MANAGING INNOVATION DISRUPTIONS: SPILLOVERS ACROSS USERS, FIRMS, AND INVENTORS IN R&D OPERATIONS

Pierre Gautreau

A Dissertation Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

Graduate Program in Business Administration
Schulich School of Business, York University
Toronto, Ontario

May 2026

Abstract

Research and development (R&D) operations commonly experience innovation disruptions, such as project discontinuations, alliance terminations, or new product launches. In healthcare, innovation disruptions are susceptible to spillover consequences because R&D typically depends on other organizations' activities, which creates an interdependence that may require considerable managerial coordination. This dissertation focuses on how innovation disruptions in healthcare lead to unintended consequences that can benefit some of the parties involved and harm others. A multi-level perspective is adopted to develop three empirical studies that examine these consequences.

The first study examines how the implementation of an AI-enabled exercise monitoring technology disrupts traditional physiotherapy service delivery in a real-world clinical setting. This study suggests that AI-enabled monitoring helps sustain patient participation to home-based exercise by reducing uncertainty about exercise performance and reinforcing accountability in service co-production. Rather than replacing the physiotherapist from the co-production process, the technology appears to strengthen the patient-provider relationship. The second study examines whether a successful drug launch by a pharmaceutical firm affects the stability of its other alliances with biotechnology partners. Using alliance and drug development data, the study shows that successful launches by a pharmaceutical firm are associated with a higher likelihood of alliance termination for biotechnology partners. Therefore, positive innovation outcomes for one firm can be associated with a vulnerability for their partner. The third study examines how drug development project discontinuations affect subsequent inventor-level innovation activities (e.g., patent outputs, patent novelty). Using matched inventor-project panel data, the study demonstrates that discontinuations are associated with increased formation of new collaborations, lower patent output, and higher patent novelty over time, which suggests that negative project events like discontinuations can redirect search toward new technological directions albeit with productivity costs.

Collectively, the three studies show that disruptions in R&D operations generate spillover consequences that extend beyond the focal event and differ across contexts and levels of analysis. The dissertation contributes to research on managing R&D by showing that disruptions are recurrent and have important implications for firm developing innovation, users that participate in or benefit from innovation, and inventors who create it.

Acknowledgements

I would first like to thank my supervisor, Moren, for your kindness and support throughout this journey, and for making my doctoral experience so enjoyable. I am so grateful for the experiences we have shared around the world and for all the ways you have helped me learn and grow. I also thank my wife, Caitlyn, for your patience during the long days and nights I devoted to this work, and for always believing in me. I am grateful to Cari and the entire team at Sunnybrook Hospital for welcoming me so warmly. I am thankful that our collaboration led to a revise-and-resubmit decision at *Production and Operations Management* for the second chapter of this dissertation. I would also like to thank Annapoorima, Vareska, and Adam for their support on the third chapter, which is also being revised for resubmission to *Production and Operations Management*. Finally, I thank the friends and family who have been present for me during this journey, as well as my dissertation committee and examiners for the time and feedback they gave to this work.

Table of Contents

List of Tables

List of Figures

Chapter 1:
Introduction

Research and development (R&D) is a complex operational process. Like other operational processes, R&D requires coordination, involves repeated activities, and depends on effective resource allocation, process control, and decision-making (Shavinina, 2003). Early models conceptualized managing innovation as a standard and predictable sequence of stages (Utterback, 1971), but subsequent research shows that innovation processes vary widely in structure and objectives (Salerno et al., 2015; Shenhar, 2001). Furthermore, individual innovation projects are only one component of a firm's innovation ecosystem and must be managed accordingly when facing external pressures and interdependencies (Hansen & Birkinshaw, 2007). These interdependencies make innovation processes inherently unstable: disruptions are frequent and often endogenous to the innovation process itself. As with any operational context, managerial intervention becomes necessary when a steady state is disrupted (Bessant et al., 2005). Understanding managerial responses to innovation disruptions is particularly important in healthcare where there are concerns about declining R&D productivity (Pammolli et al., 2011; Salerno et al., 2015) and an ageing global population is increasing demand for new medical innovations.

Innovation disruptions arise from various sources. For instance, both past successes and failures can disrupt R&D operations (Lévesque et al., 2022). As well, strategic decisions like reallocating efforts from exploration to exploitation, can lead to disruptions (Uotila et al., 2009). Regardless of their origin, disruptions require managerial interventions to return operations to a functioning steady state. Because innovation processes are interdependent with other organizational systems, disruptions can generate spillover effects that extend beyond their immediate locus. Understanding these spillover consequences is essential for advancing research on managing R&D operations.

This dissertation examines how disruptions related to healthcare innovation create unintended consequences across different levels of analysis. These disruptions can have broad operational impacts because R&D involves interdependencies in which end-users, firms, and inventors are connected through shared resources, coordinated activities, or common objectives. Accordingly, the dissertation adopts a multi-level approach. Chapter 2 evaluates how implementing a new AI-enabled technology

designed to support physiotherapy exercises—a disruption to physiotherapy service delivery—influences service co-production and creates unintended spillover benefits, including an improvement to the patient–physiotherapist relationship. Chapter 3 is set in the biopharmaceutical industry to examine how a pharmaceutical firm's drug launch affects the likelihood of terminating R&D alliances with biotechnology partners and shows that a positive disruption for the focal firm can generate negative spillovers for its partners. Chapter 4, on the other hand, investigates how the discontinuation of a drug development project affects the subsequent innovation outputs of the affected inventors for years after the discontinuation, and highlights a tradeoff whereby inventors tend to form new collaboration and increase the novelty of their innovations albeit with a lesser innovation output.

Taken together, these studies position the dissertation within a research agenda on managing R&D operations facing disruptions. Figure 1 summarizes the disruption type, context, unit of analysis, outcome, and unintended consequence examined in each chapter. The three chapters are linked by a focus on how disruptions are associated with spillover effects beyond the focal event or focal actor. Examining disruption at the user (Chapter 2), firm (Chapter 3), and inventor (Chapter 4) levels, the dissertation shows that such effects can impact relationships, behaviors, and innovation trajectories over time. While this work does not capture all possible disruptions or consequences in R&D, it contributes to managing R&D operations by showing that disruptions can generate both positive and negative spillover effects across multiple levels of analysis.

	Chapter 2: User level	Chapter 3: Firm level	Chapter 4: Inventor level
Disruption type:	Implementation of AI-enabled innovation	Launch of a new drug	Discontinuation of a drug project
Context:	Physiotherapy service delivery	Biopharmaceutical drug development	Biopharmaceutical drug development
Unit of analysis:	Users involved in service co-production	Inter-firm alliances	Individual inventors
Outcome:	Change in exercise behavior	Alliance termination risk	Changes in patent activities
Unintended consequence:	Spillover within service interaction	Spillover to alliance partners	Spillover to innovation trajectories

Figure 1.1. Overview of the three dissertation studies

Next, concise overviews of Chapters 2, 3, and 4 are presented, to outline the research questions, empirical settings, and main findings. Specifically, Chapter 2 focuses on answering: *To what extent can*

AI-enabled monitoring and feedback improve the quantity and quality of co-production inputs? And what behavioral factors drive these improvements? In professional services (e.g., physiotherapy), service co-production is common whereby customers are active participants in the service process, often without the service provider's supervision (e.g., patients exercise at home). This chapter investigates whether AI-enabled monitoring-and-feedback mobile technology can relieve the service provider from supervision of off-site co-production, in the context of physiotherapy. A 12-week field study was conducted with 124 patients, of which a cohort was delivered exercise technique feedback and progress tracking in real time (intervention participants) while another cohort was not (control participants). The dataset comprises 9,310 patient-day datapoints and 2,093 hours of movement data, complemented with semi-structured interviews with participants. The findings show that both cohorts began the study with similar exercise participation—in terms of exercise duration and exercise frequency—but the intervention cohort sustained participation over time while control-cohort participation declined. Exercise technique patterns show that sustained participation stems from more than AI feedback alone. Although the value of feedback diminishes as technique proficiency rises, sustained participation remains high in the intervention cohort. The qualitative analysis uncovers three behavioral mechanisms that explain this sustained participation: uncertainty reduction from real-time feedback; intrinsic accountability gained from progress tracking; and extrinsic accountability gained from data sharing with physiotherapists. The findings support human-AI complementarity by showing that service co-production with AI-enabled mobile technology is most effective when augmenting rather than replacing provider-customer relationships in this context. Additionally, Chapter 2 offers practical design principles for managers to leverage AI by making invisible work visible, reducing execution uncertainty, and improving service relationships.

Chapter 3 concentrates on addressing the following research question: *Do drug launches by a pharmaceutical firm increase the risk of alliance terminations for its biotech partners, and can such termination be distinguished from other types of terminations with information observable when alliances are initially formed?* Biotech-pharmaceutical alliances often involve asymmetrical dependencies, with biotech firms being more reliant on their pharmaceutical partners than *vice versa*. This imbalance poses a previously overlooked vulnerability to biotech firms: an increased likelihood of

alliance termination following a pharma partner's drug launch, even when the launch is unrelated to the alliance. Analyzing 1,649 terminated biotech-pharma alliances over two decades suggest that terminations disproportionately occur after drug launches. A survival analysis of 3,496 alliances with known outcomes shows that the risk of termination increases significantly during the three-year period following a drug launch. Using machine learning (ML) shows that this risk is predictable based on features observable at alliance formation and that these events are not strategically timed to coincide with the negative news of an alliance termination with the positive news of a drug launch. Rather, the findings reveal a hidden cost of open innovation in asymmetric partnerships. Hence, a data-driven approach demonstrates how ML can uncover management phenomena and inform biotech partner selection by identifying post-launch termination risks that were previously unknown.

Chapter 4 focuses on answering: *Does project discontinuation influence subsequent inventor-level innovation outcomes by changing future collaboration opportunities?* Inventors drive innovation. Consequently, drug development discontinuations are expected to impact subsequent inventor-level innovation outcomes through the formation of new collaborations. This chapter examines this impact by focusing on the biopharmaceutical context, where discontinuations are common and innovation can be tracked through patent activities. A comparison of inventors associated with discontinued projects against matched inventors on projects that remain active is performed using an inventor–project panel data constructed by linking Cortellis Drug Development Intelligence data with patent records. The sample includes 6,723 discontinued projects, 14,732 matched non-discontinued projects, and 164,129 inventors. The findings suggest that drug development discontinuations lead inventors to form more new collaborations over time. Discontinuations are also associated with lower patent output but higher patent novelty. Mediation results show that new collaborations help explain the increase in patent novelty, but not the decline in patent output. Furthermore, the findings demonstrate that discontinuations for scientific reasons (like safety concerns) are associated with a stronger increase in new collaborations than non-scientific discontinuations (like those due to market condition changes). However, the indirect effect through collaboration is more negative for patent output and less positive for patent novelty following scientific discontinuations. These findings are consistent with problemistic search theory and contribute to research on innovation and collaboration by showing that negative project events can

redirect inventors toward new collaborations and more novel technological directions, albeit with a cost to innovation productivity.

This dissertation advances the central claim that managing innovation requires anticipating indirect consequences of disruption. Disruptions influence motivations, reallocate attention, and alter future trajectories by inducing responses beyond the focal event. The three main chapters show that these responses are systematic within the context studied. An AI-enabled intervention can sustain user participation intrinsic and extrinsic motivators. A successful drug launch can destabilize alliance relationships for dependent partners. A project discontinuation can redirect inventors toward new collaborations and more novel technological directions, while reducing short-term output. These findings highlight spillovers as an important consideration when managing innovation processes. By considering evidence from user, firm, and inventor levels, this dissertation contributes to a multi-level perspective of how disruptions propagate through R&D operations.

Chapter 2:
AI-enabled Mobile Technology for Service Co-production: A Physiotherapy Comparative Cohort Study

2.1. Introduction

Service co-production encompasses value-creating tasks performed by customers outside the provider's direct supervision (Roels, 2014). These customer-provided inputs, whether patients completing rehabilitation exercises, clients filling in documents, or students engaging in preparatory coursework, can vary in both quantity and quality, and this variability strongly influences service performance. Managing these inputs is challenging since service providers cannot directly observe or control them, yet they are essential to the effectiveness and efficiency of the service process. Proper design and management of AI technologies can address many service challenges (Dai & Tayur, 2022; Gnewuch et al., 2024; Spring et al., 2022; L. Wang et al., 2023), including co-production (Angelopoulos et al., 2023; M. M. Li et al., 2024).

Service co-production challenges are particularly prominent in the context of physiotherapy, where low adherence to prescribed home exercise programs is a longstanding operational problem, with adherence rates typically below 50% (Bonnechre et al., 2016; Jack et al., 2010; McLean et al., 2010). Although exercises are essential for recovery (Boyer et al., 2025), patients often fail to complete them consistently or correctly outside the clinic. This non-adherence undermines recovery, prolongs treatment, and increases reliance on in-person supervision (Kılıç & Güneş, 2024). Ineffective co-production strains already overburdened healthcare systems and force operational trade-offs between quality and throughput (V. Ahuja et al., 2020). While the technical feasibility of co-production via home-based rehabilitation is well established, patients frequently express uncertainty about proper exercise technique without direct supervision (Argent et al., 2018), limiting the effectiveness of this model. In this article, we examine the impact of providing AI-based quality monitoring in physiotherapy co-production to explore whether, when, and why it can improve adherence to prescribed at-home exercises, the primary co-production input in this service process.

Despite technological advances and the importance of co-production to service performance, empirical evidence on how AI can be integrated into service processes to manage uncontrollable off-

site customer inputs remains scarce (M. Li et al., 2025), though interest on the impact of AI on service design and delivery is growing (Cohen et al., 2024). Moreover, research on AI-enabled hybrid intelligence and digital transformation highlight the challenge of sustaining human contributions over time (Angelopoulos et al., 2023; M. M. Li et al., 2024). In healthcare, systematic reviews of digital rehabilitation technologies find that very few interventions measure co-production (exercise) adherence, even fewer provide real-time feedback, and none benchmark against a control group (Sassi et al., 2024). Related work in the technology adoption literature has focused primarily on determinants of initial AI adoption, with limited attention to sustained user behavior over time (Soulami et al., 2024). Together, these literatures point to a lack of evidence on the impact of AI-enabled systems on customer contributions in co-production settings. This gap is critical in professional services like physiotherapy, where long-term adherence to prescribed inputs is essential to achieving the intended service outcomes (Boyer et al., 2025; Burns et al., 2021). This study addresses this gap by evaluating a novel digital technology in the physiotherapy service process to investigate: *To what extent can AI-enabled monitoring and feedback improve the quantity and quality of co-production inputs? And what behavioral factors drive these improvements?*

From the perspective of unified services theory (Sampson & Chase, 2022; Sampson & Froehle, 2006; Spring et al., 2022), home exercise adherence represents a primary input to the physiotherapy service process. Because this input is generated outside the clinic, in-process quality control is difficult, and variability in input quantity and quality directly affects final service outcomes (Roels, 2014). Technology-augmented service designs have been proposed to improve home care (e.g., Bavafa et al., 2021), but most digital health technologies for physiotherapy focus on reminders or passive tracking, offering no timely feedback on exercise quality (Sassi et al., 2024). We developed SPARS/PT+, an AI-enabled monitoring and feedback system that combines smartwatch-based motion sensing AI (SPARS or Smart Physiotherapy Activity Recognition System) with a mobile application (PT+) to deliver real-time feedback on exercise technique and track participation over time. This design aligns with the human–AI complementarity perspective (Farhan et al., 2025; Nguyen & Elbanna, 2025; Vaccaro et al., 2024) by allocating off-site feedback to the AI system while preserving clinical tasks for the therapist.

We conducted a 12-week comparative cohort study comprising 124 adult participants with

diagnosed shoulder injuries (100 control, 24 intervention). Two control participants withdrew before their second visit and were excluded from the analysis, leaving 122 participants (98 control, 24 intervention). Exercise patterns were collected by smartwatch inertial sensors, though we complemented these quantitative data with qualitative data from interviews. This mixed-method design enables us to further strengthen the empirical grounding of our study as well as uncover the behavioral mechanisms underlying the observed effects. Such a design supports de Treville et al.'s (2023) call to develop theoretical insights that are firmly grounded in empirical evidence. Our results show that intervention participants completed significantly more exercises than control participants, although the difference was not consistent over the 12-week period. At the onset of treatment, both groups exhibited similar behavior; however, intervention participants maintained higher participation both within and across weeks throughout the study. Qualitative analysis indicates that the intervention's effectiveness was driven by three behavioral mechanisms: real-time feedback reduced uncertainty about technique, which increased participants' confidence in exercising; progress tracking supported intrinsic accountability by visualizing goal attainment; and data sharing with providers created extrinsic accountability through participants' desire to avoid shame for not exercising.

By quantifying both short- and longer-term behavioral impact of our AI intervention, we advance three streams of literature. First, in service operations, we demonstrate how AI-enabled monitoring and feedback can be integrated into the service process to manage uncontrollable customer inputs in co-production settings, a challenge common across many professional services (Larsson & Bowen, 1989; Roels, 2014; Sampson & Froehle, 2006). Second, in behavioral operations, we provide empirical evidence that reallocating specific quality-control tasks to AI can digitally encourage accountability and sustain engagement. This also extends prior work on technology adoption, which has focused largely on initial deployment (Blut & Wang, 2020; Davis, 1989; Venkatesh et al., 2003), by showing how behavioral mechanisms drive sustained use of AI-enabled systems. Third, in technology management, we contribute to the debate on whether AI should replace or complement humans in professional service processes (Angelopoulos et al., 2023; Bianchi & Brière, 2024; Koehler & Sauermann, 2025), by providing evidence that AI can effectively complement human expertise. While grounded in physiotherapy, our findings have broad implications for service contexts where performance depends

on the quality and consistency of customer-provided inputs generated outside direct provider supervision.

2.2. Background and Hypothesis Development

In several professional services, the customer is not only a passive recipient of value but an active participant in its creation, a concept known as service co-production (Roels, 2014). According to unified services theory (Sampson & Froehle, 2006), customers function as essential, often unsupervised, productive resources whose inputs are integral to the service process itself. This can increase a firm's productive capacity or can be a compulsory part of the service, yet it introduces a core operational tension: a critical input of the process is executed beyond the firm's control. This ceding of direct control over a key input leads to significant variability in both the quantity and quality of customer contributions, which is a primary driver of overall service performance (Frei, 2006; Sampson, 2000).

In contrast to internal processes where managers can deploy tools to monitor and control quality, off-site service production creates information asymmetry (Akan et al., 2011); the provider often has limited or delayed visibility into how, when, or how well external tasks are performed. This lack of real-time process information means that variability introduced by the customer can propagate unchecked, leading to downstream disruptions such as the need for rework, prolonged service encounter times, and inefficient use of the professional's expertise. A principal concept of operations management is to design systems that can effectively monitor and manage the variability originating from essential inputs (Klassen & Menor, 2007).

Digital technologies offer opportunities to address the core operational challenges of service co-production, most notably, managing control over customer-provided inputs (Froehle & Roth, 2004; Larivière et al., 2017). In the operations management literature, well-designed technological interventions have been shown to improve process control by enabling remote monitoring (Choi et al., 2022; Palmarozzo et al., 2025) or providing real-time feedback (Margolin et al., 2025), thereby extending the provider's reach beyond the physical service facility. The design of these technologies in professional services is best guided by the principle of human–AI complementarity, which emphasizes combining human expertise with machine capabilities rather than replacing labor (Farhan et al., 2025; Nguyen & Elbanna, 2025; Vaccaro et al., 2024). As Spring et al. (2022) demonstrate, AI can automate

certain routines, but it is generally valuable in professional practice by working alongside humans. Beyond technical capabilities, successful integration also requires trust and acceptance among stakeholders. In the healthcare context, Dai & Tayur (2022) find that securing buy-in from both providers and patients is essential for successful deployment and argue that AI systems should be designed as complementary tools that build trust through transparency, user control, and alignment with service workflows. These principles directly affect user behaviors. Gnewuch et al. (2024) find that higher human involvement in AI-enabled processes can increase customer trust and perceived quality, while more AI automation increases perceived efficiency.

While research on AI in service operations is growing (M. Kim, 2023), empirical studies examining its role in service co-production remain limited. Evidence from contexts like telemedicine (Gomez et al., 2024) and customer service (Brynjolfsson et al., 2025) indicates that AI can help reduce process variability. However, most studies examine short-term effects, often in controlled settings, rather than tracking sustained impacts in real world use. Prior research has largely focused on adoption and initial engagement, with limited attention to how AI-enabled systems influence the sustained quantity and quality of customer inputs. Physiotherapy provides a particularly suitable setting to examine this gap since effective treatment depends on adherence to prescribed home exercises for extended periods (McLean et al., 2010; Wright et al., 2014). Taken together, effective technological support for physiotherapy service co-production requires (1) mechanisms for remote monitoring and real-time feedback, (2) designs grounded in human–AI complementarity, and (3) features that build trust through transparency, useability, and usefulness.

As mentioned earlier, low exercise adherence is a persistent coproduction challenge in physiotherapy (Bonnechre et al., 2016; Jack et al., 2010; McLean et al., 2010). Variability in both the quantity of exercises performed and the quality of their execution relate directly to core operations management concerns with input variability, process control, and service performance. Because adherence must be sustained across an extended treatment horizon, physiotherapy also provides a natural context for studying the short- and long-term influence of AI on managing co-production behavior.

In response to the problem of low exercise adherence in physiotherapy, a variety of digital health technologies have been developed to support remote rehabilitation (Argent et al., 2018; Burns et al., 2021; Sassi et al., 2024). These technologies often take the form of mobile applications with exercise videos, automated reminders, or wearable sensors that track general activity levels (Sassi et al., 2024). While such tools can help address the quantity problem of co-production (i.e., whether an exercise was done), they typically lack the capability to monitor and provide feedback on input quality (i.e., whether the exercise was done correctly). As a result, the critical issue of performance uncertainty remains largely unaddressed, which stresses the need for more advanced interventions that can provide remote real-time feedback on technique. Building on unified services theory, which stresses managing variability introduced by customer inputs (Sampson & Froehle, 2006), we focus on whether extending process control off-site through an AI-enabled monitoring and feedback system (SPARS/PT+) can improve physiotherapy co-production. We develop hypotheses regarding how extending process control off-site through objective, real-time feedback and performance tracking can influence co-production behavior.

Because service co-production requires providers to cede control over critical inputs, outcomes depend heavily on unsupervised execution. In physiotherapy, this lack of oversight can translate into patient uncertainty about whether exercises are performed correctly (Argent et al., 2018). This uncertainty creates process friction that can discourage patients from initiating or completing their prescribed tasks. From an operations perspective, an intervention that reduces such friction should increase the output of the co-production process. The SPARS/PT+ system is designed as an alternative to direct supervision by providing real-time validation of exercise quality at home. By confirming correct technique, AI feedback is expected to reduce uncertainty and increase patient confidence in performing their co-production tasks. With this uncertainty addressed, the intervention is expected to increase the number of exercises patients complete. Thus, we posit:

> ***H1.*** *Participants who receive off-site co-production monitoring with real-time quality validation will produce a higher quantity of inputs than those who do not.*

The service process is designed to be delivered over a specified period (e.g., a 12-week treatment plan). In co-production settings, however, customer contributions often decline over time, leading to a

gradual decay in critical service inputs. This pattern is well documented in physiotherapy, where adherence to home-based exercises decreases significantly as treatment progresses (McLean et al., 2010), and more broadly in digital contexts, where continued engagement with self-service technologies often diminishes after initial use (Blut et al., 2021). SPARS/PT+ introduces continuous process visibility through performance tracking and progress monitoring. By providing patients with a consistent feedback loop and a clear view of their contribution, we expect a reduction in the typical decline in participation over the treatment period. Therefore, we hypothesize that:

> ***H2.*** *Participants who receive off-site co-production monitoring with real-time quality validation will maintain a higher quantity of inputs over time than those who do not.*

Unified services theory identifies variability in customer input quality as a key contributor to poor service outcomes. In physiotherapy, the absence of in-process quality control for at-home exercises allows poor technique to persist uncorrected. SPARS/PT+ addresses this gap by providing immediate feedback as an off-site quality assurance loop allowing patients to detect and correct errors as they occur. With practice guided by feedback, technique should improve and then stabilize once proficiency is reached. As such:

> ***H3.*** *Participants who receive off-site co-production monitoring with real-time quality validation will exhibit a learning-curve effect in input quality, with quality increasing until it plateaus.*

We formulate ***H3*** as a within-intervention hypothesis about quality correction since our focus is on how co-production with in-process quality control influences input quality over time. Because precise accuracy scores are available only for intervention participants, a between-group comparison would require non-equivalent measures of technique (e.g., self-reports, clinician ratings, or *ex post* SPARS analysis), which would introduce significant measurement errors. A within-intervention design provides the cleanest and most valid test of this hypothesis.

Moreover, while hypotheses ***H1*** to ***H3*** describe the expected effects of the intervention, certain effects may counteract one another. For example, a learning-curve effect (***H3***) implies that the value of AI feedback diminishes as participants become more proficient in exercise technique. Yet, our expectation of sustained participation (***H2***) suggests that patients will continue to engage with the system even after the marginal benefit of feedback has declined. Should both hypotheses be supported, this would suggest that the technology's operational role is more complex than simply a quality control

tool. Initially, it may serve primarily as a feedback mechanism that facilitates learning. Over time, as technique is mastered, its role could shift into a process management tool that guides and ritualizes patient engagement. Our quantitative data alone cannot fully capture these mechanisms. Therefore, to complement our statistical analysis, we conduct follow-up qualitative interviews to probe how the intervention's role in the co-production process affects inputs over time.

