			
Sample:	Low Patient Enrollment		High Patient Enrollment	
	(1)	(2)	(3)	(4)
Post × Clinical Trial (Spillover)	0.0275***	0.0304***	0.0325***	0.0362***
	(0.000883)	(0.00107)	(0.000892)	(0.000954)
... × Failure		-0.00834***		-0.0353***
		(0.00185)		(0.00260)
Protein-disease FE	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	126,334,781	126,334,781	126,240,845	126,240,845

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. The sample in columns (1) and (2) includes only clinical trials with a below median number of patients enrolled (i.e., fewer than 50), while columns (3) and (4) include clinical trials with an above median number of patients enrolled (i.e., more than 50). *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase II or III clinical trial targeting a genetically related protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

Table C5: Spillover Effects of Clinical Trial Failures by Underlying Merit (Split Sample Regressions).

	Firm Patents Targeting a Protein-Disease Pair							
Sample:	High Merit Protein-Disease Pairs				Low Merit Protein-Disease Pairs			
Protein-Protein Distance:	1 Quartile (1)	2 Quartile (2)	3 Quartile (3)	4 Quartile (4)	1 Quartile (5)	2 Quartile (6)	3 Quartile (7)	4 Quartile (8)
Post × Clinical Trial (Spillover)	0.0244	-0.0326***	-0.0277**	-0.0105	0.00299*	-0.00348***	-0.00290***	-0.00497***
	(0.0137)	(0.00911)	(0.00937)	(0.00755)	(0.00147)	(0.000954)	(0.000826)	(0.000707)
... × Failure	-0.0587**	-0.0283*	-0.0100	-0.0177	-0.0100***	-0.00448***	-0.00111	0.000882
	(0.0195)	(0.0124)	(0.0131)	(0.0116)	(0.00269)	(0.00136)	(0.00140)	(0.00104)
Protein-disease FE	YES	YES	YES	YES	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	1,714,370	1,265,780	1,077,699	941,659	9,535,511	9,540,299	9,659,334	9,546,778

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level corresponding to those reported in Panel (b) of Figure 3. The table reports standardized beta coefficients to enable comparison across split samples. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. Columns (1)-(4) include protein-disease pairs with a positive Open Target Score, while columns (5)-(8) include protein-disease pairs with an Open Target Score equal to zero. *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase II or III clinical trial targeting a genetically related protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

Table C6: Spillover Effects of Clinical Trial Failures by Level of Firm Experience.

	Firm Patents Targeting a Protein-Disease Pair							
Sample:	Firms with High Level of Experience				Firms with Low Level of Experience			
Protein-Protein Distance:	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post × Clinical Trial (Spillover)	0.00763**	-0.00426**	-0.00238	-0.00228**	0.00980***	-0.00823***	-0.00733***	-0.00689***
	(0.00286)	(0.00151)	(0.00133)	(0.000863)	(0.00267)	(0.00157)	(0.00139)	(0.00128)
... × Failure	-0.00716	-0.000851	0.000714	0.000717	-0.0296***	-0.0155***	-0.00722***	-0.00344*
	(0.00444)	(0.00190)	(0.00196)	(0.00128)	(0.00317)	(0.00197)	(0.00173)	(0.00162)
Protein-disease FE	YES	YES	YES	YES	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	10,772,506	10,200,416	10,370,675	10,488,437	10,772,506	10,806,079	10,737,033	10,488,437

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level corresponding to those reported in Figure C3. The table reports standardized beta coefficients to enable comparison across split samples. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. Columns (1)-(4) include patent applications by firms with a high level of experience, while columns (5)-(8) include patent applications by firms with a low level of experience. Q1 denotes the closest quartile in the protein-protein association network, while Q4 denotes the most distant quartile. *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase II or III clinical trial targeting a genetically related protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

Table C7: Neighborhood-Level Effects of Clinical Trial Failures by Landscape Ruggedness.

	Firm Patents in Neighboring Protein-Disease Pairs			
	High Merit Pairs		Low Merit Pairs	
	Smooth	Rugged	Smooth	Rugged
	(1)	(2)	(3)	(4)
Post	226.0***	117.5***	76.05***	96.42***
	(12.99)	(17.09)	(5.936)	(15.92)
$\cdots \times$ Failure	-73.82***	2.142	-33.27***	26.09
	(21.66)	(32.38)	(9.130)	(39.07)
Neighborhood FE	YES	YES	YES	YES
Gene FE	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	33,998,638	32,690,773	33,998,638	32,690,773

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the neighborhood-year level. Std. err. clustered at the neighborhood level. All models include neighborhood, gene, disease, and year fixed effects. Columns (1) and (3) use the subsample of smooth landscapes, defined as neighborhoods with below-median fitness-distance correlation. Columns (2) and (4) use the subsample of rugged landscapes, defined as neighborhoods with above-median fitness-distance correlation. Columns (1) and (2) use patenting on high-merit neighboring genes (OT > 0) as the dependent variable, while columns (3) and (4) use patenting on low-merit neighboring genes (OT $= 0$). *Firm Patents in Neighboring Protein-Disease Pairs*: yearly count of USPTO patent applications in the neighborhood of the focal protein-disease pair; *Post*: 0/1 = 1 in all years after conclusion of the focal clinical trial; *Failure*: 0/1 = 1 for focal clinical trials that are terminated before their natural completion. See text for details.