2.3. Data and Methods

2.3.1. Trial Design

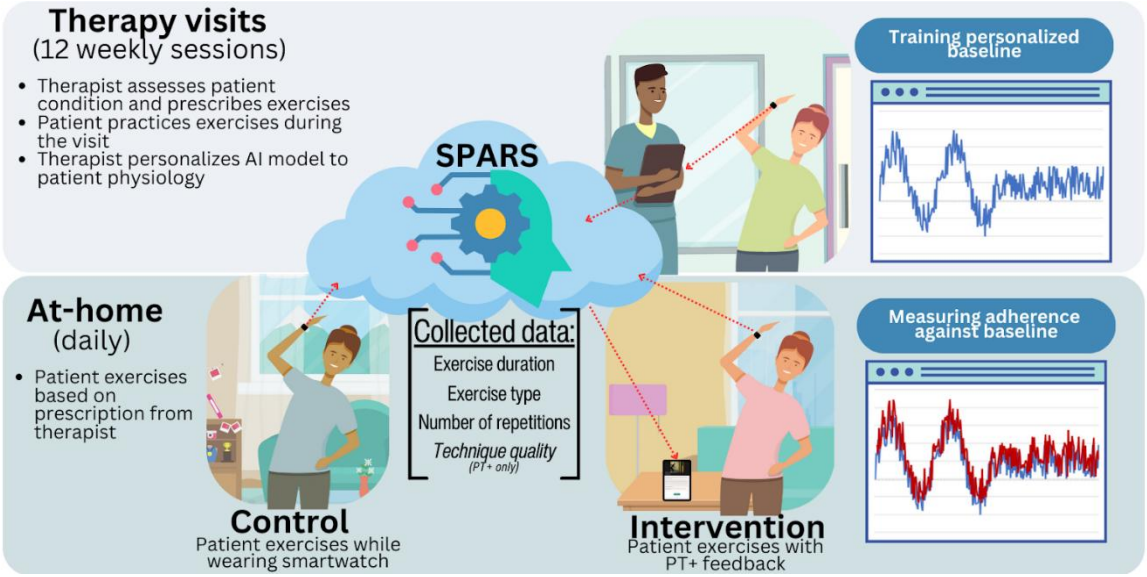
We evaluated the impact of SPARS/PT+ in a comparative sequential cohort study of 98 control and 24 intervention participants, all adults with diagnosed shoulder injuries who were scheduled to receive physiotherapy. The control cohort participated between May 2019 and April 2023 while the intervention cohort participated between January 2022 and February 2024. The sequential design was necessary because the control cohort data were needed to pre-train and validate the SPARS AI model, making concurrent enrollment with the intervention cohort infeasible.[1]

Both cohorts followed the same physiotherapy protocol consisting of 12 weekly one-on-one physiotherapy sessions, with prescribed daily exercises monitored via a smartwatch, generating 9,310 patient-days of records and 2,093 hours of movement data. For the intervention group, smartwatch data was also transmitted and processed in the cloud in real time to provide feedback on input quantity and quality via the PT+ application. Figure 2.1 outlines the study workflow and its interactions with SPARS/PT+.

All physiotherapy sessions took place at Sunnybrook Hospital's Holland Orthopaedic and Arthritic Centre in Toronto, Canada. Following the trial, semi-structured interviews were conducted with select participants to understand their experience and gain additional insights on ensuring the responsible and ethical integration of AI in healthcare (Boyer et al., 2022; Donia et al., 2024). The study protocol was approved by the Sunnybrook Research Institute ethics committee (Toronto, Canada; REB#353-2018), each cohort was preregistered with ClinicalTrials.gov (#NCT04629417 and #NCT05139173) before

1 Due to COVID-19–related delays. Seventy patients in the control cohort and eight patients in the intervention cohort enrolled during the same period.

participant enrollment, and written informed consent was obtained from all participants. The exercises used for this study were developed by two senior physiotherapists for patients seen at the Rapid Access Shoulder Program at Sunnybrook's Holland Centre.

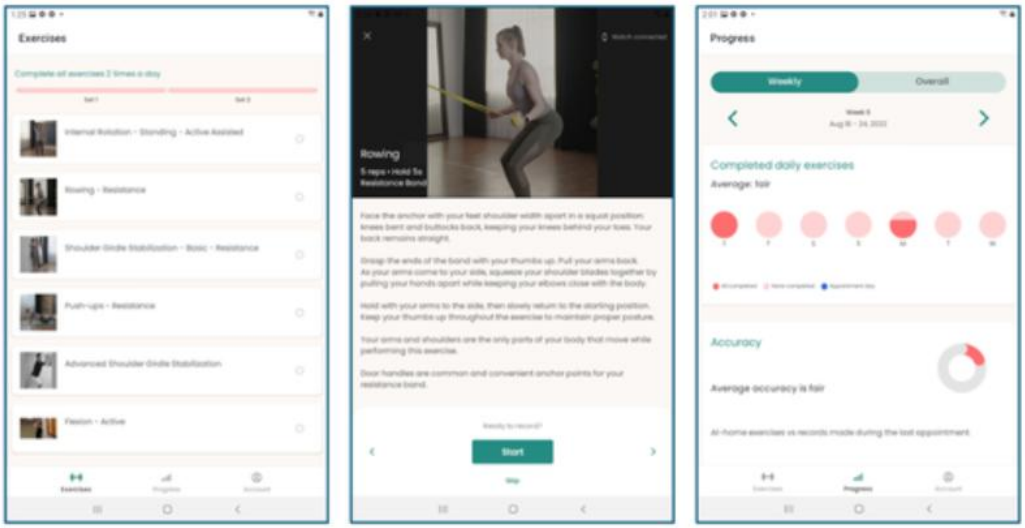


Note: All patients were equipped with smartwatches to track exercise patterns. Personalized baseline measures were recorded weekly during physiotherapy sessions, and data on at-home exercises were collected digitally. Patients in the PT+ intervention group also received a tablet with the PT+ application, which provided real-time feedback and tracking.

Figure 2.1. Workflow of Physiotherapy Treatment and At-home Exercises Monitored by SPARS

2.3.2. PT+ Mobile Application (Intervention)

In addition to providing accurate measures for analysis, SPARS outputs instant performance metrics that enabled the development of PT+, an android mobile application companion to SPARS. PT+ was designed in consultation with patients and practitioners to track the completion of prescribed exercises and display real-time feedback of the quality of each movement. The mobile tablet provided to participants connects via Bluetooth with the accompanying smartwatch so that participants have access to live performance feedback as they complete exercises at home. The application's interface, illustrated in Figure 2.2, allows participants to find their list of prescribed exercises, access exercise specific instructions and follow-along videos, view technique accuracy metrics and exercise feedback.



Note: List of exercises prescribed to the patient (left). Individual exercise where the patient can review the instructions for an exercise, initiate recording, and follow along with the instructional video (center). Patient performance and technique accuracy metrics (right).

Figure 2.2. SPARS PT+ Rehabilitation Mobile App

2.3.3. Participants

In total, the study recruited 141 participants. After some immediate abandonment, data was collected on a control group of 100 participants and an intervention group of 24 participants. All recruited participants were patients with rotator cuff conditions, including tendinosis, shoulder impingement syndrome, or degenerative or traumatic rotator cuff tear. Appendix A outlines the inclusion and exclusion criteria, which were determined in consultation with orthopedic surgeons and physiotherapists. The cost of 12 weeks of once-weekly supervised physiotherapy sessions was covered by the study, or, for injured workers participating in the Sunnybrook's Working Condition Program, through the Workplace Safety and Insurance Board of Ontario. Patients were referred to the study by orthopaedic surgeons at Sunnybrook Hospital or applied to participate in the study through the hospital's Holland Centre. Recruitment procedures were identical for both cohorts.

Both cohorts are representative samples of patients seeking physiotherapy for rotator cuff conditions at Sunnybrook's Holland Centre and are comparable to the demographics of physiotherapy practices in Canada. Consistent with clinical practice, the groups are relatively balanced by gender but contain slightly more female participants. The median age is 64 years, with ages ranging from 20 to 80 years old. Participants in both cohorts varied in physiological characteristics (injury severity, symptom duration, body mass index) and socio-economic characteristics (education, marital status, income, work

status). As shown in Table 2.1, these attributes are balanced across the intervention and control groups, suggesting that the samples are comparable and representative of the target population. Demographic controls are nonetheless included in our analyses.

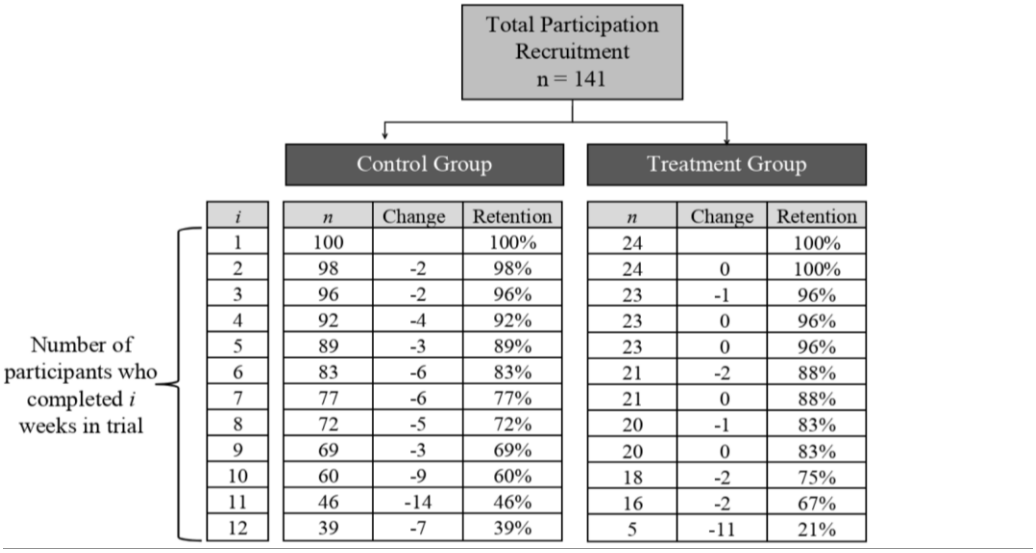
	Control Group (n = 100)		Intervention Group (n = 24)	
Age, mean (SD)	*n* = 100	62.8 (12.9)	*n* = 23	57.1 (21.6)
Baseline physical activity (hours/week), mean (SD)	93*	2.1 (4.0)	21	1.8 (3.2)
Body Mass Index (BMI), mean (SD)	96	25.4 (6.4)	20	30.0 (7.5)
Dominant side injured	100	24% (24/100)	22	36% (8/22)
Duration of symptoms (months), mean (SD)**	87	39.5 (55.3)	18	14.4 (14.0)
Sex (female)	99	59% (58/99)	23	78% (18/23)
Highest Level of Education, *n* (%)	**89**		**20**	
No high school diploma		3% (3/89)		5% (1/20)
High school diploma		14% (12/89)		20% (4/20)
Some college/university		11% (10/89)		15% (3/20)
College diploma / professional degree		34% (30/89)		40% (8/20)
University degree		38% (34/89)		20% (4/20)
Income (per annum), *n* (%)	**81**		**13**	
Less than $40,000		35% (28/81)		33% (4/12)
$40,000 to less than $80,000		28% (23/81)		33% (4/12)
$80,000 or more		37% (30/81)		41% (5/12)
Married/common law, *n* (%)	96	59% (57/96)	19	84% (16/19)
Severity of Injury, *n* (%)	**100**		**24**	
No tear		39% (39/100)		33% (8/24)
Partial tear		23% (23/100)		21% (5/24)
Full thickness tear		38% (38/100)		46% (11/24)
Smoking status, *n* (%)	**50**		**14**	
Non-smoker		80% (40/50)		71% (10/14)
Ex-smoker		14% (7/50)		21% (3/14)
Current smoker		6% (3/50)		7% (1/14)
Traumatic injury, *n* (%)	99	45% (45/99)	21	29% (6/21)
Work status (working), *n* (%)	91	58% (53/91)	18	33% (6/18)

*The baseline physical activity data from one participant was removed due to an extreme outlier (99 hours per week), which reduced the response from 94 to 93 participants.
**Control group has longer symptom pre-treatment, due to COVID-19 delays and hesitancy with seeking treatment during COVID-19. Injury severity was comparable across groups.

Table 2.1. Baseline Demographics Data

Power calculations were conducted to determine the appropriate sample size, as outlined in our preregistered research protocol (Burns et al., 2020). For the intervention group, power analysis specific to this study indicated that 20 participants were sufficient to detect a 15 percentage-point difference in participation rates at a two sided $\alpha = 0.05$. We enrolled 24 participants in the intervention group, exceeding this requirement. For the control group, a related clinical validation study (Boyer et al., 2025) enrolled 100 control participants to achieve 80% power to detect weak correlations (R > 0.25) between physiotherapy participation and clinical outcomes at a two-sided $\alpha = 0.05$. We employed this cohort

as our control group ($n = 100$), although 20 participants would have also been sufficient. Power calculations are based on participant counts (not repeated observations). The longitudinal data (9,310 patient-days) increases the precision of estimates for within-person trajectories and treatment-by-time interactions in our mixed-effects models, but it does not affect the a priori power for detecting the between-group main effect. Figure 2.3 presents the patient flow diagram. From the 141 participants recruited (to account for expected attrition), 17 withdrew before their first therapy session, and 92 completed at least eight weeks of treatment.[2]



i	*n*	Change	Retention	*n*	Change	Retention
1	100		100%	24		100%
2	98	-2	98%	24	0	100%
3	96	-2	96%	23	-1	96%
4	92	-4	92%	23	0	96%
5	89	-3	89%	23	0	96%
6	83	-6	83%	21	-2	88%
7	77	-6	77%	21	0	88%
8	72	-5	72%	20	-1	83%
9	69	-3	69%	20	0	83%
10	60	-9	60%	18	-2	75%
11	46	-14	46%	16	-2	67%
12	39	-7	39%	5	-11	21%

Figure 2.3. Flow Diagram of Enrolled Participants in the Study

2.3.4. Data Collection

SPARS is an in-house technology designed to provide greater analytical depth for exercise research (Burns et al., 2018). It uses machine learning to analyze inertial data from a smartwatch's sensor to detect when patients perform prescribed exercises from a predefined list (Boyer et al., 2021). Once an exercise is identified, SPARS records its duration and time stamps the activity, which provides the *dependent variables* for testing ***H1*** and ***H2***. The time stamps also allow us to capture within-week changes in co-production, measured by the days since the last visit, and between-week changes, measured by the number of weekly physiotherapy session attended. These variables serve as the *independent variables* for ***H1*** and ***H2***. In addition, SPARS evaluates exercise quality by comparing accelerometer and gyroscope data against personalized baselines established during each patient's most

2 Field experiments in operations management vary widely in sample size, with published studies ranging from fewer than 50 participants to several hundred or more (Gao et al., 2023, Table 2), placing our study within established norms.

recent clinical visit (Burns et al., 2022). A deep learning model generates a percentage similarity score that quantifies how closely each movement matches the patient's personalized baseline. This generates the *dependent variables* for ***H3*** and the quality feedback used in the intervention. Prior adherence studies have typically relied on self-reported exercise logs (Bassett, 2003; Bonnechre et al., 2016; Forkan et al., 2006; Henry et al., 1999; Wright et al., 2014), which are prone to self-reporting bias (Nicolson et al., 2018), and limits the granularity of data. We have validated SPARS on the participants in our control group and achieved an area under the receiver operating characteristic curve (AUROC) of 98.5% for exercise identification (Boyer et al., 2023; Burns et al., 2021).

Each participant was provided with a smartwatch to be worn on their affected extremity when performing all prescribed physiotherapy exercises. This includes all supervised physiotherapy sessions within clinics and any physiotherapy exercise sessions conducted by the patient independently at home. The smartwatches purchased for this study have a 6-axis inertial measurement unit (3-axis accelerometer, 3-axis gyroscope). The inertial sensors sampled at 50 Hz and recorded to the internal memory storage on the smartwatch. The smartwatch permits up to 150 hours of sensor data recording to its 4GB internal memory. The sensor data was transmitted via Bluetooth to an Android tablet and automatically uploaded to a secure Google Cloud server. For the control group, data was retrieved during weekly physiotherapy visits. For the intervention group, data was collected in real time using the tablet equipped with PT+ provided to each patient.

Demographics attributes have been found to be potentially significant predictors of exercise adherence (Bassett, 2003; Chester et al., 2013, 2018; Jack et al., 2010; Jin et al., 2008; Raman et al., 2017; Razmjou et al., 2017), for which we collected data to be used as *control variables*. Such controls ensure that differences between cohorts can be attributed to the intervention rather than to confounding factors. The variables measured are: age; sex; body mass index (BMI); dominant side involvement; symptom duration; mechanism of injury (traumatic or degenerative); rotator cuff tear thickness and size; type of operative procedures; comorbidity (cumulative illness rating scale) (de Groot et al., 2003); smoking status; alcohol intake; opioid intake; cannabinoid intake; baseline level of physical activity; education; marital status; job demands; socioeconomic status (current income); perceived social support (Enhancing Recovery in Coronary Heart Disease Social Support Inventory) (Mitchell et al., 2003);

patient self-efficacy (2-item Pain Self-Efficacy Questionnaire) (Nicholas et al., 2015); Patient Expectation Questionnaire (Razmjou et al., 2008); and Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983). This data was collected for each participant pretreatment. Appendix B details our selection of control variables.

We completed qualitative interviews with patients and a physiotherapist to better understand the lived experience of those interacting with SPARS/PT+ through this study. The interviews were guided to cover individual experiences throughout the trial, challenges that arose during their participation, and ethical concerns or considerations regarding future commercial rollouts. In total, 5 interviews were used to interpret the results of this study. They were recorded and transcribed, totaling 124 pages of double-spaced transcripts. Appendix C provides details of the qualitative data collection.

2.3.5. Analysis Method

2.3.5.1. Analysis of Input Quantity

The input quantity analysis evaluates the relationship between the intervention, time in the trial, and the quantity and consistency of exercises performed, thereby testing ***H1*** (input quantity) and ***H2*** (sustained inputs over time). Exercise participation was quantified using two measures: *exercise frequency*, representing the proportion of days when exercises were completed; and *exercise duration*, representing the number of minutes of exercise per day. Both outcomes are analyzed using mixed-effects regression models, each controlling for month and year fixed-effects, participant-level random effects, and our selected demographic and psychometric control variables (Appendix B). The models employed can be expressed as:

$$y_i = \beta_{0i} + \beta_1 T_i + \beta_2 W_i + \beta_3 (T_i \times W_i) + \beta_4 D_i + \beta_5 Y + \beta_6 M + \boldsymbol{\theta C}_i + \epsilon_i. \quad (1)$$

The fixed effects include: β_1, the effect of the intervention on baseline participation adherence (T_i); β_2, the effect of the number of weekly physiotherapy session attended (W_i); β_3, the interaction effect between the intervention (T_i) and number of weekly physiotherapy session attended (W_i), capturing the effect of the intervention over time; and β_4, the main effect of the number of days since the last visit (D_i). Year fixed effects ($\beta_5 Y$) were included because the control and intervention cohorts were enrolled sequentially, with many control participants completing the trial before the intervention participants

began. However, due to COVID-19 related delays, there was a 16-month period of overlap during which 78 participants (70 control and 8 intervention patients) were enrolled concurrently. This overlap allows us to control for time fixed effects in our model. Month fixed effects ($\beta_6 M$) were also included to account for potential seasonal influences on exercise behavior (e.g., holiday season, summer vacation, etc.). Additionally, the model includes a vector of control variables ($\boldsymbol{C}_i$) with corresponding coefficients ($\boldsymbol{\theta}$). The model includes random intercepts to account for variability between participants and exercises. Specifically, $\beta_{0i} = \gamma_{00} + u_{0i}$, where u_{0i} represents the random intercept for each participant and γ_{00} represent the average intercept for all participants. Finally, ϵ_i denotes the residual error for each participant, assumed to be normally distributed with $\epsilon_i \sim N(0, \sigma^2)$.

2.3.5.2. Analysis of Input Quality

The input quality analysis examines the relationship between the number of exercise repetition and technique accuracy scores using log-linear mixed-effects modeling, thereby testing ***H3***. We began with exploratory analysis (see Appendix D) to inform our model specification, which revealed significant variation in accuracy scores between participants and across exercise types, supporting the inclusion of random effects to account for nested data structures. Specifically, individual accuracy measurements recorded during exercise set *repetition* (Level 1) are grouped by *participant ID* (Level 2) and by *exercise type* (Level 3). Measurements from the same participant are likely correlated due to repeated observations within individuals while measurements from the same exercise type are likely correlated because they share similar movement difficulty. A random slope for *injury severity of injury* per *exercise type* is also added since the effect of *injury severity* on technique accuracy may differ across exercises. Because learning in repetitive tasks typically follows a curve with rapid initial improvement and diminishing gains over time, prior research recommends log-linear modeling to capture this relationship (Anzanello & Fogliatto, 2011). Accordingly, we applied a log transformation to the exercise *repetition* variable in specifying our mixed-effects model. We model the relationship between technique accuracy (y_{ij}) and the number of sets of exercise *repetition* (x_{ij}) completed by participant (i) for a given exercise type (j) based on Eq. (2).

$$y_{ij} = \beta_{0ij} + \beta_1 log(x_{ij}) + \beta_{2j} S_i + \boldsymbol{\theta C}_i + \epsilon_{ij}. \quad (2)$$

The fixed effects include: β_1, the effect of the log-transformed number of exercise *repetition* (x_{ij}). To model the effect of *injury severity* (S_i), we specify $\beta_{2j} = \beta_2 + u_{2j}$, where β_2 is the average fixed effect of *injury severity* and u_{2j} is a random slope capturing the variable effect of *injury severity* across exercise types. Additionally, the model includes a vector of control variables ($\boldsymbol{C}_i$) with corresponding coefficients ($\boldsymbol{\theta}$), accounting for demographic and psychometric factors. The model also includes random intercepts to account for variability between participants. $\beta_{0ij} = \gamma_{00} + u_{0i} + u_{0j}$, where u_{0i} represents the random intercept for participant (i), u_{0j} represents the random intercept for exercise (j); and γ_{00} represents the average intercept for all participants. Finally, ϵ_{ij} denotes the residual error for each observation, assumed to be normally distributed with $\epsilon_{ij} \sim N(0, \sigma^2)$.

As highlighted earlier, this analysis is only performed for the intervention cohort due to several methodological challenges. First, the control group data was used to pre-train and validate the technique adherence model in PT+, introducing potential data leakage bias. Second, the PT+ application provides follow-along video instructions, allowing for precise classification of each recorded movement by matching it to the video the participant was watching. While SPARS can retrospectively classify exercise types for control group data, some movements are too similar to distinguish accurately, requiring grouping of movement patterns and introducing potential errors (Boyer et al., 2023). This limitation does not affect the input quantity measures where identifying the precise exercise type is not required, but it introduces potential errors in the input quality measures.

2.3.5.3. Analysis of Semi-structured Interviews

We gain deeper insight into participants' experiences with the intervention by qualitatively analyzing semi-structured interviews on four patients and the physiotherapist who delivered most of the clinical sessions in the study. Three patients were selected from those in the intervention group who completed all 12 weeks of physiotherapy, and one patient exited the program after the 9th physiotherapy session. All interviews were conducted in person by a member of the research team formally trained in qualitative interviewing and qualitative research methods. Each session lasted approximately one hour and followed a semi-structured guide that invited participants to discuss their experience with the SPARS/PT+ technology, its perceived usefulness, and their concerns or opinions regarding privacy, data

use, and the ethics of AI-enabled health monitoring. Interviews were audio-recorded and transcribed verbatim.

A second research team member, also trained in qualitative methods, who had no prior relationship with participants and was not involved in the technology's development nor the trial's implementation, conducted the qualitative analysis to reduce potential bias. This team member followed a grounded theory approach, beginning with open coding of the transcripts to identify recurrent concepts and patterns. These initial codes were then refined into focused codes and consolidated through an iterative process into emergent themes. Peer debriefing sessions were used to validate and finalize the thematic structure.

Throughout the analysis, qualitative insights were triangulated with quantitative data, particularly patterns observed in adherence behavior and clinical outcomes. For example, themes related to patient motivation and accountability were interpreted alongside adherence metrics from the smartwatches' data. This integration helped to contextualize and validate our interpretations. Appendix C provides a detailed description of the qualitative analysis, including sample interview questions and the development of thematic codes.

2.4. Results

2.4.1. Input Quantity

H1 posits that participants who receive off-site co-production monitoring with real-time quality validation will produce a higher quantity of inputs than those who do not; in other words, participants in the intervention group will complete more home exercises than those in the control group. Co-production quantity of inputs was measured using two dependent variables: (1) *exercise frequency*, defined as the proportion of days participants completed any exercise; and (2) *exercise duration*, defined as the average number of minutes of exercise participants completed per day. We first analyzed univariate results for the continuous variable (*exercise duration*) to estimate an effect size, followed by mixed-effects models for both measures of input quantity.

Univariate analysis comparing average daily *exercise duration* per participant throughout the 12-week study using Welch's t-test shows that the intervention group did exercise significantly more than

the control group. On average, the intervention participants exercised for 7 minutes and 51 seconds per day, compared to 4 minutes and 3 seconds for the control group (t(29.752) = 4.06, $p < 0.001$). This difference represents a large effect size (Cohen's d = 1.11 [95% CI: 0.64, 1.58]) and provides support for ***H1***. However, since patients were assigned to the intervention group after control participants were enrolled, the results may be affected by time-related confounding.

	Exercise frequency			**Exercise duration**		
Models	(1) **All**	(2) **Control**	(3) **Intervention**	(4) **All**	(5) **Control**	(6) **Intervention**
Intervention (baseline effect)	-0.369 (0.473)			-0.701 (1.206)		
Days since last physio visit	-0.055*** (0.004)	-0.059*** (0.004)	0.006 (0.017)	-0.030*** (0.003)	-0.036*** (0.004)	0.155** (0.052)
Number of visits	-0.195*** (0.012)	-0.195*** (0.012)	-0.117*** (0.034)	-0.188*** (0.026)	-0.192*** (0.025)	0.640*** (0.122)
Intervention × Number of visits (time varying effect)	0.071** (0.026)			0.687*** (0.058)		
Age	0.016+ (0.009)	0.027* (0.012)	0.013 (0.011)	0.043+ (0.023)	0.048+ (0.029)	0.062 (0.050)
Sex (Female)	-0.249 (0.267)	-0.327 (0.303)	-0.586 (0.630)	-0.949 (0.702)	-0.813 (0.737)	-1.614 (2.849)
BMI (Body Mass Index)	0.002 (0.020)	0.005 (0.024)	0.037 (0.028)	0.031 (0.053)	0.025 (0.058)	0.174 (0.132)
PSEQ2 (Pain Self-efficacy)	0.032 (0.041)	0.017 (0.047)	0.072 (0.076)	0.141 (0.108)	0.074 (0.114)	0.303 (0.356)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects			**Variance**			
Participant	1.637 (1.279)	1.826 (1.351)	0.587 (0.766)	11.62 (3.408)	11.07 (3.327)	14.34 (3.787)
Residual				28.94 (5.380)	24.92 (4.992)	49.27 (7.019)
Num.Obs	9,310	7,946	1,364	9,310	7,946	1,364
AIC	9,357	7,733	1,583	58,201	48,492	9,252
BIC	9,542	7,894	1,693	58,394	48,660	9,367

$+ p < 0.1$, $* p < 0.05$, $** p < 0.01$, $*** p < 0.001$

Table 2.2. Regression Coefficient Estimates for Exercise Frequency and Quantity

Next, we conduct multivariate analyses using mixed-effects models, which account for this potential bias, offering a more robust test for ***H1***. Results are summarized in Table 2.2. Models 1–3 analyze *exercise frequency* using mixed-effects logistic regressions to model the binary outcome of whether exercise was performed on a given day. Models 4–6 analyze *exercise duration* using mixed-effects linear regressions to model exercise duration per day as a continuous variable. Models 1 and 4 include all participants, while Models 2, 3, 5, and 6 show separate regression estimates for the control

and intervention groups.

For both measures of co-production input quantity, the baseline effect of receiving the intervention was not statistically significant (Model 1 [*exercise frequency*]: $\beta = -0.369,\ p = 0.44$; Model 4 [*exercise duration*]: $\beta = -0.701,\ p = 0.10$). This indicates that both cohorts produced a similar quantity of inputs at the start of the trial. While univariate analysis supports ***H1***, the multivariate models suggest that between-group differences in exercise duration only emerges over time. Thus, ***H1*** receives partial support as we had expected an immediate intervention effect.

Since differences in exercise behavior can only have emerged over time, we next examine ***H2***, which posits that participants who receive off-site co-production monitoring with real-time quality validation will maintain higher quantity of input over time than those who do not. That is, intervention participants will sustain higher exercise participation throughout the treatment period compared to those in the control group. For both measures of co-production input quantity, the significant positive interaction between intervention and number of visits (Model 1 [*exercise frequency*]: $\beta = 0.071, p = 0.006$; Model 4 [*exercise duration*]: $\beta = 0.687,\ p < 0.001$) indicate that the intervention reduces the decline in participation over time, providing support for ***H2***.

Comparing the regression coefficients for *days since last physio visit* for the control group (Models 2 and 5) and the intervention group (Models 3 and 6), we observe that the within-week decline detected in the control group (Models 2 [*exercise frequency*]: $\beta = -0.059,\ p < 0.001$; Model 5 [*exercise duration*]: $\beta = -0.036,\ p < 0.001$) is not present in the intervention group (Models 3 [*exercise frequency*]: $\beta = 0.006,\ p < 0.73$; Model 6 [*exercise duration*]: $\beta = 0.155,\ p = 0.03$). Hence, intervention participants sustained their participation throughout the week, while control participants decreased their participation. Contrary to expectations according to ***H2***, whereby intervention participants would sustain their exercise duration over time, those participants actually increased the time spent exercising as more days passed since their last appointment. This unexpected positive effect could indicate a tendency to increase co-production inputs in the days leading up to their next visit, possibly motivated by a sense of accountability associated with the intervention.

Comparing the regression coefficients for *number of visits*, we observe that both cohorts experience some week-over-week decline in *exercise frequency*, albeit 40% less for the intervention group (Models 3: $\beta = -0.117$, $p < 0.001$) than for the control group (Models 2: $\beta = -0.195$,, $p < 0.001$). This signifies that both cohorts had a tendency of skipping more exercise days week-over-week, but the decline was considerably smaller among intervention participants. Regarding *exercise duration*, the control cohort tended to decrease their time spent exercising week-over-week (Models 5: $\beta = -0.192$, $p < 0.001$) while the intervention group increased their time spent exercising (Models 6: $\beta = 0.640$, $p < 0.001$). The positive between-week coefficient reflects expected adherence, since patients are prescribed progressively more exercises at each physiotherapy session.

Altogether, the results within-week and across-weeks indicate that the intervention successfully moderates the decline in co-production inputs over time observed in the control group, providing strong support for ***H2***. Appendix E shows that these findings remain consistent across specifications with and without demographic controls, with and without year and month fixed effects, when restricting the sample to the 78 participants enrolled during the overlap period between cohorts[3], and when using pairwise matching to address the fact that intervention assignment was not perfectly randomized. However, the lower AIC and BIC values confirm that control variables and time fixed effects are appropriate for improving model fit without overfitting. Additionally, the unexpected positive effect of *days since last physio visit* on *exercise duration* warrants further investigation, which we later address with interviews.

2.4.2. Input Quality

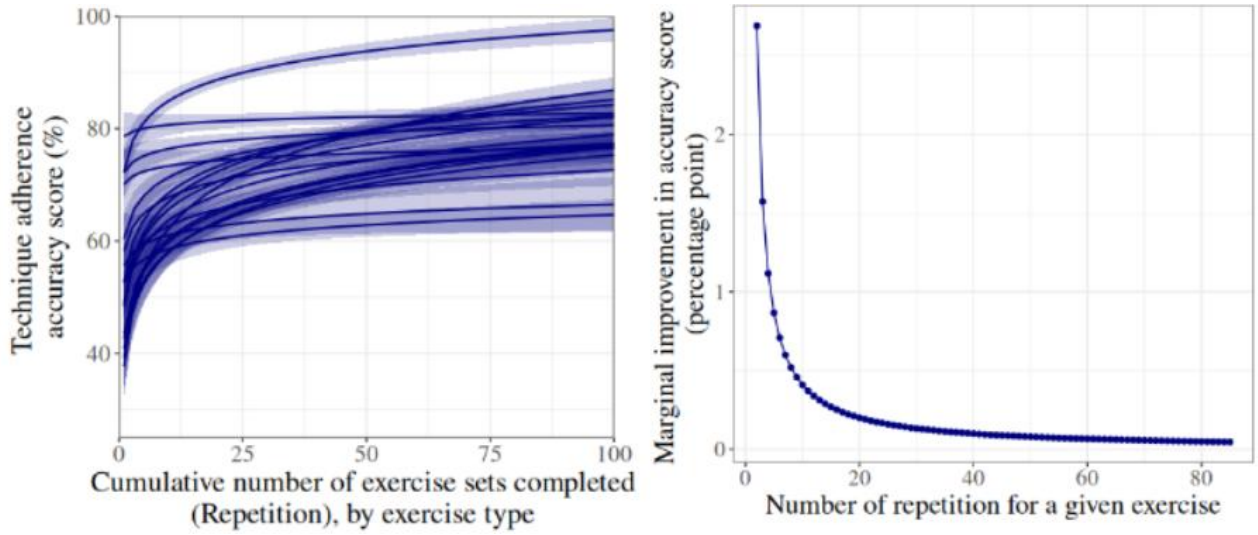
H3 posits that participants who receive real-time quality validation during co-production will exhibit increases in input quality with practice following a learning-curve. We measure co-production input quality in the intervention cohort (24 participants performing 18 distinct shoulder exercises) as the average *accuracy* score for each repetition within an *exercise set* (typically 10 repetitions per set). In total, our dataset contains 15,032 records of *exercise sets*. To test ***H3***, we model the relationship between

[3] The exercise frequency model did not converge in this subsample due to sample size limitation (only 8 intervention participants); therefore, only exercise duration results are reported.

the number of sets completed and technique *accuracy* using log-linear mixed-effects regressions. Results are presented in Figure 2.4, which shows (a) learning-curve trajectories by *exercise type*, and (b) the average marginal improvement in *accuracy* as a function of repetitions. As panel (b) indicates, improvements are most pronounced during the first 10 *exercise sets*, after which gains diminish to approach a plateau.

The data substantiates our learning curve hypothesis (***H3***): repetition is a primary factor in input quality. Table 2.3 presents the mixed-effects regression results, which confirm the statistical significance of this effect. While covariates such as *injury severity*, *age*, *body mass index (BMI)*, *gender*, and *pain self-efficacy questionnaire* (*PSEQ2*) were included to control for potential confounders, none showed statistically significant effects. The high negative correlation between the *exercise type* intercept and *injury severity* slope (-0.87) suggests that exercises with high baseline accuracy are less affected by *injury severity*. The logarithmic relationship between *exercise sets* completed and technique *accuracy* ($\beta = 3.88$, $p < 0.001$) indicates that marginal increase in quality is greatest early in the trial when participants struggle to perform movements accurately. Thus, the benefits of receiving real-time quality validation is primarily beneficial in the early stages of co-production. For example, from Figure 2.4(b), after completing seven sets of a given exercise, participants experience a marginal improvement of less than 0.5 percentage point compared to a 2.67 percentage point improvement between the first and second exercise set, amounting to an 81% decrease in learning rate within the first week. Because technique *accuracy* plateaus while exercise participation continues to rise week-over-week in the intervention group (Table 2.2), the higher input quantity cannot be explained by corrective feedback alone. This suggests that the real-time feedback feature alone does not account for the higher participation observed in later stages of co-production among intervention participants.

For robustness, we compared mixed-effects regression models with and without a log transformation of the repetition variable to validate the use of a log-linear model. The results, presented in Appendix F, indicate that the logarithmic model provides a better fit, as demonstrated by lower values for the REML, AIC, and BIC criteria (Verbyla, 2019).



(a) Learning curve by exercise type (18)

(b) Average marginal improvement in accuracy score as a function of exercise repetitions

Figure 2.4. Input Quality Learning Curves

Fixed Effects	Estimate (β)	SE	df	t-value	p-value
(Intercept)	50.88	25.87	13.4	1.97	0.0703
log(exercise sets)	3.88	0.19	15.04	20.34	<0.001***
Injury severity	-4.94	4.82	15.45	-1.03	0.3213
Age	-0.02	0.18	12.92	-0.12	0.9044
BMI	0.36	0.51	12.87	0.71	0.4924
Sex	-4.76	9.81	12.85	-0.49	0.636
PSEQ2	1.36	1.4	12.96	0.98	0.346
Random Effects		**Variance**		**Std. Dev.**	**Correlation**
Participant ID (Intercept)		220.41		14.85	-
Exercise type (Intercept)		251.02		15.84	-
Injury severity (Slope) × Exercise type		36.67		6.06	-0.87
Residual		491.49		22.17	-
Metric					**Value**
Number of Observations					15,032
Number of Participants					24
Number of Exercises					18

$+ p < 0.1$, $* p < 0.05$, $** p < 0.01$, $*** p < 0.001$

Table 2.3. Mixed-effect Regression Model of Exercise Repetition on Technique Accuracy

2.4.3. Mechanisms Driving Behavioral Change

A qualitative analysis of our semi-structured interviews enables us to identify the mechanisms that help explain why the intervention sustained higher co-production input quantity over time, even after intervention participants had reached technique proficiency. From this analysis, three themes emerged from this analysis: patients gained (1) uncertainty reduction from real-time feedback; (2) intrinsic accountability from progress tracking; and (3) extrinsic accountability from data sharing with physiotherapists. We next elaborate on these themes, each of which representing behavioral mechanisms through which the intervention influences co-production behavior.

Uncertainty reduced from real-time feedback: Participants valued the immediate feedback on input quality provided by PT+, which helped them to adjust their technique. They noted that the real-time indicators (categorized as "green", "yellow" or "red" zones) helped them self-correct and refine their technique. As one participant described, the indicators provided clarity that was previously lacking: "*I like that aspect, it's giving immediate feedback. Before, when you're doing the exercises, well, are you doing it quite correctly? Think so, but don't know"* (Patient A).

While AI-driven feedback was said to be valuable, participants stressed the need to balance technology with human interaction. They found AI feedback helpful for reducing uncertainty but appreciated supplementing it with explanations from their physiotherapist. As Patient B explained, *"The technology is good, and it certainly has acted as a tool to motivate. But personal interaction is extremely important"*. This shows that while AI feedback reduced uncertainty, sustained engagement depended on complementing technology with professional oversight. For example, pain and discomfort often acted as barriers to exercise, but the physiotherapist's assessment of both technique accuracy data and in-clinic progress reassured patients that continuing was safe and worthwhile, particularly when they felt uncertain about their technique. As one participant described: *There's only one [exercise] that I'm having a hard time [with], and I told [the physiotherapist], 'Can I just skip this one?' 'No, you can't skip that one.' Just doing it [hurts] me. But he looked at me and said, 'You're [making] progress there'"* (Patient C).

The digital technology also facilitated retrospective review with the physiotherapist to assess technique accuracy together. With it, patients felt engaged with their physiotherapist throughout the week rather than only during visits. As one participant described, the technology created a sense of continuous oversight: *"It's like you have somebody overseeing what you're doing on a daily basis.[...] With this, they're able to see how correctly I'm doing the exercise"* (Patient D). Participants' descriptions of improved technique over time are consistent with observed trends in Table 2.3. As one participant noted, "*I like the fact that my exercise improves, [and] my methodology of doing that exercise improves from one week to the next because of the way that I'm monitored"* (Patient D).

Intrinsic accountability gained from progress tracking: The tracking of co-production input quantity was a second important feature mentioned in the interviews. Participants noted that this aspect

of the PT+ application fostered a sense of accountability, which in turn increased their motivation. For many, the technology served as a reinforcement tool to help them stay on track despite busy schedules. One patient stated: *"There's been times [when] I don't feel like doing this today, but I end up doing it anyway because I made the commitment. I've got the equipment here. It has acted as a motivator"* (Patient B). Unlike their past physiotherapy experiences where competing priorities led to inconsistent exercise adherence, the PT+ application gave them a tangible nudge to stay engaged. This explains why participants in the intervention cohort consistently maintained higher co-production inputs throughout the 12-week trial while participation gradually declined in the control cohort. As one participant explained, "*I'm one of these personalities that might need to be monitored a bit, because it increases my competitiveness and increases my desire to achieve"* (Patient D).

Extrinsic accountability gained from data sharing: Both patients and the physiotherapist noted that sharing co-production performance data with the physiotherapist provided additional benefits. Since exercises had to be completed to be recorded, patients felt a sense of duty and pride in sharing their results. The physiotherapist observed that the desire to avoid disappointing them increased the quantity of co-production inputs: *"That sense of [the physiotherapist] watching is a very good motivation for them to exercise. As a physiotherapist, I see very good results because [patients] are doing the exercises"* (Physiotherapist). The expectation of reporting back to the physiotherapist served as an external motivator. As one participant explained, *"I can see my progress. And also, it's reminding me that I have no excuse—I have to do it because if I skip it, I have to tell [my physiotherapist], 'Sorry [physiotherapist], I did not do [it] for how many days, I was sick'"* (Patient C).

2.5. Discussion and Conclusion

This study examines how an AI-enabled monitoring system influences co-production in physiotherapy, a professional service where critical inputs (prescribed home exercises) are performed outside provider oversight. We find that this digital technology increases input quantity by sustaining exercise inputs over time, but contrary to our initial expectation, intervention participants did not immediately begin with higher input quantity. Instead, the intervention helped sustain the frequency and duration of inputs while the control group's participation declined over time. With respect to input quality, technique

improved with practice and the marginal benefit of feedback diminished as mastery was approached, consistent with a learning-curve effect. Although direct comparison with the control cohort was not possible, observed improvements exceeded those typically reported in the literature (Faber et al., 2015). From a service operations perspective, these findings show that AI-enabled monitoring can extend process control to improve off-site co-production. Nonetheless, establishing the intervention's effectiveness falls short on clarifying the behavioral mechanisms through which these effects arise.

Our qualitative interviews provide insights into these mechanisms, which align with established behavioral theories of self-determination, and accountability, by suggesting that the digital technology increased co-production inputs not only through automated feedback but also by reinforcing provider–client relationships. We uncover three distinct behavioral mechanisms that influenced participants' engagement throughout the study. First, uncertainty reduced through real-time feedback helps explain the improvement in technique and partially account for the higher exercise completion rates. Consistent with feedback-based frameworks of self-regulation (Carver & Scheier, 1981, 1982), the AI's real-time indicators acted as a feedback loop, allowing patients to compare their performance against a standard and correct errors in the moment. This in-process quality control mechanism reduced uncertainty and made goals feel attainable. Second, intrinsic accountability from progress tracking fostered sustained participation by visualizing progress and habitualizing daily routines. This aligns with self-determination theory (Ng et al., 2012; Ryan & Deci, 2000), which posits that motivation is sustained when activities support autonomy (control over one's actions) and competence (feeling effective). Progress tracking encouraged goal commitment and transformed co-production into a self-regulated process. Third, the sharing of data with physiotherapists demonstrates how AI-enabled monitoring with human oversight can function as a powerful hybrid control system. The desire to show good results to therapists aligns with social accountability theory (Lerner & Tetlock, 1999), which suggests that individuals are influenced by anticipating the need to justify their actions to others. This explains why participation remained high even after corrective feedback benefits diminished and why some patients intensified their efforts before appointments. Additionally, the digital technology reinforced the patient-therapist relationship, fulfilling the relatedness need described in self-determination theory (Ryan &

Deci, 2000). The technology, therefore, improved co-production not just through automated feedback but by strengthening the human-centric aspects of the service.

Combining these qualitative insights with our mixed-effect regression results, this study advances operations management theory in three domains: service operations, behavioral operations management, and technology management in services. First, we contribute to service operations by showing how digital technologies can extend process control into the off-site domain where customer inputs are typically invisible and difficult to manage. Prior research stresses that variability in co-production inputs is a major source of operational risk (Roels, 2014; Sampson & Froehle, 2006). Our findings demonstrate that AI-enabled monitoring can reduce this variability, thereby improving service performance and outcomes.

Second, we extend behavioral operations management research by illustrating how accountability can be digitally encouraged. Prior work highlights the importance of behavioral factors such as monitoring and accountability in operational contexts (Croson et al., 2013; Donohue et al., 2020). In line with this perspective, our study shows how digital monitoring and goal tracking mechanisms can sustain engagement. Our quantitative and qualitative evidence show that intrinsic accountability (progress tracking) and extrinsic accountability (anticipation of reporting to a provider) both sustain engagement over time. This suggests that technology can complement traditional social mechanisms of control in co-production processes.

Third, we contribute to the technology management literature by adding to the debate on whether AI should function as an independent agent or as an assistant to humans (Angelopoulos et al., 2023; Bianchi & Brière, 2024; Koehler & Sauermann, 2025). While SPARS/PT+ was originally conceptualized as a quality-monitoring tool to reduce errors and accelerate learning (as an independent agent), the lack of baseline participation differences indicates that the intervention's benefits did not stem from reducing input quality uncertainty alone but instead from multiple behavioral mechanisms, including supporting the therapist–patient relationship. In our study, AI could deliver continuous monitoring, progress tracking, and real-time feedback, while physiotherapists provided judgment, reassurance, and accountability support. This human–AI complementarity contrasts with perspectives that envision AI systems as autonomous agents (Angelopoulos et al., 2023). Together, these insights

improve our understanding of how AI technologies can reconfigure task allocation in professional service operations.

Beyond theoretical contributions, our findings enable us to suggest design principles for service managers seeking to sustain customer contributions in co-production settings. These principles are grounded by the behavioral mechanisms identified in this study. We outline three principles:

- *Design Principle 1: Make invisible work visible.* A common challenge in professional services is that off-site customer contributions remain unobservable until the next interaction. Our intervention shows that even simple tracking can generate accountability in the days leading up to the next service visit. For managers in education, consulting, healthcare, or financial services, systems that track and bring awareness to customer progress can transform invisible work into visible commitments that reinforce accountability without requiring constant provider interventions.
- *Design Principle 2: Provide real-time quality feedback to reduce uncertainty.* Uncertainty about whether tasks are being performed correctly can lead to dropout or reduced effort. Participants reported that AI feedback reassured them that they were "doing it correctly," lowering psychological barriers to participation. Service managers should identify moments of uncertainty in their customer journey and introduce timely feedback mechanisms, whether algorithmic signals, structured checklists, or human interventions, that reduce uncertainty and build confidence in task execution.
- *Design Principle 3: Leverage technology to improve human relationships.* While AI feedback was valued, participants emphasized that the expectation of reporting to a physiotherapist was a factor sustaining co-production efforts. For service managers, this highlights the importance of designing digital tools that augment rather than replace human touchpoints. Value emerges when automation handles routine tasks that allows for better provider–client interactions, enabling more personalized and engaging service experience.

Notwithstanding its theoretical and practical implications, our study also has limitations that open doors for future research. One limitation concerns the sequential cohort study design. While necessary for AI training and validation, it raises the possibility of unobserved time effects. We mitigate this concern with year and month fixed effects, but future research should replicate findings in a randomized trial design. Another limitation is that AI systems require substantial training and validation data, which resulted in larger control than intervention cohort. However, sensitivity checks (Appendix E) indicate that results are robust to this asymmetry. Conducting the study at a single site may also limit generalizability, though insights derived here are likely relevant to other professional service contexts.

Similarly, while our qualitative interviews provide valuable insight into the behavioral mechanisms driving higher rates of co-production, their generalizability may be limited by the modest sample size and potential overrepresentation of engaged participants. Nevertheless, the quantitative results show significant outperformance among most intervention patients relative to the control cohort, not only those we interviewed, supporting the validity of the findings. Also, an interview with the physiotherapist who guided most of the 124 patients corroborated the perspectives shared by the interviewed patients.

In addition, our study focused on a 12-week rehabilitation period. Discussions with physiotherapists indicate that the first eight weeks of treatment typically deliver the greatest impact, suggesting that our study period already exceeds practical standards. Nevertheless, the longer window does not establish whether co-production benefits persist once formal rehabilitation ends, an important direction for future medical research. Establishing motion tracking technology in medical practice will also require broader validation across populations and medical conditions, as SPARS/PT+ remains an early-stage system trained and validated for a single injury type (shoulder injury). To facilitate such progress, much of our technology has been open-sourced, with training data available at https://github.com/dmbee/SPAR-dataset and code available at https://github.com/dmbee/seglearn. Finally, adoption of technologies such as SPARS/PT+ will also depend on structural and economic incentives. For example, in fee-for-service healthcare systems, providers may not be rewarded for efficiency gains from reductions in billable in-person care (Andritsos & Tang, 2018). These broader institutional factors represent important considerations for scaling and sustaining AI-enabled co-production technologies.

Our findings are applicable beyond physiotherapy to a wide range of professional services where customer inputs are generated off-site and remain difficult to monitor. For example, prior research highlights similar challenges in other domains: in tax preparation, service delivery cannot begin until customers provide the required information (Sampson & Froehle, 2006); in auditing, audit quality suffers when clients delay or withhold requested evidence (Hatfield et al., 2025); in consulting, project outcomes depend heavily on the quality of client contributions to solution development (Fu et al., 2023); and in education, student learning is directly tied to the effort they invest outside the classroom (Sampson & Froehle, 2006). Across these contexts, service quality is compromised when customer-

provided inputs are late, incomplete, or inaccurate. By demonstrating how AI-enabled monitoring can make invisible work visible, provide timely feedback, and establish accountability structures, our study suggests design principles for reducing input variability in these settings. In operations management terms, this represents an opportunity to extend process control to stages of the service process traditionally beyond managerial reach. Future research should examine the scalability and cost-effectiveness of such systems, as well as the design trade-offs between automation and human oversight. A central question is how to balance the efficiency of algorithmic monitoring with the relational and motivational value of human interaction. Comparative studies across industries could identify the conditions under which hybrid human–AI processes produce the greatest operational benefits.

This study provides empirical evidence that AI-enabled monitoring can improve the quantity and quality of co-production inputs in professional service operations. Using a sequential cohort field study in physiotherapy, we show that the AI-enabled system we developed (SPARS/PT+) sustained higher co-production inputs over time and input quality improved following a learning-curve effect. By combining quantitative and qualitative evidence, we identified the mechanisms—uncertainty reduction, intrinsic accountability, and extrinsic accountability—that sustained engagement over the rehabilitation period. These findings extend service operations theory by showing how digital technology is a feasible solution to the problem of managing off-site customer inputs, a long-standing challenge in co-production contexts. For managers, our results offer practical design principles for leveraging AI to make invisible work visible, reduce execution uncertainty, and improve provider–customer relationships. Taken together, this research contributes to operations management by providing both theoretical insights and actionable guidance for designing hybrid service systems in which human and AI complement one another. As AI-enabled monitoring becomes more prevalent across industries, understanding how to integrate such tools into co-production processes will be central to improving service reliability, efficiency, and customer experience.

Chapter 3:
Exploring biotech-pharma alliance breakups following drug launches

3.1. Introduction

Biotech firms often rely on partnerships with large pharmaceutical firms that provide capital, specialized knowledge, and proprietary assets to navigate the resource-intensive and high-risk process of drug development (Hora & Dutta, 2013; Taneri & De Meyer, 2017; Tian et al., 2021). For biotech firms, alliances are essential for their financial viability and sustained operations. In contrast, pharma firms treat alliances as one component of a broader portfolio that hedges against the revenue shocks created by the "patent cliff", the predictable loss of exclusivity on drugs after patent expiry (Fisher et al., 2014). Recent evidence confirms that firms actively rebalance such collaborations, redirecting resources when the mix no longer matches strategic needs (Cooper et al., 2004; Krieger et al., 2022; Lugovoi et al., 2022). Consequently, pharma firms may terminate previously formed alliances to adapt to changing requirements. While alliance terminations are often studied as responses to external threats, such as shifting market conditions, misaligned goals, or competing internal projects (Asgari et al., 2018; A. Ghosh & Klueter, 2022; Pangarkar et al., 2017), less attention has been given to how a firm's successes influence alliance termination decisions. A drug launch by a pharma firm is one such example that alters the risk–return balance of its R&D portfolio. That is, it can prompt the termination of alliances, even in unrelated therapeutic areas, as firms reallocate resources in response to strategic reprioritization triggered by their own success. For the pharma partner, for whom each alliance is only one asset in a deliberately diversified R&D portfolio, such agility can entail withdrawing funding and clinical expertise from resource-dependent biotech firms. For biotech partners, however, alliances are a lifeline, meaning termination decisions can be devastating.

Research on asymmetric innovation alliances, such as biotech-pharma collaborations, has uncovered risks, like the opportunistic behavior of knowledge misappropriation, which can lead to alliance failure (Diestre & Rajagopalan, 2012; Gopalakrishnan et al., 2008; Hora & Dutta, 2013; Kalaignanam et al., 2007; Katila et al., 2008; McCutcheon & Stuart, 2000; Ramasubbu et al., 2019). However, there is a lack of empirical evidence on whether a pharma firm's success in an unrelated drug

can trigger an alliance termination with biotech firms it partners with. Public disclosures hint that such terminations may occur. Pfizer, for instance, decreased early-stage rare-disease partnerships after launching its COVID-19 vaccine, terminating several alliances, selling related assets, and shifting away from its prior emphasis on developing partnerships in the orphan drug market (Nawrat, 2019; Pfizer, 2015; Stynes, 2010; Taylor, 2023). For biotech startups, such termination can disrupt operations, tarnish reputation, and weaken financial performance (Eslami et al., 2023; Havenaar & Hiscocks, 2012).

This study addresses the following research question: *Do drug launches by a pharmaceutical firm increase the risk of alliance terminations for its biotech partners, and can such termination be distinguished from other types of terminations with information observable when alliances are initially formed?* Addressing this question allows us to examine whether the timing of termination contains a predictable signal. It also addresses the challenge of partner selection under uncertainty: if post-launch terminations are distinguishable using information available at alliance formation, then biotech firms may be able to anticipate which partners are more likely to deprioritize the alliance if success (i.e., launching a drug) occurs. Finally, it offers an evaluation of the structural risk that a partner's internal success poses to alliance stability.

A pharmaceutical firm's decision to terminate unrelated alliances with biotech partners after a drug launch can stem from several independent or interactive factors.[4] Rather than identifying specific mechanisms underlying such terminations, our focus is on detecting observable predictors of termination risk to help biotech firms anticipate and mitigate these vulnerabilities. Using a dataset of 1,649 terminated biotech–pharma alliances spanning 20 years from Clarivate's Cortellis database, we first confirm that terminations occur significantly more often after a drug launch than before. We then use survival analysis based on 3,496 alliances with known outcomes to show that a pharma partner's drug launch significantly increases the hazard of alliance termination. Next, we apply supervised machine learning (ML) to demonstrate that post-launch terminations are predictable from features observable at the time the alliance is formed. We also show that key statistical assumptions required for

4 Examples of factors include: (a) internal strategy to reconfigure its portfolio and optimize resource allocation; (b) evaluation of the opportunity costs associated with discontinuing alliances with biotech firms possessing certain capabilities; and (c) dynamics of the relationship between the two parties and the broader implications for the firms' network cohesion.

inference-based methods are violated, making a flexible data-driven approach more appropriate. Finally, a post-hoc test indicates that these terminations are unlikely to reflect opportunistic timing but instead reveal a distinct structural vulnerability.

Our first contribution is to the operations management literature on alliance formation and termination in innovation-intensive industries. Prior research highlights how alliance portfolio configurations, governance structures, and partner selection decisions influence collaboration outcomes under uncertainty (Dutta, 2013; Taneri & De Meyer, 2017; Tian et al., 2021). We extend this stream by showing that a partner's unrelated success can introduce a risk of termination for biotech firms. This highlights a potential dark side of social capital (Faems et al., 2010; Villena et al., 2011). Second, we contribute to the innovation alliance literature by demonstrating that terminations following drug launches are linked to partner characteristics that are observable at the time an alliance is formed. In particular, interactions among pharma partner attributes help explain which alliances are more vulnerable to a potential post-launch termination. Hence, scholars must consider the interdependencies of predictors surrounding biotech firms' partnership decisions (Subramanian et al., 2013). Third, we contribute to the innovation management literature (Dziallas & Blind, 2019; MacCormack & Mishra, 2015; Savva & Scholtes, 2014; Taneri & De Meyer, 2017) by showing that drug launches are a unique vulnerability for biotech firms. Specifically, we find that drug launches increase the hazard of alliance termination, with no evidence indicating that pharmaceutical firms strategically delay reporting terminated partnerships to coincide with the positive news associated with launching a new drug.

Calls for the increased use of ML models and explainable AI techniques in management research (Chou et al., 2023; Choudhury et al., 2021) prompted us to explore a problem that necessitates a prediction-focused solution methodology. While these tools are well-suited for addressing predictive questions, their application to theory-building research has been limited. To this end, we provide a rigorous demonstration of how ML can be used to analyze management problems characterized by high-order interactions, nonlinear effects, and the need to accurately forecast uncertain outcomes.

3.2. Background

3.2.1. Symbiotic Relationship in Biotech-pharma Alliances

Both biotech and pharma firms rely on drug development alliances to achieve innovation objectives (Hora & Dutta, 2013). The former typically engage in risky, long-term research and often depend on alliances with pharma firms to sustain their R&D operations. This dependence stems from their need for capital, regulatory expertise, and clinical trial management, resources that established pharma partner can provide (Baum et al., 2000). In line with the resource dependency perspective (Pfeffer & Salancik, 1978), biotech firms seek large pharma partners because they are better equipped to offer these essential resources. This is especially evident at the onset of clinical trials, when resource demands spike and alliance formation rates increase accordingly (D'Arville, 2004; Tyebjee & Hardin, 2004).

While biotech firms can seek capital from various sources to fund R&D projects (venture capital, IPOs, or private equity), pharma firms tend to offer more favorable valuations (Tyebjee & Hardin, 2004). Their domain expertise allows them to better assess a biotech project's potential and assign lower risk premiums. Other sources of innovation funding, like venture capital, also have investment horizons that are too short to fund early-stage drug development projects (Havenaar & Hiscocks, 2012). For a biotech firm, this means that an alliance with an established pharma firm can be formed while retaining a greater share of value in the venture and potentially securing a stable long-term partner.

For established pharma firms, forming alliances with biotech firms is vital to maintain a steady pipeline of new drugs. This is needed to sustain operations in an industry where profit margins evaporate following the expiration of patents (Fisher et al., 2014; Tyebjee & Hardin, 2004). In other words, pharma firms do not rely on individual alliances, but instead, depend on maintaining a robust portfolio of partnerships (Whittaker & Bower, 1994; Evens, 2016). Over the years, pharma firms have increasingly become buyers of external sources of innovation, allowing them to diversify their R&D activities into more therapy areas (Lugovoi et al., 2022; Yashiro et al., 2023). Based on this trend, the success and survival of many pharma firms depend on a steady and productive alliance portfolio to maintain a diverse pipeline of new R&D opportunities (Krieger et al., 2022).

Since pharma firms are not dependent on any single alliance and typically invest broadly to hedge

the risk of new drug development, they have a tendency to terminate alliances when the future expected returns of an underlying project decrease (Min, 2017; Taneri & De Meyer, 2017). Furthermore, a large alliance portfolio allows firms to form substitutable alliances, which reduces their dependency on any single alliance (Cui, 2013). This lower dependency reduces the likelihood of escalating or maintaining commitments as their portfolio grows (Pangarkar et al., 2017). The disposable nature of alliances by pharma firms is reflected in Krieger et al. (2022); they tend to divest resources when competitors experience adverse market conditions in similar therapy areas. Intuitively then, if individual alliances are disposable, it is logical that the launch of a drug offers an opportune time to recalibrate the alliance portfolio if it alters the firm's internal resource needs, managerial attention, risk-return balance of the remaining portfolio, or the firm's strategic priorities.

The dilemma lies in the fact that while alliances are mutually beneficial, they are also driven by each partner's self-interest, which can sometimes lead to unwanted terminations from the standpoint of one of the partners (McCarter & Northcraft, 2007; Sadowski & Duysters, 2008). For instance, Gan and Korsgaard (2022) observe that opportunistic behavior, such as knowledge misappropriation, by large firms allied with smaller partners can lead to terminated alliances despite apparent resource complementarity. Madhok et al. (2015) add that even the perception of inequity in how alliance benefits are distributed between partners can lead to alliances falling apart. However, even in partnerships that seem symbiotic, alliance terminations may occur if one partner senses a better opportunity elsewhere or when the preferences of partners is different from what is expected from the partnership (Hampl, 2020; Taneri & De Meyer, 2017). Our research contributes to this discussion by exploring the specific vulnerabilities faced by biotech firms when their pharma partner launches an unrelated drug, a risk that originates outside the partnership but is not driven by external market dynamics.

3.2.2. Alliance Terminations

While some terminations may be justified, ending an alliance typically has negative consequences for the underlying project forming the basis of the partnership (Bakker, 2016). For instance, terminations based on scientific merit can enhance innovation productivity, but only 17% of biotech-pharma alliances are terminated due to poor future prospects (Pangarkar, 2003). Furthermore, there is a risk that

a project is terminated before its merits can be adequately evaluated, which would contribute to the grand challenge of low innovation productivity in the biopharmaceutical industry (Pammolli et al., 2020; Paul et al., 2010). Given the substantial investment of time and capital, external disruptions to these alliances can lead to significant waste of resources. While some projects may still progress to launch, terminating the alliance will inevitably lead to delays as a new partner must be secured.

While a biotech firm may not operate in the therapeutic area associated with the newly launched drug, alliances with highly capable biotech partners can still offer substantial long-term strategic value. Thus, the opportunity costs associated with discontinuing any relationship must be carefully considered (Asgari et al., 2024). Moreover, partnership dynamics play a critical role in a termination decision. For instance, if a biotech partner has a history of repeated collaborations, particularly those reinforced by influential third-party connections that contribute to a cohesive network, terminating the alliance can negatively impact the pharma firm's access to network-related benefits (Reagans & McEvily, 2003). Consequently, just as alliance formation is influenced by the incentives for linkage creation and the expectation of future collaboration (G. Ahuja, 2000), firms must carefully weigh the tradeoffs between the benefits and potential opportunities that are lost from its discontinuation (Asgari et al., 2024).

In light of the operational and financial gravity of such decisions, the mechanisms that drive biotech-pharma alliance terminations after an unrelated drug launch may depend on several factors. First, bringing a drug to market require substantial investments, and at a certain point, an alliance may no longer provide sufficient value (Madhok et al., 2015). Terminating these partnerships allows a pharma firm to reallocate resources and focus on supporting the newly launched drug. Second, from a real options perspective (Vassolo et al., 2004), firms that have invested in several unrelated exploratory projects may be compelled to shift attention and resources away from these initiatives to prioritize the advancement of the new drug (Hoffmann, 2007). Multiple diverse alliances compete for attention and support (Cui & O'Connor, 2012), and those exploring less strategically aligned pathways may lose relevance and no longer be prioritized by the pharma firm (Lunnan & Haugland, 2008). Third, firms may reorient their partnership strategy towards forming alliances that support the further development of the recently launched drug. Such partnership reconfiguration entails abandoning some alliances while forming new ones that help to exploit the advancement (Hoffmann, 2007).

Also noteworthy is the disparate and fragmented findings from alliance termination research regarding which factors are associated with higher or lower termination risk and the mechanisms that lead to termination, as highlighted in Rajan et al.'s (2020) comprehensive literature review. This inconsistency justifies industry-specific studies since findings often vary across contexts. For instance, Mata and Portugal (2015) study terminated Portuguese joint ventures across multiple industries and observe that larger firms are less likely to terminate partnerships relative to smaller firms. In contrast, research specific to the biopharmaceutical industry, like the study by Dan and Zondag (2016), find that larger firms have a greater propensity to terminate alliances than their smaller counterparts. Our literature review reveals similar contradictions. Table 3.1 summarizes this literature, where we categorize factors found to be statistically significant predictors of alliance terminations.[5]

Most studies in Table 3.1 focus on identifying factors that reduce the likelihood of alliance termination or extend the duration of an existing partnership. The mechanisms behind many of the factors are linked to the aforementioned strategies and, as indicated, the decision to terminate an alliance, and which alliance to terminate, is likely shaped by a combination of several of these factors. Since firms make alliance formation decisions under uncertain conditions, research based on predictive modeling of this adverse outcome is particularly important for practice. This approach aligns well with the literature, which primarily focuses on identifying alliance termination determinants rather than establishing causal inference (Rajan et al., 2020). Such research is especially important to biotech firms as they depend on alliances for their survival. Thus, from their perspective, identifying potential threats that are largely outside their control (like unrelated drug launches) is essential for effective partner selection and strategic risk management. Our research attempts to address this aim by creating ML models that use information known at alliance formation to predict the risk of post-launch termination, thereby helping biotech firms safeguard themselves against such vulnerabilities.

5 Table 3.1 is not limited to drug development alliances, but the biopharmaceutical industry dominates this literature

	Factor	Description	Reference
Firm level	**Firm size**	Size of firm measured by number of employees	Dan & Zondag (2016)$^{+}$; Mata & Portugal (2015)$^{\cup}$
	Firm failure experience	Number of past projects that ended without reaching successful completion	Pangarkar (2009)$^{+}$
	Alliance experience	Number of past alliances—measured either by number of past formations of alliances	Greve et al. (2010)$^{+}$; Koval (2021)$^{+}$; McCutchen et al. (2008)$^{+}$; Pangarkar (2003, 2009)$^{-}$
	Competing alliances	Number of competing alliances within the focal firm's alliance portfolio	Asgari et al. (2018)$^{+}$; Cui (2013)$^{-}$; Cui et al. (2011)$^{+}$; Hampl (2020)$^{+}$; Hu et al. (2021)$^{-}$
	Pipeline strength	A measure of the consistency of portfolio project distribution in different phase of development	Peck et al. (2015)$^{+}$
	Technology experience	Dichotomous variable measuring whether a firm has used the same technology on past project	Dan & Zondag (2016)$^{+}$
Alliance level	**Alliance experience**	Whether the two partner firms have had prior alliances	Koval (2021)$^{-}$; Reuer & Zollo (2005)$^{+}$; Xia (2011)$^{-}$
	Alliance centrality	Number of alliance partners in common between both partners	Greve et al. (2010)$^{-}$; Hu et al. (2021)$^{\cup}$; Polidoro et al. (2011)$^{-}$; Weng & Tang (2023)$^{-}$
	Alliance duration*	Length of time from alliance formation to termination	Dan & Zondag (2016)$^{-}$; Greve et al. (2010)$^{-}$; Lunnan & Haugland (2008)$^{-}$; Madhok et al. (2015)$^{+}$; Mata & Portugal (2015)$^{\cup}$; Shi (1998)$^{-}$
	Equity stake	Whether the alliance agreement was predicated in-part or in-whole on equity stake	Bierly & Coombs (2004)$^{+}$; Duysters & De Man (2003)$^{-}$; Pangarkar (2003)$^{-}$; Pangarkar et al. (2017)$^{+}$; Tandon et al. (2022)$^{+}$
	Alliance experience asymmetry*	Difference in number of past alliances between both partners	McCutchen et al. (2008)$^{-}$
	Inherited alliance	Whether an alliance was inherited through M&A activities	Min (2017)$^{+}$
	International distance between partners	Dichotomous variable of whether firms are based in different country or measured through scale of cultural distance	Dan & Zondag (2016)$^{+}$; Malik & Zhao (2013)$^{-}$; McCutchen et al. (2008)$^{-}$; Meschi & Riccio (2008)$^{+}$; Mohr et al. (2016)$^{+}$
	Perceived benefit of the alliance	Perception that the cost of maintaining the alliance is worth the expected returns	Klossek et al. (2015)$^{+}$; Sadowski & Duysters (2008)$^{-}$
	Alliance portfolio size	Number of current alliances within the firm's portfolio	Hohberger et al. (2020)$^{-}$; Koval (2021)$^{\cup}$; Pangarkar et al. (2017)$^{-}$; Yamanoi & Cao (2014)$^{+}$
	Alliance portfolio size asymmetry*	Difference in portfolio size between partners	Meschi & Norheim-Hansen (2020)$^{+}$; Yamanoi & Cao (2014)$^{-}$
	Alliance scope	Number of projects related to the alliance	Bierly & Coombs (2004)$^{-}$; Hu et al. (2021)$^{-}$; Reuer & Zollo (2005)$^{-}$
	Substitution option	Dichotomous variable measuring whether a substitute partner exists	Madhok et al. (2015)$^{+}$; Xia (2011)$^{+}$; Yamanoi & Cao (2014)$^{+}$
Project level	**Project stage at alliance formation**	Stage of project in the development process when alliances are formed	Bakker & Shepherd (2017)$^{-}$; Bierly & Coombs (2004)$^{\cap}$
	Project stage at alliance termination*	Stage of project in the development process when alliances are terminated	Bakker & Shepherd (2017)$^{-}$
	Project progress*	Whether the project progressed through at least one stage during the alliance	Asgari et al. (2018)$^{-}$; Min (2017)$^{-}$
	Project expected progress	Whether the partners expect the project will continue to progress	Dan & Zondag (2016)$^{-}$; Sadowski & Duysters (2008)$^{-}$
	Technology uncertainty	Lack of past success using the same technology	Perry et al. (2004)$^{+}$

Legend: $^{+}$ positive relationship; $^{-}$ negative relationship, $^{\cup}$ U-shape relationship, $^{\cap}$ inverse U-shape relationships**.**
Variables considered in our analysis are represented in bold. A star (*) is added to variables considered indirectly, or in a secondary analysis.

Table 3.1. Factors Significantly Associated with the Likelihood of Alliance Termination

3.3. Identifying the Risk of Alliance Terminations Following Drug Launches

We adopt a two-part empirical approach to assess whether drug launches by pharmaceutical partners affect the risk of alliance termination. First, we test whether terminations are disproportionately more likely to occur after a drug launch. Second, we estimate a hazard model to assess whether, and when, a partner's drug launch increases the risk of termination.

3.3.1. Data, Sample Construction, and Measurement

We use three databases published by Clarivate containing biopharmaceutical industry intelligence data. The Cortellis Deals Intelligence database provides a comprehensive list of alliances between firms engaged in the drug development process, which consists of projects in the discovery or clinical phases. We extract all records (9,412) related to alliances that were terminated between January 1, 2000, and December 31, 2019. The cutoff date is chosen to exclude alliance terminations that may have been influenced by the COVID-19 pandemic, thereby avoiding the confounding effects of an externally driven shock to the overall pharmaceutical industry. Since alliances in this database are represented as dyads, a single observation represents a biotech–pharma pair.

Mergers and acquisitions (M&As) are common in this industry and acquired firms frequently retain their name when becoming a subsidiary. The Cortellis Deals Intelligence database classifies each alliance by the name of the lower-level subsidiary instead of the parent firm. Since we analyze alliance terminations at the parent level, we link each alliance with its respective parent firm by extracting all reported M&As to produce a comprehensive repository of subsidiaries, including the date of consolidation. We then cross-reference alliance termination dates with M&A activity to identify and add the parent firms of both biotech and pharma partners to our dataset. Finally, using the Cortellis Company Intelligence database, we determine whether each partnership contains a biotech firm, pharmaceutical company, or some other organization (e.g., academic institution). This process results in the retention of 3,996 records, all of which exclusively consist of biotech-pharma alliances.

We also use the Cortellis Drug Discovery database to identify instances where a biotech-pharma partnership was terminated, and the pharma partner had launched a drug that was unrelated[6] to the underlying project in the biotech-pharma alliance. Given the high concentration of small firms in the biopharmaceutical industry, some observations represent alliances between firms that never launched a drug. To avoid confounding predictive signals between alliances that remained active post-launch and those that never experienced a drug launch, we exclude alliances in which no drug was ever brought to market. By restricting the sample to alliances with at least one drug launch, we can identify features that distinguish launches followed by a termination from those that did not. The final sample comprises 1,649 alliances in which the pharmaceutical partner launched at least one drug during the study period.

Regarding the predictors (independent variables), Table 3.1's literature review includes those identified as significant in previous empirical studies on alliance terminations that employ statistical regression methods. We use this review as a foundation for selecting variables to conduct survival analysis and train ML models that can, upon the formation of a biotech-pharma partnership, predict the likelihood that an alliance is terminated after a drug launch. In what follows, we present the variables employed in our analyses, classified based on firm-, alliance-, and project-level attributes. Appendix G offers descriptive statistics for each variable described herein.

Firm-level Attributes. For each partner, *biotech (pharma) firm size* is calculated by counting the number of subsidiaries held by the parent firm, that is, the number of acquired firms. Although these measures correlate with other firm size metrics, like the number of employees and operating revenues, they are not used due to high rates of missingness. *Biotech (pharma) R&D portfolio size* is the number of active drug development projects undertaken by the parent biotech (pharma) firm at the time the alliance was formed. This includes independent projects and those as part of an alliance. *Biotech (pharma) failure experience* is the number of drug development projects terminated for the biotech

6 We ensure unrelatedness between the focal drug project and the related drug launch by only considering drug launches that are based on different medical indications and technologies. This ensures that the alliance termination is not influenced by factors such as learning from technological failures or abandoning a parallel project with the same medical indication. We also exclude drug launches involving the biotech firm.

(pharma) firm before the formation of the partnership. For each partner, *biotech (pharma) alliance portfolio size* is the number of active alliances formed before the formation of the focal partnership. To ensure we only count R&D alliances, we include those where a drug development project was in the discovery or clinical trial phases. Alliances formed after the completion of clinical trials are excluded, because the cost of innovation is minimal at this stage, and the alliance has less impact on R&D operations (Tyebjee & Hardin, 2004). *Biotech (pharma) technology experience* measures the number of times the biotech (pharma) firm has used related technologies in past drug discovery projects, including whether the firm worked on a drug project alone. *Biotech (pharma) indication experience* counts the number of past drug development projects in which the biotech (pharma) firm has targeted the same medical indication as in the focal project; indication represents the medical condition the drug aims to address. This variable is not explicitly identified in Table 3.1's literature review, yet it represents a source of uncertainty similar to *technology experience* (Lévesque et al., 2022).

Alliance-level Attributes. *Alliance experience* is the frequency of previous alliances between the two firms. Even if the specific subsidiaries have not directly worked together, this variable captures the institutional knowledge from prior collaborations between parent companies.

Project-level Attributes. *Project stage at alliance formation* is an ordinal variable categorizing the stage of drug development prior to the formation of the alliance for the focal drug project: (1) Discovery; (2) Pre-clinical; (3) Stage 1 clinical; (4) Stage 2 clinical; (5) Stage 3 clinical. *Project progress* is a dichotomous variable indicating whether the focal drug project advanced through any stages of the drug development process between the start of the alliance and the drug launch. *Technology uncertainty* is a dichotomous variable indicating whether the technology used in the drug development project has been successfully utilized in any previously launched drugs (0) or not (1).

Unique Predictors Not Identified in our Literature Review. *Therapy area* addresses how alliance dynamics vary based on the drug class (e.g., cancer, musculoskeletal, cardiovascular) to which the focal project belongs. The premise is that therapy areas may experience different levels of market volatility, regulatory scrutiny, and competitive pressures, which can affect alliance stability and

longevity (Albrecht et al., 2018). For instance, cancer drugs, which face intense competition and rapid innovation cycles, can exhibit higher termination rates as compared to more stable therapy areas like musculoskeletal disorders. *Principal investigator* is a dichotomous variable that equals 1 if the biotech firm serves as the drug development lead, and 0 if it is the pharma firm. Serving as the principal investigator can affect a firm's reliance on the focal project, and thus, has the potential to shape strategic priorities and affect alliance termination decisions (Ambulkar et al., 2022).

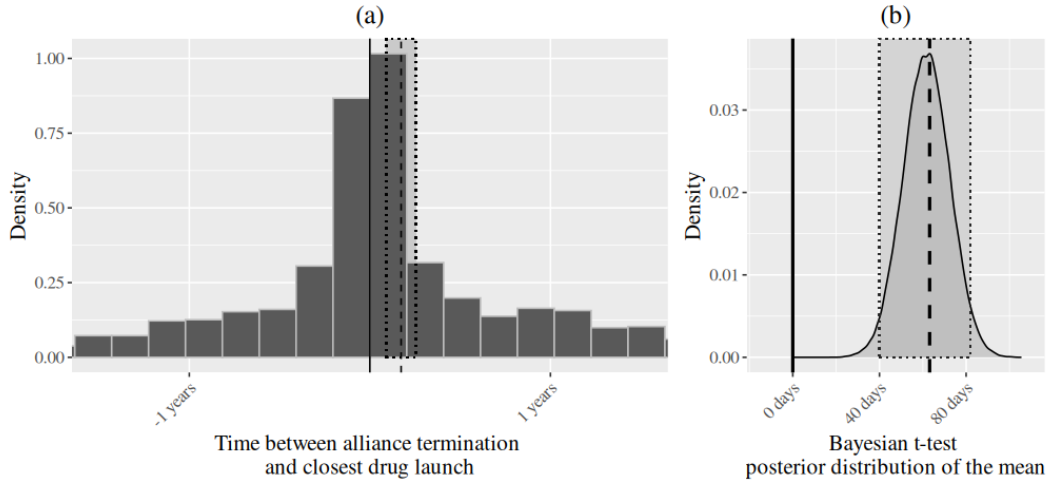
3.3.2. Frequency of Alliance Terminations Following Drug Launches

To assess whether the risk of alliance termination may increase following a drug launch, we first plot the distribution of the number of days between each termination and its nearest drug launch (Figure 3.1a). The distribution is left skewed with a mean greater than zero. A binomial test confirms that the proportion of terminations occurring after a drug launch (57%) is significantly different from 0.5 (random chance), with $p < 0.001$ and a 95% confidence interval of [0.548, 0.602]. We also perform a Bayesian one-sample t-test using the number of days between each termination and the nearest drug launch as the outcome, where negative values indicate terminations that occurred before the launch and positive values indicate those occurring after. This method is selected over a traditional t-test due to its robustness against outliers and suitability for thick-tailed data (Kruschke, 2013). To this end, we fit an intercept-only Bayesian regression estimating the mean difference in days, using a standard normal distribution as the prior, which assumes no effect and produces conservative results. The resulting posterior distribution (Figure 3.1b) shows a maximum *a posteriori* (MAP) estimate of 63.15, with 98.3% probability that the effect exceeds zero and a 95% credible interval of [40.62, 82.67].

Since the timing of a drug launch is largely dictated by regulatory approval, it is unlikely to be influenced by alliance terminations.[7] Therefore, any correlation between these events is expected to be unidirectional, whereby drug launches influence alliance terminations but not *vice versa*. If drug

7 New drug application review and approval varies by jurisdiction, but the approval process is scientifically driven and independent from the applicant. For examples of process detail, see reports from the US FDA timeline (United States Government Accountability Office, 2020) and European Medicines Agency (2024).

launches are not associated with an increase in alliance terminations, we would expect a zero-mean distribution because both events would be random and uncorrelated. Thus, the exploratory data analysis (EDA) findings in Figure 3.1 show that terminations occur more frequently after drug launches, indicating a correlation. However, this does not establish whether drug launches pose a true vulnerability to biotech firms. We address this question next using a survival analysis, which also allows us to empirically determine the duration of the post-launch termination risk window.



Legend: Positive values indicate terminations occurring after a drug launch, while the dotted region in (a) marks the magnified region shown in greater details in (b) which depicts the posterior estimate of the mean.

Figure 3.1. Distribution of Time Between Alliance Termination and the Closest Drug Launch

3.3.3. Estimating Termination Risk Over Time: Hazard Model

We assess whether a partner's drug launch increases the likelihood of alliance termination over time, by estimating a Cox proportional hazards (CPH), which accommodates right-censoring and time-varying covariates (Andersen & Gill, 1982). This model does not require specification of the baseline hazard function and avoids inference-based assumptions (Allison, 2014) like linearity, multicollinearity, or homoscedasticity. We first restrict the analysis to 3,496 alliances that have a known outcome (either completed or terminated), excluding those still marked as active to mitigate censoring bias caused by potential misclassification in the database. We then implement a model using the full Cortellis sample of 34,208 alliances. Appendix H provides additional details and descriptive statistics.

For the CPH model, the dependent variable is the hazard of termination as a function of time since

alliance formation. The primary independent variable is a time-varying covariate indicating whether the partner has launched a drug at a given time. Specifically, we segment each alliance's timeline into three intervals: (1) the period prior to the launch of a drug by the pharma partner (pre-launch); (2) the period following the launch (launch window); and (3) the period after this window (post-launch). The time-varying covariate, *drug launch window*, is coded as 1 during the launch window and 0 otherwise. Each alliance is represented by one or more data points, depending on when termination occurs. For instance, if an alliance terminates during the launch window, its timeline includes one data point covering its survival through the pre-launch period and a second data point capturing its termination during the launch window. Likewise, if termination occurs in the post-window period, the timeline includes data points for pre-launch and launch window exposure, followed by a third data point capturing termination in the post-window phase. This approach follows Andersen and Gill's (1982) approach, which accommodates time-varying covariates and repeated event intervals per alliance. We also include time-invariant controls for firm-, alliance-, and project-level characteristics observable at alliance formation.

Using the dataset of alliances that have a known outcome, we determine the appropriate launch window duration by estimating a series of CPH models using one-year rolling windows that begin at different lagged intervals following the launch date (ranging from 0 to 60 months). For each of the 61 rolling windows, we construct a dataset that captures exposure to a partner's drug launch within a one-year period starting at a specified number of months (the lag) after the drug launch. We then estimate a CPH for each dataset and extract the corresponding hazard ratios and confidence intervals. Figure 3.2 illustrates the lagged timing of the post-launch exposure windows (left) alongside the corresponding results (right). The effect of drug launches on alliance termination risk is strongest with time windows beginning within the first two years (24 months) and is not statistically significant beyond 3 years (36 months). This confirms that the higher risk of termination post-launch is concentrated within a three-year period following a drug launch and will be used as the post-launch window size for the remainder of our analyses. This three-year window of vulnerability is consistent with the typical amortization

period for R&D investments, which reflects the expected economic benefit timeframe (Cazavan-Jeny et al., 2011; Griliches, 1984). Scholars use similar timeframes to assess the impact of patents on firms (G. Ahuja, 2000; Boeing et al., 2016), and the effect of announcements on biotech stock prices (Williamson et al., 2024).

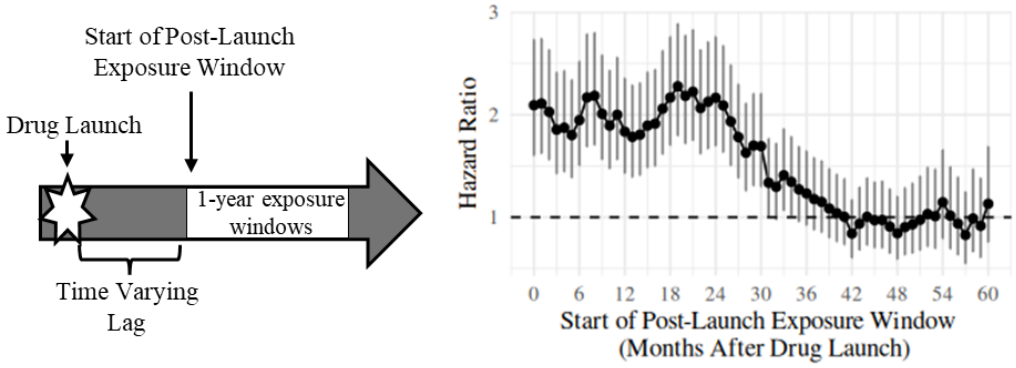


Figure 3.2. Sensitivity of Alliance Termination Risk Relative to Timing of Drug Launch (one-year rolling windows)

We next estimate CPH models with three-year launch windows and present the results in Table 3.2. These results indicate that alliances face elevated termination hazard during the three-year window following a partner's drug launch. For the dataset with known outcomes, the hazard ratio for *drug launch window* of 1.581 ($p < 0.001$) implies that the risk of termination is 58% higher during this post-launch period. In the full sample of 45,277 alliances, we find a significantly higher hazard ratio (18.988 vs. 1.581), likely due to some alliances being misclassified as active. Many of the "active" alliances ended during the study period but were never classified as such due to missing information. We find that these active alliances disproportionately lack any observed drug launches, even after accounting for alliance duration. Alliances between firms that have never launched a drug likely generate less public information, such as status updates, and are thus more prone to misclassification. This misclassification leads the CPH model to overstate the effect of drug launches on termination risk, although the estimated effect remains directionally consistent.

	Terminated or Completed Alliances				All Alliances			
Variable	coef	exp(coef)	se(coef)	p-value	coef	exp(coef)	se(coef)	p-value
Drug launch window	**0.458**	**1.581**	**0.060**	**<0.001**	**2.944**	**18.988**	**0.077**	**<0.001**
Pharma failure experience	0.010	1.010	0.001	<0.001	0.014	1.014	0.001	<0.001
Pharma firm size	0.013	1.013	0.002	<0.001	0.012	1.012	0.002	<0.001
Pharma R&D portfolio size	-0.006	0.995	0.001	<0.001	-0.007	0.993	0.001	<0.001
Pharma alliance portfolio size	-0.004	0.996	0.001	<0.001	-0.006	0.994	0.001	<0.001
Pharma technology experience	-0.183	0.833	0.157	0.242	-0.774	0.461	0.398	0.052
Biotech failure experience	0.008	1.008	0.003	0.001	0.005	1.005	0.003	0.072
Biotech firm size	0.022	1.022	0.003	<0.001	0.033	1.033	0.003	<0.001
Biotech R&D portfolio size	-0.003	0.997	0.003	0.259	-0.006	0.995	0.003	0.041
Biotech alliance portfolio size	-0.009	0.991	0.002	<0.001	-0.001	0.999	0.002	0.692
Biotech technology experience	-0.601	0.548	0.295	0.042	-0.280	0.756	0.266	0.294
Alliance experience	-0.002	0.998	0.017	0.906	0.016	1.016	0.018	0.385
Principal investigator	-0.134	0.874	0.042	0.001	0.333	1.395	0.043	<0.001
Technology novelty	-0.227	0.797	0.155	0.143	-0.517	0.596	0.173	0.003
Indication novelty	-0.559	0.572	0.145	<0.001	-1.419	0.242	0.152	<0.001
	Concordance = 0.64 (se = 0.006). Number of alliances = 3,496				Concordance = 0.701 (se = 0.005) Number of alliances = 34,208			

Table 3.2. CPH Model Results

3.4. Predicting Post-Launch Terminations

We determine whether post-launch terminations are distinguishable from other terminations by training classification-based ML models to predict, at the time of alliance formation, which alliances are more likely to be terminated following a partner's drug launch rather than for unrelated reasons. In addition to theory building, we aim to develop classifiers and interpretable models that help guide biotech managers in making alliance decisions under uncertainty. Since the research question is inherently a predictive one, regression analysis and a reliance on in-sample predictions is problematic (Shimao et al., 2025). Moreover, while statistical inference methods like logistic regression yield interpretable results, they rely on key assumptions: no multicollinearity; a linear relationship between the independent variables and the log odds of the dependent variable; independent errors; and no influential outliers (Stoltzfus, 2011). In Appendix I, multiple statistical tests show that these properties do not hold for our dataset. In addition, the literature on alliance terminations in the biopharmaceutical sector reveals varied statistical relationships—some studies report positive correlation, others are negative, and some are nonlinear (see findings from Asgari et al., 2018; Cui et al., 2018). This suggests context-specific relationships and complex statistical associations.

To align the empirical strategy with our prediction-focused research question, and to address the

methodological challenges discussed above, we train a series of ML models and compare their performance. Where appropriate, hyperparameters are tuned using five-fold cross-validation. Following best practices (Sheetal et al., 2023), we include a Naïve Bayes classifier as a baseline classifier despite it not accounting for the interaction between feature variables. We also train a Bayesian Additive Regression Tree (BART) model (Chipman et al., 2010), which captures complex relationships while still being interpretable, and a Bayesian Neural Network (BNN) as a 'black box' model to evaluate the tradeoff between prediction accuracy and interpretability (Sharaf et al., 2020). Additionally, we train frequentist ML counterparts to the above models, including Random Forest, XGBoost, Ridge, and Lasso regression models, to compare out-of-sample performance. We select the best-performing model, whether Bayesian or frequentist, for subsequent analyses and to assess the robustness of our findings.

3.4.1. Data preparation and dependent variable

We augment our previously used dataset of 1,649 terminated alliances by coding whether a drug launch by the pharma partner occurred within a 3-year window before or after an alliance termination. We also perform a sensitivity analysis with other window sizes (see Appendix J). Since pharma partners could have launched multiple drugs within the study period, each observation represents a termination–drug launch pair, as shown in Figure 3.3, where a gray dot represents an alliance termination. This process results in a final sample of 8,790 data points, each representing a terminated biotech-pharma partnership where the pharma firm launched at least one drug during the 20-year study period.

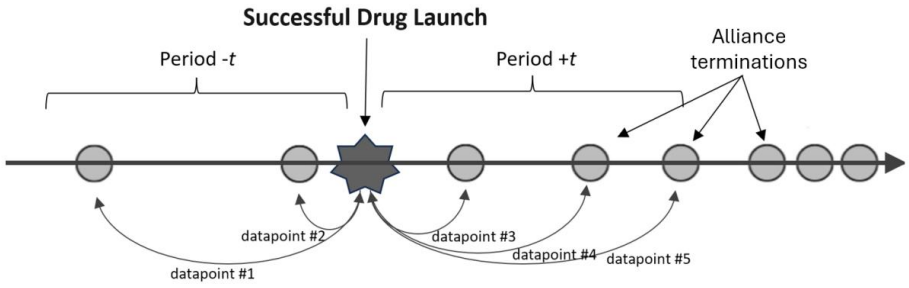


Figure 3.3. Constructing Alliance Termination–Drug Launch Pairs

We dummy code each observation based on Eq. (3), where θ_a represents the date of alliance

termination and θ_d represents the date the pharma partner launched a drug (*note:* we omit data points where $\theta_a = \theta_d$ because the trained model with and without these instances is identical).

$$Y = \begin{cases} 1, & \theta_a - \theta_d > 0 \text{ and } |\theta_a - \theta_d| \leq 3 \\ 0, & \text{otherwise} \end{cases}. \quad (3)$$

We intentionally restrict the data sample to alliances that ultimately terminated, regardless of whether the termination followed a drug launch. This conditional modeling approach directly addresses our central research question: *Can features observable at the time of alliance formation distinguish post-launch terminations from those that are independent of drug launches?* By focusing only on terminated alliances, we eliminate the uncertainty introduced by right-censored observations (ongoing alliances with unknown outcomes). If post-launch terminations are distinguishable based on *ex ante* features, this suggests that these events are not purely random. Thus, identifying these patterns can enable biotech firms to anticipate and potentially mitigate the risk of post-launch terminations when selecting partners.

3.4.1.1. Data Segmentation

The ML models are trained using a dedicated training set, and a classification threshold is determined using a validation set before final testing is performed on a held-out set. Following Chipman et al. (2010), 20% of the total data is assigned to a holdout test set (n = 1,658). The remaining 80% of the data is then split, with 75% (n = 5,412) used for training and 25% (n = 1,720) used for validation. To prevent data leakage, observations are split based on the name of the focal drug project, ensuring that the same project does not appear in both the training and held-out datasets. For robustness, we also use a temporal train-test split, assigning the most recent 20% of the data to the holdout set. Appendix J summarizes the results and confirms the feasibility of predicting post-launch termination risk.

3.4.1.2. Classification Threshold

Probabilistic classifiers output a likelihood that an instance belongs to a given class. Thus, to assess predictive performance, a classification threshold must be set to convert this continuous output into a

binary prediction. This is given in Eq. (4) where $L(X_i)$ is the likelihood estimate given the observed set of features X_i of alliance i and $\tau \in (0,1)$ is the classification threshold.

$$\hat{y}_i = \begin{cases} 0, & L(X_i) < \tau \\ 1, & L(X_i) > \tau \end{cases}. \tag{4}$$

The optimal classification threshold can be determined via hyperparameter tuning using the validation set, however, the appropriate objective for selecting τ depends on the specific problem context (He & Garcia, 2009). In our case, we find the minimum classification threshold (τ^*) that maximizes the positive prediction value (PPV) (Ghanem et al., 2023). This approach is consistent with established practices for handling imbalanced datasets, where the accurate detection of the minority class requires careful calibration to minimize the number of false positive predictions (Caton & Haas, 2024; Gu et al., 2009; Jui & Rivas, 2024; Leevy et al., 2018; Rajkomar et al., 2018). To this end, we aim to select τ^* which represents the solution to the following optimization problem:

$$\min_{\tau \in (0,1)} \max_{\boldsymbol{\theta}} \left[P\left(y = 1 | \hat{y} = 1;\, \boldsymbol{\theta}\right) \right] \text{ subject to } P(\hat{y} = 1 | \mathrm{y} = 0; \boldsymbol{\theta}) > 0, \tag{5}$$

where $P(.)$ is the probability induced by the chosen classification function and $\boldsymbol{\theta}$ is a vector of parameters. In Eq. (5), we select τ^* such that it is the minimum threshold that maximizes the proportion of correctly predicted positives while ensuring that at least some false positives are present to avoid trivial solutions.[8] Since we cannot optimize Eq. (3) directly due to the iterative nature of training and testing in ML models, we rely on validation-based metrics as proxies to determine τ^* using sampling-based, hyperparameter tuning algorithms as in Akiba et al. (2019).

In Table 3.3, we report the detection rate, the proportion of positive predictions ($\hat{y} = 1$), to assess the models' capacity to identify alliances at risk. We use PPV to compare model performance with a particular focus on those with the highest values. Finally, we report the negative prediction value (NPV), which, by design, is expected to approximate random chance (50%). This is because

8 Without the constraint to allow false positives, the optimal threshold would be set as to never make positive prediction, thus scoring a perfect PPV.

terminations unrelated to a drug launch (the negative class) are equally likely to occur before or after the launch date. As a result, even a perfect classifier would achieve an NPV of 0.5.

3.4.2. Classification Results

In Table 3.3, we present the prediction results for each model using the validation dataset used to determine τ^*, along with performance on a held-out test set.[9] Comparing PPV between the validation and test sets helps to detect overfitting, as models often perform better on validation data. Conversely, underfitted models tend to have poor PPV on both datasets. The baseline Naïve Bayes classifier shows signs of underfitting, with a PPV of 52.3% on the validation set and 51.5% on the test set. The low PPV and inflated detection rate suggest that the model is too weak to reliably identify terminations linked to drug launches. This is likely due to Naïve Bayes' assumption of predictor independence and equal weighting of all predictors, including irrelevant ones. The Ridge and Lasso models perform better, with out-of-sample PPVs of 90.0% and 81.8%, respectively, as regularization helps reduce multicollinearity, minimize irrelevant features, and mitigate overfitting. However, their low detection rates are indications of underfitting. Their inflated PPV values stem from rarely predicting positive classifications, which is likely due to their inability to capture nonlinear effects and higher-order interactions. In contrast, the BNN and Random Forest models show signs of overfitting. BNN scores a high PPV on the validation set (92.3%) but significantly lower out-of-sample performance (73.1%). Random Forest scores a low PPV on both datasets, but its out-of-sample detection rate decreases considerably (from 8.5% to 0.6%). These models likely overfit to noise and irrelevant predictors, reducing generalizability. This overfitting also contributes to their low detection rates, as they produce few positive out-of-sample predictions.

The gray cells in Table 3.3 highlight BART and XGBoost, which provide the best model fit. These models have the highest detection rates and their PPVs on the held-out set are near or exceed 90%. As ensemble methods, they effectively capture complex relationships, reduce the influence of

9 Through random draw, the proportion of terminations after a drug launch is higher in the held-out sample data. Thus, some models appear to perform better in the held-out sample than the in-sample data. The results are unimpacted by this random draw.

irrelevant predictors, and control overfitting through regularization. Nevertheless, their detection rates remain below the observed post-launch termination increase (approximately ~7% as in Section 3.2), highlighting the tradeoff between precision and recall. Because the models rely solely on information available at alliance formation, unobserved factors likely limit their performance.[10] However, their high PPVs and detection rates suggest that these models can be used to predict whether a newly formed alliance is more likely to be terminated following a drug launch as opposed to other potential reasons.

		Bayesian Models			Frequentist Models			
		BART	BNN	Naïve Bayes	Random Forest	XGBoost	Ridge	Lasso
	Threshold	0.787	0.610	0.994	0.994	0.929	0.690	0.732
Prediction on validation data	**PPV** (Positive prediction value)	0.9714	0.923	0.523	0.7298	0.972	0.917	0.875
	NPV (Negative prediction value)	0.486	0.439	0.563	0.5128	0.488	0.481	0.480
	Detection Rate	0.0207	0.0066	0.5180	0.0849	0.0213	0.0067	0.0043
Prediction on testing data	**PPV**	**0.875**	0.731	0.515	0.7836	**0.974**	0.900	0.818
	NPV	**0.503**	0.446	0.474	0.5000	**0.505**	0.488	0.486
	Detection Rate	**0.0425**	0.0084	0.5099	0.0064	**0.0419**	0.0099	0.0050

Legend: Gray cells represent the best performing model from the Bayesian (left) and Frequentist (right) algorithms.

Table 3.3. Comparisons of Predictive Performance Results

Using the BART model, we generate a posterior distribution for each prediction, allowing us to quantify uncertainty. This is achieved by aggregating predictions across multiple decision trees and Markov Chain Monte Carlo iterations (J. L. Hill, 2011). In Figure 3.4, we present the distribution of predictions for each observation, which is over 2,000 iterations of the BART algorithm. In Figure 3.4a, we observe that the cumulative posterior distribution for a true positive prediction exceeds the 0.5 threshold, indicating high confidence. Conversely, the distribution in Figure 3.4b reflects true negative predictions and is symmetric with a mean near 0.5, indicating performance that is close to random chance.

10 In a subsequent section, we incorporate information collected up to when the alliance is terminated (e.g., progress on the focal project and timing of termination relative to the drug launch). This additional information improves the detection rate.

Taken together, the results in Table 3.3 and Figure 3.4 demonstrate that a prediction model can reliably be used to identify alliances at risk of termination after an unrelated drug launch using only information available at the time of formation. To test the robustness of our findings, we vary the time window used when constructing alliance termination–drug launch pairs from three years to two and four years, as reported in Appendix J. We also reanalyze the data after incorporating a washout period. That is, we exclude cases where alliance termination and drug launch occur within 30 days or 60 days of each other, to account for possible systemic timing errors. Results remain consistent across all sampling distributions.

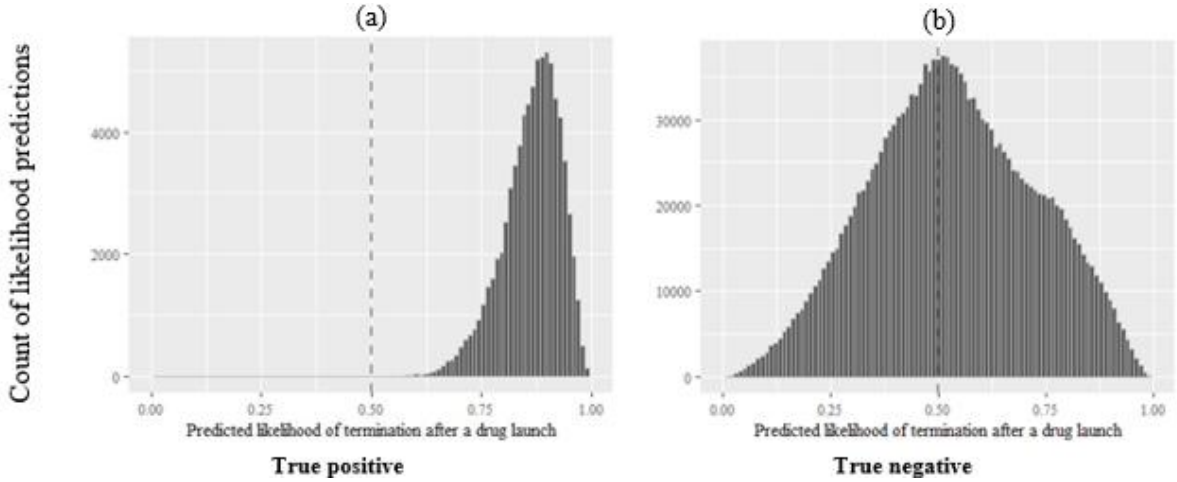


Figure 3.4. Sum of Posterior Probability Predictions

3.4.3. Exploring Predictors of Post-Launch Termination

We next identify key predictors influencing the model's predictive accuracy, with a focus on identifying interactions and nonlinear effects. This approach aligns with the principles of explainable AI, which aim to provide transparency to complex models to improve our confidence in their predictions (Chaddad et al., 2023; Chen et al., 2022; Gilpin et al., 2018). Identifying important predictors, interactions, and nonlinear effects also ensures practical relevance by confirming that the model's predictions are based on factors that management can directly control (Xu et al., 2019).

3.4.3.1. Predictor Selection

We perform a feature importance test by analyzing the inclusion proportion, which indicates how

frequently a variable is used across the decision trees in the BART model.[11] Because some variables may be used to consistently overfit to noise while others may be rarely selected but have a significant predictive effect, we use a permutation-based inferential method suggested by Bleich et al. (2014).[12] This involves training new BART models with the original variables (X), and generating random, uncorrelated values where no real association between variables and outcomes exists. We generate 100 such datasets and train 50 BART models on each to compute the average inclusion proportion for each variable, represented by the thinner (light gray) histogram bars in Figure 3.5. Variables with actual inclusion proportions (dark gray bars) that are significantly higher than their null inclusion proportions—based on a t-test ($p < 0.05$)—are identified as true predictors meriting further investigation. We visualize these t-test results in Figure 3.5 by displaying 83% confidence intervals, where non-overlapping intervals closely correspond to significant t-test results.[13]

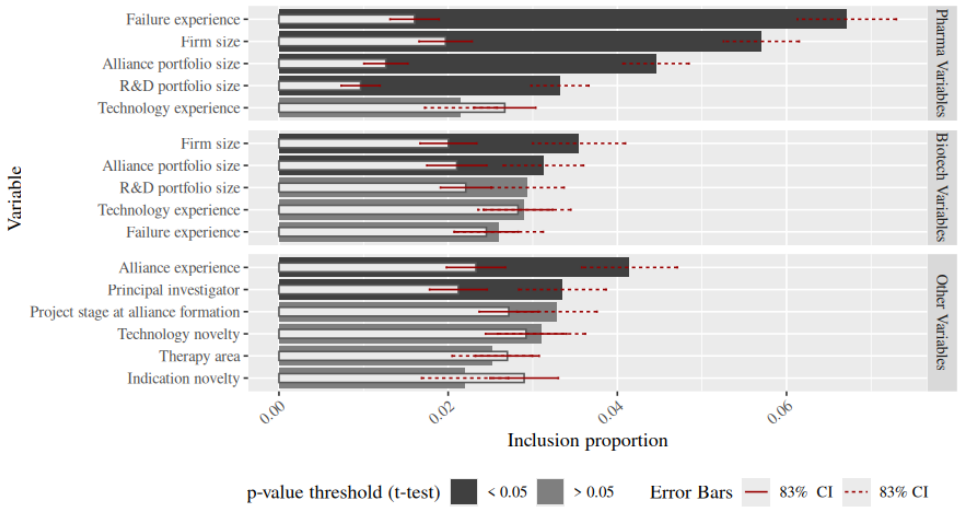


Figure 3.5. Variables of Importance Using a Permutation-Based Method from Bleich et al. (2014)

11 Since BART learns from past iterations, the inclusion proportions of the first series of trees are mainly based on random chance until the model performs sufficient iterations to learn the prior probability distributions. As recommended by Chipman et al. (2007), we discard the first 200 iterations from the inclusion proportion analysis to avoid decision trees built without strong prior learning.

12 Bleich et al. (2014) proposes two other methods to test BART's inclusion proportion. We use the method that rules out the fewest number of predictors.

13 83% confidence intervals of two means are approximately equivalent to a 2-sample t-test results with a p-value of 0.05. The error bars in Figure 5 show a visual approximation of the t-test results for illustrative purposes. The actual statistical significance was determined using formal t-tests.

The analysis indicates that four attributes of the pharma partner are important predictors of post-launch alliance termination, whereas only two attributes of the biotech firm are relevant. The pharma firm's *failure experience*, *size*, *alliance portfolio* size, and *R&D portfolio size* are relevant as well as the *biotech firm's size* and *alliance portfolio size*. The other predictors deemed important are the frequency of previous alliances between the two firms labelled *alliance experience*, and which firm serves as the *principal investigator* on the focal project.

3.4.3.2. Model Interpretation

To derive actionable insights and a deeper understanding of which alliances may be vulnerable to post-launch terminations, we apply Shapley Additive Explanations (SHAP), an interpretable AI technique. Although simpler models are generally preferred for interpretability (Rudin, 2019), they must also achieve acceptable predictive performance. We therefore retrain each ML model from Table 3.3 using only the important predictors identified in Figure 3.5 and evaluate their performance on the validation and held-out test sets. As shown in Appendix K, XGBoost and BART continue to outperform non-tree-based models while an inherently interpretable decision tree fails to achieve sufficient accuracy. Among the two, XGBoost yields a slightly higher PPV and detection rate (95.3% and 8%) as compared to BART (93.6% and 6.7%). Hence, we use SHAP to evaluate the impact of each feature on the XGBoost model's predictions (Lundberg & Lee, 2017). Unlike feature importance measures, SHAP values capture both the *magnitude* and *direction* of each feature's impact on the predicted likelihood of termination.

We calculate SHAP values for each out-of-sample prediction and aggregate the results in Figure 3.6. To interpret this figure, we consider two aspects of each plot: (1) a concentration of the same color (blue or red) and their position relative to 0.0 on the X-axis; and (2) the value on the X-axis, which positively relates to each predictor's influence on the likelihood of post-launch alliance termination. Red values concentrated on the right of 0.0 indicate a positive relationship with the outcome, while blue values indicate a negative relationship. Figure 3.6a illustrates that this analysis substantiates the feature importance test in Figure 3.5 by showing that *pharma failure experience* and *pharma firm size*

have the strongest influence on predictions. This is demonstrated by the heavy concentration of the SHAP values away from 0.0 and by the distinct separation of color. We observe a negative relationship between *pharma failure experience* and the likelihood of alliance termination post-launch, indicated by a blue concentration to the right of 0.0. *Pharma firm size* is also negatively associated with this likelihood. *Alliance experience* appears to have a positive relationship demonstrated by a red concentration to the right of 0.0. The role of *principal investigator*—dummy coded to 1 (red) for the biotech firm—shows a clear boundary, whereby a biotech principal investigator is associated with a lower likelihood of termination post-launch. However, the influence of this variable is weak (small SHAP values) as compared to the other predictors. Finally, *Pharma R&D portfolio size*, *pharma alliance portfolio size*, *biotech firm size*, and *biotech alliance portfolio size* do not have a distinct color separation, suggesting that their effects likely stem from interactions with other predictors.

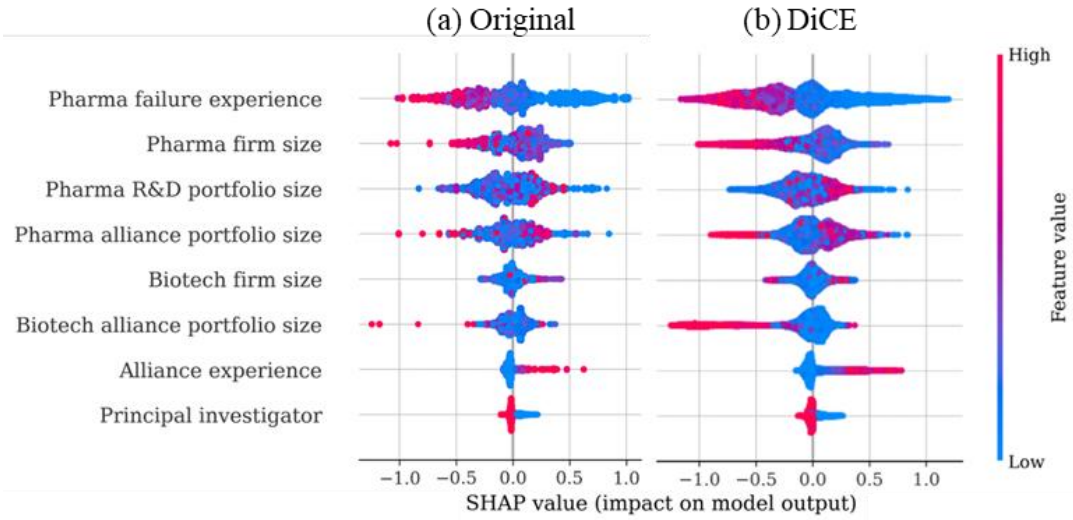


Figure 3.6. SHAP Values Using the XGBoost Model with Important Predictors from Figure 3.5

SHAP values provide global explanations that offer general insight into which predictors have the most consistent influence on a model's predictions. However, global explanations can overlook case-specific interactions and fail to capture the local impact of predictors on individual alliances, particularly those near the decision boundary, where small changes can significantly influence predictions. To address this limitation, we conduct a Diverse Counterfactual Explanation (DiCE)

analysis, which generates scenarios that span the decision boundary by identifying the minimal changes needed to flip each prediction (Mothilal et al., 2020). These scenarios help uncover tipping points for reducing the risk of post-launch termination, since the original data does not necessarily cluster near the decision boundary. Given that multiple changes can result in a flipped prediction, we generate three counterfactual scenarios for each data point, optimizing for diversity. That is, each of the three scenarios is selected to maximize its distance from the others using multiple distance metrics, as defined in Mothilal et al. (2020). Appendix L presents a single example of the counterfactuals generated using DiCE, based on a randomly selected sample from the held-out test set. This example demonstrates that certain features, such as *pharma failure experience*, are more influential in changing predictions, while in other cases, multiple feature modifications can produce the same effect.

After generating the counterfactual scenarios, we calculate SHAP values and present the aggregated results in Figure 3.6b. The results are consistent with those in Figure 3.6a, indicating that the direction and magnitude of feature influence at the decision boundary are consistent with those observed in the original data. However, global SHAP values do not reveal how features interact or how their influence on predictions changes for different feature values. To address this, we generate SHAP dependence plots, which visualize the marginal influence of one feature while capturing interactions with a second feature through color coding. This allows us to examine how feature contributions vary conditionally. We apply this analysis to DiCE-generated counterfactuals because they represent cases where small changes in feature values can result in significantly different predictions. This provides better coverage of critical regions, especially where the original data is sparse. We systematically inspect every two-way interaction plot and present those with a clear interaction effect in Figure 3.7. Within each plot we: (1) identify a clear separation between blue and red areas, where one area is distinctly higher in the graph than the other; (2) interpret a downward shift from red to blue as a positive moderating effect of the corresponding feature (indicated on the right side of the graph); and (3) consider the interaction to be heterogeneous if the color separation is inconsistent across the X-axis.

Beginning with Figure 3.7a, we first observe an evident separation of the two colors, particularly

for values greater than 100 on the X-axis. This separation means that an interaction exists between *pharma failure experience* (X-axis) and *pharma R&D portfolio size* (color scale).[14] Specifically, for pharma firms with more than 100 past failures, those with a larger R&D portfolio (represented by the red area) tend to increase the likelihood of post-launch alliance termination given the move from blue to red upward (i.e., the red area appears above the blue one). A similar two-way interaction emerges in Figure 3.7b between *pharma R&D portfolio size* (X-axis) and *pharma firm size* (color scale), showing a positive moderating effect given the color transition from blue to red upwards. This positive moderating effect for pharma firm size is prevalent for firms with large R&D portfolios (greater than 350 drug projects). A more subtle moderating effect appears in Figure 3.7c, concerning the interaction between a pharma firm's *alliance portfolio size* and its overall *firm size*. Specifically, among pharma firms with large alliance portfolios, those with a medium number of subsidiaries (around 60) are less likely to terminate alliances after a drug launch. The fourth significant interaction is between *pharma R&D portfolio size* and *pharma alliance portfolio*. While the predictors are correlated, instances where a pharma firm has a large R&D portfolio and few alliances show a lower likelihood of post-launch alliance termination.

[14] A similar moderating effect was observed between pharma failure experience and the size of the pharma firm's alliance portfolio. This figure was excluded as it is nearly identical to Figure 3.7a.

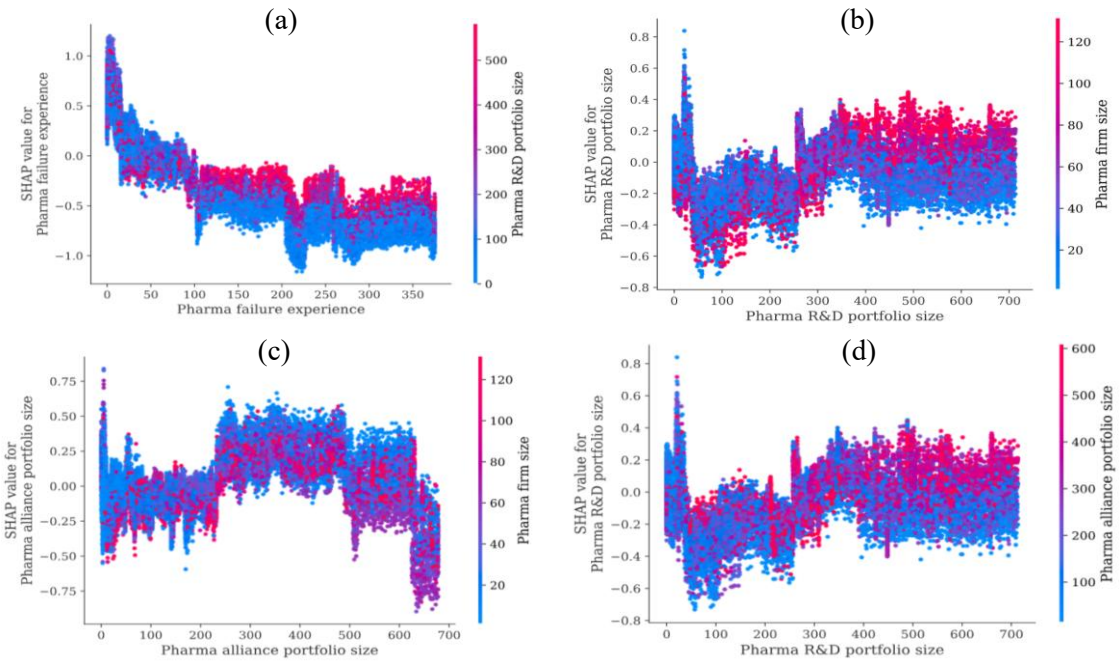


Figure 3.7. SHAP Values Interactions

3.5. Testing Strategic Timing as an Alternative Explanation

We provide a post-hoc analysis to test whether the observed increase in post-launch terminations could be explained by strategic delays: cases in which alliances that could have ended earlier are instead retained and terminated shortly after a drug launch to soften the impact of the announcement in the shadow of the news of a drug launch. While the presence of a vulnerability window extending for three years beyond the drug launch weakens this alternate explanation, we include an additional test to further refute this possibility.

If post-launch terminations were opportunistic timing rather than present a unique vulnerability, these terminated alliances would involve weaker drug projects that had already reached the limits of their potential, and thus, would be less likely to progress to the next stage of drug development. To test this hypothesis, we train an ML model to predict, *post-alliance drug progress*, a binary outcome variable that equals 1 if the project advances to the next stage of drug development, and 0 if it does not. We include the same firm-, alliance-, and project-level characteristics, now observable at drug launch, and add an additional feature, *post-launch alliance termination*, which is a dichotomous variable that

reflects whether an alliance termination is predicted[15] to occur after a drug launch. That is, it equals 1 if the analysis from Section 3.4 predicts that an alliance will be terminated after a drug launch, and 0 otherwise. If strategic timing were driving post-launch terminations, then the presence of the predictor should be associated with a higher probability of non-progression.

In addition to the key predictors identified earlier (Figure 3.5), we include five dynamic features that capture alliance-level developments over time.[16] *Biotech (pharma) R&D portfolio growth* assesses the difference in the number of active R&D projects for the biotech (pharma) firm between the formation and termination of the alliance. Positive values indicate portfolio growth during this period, while negative values represent a reduction in the number of active projects. *Time between alliance formation and drug launch* measures the number of days between the formation of the alliance and the drug launch. *Progress during alliance* is a dichotomous variable that equals 1 if the focal project progressed by at least one stage in the drug development process during the period of the alliance, and 0 otherwise.

Following Shimao et al. (2025), and consistent with our earlier analysis (see Section 3.4), we train two XGBoost classification models to predict *post-alliance drug progress*: one with the dichotomous variable *post-launch alliance termination* and one without. We withhold 20% of the data for out-of-sample evaluation, tune hyperparameters using five-fold cross-validation, and asses predictions based on a threshold of 0.5.[17] The out-of-sample performance of the XGBoost model with the *post-launch alliance termination* predictor has a PPV of 1.0, indicating that all drugs predicted to progress to the next development stage are correctly identified. We also observe a NPV of 0.992, demonstrating that the model predicts non-progressing drugs with few false negatives, confirming the model's strong

15 We use predicted values to better isolate terminations that are linked to drug launches. While models using actual values produced similar results, using predicted values helps to reduce noise in the signal, leading to more accurate and conservative estimates.

16 Some attributes (i.e., firm size and failure experience) did not have sufficient differences between the start and end of alliances to be considered relevant.

17 Unlike in the main analysis where we aim to predict the data points related to the *increase* in termination after a drug launch, here we follow a typical classification problem where we aim to accurately classify each data point (NPV does not tend towards 0.5).

classification performance.[18]

With the model accurately classifying future project progress (PPV = 1.0; NPV = 0.992), we assess whether the addition of *post-launch alliance termination* adds predictive value. To facilitate this, we perform Monte Carlo cross-validation by training and evaluating each model on 100 random train-test splits to compare the difference in R^2 (one including the *post-launch alliance termination* and the other excluding it) with a Student's t-test.[19] Table 3.4 shows no statistically significant difference in predictive performance between the two models (p = 0.4065), indicating that *post-launch alliance termination* does not improve the prediction of *post-alliance drug progress*.[20] These findings offer no support for the strategic timing explanation. If firms systematically delayed terminating weak drug projects, we would expect those alliances to exhibit lower rates of post-alliance progression. Conversely, if post-launch terminations reflected premature decisions that ended alliances with unrealized potential, we would expect higher rates of subsequent progression. We find no evidence of either pattern.

Predicting post-alliance drug progress	Mean R^2	Std. dev.	df	t-statistic	p-value
With the *post-launch alliance termination* predictor	0. 8884	0. 0135	198	-0.83184	0. 4065
Without the *post-launch alliance termination* predictor	0. 8869	0. 0122			

Table 3.4. Effect of Post-launch Alliance Terminations on Post-alliance Drug Progress

3.6. Discussion and Conclusion

Using 20 years of biotech–pharma alliance data, we find that alliance termination risk increases after the pharma partner launches an unrelated drug. This finding is confirmed by survival analysis, which shows a significant rise in termination risk within three years post-launch. We rule out opportunistic

18 While the main analysis prioritizes PPV, predicting post-alliance progress yields a balanced prediction between PPV and NPV, as shown by an Area Under the Receiver Operating Characteristic Curve (AUC-ROC) score of 0.954 (Provost et al., 1998).

19 Student t-test is used because a Fisher F-test indicates equal variance between the two groups ($F = 0.9586$, numerator degrees of freedom = 99, denominator degrees of freedom = 99, $p = 0.8338$).

20 We repeated the analysis using three alternative measures of progress: (i) advancement through two development stages; (ii) completion of the final clinical trial stage; and (iii) whether the terminated project ultimately resulted in a drug launch. We found similar results although the accuracy of the prediction models was limited by increasingly unbalanced datasets.

timing as a mechanism and, using supervised machine learning (ML) and explainable AI, show that post-launch termination risk is predictable based on features observable at alliance formation.

The ML models identify several key predictors of post-launch termination risk: the pharma firm's failure history and R&D portfolio size; both firms' size and the extent of their alliance portfolios; the frequency of prior collaborations; and whether the biotech firm leads development. Notably, a pharma firm with many past failures may not necessarily be a poor partner, rather, such firms may be more adept at evaluating the potential of a new partnership (Love et al., 2023a). This supports earlier findings that firms with more failures tend to sustain longer alliances (Pangarkar, 2009), although prior studies did not examine this in the context of drug launches. Surprisingly, fewer prior alliances with the same pharma partner are predicted to have a lower post-launch termination risk, which contradicts Reuer and Zollo (2005). Thus, in the context of drug launches, a history of collaboration does not necessarily signal future stability. For biotech firms, serving as the principal investigator and maintaining a large alliance portfolio can reduce termination risk following the launch of a drug by one of its pharma partners. When biotech firms lead, pharmaceutical firms likely conduct more thorough due diligence (Tyebjee & Hardin, 2004). This has the potential to foster a more well-aligned partnership in which the resources required from the pharma firm coincide with its expected ROI. Finally, we find that project-level characteristics, such as the current drug development phase, therapeutic area, or underlying technology, do not significantly help to predict post-launch termination.

Counterfactual and SHAP analyses shed further light on how these attributes affect post-launch termination risk. In particular, the risk of post-launch termination is reduced when the pharma partner is a large firm and maintains a moderate alliance portfolio size *relative* to its R&D portfolio. Moreover, the pharma partner's failure experience is moderated by R&D portfolio size: firms with more past failures relative to the breadth of their R&D portfolio are even less likely to terminate alliances post-launch. For biotech firms seeking stable partners, these results suggest that they should prioritize large pharma firms with extensive R&D operations and experience from multiple terminated drug projects. When evaluating partners with many active alliances, biotech firms should favor those with mid-sized

operations, as they are less likely to end alliances after a launch. This finding contrasts with Dan and Zondag (2016), who report that smaller pharma firms are generally less prone to terminate alliances, indicating that such patterns may not extend to the post-launch context. A common theme is that biotech firms can use predictive models to help guide pharma partner selection and reduce the risk of post-launch alliance termination. Alternatively, contracts can include specific opt-out clauses or extended notification periods to protect against unexpected partner withdrawal (Savva & Scholtes, 2014).

From a theoretical perspective, the implications from the aforementioned findings are twofold. First, while prior research has highlighted the important of collaborative biotech-pharma partnerships (e.g., Baum et al., 2000; Polidoro & Yang, 2021; Powell et al., 1996; Rothaermel, 2001; Stuart et al., 2007), and the risks of opportunistic behavior in asymmetric partnerships (Kalaignanam et al., 2007; Mishra et al., 2015; Villena et al., 2011), we show that a partner's success can also create a vulnerability for the more dependent party. Our work thus expands the literature on the "dark side" of social capital by highlighting how alliance instability can emerge not only from failure or conflict, but also in the wake of successes. Second, our results suggest that alliance termination decisions are shaped by multiple interacting factors. From the pharma firm's perspective, launching a drug can drive a reallocation of managerial attention and capital toward commercialization. Such a shift from exploration to exploitation is known to terminate R&D initiatives (Koval, 2021; Min, 2017; Tandon et al., 2022), but it does not automatically lead to alliance termination. Our analysis suggests that projects terminated after a drug launch are just as likely to progress to the next stage of development as compared to other terminated projects, thereby ruling out the possibility that pharma firms hold on to alliances just to end them after a drug launch. Additionally, the timing of post-launch terminations, typically extending over three years after a launch, aligns with the period of intensified commercialization efforts. Thus, it is more likely that pharma firms reevaluate their portfolio pipeline by assessing the expected returns and strategic complementarity of remaining alliances according to their new resource availability.

While our study provides actionable insights, some limitations must be acknowledged. We rely on data from a leading pharmaceutical intelligence platform, which ensures practical relevance but does not capture the full causal structure behind termination decisions. Our main objective is not to estimate causal effects but to uncover predictive patterns using only features available at the time of alliance formation. We do however provide evidence that drug launches pose a risk to the alliance's survival, but we do not exploit exogenous variation in launch timing or alliance structure, meaning unobserved confounding may still influence the results. That said, we strengthen internal validity by applying out-of-sample validation, using only pre-formation features, and directly ruling out strategic timing as an alternative explanation. Future research could build on our findings by incorporating stock market data, regulatory disclosures, or internal documents to better uncover underlying mechanisms.[21]

Even though our analysis focuses on the biopharmaceutical industry, where data is readily accessible and advanced analytics are used in practice (Banner et al., 2021; Yadav et al., 2024), we believe the insights are applicable to other industries. Firms in diverse industries form exploratory alliances that are terminated when exploitation opportunity emerges (Koval, 2021). For instance, consider GoPro's partnership with Red Bull to cross-promote their products (GoPro, 2016). This alliance ended shortly after GoPro launched new live-streaming camera features that drew significant attention (GoPro, 2020), illustrating how even promising alliances can be vulnerable to a new product launch. Our study thus offers a starting point for future research to examine similar risks in other industries and to advance a more general theory of alliance management following partner success.

ML models are widely applied to optimize operations and forecast demand, and there is growing use for evaluating management and policy interventions (Boutilier et al., 2022; Chan et al., 2023; von Zahn et al., 2024; G. Wang, 2022). We contribute to this emerging area by applying predictive models

21 While we relied on prior studies to identify factors that would predict the likelihood of alliance termination between biotech and pharma firms after the launch of a drug by the pharma firm, our study did not include all these factors, albeit for justifiable reasons. The excluded predictors, along with justifications, are presented in Appendix M. Nonetheless, future research could explore whether the omitted predictors, especially the network-level factors, might play a significant role in influencing alliance termination decisions in other contexts

to uncover insights into a previously unexplored organizational dynamic. This approach aligns well with calls for the adoption of AI for data-driven theory development (Chou et al., 2023; Choudhury et al., 2021; Shrestha et al., 2021). Mirroring the biopharmaceutical industry's widespread adoption of ML and analytics (Yadav et al., 2024), we advocate for using such tools to inform alliance partner decisions and suggest accounting for the potential destabilizing effects that future drug launches may bring.

Chapter 4:
Drug development discontinuations and innovation outcomes: The mediating role of new collaborations

4.1. Introduction

Inventors in science-based industries face long development cycles, high uncertainty, and frequent project discontinuations (Krieger, 2021; Pisano, 2006). Projects can be discontinued for scientific, regulatory, or strategic reasons, and these discontinuations are essential for managing research portfolios (Paul et al., 2010; Subramanian et al., 2020). In biopharmaceutical research and development (R&D), discontinuations can impact the subsequent innovation output of the inventors involved. Although prior research has examined the organizational and financial consequences of discontinuing R&D projects (DiMasi et al., 2016; Subramanian et al., 2020), as well as the psychological responses to negative project outcomes (Shepherd et al., 2011; Ucbasaran et al., 2013), much less is known about how such project discontinuations affect the subsequent innovative activity of the inventors themselves. This is problematic because inventors are at the heart of innovation.

Innovation in this sector requires collaboration to access complementary expertise (Rothaermel & Deeds, 2004; Singh & Fleming, 2010), and such collaborations are associated with greater innovation output and more novel innovations (G. Ahuja, 2000; Singh & Fleming, 2010). A discontinued project can influence inventors' subsequent collaboration through multiple mechanisms. On the one hand, collaborators can interpret a discontinuation as a negative signal about the capabilities of the inventors involved, particularly in settings where individual contributions and project progress are difficult to observe directly (J.-Y. Kim & Miner, 2007). Consistent with attribution theory (Kelley, 1973; Weiner, 1985), such attribution could discourage forming new collaborations with inventors associated with discontinued projects. In contrast, problemistic search theory suggests that discontinuations can provide feedback that encourages inventors towards new collaborators to explore new technological directions (Greve, 2003). Because collaborations provide access to complementary expertise and knowledge,

changes in collaboration opportunities can shape subsequent innovation outcomes. To resolve this puzzle, we therefore ask: *Does project discontinuation influence subsequent inventor-level innovation outcomes by changing future collaboration opportunities?* We also examine whether this mediation pathway varies with the reason for discontinuation.

Biopharmaceutical R&D provides an ideal setting for studying how project discontinuations shape collaboration since inventors in this sector are highly specialized and depend on complementary expertise from collaborators (Rothaermel & Deeds, 2004). Additionally, project discontinuations are publicly observable through clinical trial registries, company disclosures, and regulatory filings (Sukhatme, 2007; Zarin et al., 2011). These disclosures can influence how scientific contributions are evaluated (Jamieson et al., 2019), and reputational signals may affect inventors' ability to attract new collaborators in pharmaceutical R&D (Wang & Jiao, 2022). Despite the importance and visibility of discontinuation events in this sector, scholars have largely examined the firm- or project-level consequences of discontinuation decisions, including strategic portfolio adjustment and organizational learning (e.g., Kim & Miner, 2007; Krieger, 2021; Love et al., 2023; Ucbasaran et al., 2013), leaving the individual inventor as a relatively understudied unit of analysis. It remains unclear whether and how project discontinuations affect the formation of new collaborations and whether the resulting changes affect subsequent innovation outcomes. Understanding these consequences is important because individual inventors drive knowledge production and changes in their outputs following a discontinuation can inform how to manage R&D talent and how policymakers should design disclosure policies.

In this study, we evaluate the effect of project discontinuations on subsequent new collaborations and their mediating effect on inventor-level patent outputs (i.e., productivity) and patent novelty (i.e., new technological directions that are new to the inventor, not to the world). We also examine whether the impact of project discontinuation varies with the reason for discontinuation. To do so, we construct an inventor–project panel dataset that links drug development projects to the inventors involved in their associated patents. Using information from the Cortellis Drug Development Intelligence database and

patent records, we identify inventors whose projects were discontinued and compare their subsequent innovative activity to that of inventors working on matched projects that were not discontinued. We then use difference-in-differences models to examine how discontinuation events affect collaboration formation and subsequent innovation outcomes. Our results show that project discontinuations lead inventors to form more new collaborations, suggesting that inventors respond to discontinuation by accessing new knowledge sources. This results in more novel patents but fewer patents overall. Mediation analyses indicate that these new collaborations mediate subsequent innovation outcomes, particularly patent novelty. We also find that these effects vary with the reason for discontinuation. The increase in new collaborations is stronger following scientific discontinuations (e.g., safety concerns) compared to non-scientific ones (e.g., market condition changes). However, relative to non-scientific discontinuations, the mediating role of new collaborations following scientific discontinuations has a stronger negative association with patent output and a weaker positive association with patent novelty.

This study contributes to three streams of research. First, it contributes to the literature on innovation and collaboration by showing how negative project outcomes reshape subsequent collaboration formation and innovation performance. Expanding on prior studies that state benefits of collaboration for innovation (G. Ahuja, 2000; Singh & Fleming, 2010), our findings show that new collaborations formed after project discontinuations increase novelty of innovation but reduce the inventor's productivity. Second, our study contributes to research on R&D project discontinuation and failure by uncovering the mediating role of new collaborations. Although project discontinuations could discourage collaboration through negative attribution, our results suggest that inventors respond to discontinuation by forming more new collaborations, leading to more novel technological directions, consistent with problemistic search theory. Third, responding to a call to study problemistic search as a temporal process (Posen et al., 2018), we evaluate how inventors' collaboration and patent trajectories evolve over time following discontinuation events. The observed lagged increases in new collaborations and patent novelty suggest that reorientation occurs gradually, consistent with coordination costs from forming and integrating new collaborations (Bercovitz & Feldman, 2011).

4.2. Theoretical framework and hypotheses

Project discontinuations represent shocks in the research process that can influence the subsequent innovation activities of the inventors involved. Prior research shows that scientists and inventors adjust their research trajectories following such shocks, often redirecting their efforts toward new technological directions (R. Hill et al., 2025; Madsen & Desai, 2010; Posen et al., 2018). However, these shifts can carry performance costs. Evidence from recent work on scientific careers suggests that researchers who pivot toward new research directions following shocks often experience a temporary decline in productivity (R. Hill et al., 2025). As such, we expect that inventors will produce fewer patents but pursue more novel technological directions following a project discontinuation, *ceteris paribus*.

However, drug development research relies heavily on collaboration (Tian et al., 2021), and project discontinuations may influence the formation of subsequent new collaborations. As collaborations provide access to complementary expertise and knowledge, changes in collaboration formation can influence inventors' subsequent patent activities. Project discontinuations can affect collaboration through multiple mechanisms. On the one hand, collaborators may interpret discontinuation as a negative signal about the capabilities of the inventors involved, which could reduce the formation of new collaborations. On the other hand, discontinuation could lead inventors to seek new collaborators to explore alternative technological directions. We next examine how project discontinuations influence the formation of new collaborations and how these collaborations mediate subsequent innovation outcomes. Figure 4.1 summarizes the theoretical framework that integrates the mechanisms and hypotheses developed below.

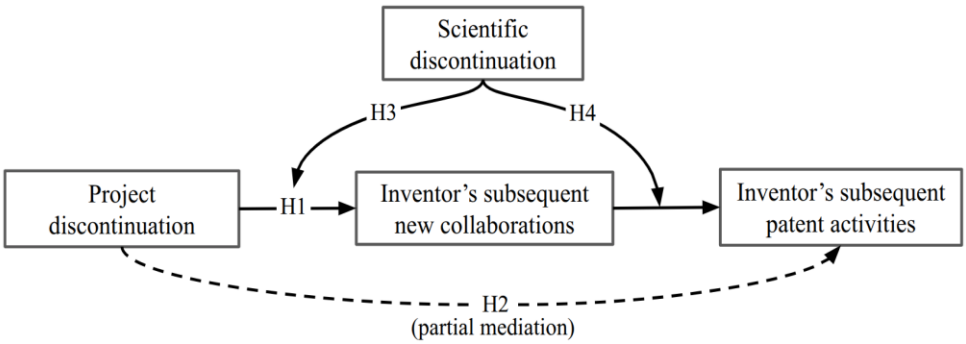


Figure 4.1. Theoretical Framework[22]

4.2.1. Effect of project discontinuation on subsequent new collaborations

The discontinuation of drug development projects is routinely disclosed through clinical trial registries and regulatory filings (Rathee et al., 2025; Tyagi et al., 2025; Zarin et al., 2011). Research shows that such disclosures are used by external audiences to assess scientific progress (Light & Lexchin, 2021; Neuberger et al., 2017). Because individuals are closely associated with the outcomes of the projects to which they contribute (J.-Y. Kim & Miner, 2007), discontinuation events can be interpreted as reflections of inventors' scientific contributions and abilities.

Attribution theory posits that observers infer causes from observable outcomes when underlying processes are opaque (Heider, 1958; Kelley, 1973; Weiner, 1985). In the context of biopharmaceutical R&D, where scientific progress and individual contributions are difficult to observe directly (Sukhatme, 2007), publicly disclosed project discontinuations can therefore serve as signals about inventor capability. Because collaboration depends on perceptions of inventor competence, such attribution could reduce potential collaborators' willingness to initiate new ties following a project discontinuation. Consequently, we hypothesize that:

H1a. *Project discontinuation will reduce an inventor's subsequent formation of new collaborations.*

[22] We do not estimate the direct association between new collaborations and patent activities because both measures are constructed from the same patent records, which introduces endogeneity. Moreover, prior research has already established a positive relationship between collaboration and innovative output (Wuchty et al., 2007).

While project discontinuations could discourage collaboration through negative attribution, they can also trigger problemistic search and adaptation behavior by the inventors involved. Problemistic search theory suggests that negative performance feedback can stimulate search for alternative solutions when existing approaches fail to meet expectations (Cyert & March, 1963; Greve, 2003; Posen et al., 2018). Recent research suggests that innovation failures can generate learning opportunities that shape subsequent innovation strategies and experimentation (Baxter et al., 2023; T. Wang, 2023), and negative signals can encourage adaptive exploration of new domains (Zeijen et al., 2025). In the context of drug development, project discontinuation represents a signal that a particular approach is not yielding the desired result. Such feedback can redirect research efforts and lead to new collaborations for complementary expertise. If inventors respond to project discontinuation through problemistic search, discontinuation events should lead to an increase in the formation of new collaborations. Consequently, we propose the following alternative hypothesis:

H1b. *Project discontinuation will increase an inventor's subsequent formation of new collaborations.*

4.2.2. Collaboration as a mediator linking discontinuation and innovation

A change in the formation of new collaborations following a discontinuation can mediate the relationship between project discontinuation and subsequent innovation outcomes. Research shows that collaborations provide access to diverse expertise and are associated with both higher innovation output and more novel innovations (G. Ahuja, 2000; Singh & Fleming, 2010). If project discontinuations reduce the formation of new collaborations (H1a), inventors will have fewer opportunities to access complementary knowledge and expertise, which can limit both the quantity and novelty of subsequent innovations (Rhee & Haunschild, 2006; Shepherd & Haynie, 2011). Conversely, if project discontinuations increase the formation of new collaborations (H1b), inventors gain access to additional sources of expertise that can support higher patent output and exploration of new technological directions (Greve, 2003; Love et al., 2023b; Madsen & Desai, 2010; Shaik et al., 2023; Tsinopoulos et al., 2019). In either case, changes in the formation of new collaborations provide a mechanism through

which project discontinuations can influence subsequent innovation outcomes. Therefore, we hypothesize that:

H2. *The formation of new collaborations will be positively associated with an inventor's subsequent (a) patent output and (b) patent novelty and will mediate the relationship between project discontinuation and these subsequent innovation outcomes.*

4.2.3. Reason for discontinuation as a moderator

The effect of project discontinuation on subsequent collaboration formation may vary depending on the reason for the discontinuation. This reason can influence how inventors and prospective collaborators interpret the event and adjust their subsequent behaviors (Kelley, 1973). Discontinuations attributed to scientific causes, such as lack of efficacy or safety concerns, provide direct feedback about the limitations of the underlying scientific approach. From an attribution theory perspective, they may be interpreted as negative signals about an inventor's capabilities which could discourage the formation of new collaborations (Shepherd & Haynie, 2011). From a problemistic search perspective, this feedback can signal that the current research trajectory is unlikely to succeed, leading inventors to adjust their research direction and form collaborations to access complementary expertise (Greve, 2003). In contrast, discontinuations attributed to external factors, such as portfolio reprioritization or changes in market conditions, provide less information about the scientific viability of the project. They are thus less likely to influence collaboration formation or research direction.

As a result, the effect of project discontinuation on the formation of new collaborations should depend on the reason for discontinuation. Discontinuations attributed to scientific reasons provide stronger informational signals and therefore should generate stronger effects on new collaboration formation, either positive or negative, than discontinuations not attributed to scientific reasons. If scientific discontinuations are interpreted as reflecting inventor capability, the discontinuation–collaboration relationship should become more negative. If scientific discontinuations trigger problemistic search, the discontinuation–collaboration relationship should become more positive. Therefore, we hypothesize that:

H3. *When discontinuations are attributed to scientific reasons, the discontinuation–*

collaboration effect will be stronger (more negative in H1a or more positive in H1b) than when discontinuations are attributed to non-scientific reasons.

While scientific failures signal that the current technological approach is not succeeding, inventors may respond by forming collaborations with specialists who can help diagnose and resolve specific technical problems. Such collaborations are likely to focus on incremental improvements to existing approaches through local search, which can require substantial effort without necessarily producing new patentable inventions (Fleming, 2001; March, 1991; Posen et al., 2018; Zeijen et al., 2025). When new patents are produced from local search, they are also more likely to be within the same technological domain as an inventor's prior patents. In contrast, non-scientific discontinuations, like strategic reprioritization of projects, can redirect inventors toward new technological domains through more exploratory collaborations. Exploratory search can lead to additional patent activities (Rosenkopf & Nerkar, 2001) and is more likely to span technological domains novel to the inventor. Therefore, we hypothesize that:

H4. *The indirect effect of project discontinuation on (a) patent output and (b) patent novelty, via new collaborations, will be reduced for scientific discontinuations compared to non-scientific discontinuations.*

4.3. Method

4.3.1. Data sources and sample construction

We construct a biopharmaceutical inventor–project panel dataset by merging drug development project records with patent and inventor data. Information is drawn from the Cortellis Drug Development Intelligence Database, which contains comprehensive coverage of pharmaceutical R&D projects, including discontinuation events, and links to associated patents. Inventors are identified through inventorship information on patents associated with each project, which allows us to link individual inventors to their collaborators and to specific drug development efforts. We focus on projects discontinued between January 1, 2000 and January 1, 2019. We end the discontinuation window in 2019 to ensure five years of post-discontinuation patent activities can be observed for each inventor.

For each discontinued project, we extract information on project development phase, targeted

medical indication, discontinuation year, discontinuation reason, and firm characteristics. We filtered duplicate discontinuation records from multiple announcements regarding a single event by preserving the earliest announced discontinuation date. We collected a total of 6,723 discontinued projects.

We construct a matched sample of drug development projects that do not experience a discontinuation during the study window to use as a counterfactual comparison group. For each discontinued project, we identify related projects that are active in the same calendar year and share similar observable characteristics prior to the discontinuation event. We perform exact matching on projects that share the same medical indication, target medical action, and inception calendar year as the discontinued project. When more than five exact matches are available, we randomly select five control projects. If fewer than five control projects meet this threshold, we retain all available matches. Matched projects are required to remain active (i.e., not discontinued) for at least the length of the study window to ensure no discontinuation treatment affects patent trajectories. Each related project is assigned a pseudo-event year corresponding to the discontinuation year of its matched discontinued project. In total we collected 14,732 matched projects.[23]

We construct the inventor–year panel by linking the patents that Cortellis associates with each drug project to patent filing records that identify the inventors. In total we identified 58,138 inventors who experienced a discontinuation and 105,991 matched inventors who did not experience a discontinuation. We then extract all patent publications corresponding to the 164,129 inventors and identify those granted within five years of the discontinuation event (or pseudo-event for the matched sample). Figure 4.2 illustrates this process for *patent output*. The same process applies to *patent novelty* and *new collaborations*.

[23] Among the discontinued projects, 201 have one match, 185 have two matches, 140 have three matches, 115 have four matches, and 5,678 have five matches. Also, 404 discontinued projects have no matches. We compared results including and excluding these unmatched projects with similar results.

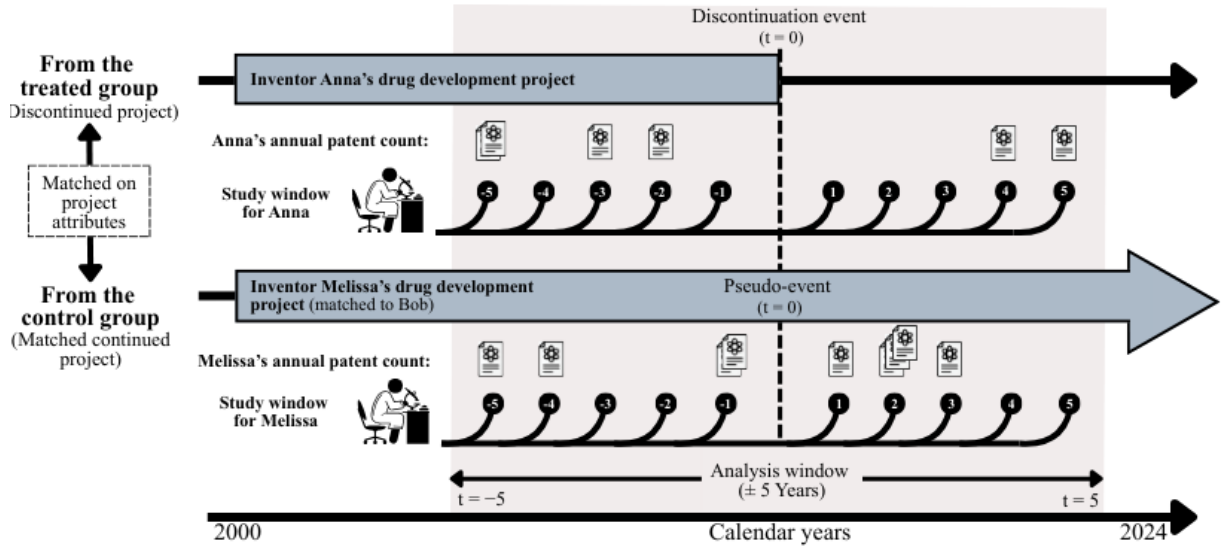


Figure 4.2. Example of Patent Output for One Focal and One Matched Inventor

Many innovations generate multiple patent publications (e.g., filings in different jurisdictions) that belong to the same patent family (i.e., they represent the same innovation). To avoid double counting, we retain only the first publication in each family and assign the earliest observed publication year. For each family, we retain co-authors as a measure of collaborators, and International Patent Classifications (IPC) 4-character prefix codes as a measure of *patent novelty*. To ensure that new collaborations are based on new co-authors and that new technological directions are based on new IPC codes, we construct pre-event baselines of co-inventors and IPC codes using observations more than five years prior to the discontinuation (from the inventor's first recorded patent to five years prior to the event). We do so by extracting co-inventors and IPC codes from every patent related to an inventor that were granted more than five years before the discontinuation (or pseudo-discontinuation) event. These baselines are used to account for past collaborators and IPC codes when constructing panel data.

Using the linked project–inventor–patent data, we construct an inventor-level event-time panel centered on project discontinuations. The unit of analysis is 'the inventor × year relative to the event'. For inventors associated with a discontinued project, we define the event year as the calendar year in which the focal project is discontinued. For inventors in the matched sample, we assign a pseudo-event year equal to the discontinuation year of the project to which they are matched. We then retrieve all

patents for each *inventor* × *relative year* to compute the outcome measures described next.

4.3.2. Outcome, mediator, and moderator measures

We employ outcomes that capture the quantity and novelty of innovation. We also measure a mediator that captures the number of new collaborations each year relative to the discontinuation and a moderator that captures whether the discontinuation was due to a scientific reason. The moderator is defined at the project level. The outcome and mediator variables are defined at the inventor–year level and indexed in event-time relative to the discontinuation year. Although the mediator is measured in the same year as the outcomes, collaborations observed through patent co-authorship must have been formed before the patent is granted. Therefore, the collaboration necessarily occurs prior to the patent activities. For each measure, inventor–years with no patent activity are coded as zero. Table 4.1 summarizes these measures.

	Variable name	Description
Outcomes	**Patent output**	The number of distinct patent families which the focal inventor is granted a patent in year i.
	Patent novelty	The number of new IPC code prefixes in year i relative to the focal inventor's pre-event IPC portfolio, divided by the number of patents in year i. IPC codes are operationalized at the 4-character subclass level (e.g., "G16H").
Mediator	**New collaboration**	The number of new co-inventors in year i relative to the focal inventor's pre-event collaborator set, divided by the number of patents in year i.
Moderator	**Scientific reason**	An indicator equal to 1 if the focal project was discontinued for a scientific reason (e.g., lack of efficacy or safety concerns) and 0 otherwise. Non-scientific discontinuations include commercial, strategic, or financial decisions.

Table 4.1. Definitions of Outcome Measures

We define *patent novelty* and *new collaborations*, as formalized in Eq. (6) below, as the number of distinct attribute values that appear in year τ but were not observed in prior years. We denote A_τ as the set of distinct values for attribute A in year τ (e.g., IPC 4-character prefix or collaborator name). We denote novelty by $n_{A,\tau}^{new}$ as the count of elements in A_τ that do not belong to any previous sets A_t for $t < \tau$. Formally,

$$n_{A,\tau}^{new} = |A_\tau \notin A_{t<\tau}|. \quad (6)$$

The moderator is based on the stated reason for project discontinuation reported in the Cortellis

Drug Development Intelligence Database. Cortellis reports the primary reason associated with each discontinuation using standardized categories. We construct an indicator variable equal to 1 if the project was discontinued for lack of efficacy or safety concerns (scientific reason), and 0 otherwise (non-scientific reason). Table 4.2 presents the descriptive statistics for these outcome measures.

Variable	Mean (Treated)	SD	N	Mean (Control)	SD	N
Cumulative number of patents before event window (relative years ≤ −5)	5.334	54.8458	4,352,502	4.4069	70.5521	7,778,991
Annual patent output	5.5790	69.5802	4,352,502	5.4871	88.6489	7,778,991
Annual patent novelty (new IPC 4-character subclasses)	4.2408	15.6488	4,352,502	3.4665	17.0543	7,778,991
Annual new collaborations	15.6818	61.2594	4,352,502	13.6622	71.9456	7,778,991

Table 4.2. Descriptive Statistics of Outcome Measures

Because *patent output* follows a right-skewed distribution common in patent data (Silverberg & Verspagen, 2007), we apply an inverse hyperbolic sine (IHS) transformation to the *patent output* measure to reduce skewness while retaining zero observations (Bellemare & Wichman, 2020). We scale *patent novelty* and *new collaborations* by *patent output* in the same year by constructing ratios that represent novelty and collaborations per patent. This reduces the correlation between these measures and *patent output* and instead represents the intensity of novelty and collaborations per patent.

4.3.3. Analysis approach

We estimate the effect of project discontinuation using an event-time difference-in-differences approach centered on the discontinuation year of drug development projects. Treated (discontinued) and control (not discontinued) projects are matched on pre-event characteristics, which ensures that they are comparable based on observable factors before the discontinuation event. In addition, assigning pseudo-event years to control projects and including calendar-year fixed effects ensures that both groups are exposed to the same time conditions. Together, these account for project-level selection on observables and timing conditions. A key assumption is that the outcomes of treated and control

inventors would follow similar trends in the absence of project discontinuations. We assess this assumption by examining pre-discontinuation trends for each model.

All models include inventor fixed effects, which control for time-invariant differences in inventors' baseline activity, and calendar-year fixed effects, which control for shocks common to all inventors in a given year. Standard errors are clustered at the inventor level to account for within-inventor serial correlation, and all statistical inference is based on these clustered standard errors. Eq. (7) represents our event-study difference-in-differences design with

$$Y_{it} = \alpha_i + \gamma_t + \sum_{k=-5, k\neq 0}^{5} \beta_k \left(T_i \times D_{it}^k\right) + \epsilon_{it}. \quad (7)$$

Y_{it} is the outcome variable, α_i are inventor fixed effects, and γ_t are calendar-year fixed effects. T_i is a treatment indicator equal to 1 if inventor i experiences a discontinuation and 0 otherwise. D_{it}^k is an event-time indicator equal to 1 if inventor i is k years from the (real or pseudo) discontinuation event in year t, and 0 otherwise. The coefficients β_k are the average treatment effects on the treated at time k relative to the year of the discontinuation ($k = 0$).

4.3.3.1. Mediation analysis

We also examine whether changes in collaboration account for the effect of project discontinuation on inventors' subsequent patent activities using an event-study difference-in-differences mediation analysis. We first estimate an event-study model for the mediator, as shown in Eq. (8) with

$$M_{it} = \alpha_i + \gamma_t + \sum_{k=-5, k\neq 0}^{5} \pi_k \left(T_i \times D_{it}^k\right) + u_{it}, \quad (8)$$

where M_{it} is the mediator variable (i.e., new collaborations per granted patents). The coefficients π_k are the effect of discontinuation on the mediator at each event time k. Hence, Eq. (8) will be used to test **Hypothesis 1**. We then estimate an event-study model for the outcome conditional on the mediator, as shown in Eq. (9) with

$$Y_{it} = \alpha_i + \gamma_t + \sum_{k=-5, k\neq 0}^{5} \beta_k \left(T_i \times D_{it}^k\right) + \theta M_{it} + \epsilon_{it}, \quad (9)$$

where, again, β_k are the average treatment effects on the treated at time k relative to the year of the discontinuation ($k = 0$), and θ is the relationship between the mediator and the outcome conditional

on treatment, event time, and fixed effects. Hence, Eq. (9) will be used to test **Hypothesis 2**. Both models include inventor fixed effects (α_i) and calendar-year fixed effects (γ_t). Standard errors are clustered at the inventor level.

4.3.3.2. Moderator analysis

We examine the heterogeneity in the effect of discontinuation on collaboration with moderator analyses to test **Hypothesis 3**. This is represented by Eq. (10) with

$$M_{it} = \alpha_i + \gamma_t + \sum_{k=-5,k\neq 0}^{5}\left[\pi_k\left(T_i \times D_{it}^k\right) + \lambda_k\left(T_i \times D_{it}^k \times Z_i\right)\right] + u_{it}. \quad (10)$$

Z_i represents the moderator for inventor i, and λ_k captures whether the effect of a discontinuation on new collaborations varies for different moderator values. We assess statistical significance of the moderator using a joint Wald test of the interaction coefficients between the treatment indicators, event-time indicators, and the scientific-reason moderator (λ_k). The tests evaluate whether the interaction coefficients are jointly different from zero across event years, which allows us to determine whether the moderator has an overall effect even if individual event-year coefficients are not statistically significant.

To examine whether the mediation effect is moderated by the reason for discontinuation, we estimate moderated mediation models to test **Hypothesis 4**. This is represented by Eq. (11) with

$$Y_{it} = \alpha_i + \gamma_t + \sum_{k=-5,k\neq 0}^{5}\left[\beta_k\left(T_i \times D_{it}^k\right) + \lambda_k\left(T_i \times D_{it}^k \times Z_i\right)\right] + \theta M_{it} + \rho(M_{it} \times Z_i) + \epsilon_{it}, \quad (11)$$

where β_k is the event-time treatment effect for non-scientific discontinuations in year k. The coefficient λ_k represent the additional effect when the discontinuation occurs for scientific reasons. Therefore, the total treatment effect for scientific discontinuations is $\beta_k + \lambda_k$. For the mediator terms, θ is the relation between the mediator and outcome for non-scientific discontinuations. The coefficient ρ is the additional effect when the discontinuation occurs for scientific reasons. Therefore, $\theta + \rho$ represents the mediator effect when a discontinuation is for scientific reasons.

4.4. Results

4.4.1. Trends and total effects

This subsection provides descriptive trends and the total effects of project discontinuation on patent output and novelty before testing the hypotheses with the mediator and moderator in the next subsections. Figure 4.3 presents the event-time trends in annual patent outputs for inventors whose projects are discontinued and for inventors in the matched comparison sample. *Patent output* is presented as percentage changes relative to the event year to facilitate comparison across samples. In the five years preceding the discontinuation event, the two groups follow closely aligned trajectories. Over this period, inventors in both groups experience a gradual increase in annual patent publications. Following a discontinuation, however, treated inventors deviate from this trajectory. Their output declines relative to the pre-event trend for a few years before returning to its pre-event trend. In contrast, matched control inventors remain on a similar growth path after their pseudo-event. While Figure 4.3 does not control for covariates or fixed effects, it provides visual evidence of a parallel pre-event trend between treated and matched control inventors. It also motivates the regression-based event-study analyses that follow.

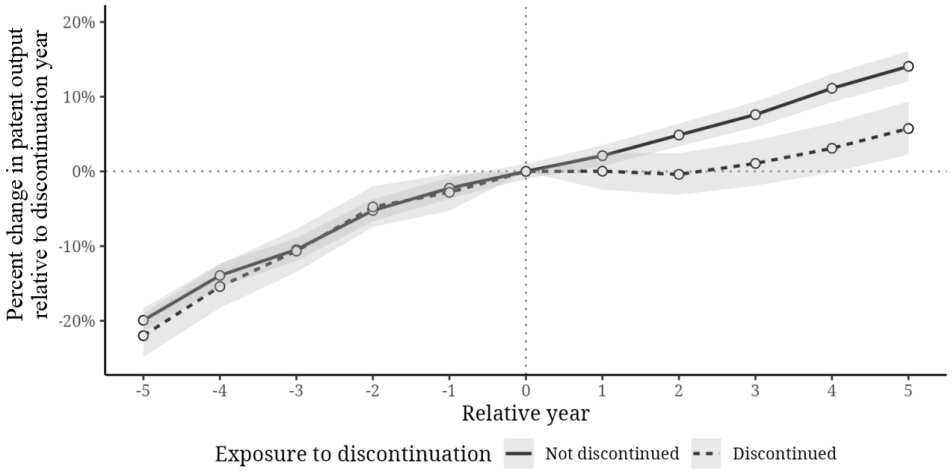


Figure 4.3. Patent Output Trajectories for Treated and Control Groups

Figure 4a and Figure 4b report event-study difference-in-differences estimates of the effect of project discontinuation on (a) *patent output* and (b) *patent novelty.* Each point represents the average treatment effect in a given year relative to the discontinuation. Error bars represent 95% confidence

intervals based on inventor-level clustered standard errors. In Figure 4a, the coefficients for *patent output* follow parallel trends in the pre-discontinuation period since the coefficients prior to discontinuation ($t < 0$) are indistinguishable from zero. There appears to be some decline in *patent output* after discontinuations, though this trend is only statistically significant in years 4 and 5. Figure 4b also shows a parallel trend leading up to the discontinuation, then *patent novelty* increases after the discontinuation to become statistically significant from year 2 until year 5.

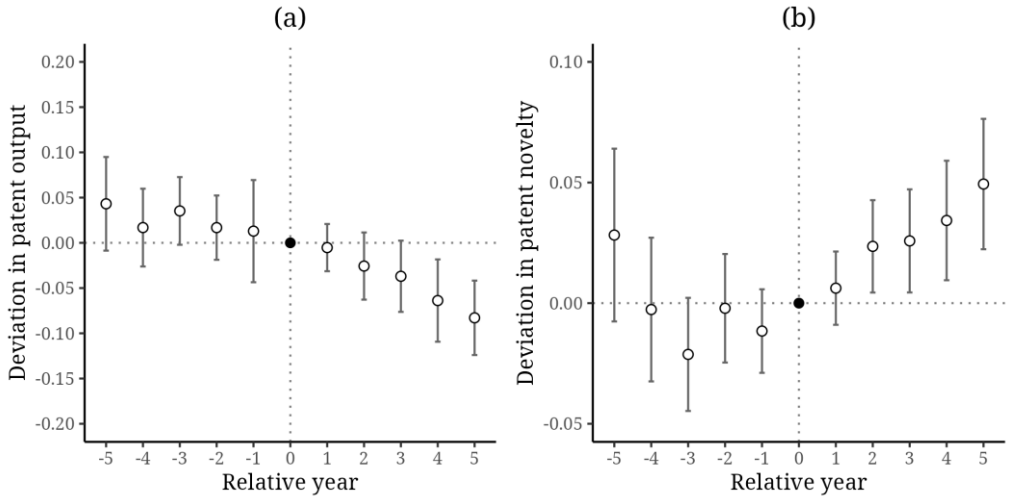


Figure 4.4. Event-time Regression Coefficients Representing Total Treatment Effects for (a) patent output and (b) patent novelty

4.4.2. Effects on forming new collaboration (mediator)

Figure 4.5 presents event-study estimates of changes in *new collaborations* measured by the number of distinct co-inventors per patent who have not previously appeared on patents together, corresponding to the competing **Hypotheses 1a** and **1b**; the regression table is reported in Appendix N. Figure 4.5 shows no difference in *new collaborations* between the two samples prior to discontinuation, also consistent with the parallel-trends assumption. In the post-discontinuation period, new collaborations formation increases compared to the matched sample. Therefore, Hypothesis 1b is supported and Hypothesis 1a is not.

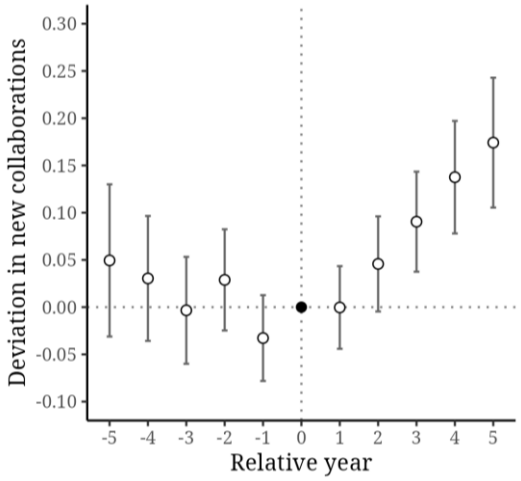


Figure 4.5. Event-time Regression Coefficients Representing Effect of Discontinuations on New Collaborations (Mediator)

4.4.3. Mediation Analysis

We assess whether changes in *new collaborations* contribute to the observed change in patent activities following project discontinuation by examining the role of *new collaborations* as a mediator, corresponding to **Hypothesis 2**. Table 4.3 presents the results. Models 1 and 3 report the baseline event-study estimates for patent output and patent novelty, respectively, without including the mediator. Models 2 and 4 are extensions that include *new collaborations* as mediator. Accordingly, *new collaborations* is negatively associated with *patent output* (Model 2) but positively associated with *patent novelty* (Model 4). When we include the mediator in the patent output model (Model 2), the treatment-effect estimates remain unchanged compared to the baseline model (Model 1). This indicates that although *new collaborations* are negatively associated with *patent output*, they do not appear to mediate the effect of discontinuation on *patent output*. For *patent novelty*, including new collaborations as a mediator in Model 4 reduces the magnitude and statistical significance of the treatment-effect estimates compared to Model 3. Although the treatment coefficients remain statistically significant, the reduction suggests that the increase in *patent novelty* is partially mediated through an increase in *new collaborations*. Hypothesis 2 is thus supported for patent novelty (2b), but not for patent output (2a) since we find the opposite effect.

	Patent output		Patent novelty	
	(1)	**(2)**	**(3)**	**(4)**
Treated × t = −5	0.0431	0.0442	0.0282	0.0251
	(0.0264)	(0.0263)	(0.0154)	(0.0141)
Treated × t = −4	0.0169	0.0174	−0.0027	−0.0043
	(0.0219)	(0.0221)	(0.0128)	(0.0121)
Treated × t = −3	0.0353	0.0351	−0.0213	−0.0208
	(0.0191)	(0.0188)	(0.0101)	(0.0099)
Treated × t = −2	0.0168	0.0175	−0.0021	−0.0042
	(0.0181)	(0.0178)	(0.0097)	(0.0091)
Treated × t = −1	0.0129	0.0119	−0.0116	−0.0084
	(0.0288)	(0.0286)	(0.0074)	(0.0073)
Treated × t = 1	−0.0053	−0.0055	0.0062	0.0070
	(0.0133)	(0.0131)	(0.0065)	(0.0064)
Treated × t = 2	−0.0257	−0.0246	0.0235**	0.0202**
	(0.0189)	(0.0186)	(0.0082)	(0.0078)
Treated × t = 3	−0.0370	−0.0346	0.0258**	0.0186*
	(0.0201)	(0.0197)	(0.0092)	(0.0084)
Treated × t = 4	−0.0638*	−0.0603*	0.0343**	0.0234*
	(0.0232)	(0.0229)	(0.0106)	(0.0095)
Treated × t = 5	−0.0829***	−0.0782***	0.0494***	0.0351***
	(0.0210)	(0.0212)	(0.0116)	(0.0658)
New collaborations		−0.0283***		0.0860***
		(0.0058)		(0.0043)
Inventor fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Clustered standard errors	Inventor level	Inventor level	Inventor level	Inventor level
Observations	12,131,493	12,131,493	12,131,493	12,131,493
R^2	0.9346	0.9349	0.26983	0.30944
Within R^2	0.00146	0.00682	0.00055	0.05477

*** p<0.001, ** p<0.01, * p< 0.05

Table 4.3. Mediation Effect of Subsequent New Collaborations on Patent Output and Patent Novelty

4.4.4. Moderator and moderated mediation analyses

Table 4.4 presents the moderating role of the reason for discontinuation. Model 5 examines whether scientific versus non-scientific reasons for discontinuation moderate the effect of discontinuation on *new collaborations*, corresponding to **Hypothesis 3**. Models 6 and 7 present the full moderated mediation analysis for the *patent output* and *patent novelty* outcomes, corresponding to **Hypothesis 4**. For all three models, the event-time coefficients show no evidence of pre-trends differences between treated and control inventors prior to a discontinuation.

	New collaborations (Mediator)	Patent output	Patent novelty
	(5)	**(6)**	**(7)**
Treated × t = −5	0.0311 (0.0265)	0.0408 (0.0256)	0.0291 (0.0216)
Treated × t = −4	0.0138 (0.0292)	0.0121 (0.0209)	0.0004 (0.0185)
Treated × t = −3	−0.0053 (0.0259)	0.0203 (0.0199)	−0.017 (0.0153)
Treated × t = −2	0.0243 (0.0212)	0.0072 (0.0170)	−0.0012 (0.0082)
Treated × t = −1	−0.0422 (0.0286)	0.0164 (0.0267)	−0.0092 (0.0079)
Treated × t = 1	−0.0224 (0.0275)	−0.0088 (0.0160)	0.0097 (0.0063)
Treated × t = 2	0.0346 (0.0218)	−0.0414 (0.0178)*	0.0214 (0.0084)*
Treated × t = 3	0.0722 (0.0259)**	−0.0545 (0.0181)**	0.0208 (0.0104)
Treated × t = 4	0.1047 (0.0289)**	−0.0776 (0.0239)**	0.0256 (0.0089)**
Treated × t = 5	0.1550 (0.0314)***	−0.0946 (0.0277)**	0.0373 (0.0105)**
Treated × t = −5 × Scientific reason	0.0676 (0.0995)	0.0529 (0.1293)	−0.0365 (0.0388)
Treated × t = −4 × Scientific reason	0.0683 (0.0779)	0.0728 (0.1227)	−0.0480 (0.0279)
Treated × t = −3 × Scientific reason	−0.0026 (0.0757)	0.1809 (0.1329)	−0.0450 (0.0228)
Treated × t = −2 × Scientific reason	0.0000 (0.0746)	0.1251 (0.1501)	−0.0330 (0.0215)
Treated × t = −1 × Scientific reason	0.0612 (0.0606)	−0.0507 (0.1471)	0.0113 (0.0193)
Treated × t = 1 × Scientific reason	0.1581 (0.0657)*	0.0419 (0.1256)	−0.0298 (0.0180)
Treated × t = 2 × Scientific reason	0.0545 (0.0909)	0.2001 (0.1286)	−0.0152 (0.0159)
Treated × t = 3 × Scientific reason	0.1331 (0.0928)	0.2396 (0.1273)	−0.0243 (0.0271)
Treated × t = 4 × Scientific reason	0.2493 (0.1030)*	0.2105 (0.1429)	−0.0214 (0.0228)
Treated × t = 5 × Scientific reason	0.1319 (0.1149)	0.1975 (0.1482)	−0.0248 (0.0274)
New collaborations		−0.0276 (0.0060)***	0.0867 (0.0052)***
New collaborations × Scientific reason		−0.0276 (0.0129)*	−0.0283 (0.0135)*
Inventor fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Clustered standard errors	Inventor level	Inventor level	Inventor level
Observations	12,131,493	12,131,493	12,131,493
R^2	0.352911	0.93504	0.240532
Within R^2	0.000659	0.007869	0.054959

*** p<0.001, ** p<0.01, * p< 0.05

Table 4.4. Moderated Mediation Effect Based on Reason for Discontinuation

Model 5 indicates that inventors gradually increase the formation of *new collaborations* following a discontinuation (*Treated* × *t*). The increase becomes positive and statistically significant three years after the event and continues to grow in later years (event years 3–5). The interaction terms (*Treated* × *t* × *Scientific reason*) test whether this effect varies for scientific versus non-scientific discontinuations. Although few interaction coefficients are individually statistically significant, a joint Wald test indicates that *scientific discontinuations* are associated with a stronger increase in *new collaborations* overall (Wald statistic = 7.94, $p < 0.001$). This finding supports Hypothesis 3, which predicted that the effect of project discontinuation on collaboration formation varies with the reason for discontinuation. Specifically, the discontinuation–collaboration effect is more positive (i.e., stronger) when discontinuations are attributed to scientific reasons than when they are attributed to non-scientific

reasons.

Model 6 represents the moderated mediation model for *patent output*. The baseline event-time coefficients (*Treated* × *t*) indicate that patent output declines beginning in the second year after non-scientific discontinuations, with statistically significant effects observed in event years 2–5. None of the interaction coefficients (*Treated* × *t* × *Scientific reason*) are individually statistically significant, suggesting that the decline in patent output does not differ significantly by discontinuation reason. The mediator coefficient (*New collaborations*) indicates that new collaborations are negatively associated with patent output. The interaction term (*New collaborations* × *Scientific reason*) is also negative and statistically significant, indicating that the negative association between *new collaborations* and *patent output* is stronger following scientific discontinuations. These results provide support for Hypothesis 4a since the relationship between new collaborations and patent output varies with the reason for discontinuation. Specifically, scientific discontinuations are associated with a more significant decline in patent output than discontinuations associated with non-scientific reasons.

Model 7 presents the moderated mediation model for *patent novelty*. The baseline event-time coefficients (*Treated* × *t*) indicate a gradual increase in patent novelty beginning two years after the discontinuation event, with positive and statistically significant effects in event years 2, 4, and 5. The mediator coefficient (*New collaborations*) shows that new collaborations are positively associated with *patent novelty*. However, the interaction term (*New collaborations* × *Scientific reason*) is negative and statistically significant, indicating that the positive association between *new collaborations* and *patent novelty* is weaker when discontinuations occur for scientific reasons. This result implies that new collaborations still increase novelty following scientific discontinuations, but the increase in novelty is smaller than after non-scientific discontinuations, providing support for Hypothesis 4b.

4.5. Discussion and conclusion

We examined whether and how an inventor's project discontinuation affects subsequent patent output (productivity) and patent novelty (technological directions new to the inventor) through the formation of new collaborations, and whether this mechanism varies by the reason for discontinuation. Using the

biopharmaceutical sector, where discontinuations are publicly disclosed and inventors can be linked to drug development projects and patents over time, we find that a project discontinuation is expected to be followed by more new collaborations, lower patent output, and higher patent novelty for inventors who have experienced a discontinuation. The results further show that new collaborations mediate the increase in patent novelty. We also find that if an inventor's project is discontinued for a scientific reason, then we can expect a stronger increase in new collaborations than if the discontinuation was for a non-scientific reason. However, the mediation effects of new collaborations on the inventor's patent output and patent novelty are reduced if the discontinuation is for a scientific reason.

Central to this study is whether project discontinuations reduce or increase new collaborations (Hypotheses 1a and 1b). Attribution theory suggests that publicly disclosed discontinuations could be seen as a negative signal of inventor capability, which could discourage potential collaborations. In contrast, problemistic search theory suggests that a discontinued project provides feedback that leads inventors to seek new knowledge sources through new collaborations. Notwithstanding this puzzle, our findings are overall more consistent with the problemistic search perspective since we find that inventors form more new collaborations after a discontinuation. Granted, this finding does not imply that negative attribution is absent but from our analysis, it is problemistic search that is likely the more dominant perspective explaining the net effect of project discontinuation. Project failures and discontinuations are common in the biopharmaceutical industry (DiMasi et al., 2016), and innovation depends strongly on collaboration among specialized individuals (Pisano, 2006; Rothaermel & Deeds, 2004). Negative attribution could be more pronounced in other industries, especially where failures are less common, collaboration among specialized inventors is less central, or individual responsibility for project outcomes is easier to assign.

An unanticipated finding pertains to our postulation of inventors' new collaborations (following a discontinuation) mediating the increase in patent output (Hypothesis 2a), yet we find this association to instead decline patent output. Since new collaborations appear to increase patent novelty relative to an inventor's prior work, these collaborations may expose inventors to unfamiliar knowledge domains

and enable novel recombination (Fleming, 2001; Walsh et al., 2016). At the same time, forming new collaborations can involve coordination costs and learning frictions (Bercovitz & Feldman, 2011; S. Ghosh & Wu, 2023) that reduce patent output. A positive mediation between new collaborations and patent novelty, along with a negative mediation with patent output, is consistent with this interpretation. This interpretation is also consistent with recent work showing that research pivots following a shock are associated with declines in productivity (R. Hill et al., 2025).

Moreover, the moderated-mediation results show that scientific discontinuations lead to a stronger increase in new collaborations than non-scientific discontinuations. However, those additional collaborations are associated with weaker patent novelty increases and a stronger negative association with patent output (Table 4.4), which is likely a product of local versus exploratory search (March, 1991; Rosenkopf & Nerkar, 2001). Scientific discontinuations provide specific technical feedback. In response, inventors may seek collaborators who can help diagnose and resolve the problem. This type of local search deepens expertise within an existing domain. It can require substantial effort and repeated trial and error while producing fewer patentable outputs. Additionally, patents that are produced could remain closer to the inventor's prior work. By contrast, discontinuations driven by (non-scientific) external forces, like portfolio reprioritization, could lead inventors to redirect their work toward novel opportunities through exploratory collaborations (Carnabuci & Operti, 2013; Dahlander et al., 2016). Such collaborations would be more likely to span new technological domains and generate inventions that are more novel relative to the inventor's prior work (Fleming, 2001; Rosenkopf & Nerkar, 2001).

Finally, we find that the effects of project discontinuations develop gradually over time. Patent output begins to decline in the second year after discontinuation and continues to fall through year five (Figure 4.4a). Patent novelty increases with a lag and becomes stronger over time (Figure 4.4b) and new collaborations follow a similar pattern with statistically significant increases beginning in year three (Figure 4.5). Because these trends follow similar lags, the decline in patent output may coincide with, and partly reflect, a reorientation towards more novel patents, particularly after non-scientific

discontinuations (Table 4.4). This latter finding contributes empirical evidence consistent with the recombinant knowledge literature, which suggests a potential tradeoff between productivity and novelty (Fleming, 2001; Kok et al., 2019). When firms redirect R&D resources toward new directions, they may need to accept a temporary decline in productivity that can last for several years.

The above findings contribute to three research streams. First, to research in innovation and collaboration we provide empirical evidence that negative project events can influence subsequent collaborations. Scholarly work in this stream shows that collaboration facilitates innovation (G. Ahuja, 2000; Singh & Fleming, 2010). Collaborations expose inventors to new knowledge domains, which can increase the likelihood of novel recombinations but also introduce greater uncertainty and coordination costs in the innovation process. We add an event-driven perspective by showing that project discontinuations can lead to new collaborations over several years. We also find that these collaborations are linked to more novel innovations but can also reduce the quantity of patent output. This pattern is consistent with prior findings supporting that search in unfamiliar knowledge domains can increase the novelty of innovation while reducing productivity (R. Hill et al., 2025).

Second, to research in R&D project discontinuation we contribute through a focus on the individual inventor. Scholars have primarily examined organization-level outcomes, including learning from failure (e.g., Krieger, 2021; Love et al., 2023; Madsen & Desai, 2010) and the financial consequences of R&D terminations (e.g., DiMasi et al., 2016). Related work has also examined how prior success and failures affect decisions to discontinue projects (e.g., Lévesque et al., 2022), though focusing on project-level decisions rather than on individual inventors. By examining inventors directly, we identify a mechanism through which project discontinuations affect subsequent innovation: the formation of new collaborations that mediate changes in inventors' later patent activities.

Finally, we contribute to the problemistic search research by offering inventor-level insights into how search is affected by project discontinuations. This research stream has mainly shown that firms increase knowledge recombination after negative performance feedback (Eggers & Suh, 2019; Greve, 2003). Recent studies show that performance feedback also influences firm's direction and locus of

search (Keil et al., 2024), and that failure can shift firms' innovation strategies toward exploration or exploitation depending on the context (Xiong et al., 2024; Zhang et al., 2024). Building on these insights, we show that similar patterns occur at the level of individual inventors, who ultimately carry out firms' innovation responses to discontinuations. Following project discontinuation, inventors form new collaborations that redirect their innovation activity toward new knowledge combinations. We further add that the reason for discontinuation matters, as different discontinuations can lead inventors to engage in different types of searches. Moreover, in line with calls to study problemistic search as a temporal process (Posen et al., 2018), we suggest that post-failure search develops over time, operates through collaboration, and affects patent output and patent novelty differently.

In terms of practical implications, we provide empirical evidence, particularly for R&D managers, on the association between project discontinuations and forming more new collaborations, suggesting that the social capital (or collaboration potential) of inventors does not disappear when a project ends. For R&D portfolio management, ending scientifically viable projects for strategic reasons does not prevent inventors from pivoting towards exploratory work in new domains (i.e., greater patent novelty), though a reduction in patent output should be expected. By contrast, scientific failures may lead towards collaborations that support more local searches and incremental improvements (i.e., lower patent novelty), which might also come with lower patent outputs. For policymakers, disclosure requirements around project discontinuations does not appear to stigmatize inventors. In our setting, discontinuations are publicly observable through clinical trial registries and regulatory filings, yet inventors form more new collaborations after these events. This suggests that transparency around discontinuation does not impede future collaborations.

Nevertheless, our findings should be interpreted while considering a few limitations. The most consequential for causal inference is the possibility of unobserved confounding. Although the matched event-study design controls for project-level observables and we control inventor-level fixed effects, a risk remains that inventors whose projects are discontinued are systematically different from matched control inventors in ways that affect both collaboration formation and subsequent innovation. Ruling

out unobserved confounding is a challenge since publicly available data do not contain specific information about each individual inventor. The parallel pre-trends provide reassurance for our results, but they cannot fully rule out unobserved selection on inventor characteristics. Future work employing experimental or quasi-experimental design could strengthen causal inference.

Second, patent-based measures capture only the formal codified dimension of innovation. Likewise, collaboration measured through co-inventorship does not capture informal knowledge exchange or other collaboration activities that do not lead to patent filings. Survey-based research could provide empirical evidence on more informal changes in innovation following project discontinuation. Another measurement limitation is that we rely on patent granted dates because these dates are readily available in the Cortellis patent data. Granted dates lag their underlying innovation activities because of patent assessment and processing time, whereas the date when a patent application is filed would shorten the timing gap between innovation activities and the measured outcome. Future research could retrieve patent application dates and re-estimate the models as a robustness test.

Finally, the five-year event window may be too short to fully capture inventors' long-term post-discontinuation trajectories. The patterns we observe are still evolving at the end of the window. Longer observation windows could reveal whether patent output eventually recovers and when the increase in novelty stabilizes or dissipates. We re-estimated the models using four- and six-year windows and obtained similar results. However, extending the analysis over much longer periods presents practical challenges. Inventors change employers, leave R&D, or retire, which makes it difficult to track their innovation activity using patent data alone. Linking patent records with inventor-level personal career data could provide a clearer understanding of long-term trajectories.

Work remains to be done to deepen our understanding of the influence of project discontinuation on subsequent innovation activities at the individual inventor level, and the role of both future collaborations and reasons for discontinuation. Nonetheless, we have moved a step forward by offering some resolution to the puzzle of whether project discontinuations reduce or increase new collaborations. In biopharmaceutical R&D, drug development project discontinuations appear to

increase new collaborations and to a greater extent when the discontinuation is for a scientific reason rather than for a non-scientific one. Innovation is an uncertain pursuit, with many setbacks along the way. As Canadian inventor Joseph-Armand Bombardier often asked, *"What should I invent next?"* Our results suggest that project discontinuations may serve to lead inventors towards novel innovations.

Chapter 5:
Conclusion

This dissertation examines how innovation disruptions in healthcare generate indirect consequences at three levels of analysis: the users that participate in or benefit from innovation; the firm developing innovation; and the inventors who create it. The multi-level perspective shows that unintended consequences of disruption are prevalent at each level, vary by context, and influence research and development (R&D) operations in various ways. The results also show that the consequences of disruptions go beyond the focal event and affects other parts of R&D. At the user level, AI-enabled monitoring can sustain participation in service co-production. At the firm level, a successful drug launch can increase alliance instability for dependent partners. At the inventor level, project discontinuations can redirect inventors toward new collaborations and more novel innovations. The findings suggest that disruption should be considered a recurrent feature of R&D operations with complex consequences that require managerial attention.

At the user level, Chapter 2 demonstrates that the introduction of AI-enabled monitoring and real-time feedback sustained user participation in physiotherapy co-production over time. Although both the control and intervention participants began with similar exercise behavior, those using the AI-enabled intervention maintained higher levels of participation while the control group's engagement declined. This pattern occurs because the technology reduced uncertainty about exercise performance, as expected, but unexpectedly the technology also reinforced accountability through progress tracking and data sharing with the physiotherapist. Therefore, a purpose-built technology can improve co-production through secondary means like improving the user-service provider relationship. At the firm level, Chapter 3 shows that a successful drug launch can destabilize existing inter-organizational relationships. Pharma firms are more likely to terminate R&D alliances with biotech partners following a drug launch, even when those alliances are unrelated to the launched drug. This suggests that a positive disruption for the focal firm can create adverse spillovers for dependent partners, likely by leading pharma firms to reprioritize their R&D portfolios or redirecting managerial attention toward

commercialization. Hence, drug launches can reduce alliance stability and increase termination risk for partners more dependent on the relationship. At the inventor level, Chapter 4 shows that project discontinuations redirect innovation trajectories. Inventors associated with discontinued drug development projects form more new collaborations over time, and these collaborations are associated with greater patent novelty but lower patent output. This pattern suggests that some negative disruptions lead inventors to reorient toward new knowledge sources and technological directions, with adjustment costs that reduce productivity.

5.1. Theoretical and Practical Implications

For research in operations management and innovation, the dissertation shows that disruption is a recurrent and consequential feature of innovation in the R&D contexts studied. Scholarly research has shown that innovation processes are complex and unstable, with variability arising from interdependencies and shifting conditions (Bagno et al., 2017; Benner & Tushman, 2003; Salerno et al., 2015). This dissertation contributes to this literature by showing how disruptions generate indirect consequences across multiple levels of the R&D system. Using data that spans decades in Chapter 3 and Chapter 4, it also shows that disruptions recur over time and that their consequences are also consistent over time. Chapter 2, on the other hand, studies individuals over a shorter period (12 weeks). However, the study enrolled participants over multiple calendar years, including periods spanning before, during, and after COVID-19 restrictions. The results remained consistent in robustness tests that control for time-related confounding, which provides evidence that these results are also consistent over time.

For research on collaboration and innovation, the dissertation demonstrates that responses to disruption often involve changes in relationships with other stakeholders. Relational changes from disruptions appears in different forms across each level of analysis. At the end-user level, data transparency from an AI-enabled solution can improve the relationship with service providers. At the firm level, a drug launch is shown to have the potential to destabilize alliances. At the inventor level, project discontinuations lead inventors to form new collaborations, which helps explain subsequent

changes in innovation outcomes. Broadly, the findings suggest that responses to disruptions are often relational in nature. Moreover, across the three main chapters, the consequences of disruption differ depending on the type of disruption and the context in which they occur. In Chapter 2, AI-enabled monitoring reduces uncertainty about exercise performance and reinforces accountability in service co-production. In Chapter 3, a successful drug launch is likely associated with portfolio reprioritization that can destabilize alliance relationships. In Chapter 4, project discontinuations are associated with new collaborations and movement toward new technological directions, although with lower innovation output.

In terms of practical implications, effective AI deployment in service settings should account for how the technology affects all service participants and how they interact with one another. In physiotherapy co-production, the intervention appeared especially effective because it strengthened the patient's sense of accountability to the physiotherapist over time. Moreover, AI can be effective when it complements rather than replaces human interaction as the technology improved co-production by extending provider oversight and reinforcing the patient–physiotherapist relationship. For managers, the broader implication is that AI-enabled service design should consider human–AI complementarity. As well, managers should not assume that successful product launches for a focal firm is a positive signal for its alliance partners. Instead, launches can increase the risk of alliance instability by shifting away from other existing collaborations. For biotech firms that depend on larger pharma partners, this highlights the importance of carefully selecting partners and assessing each firm's relative dependence on the alliance.

On the other hand, at the inventor level, project failures can reduce innovation activities but also redirect it toward new collaborations and more novel technological domains. For R&D managers, this implies that promoting new collaborations after a discontinuation can help inventors reorient their innovation efforts toward new technological domains. Simultaneously, this reorientation associates with lower patent output alongside greater novelty. Managers should thus be cautious about evaluating post-discontinuation performance using only immediate productivity metrics. Across all three main

chapters, a broader managerial implication is that the effects of disruption should be evaluated over long time horizons since consequences often emerge gradually instead of at the moment of the focal event. Hence, early post-disruption outcomes may provide an incomplete picture of long-term effects. Effective management therefore requires patience to fully observe how consequences evolve over time.

5.2. Limitations and Directions for Future Research

Several limitations must be acknowledged when interpreting the findings of the dissertation, which also evokes avenues for future research. First, the empirical settings are concentrated in the healthcare service contexts and in biopharmaceutical R&D. Although these settings are well suited for studying innovation disruptions because they involve high uncertainty, interdependences, and observable outcomes, the healthcare focus limits the external validity of the findings. The patterns may differ in industries with different innovation cycles, organizational structures, or regulatory environments. For instance, regulated industries with high development costs like resource extraction and aerospace, may show similar patterns, whereas software development, with the ethos "move fast and break things", may behave entirely differently because disruptions are generally expected. Therefore, future studies could perform comparative research to contrast the impact of innovation disruptions in other industries. Multi-industry datasets and multi-site studies would make it possible to assess whether the patterns identified here are specific to healthcare or reflect more general patterns of R&D management.

Second, the dissertation focuses on a few specific types of disruptions and therefore does not capture the full range of shocks that can affect R&D operations. Each study isolates a singular disruption and its impact on a specific context. Future research could examine additional types of disruption—for instance regulatory, organizational, and technological shocks—or study multiple disruptions together, to determine whether they generate similar spillovers or other distinct consequences. Future research could also examine the boundary conditions of disruption spillovers by identifying when they are more or less likely to arise, such as for different types of actors (e.g., users, firms, or inventors) or at different stages of the innovation process.

Third, the analyses rely on observable proxies to capture complex underlying phenomena. Measures like patent output, patent novelty, collaboration through co-inventorship, and exercise participation metrics provide tractable ways to study innovation with large data samples. However, they do not fully capture the underlying mechanisms that explain these phenomena. For instance, patent-based measures represent formal innovation outcomes, but they overlook informal collaborations that do not lead to co-inventorship, idea generation that fails to materialize, or successful innovation that is kept as a trade secret. Likewise, exercise participation metrics capture how much a patient performs prescribed exercises, but do not reflect whether the intervention produce spillover effects to other activities of healthy living. Future studies could investigate the underlying mechanisms of disruption more directly. Work that incorporates surveys, interviews, and behavioral experiments would help clarify why disruptions lead to certain outcomes. For instance, such work could explain to what extent a shift from exploration to exploitation explains alliance termination after a drug launch, how much a physiotherapy AI intervention increases exercise participation specifically because patients feel more accountable to their physiotherapist, and whether project discontinuation does create negative attribution toward inventors although that effect is overshadowed by problemistic search behavior.

Fourth, the dissertation adopts a multi-level perspective by examining disruption at the user, firm, and inventor levels, but these analyses are conducted separately. While this structure allows for close attention to level-specific outcomes, it does not permit direct observation of cross-level consequences within a single empirical setting. For instance, in the physiotherapy context, the dissertation does not examine whether higher exercise participation by patients leads to lower demand for physiotherapy visits or leads to higher customer retention. It also does not trace whether a drug launch by a pharma firm affects the outputs of inventors at a biotech partner whose alliance is terminated after the launch. Similarly, we do not know to what extent changes in inventor outputs after project terminations affect staffing turnover within the firms involved. Future research could develop designs that model multiple levels of analysis simultaneously to contribute to an understanding of how user-level, firm-level, and inventor-level spillovers interact. Such work could assess whether the consequences of disruption at

one level also leads to consequences at other levels over time. This would provide a more integrated view of how disruptions affect R&D operations.

Last, although the empirical designs use methods intended to strengthen inference, the possibility of endogeneity remains. The three studies rely on longitudinal analyses, which helps strengthen the evidence by establishing temporal ordering between the disruption event and the outcomes. However, since none of the studies rely on randomized control trials or quasi-experimental treatment, stronger causal claims would require additional research by using more exogenous treatments. Longer observation windows would also be valuable because some consequences of disruptions tend to lag the focal event. Such work would help determine whether the consequences observed produce lasting impacts or whether short-term effects eventually dissipate.

Notwithstanding the abovementioned future research agenda, the three main chapters collectively support that innovation disruptions generate consequences that extend beyond the focal event and affect other actors in the R&D system. Across the user, firm, and inventor levels, disruptions reshape behavior, relationships, and outcomes even for a period after the event. The shared pattern across the dissertation is therefore that disruption produces indirect effects, even though these effects differ by context and level of analysis.

References

Ahuja, G. (2000). The duality of collaboration: Inducements and opportunities in the formation of interfirm linkages. *Strategic Management Journal*, *21*(3), 317–343. https://doi.org/10.1002/(SICI)1097-0266(200003)21:3%3C317::AID-SMJ90%3E3.0.CO;2-B

Ahuja, V., Alvarez, C. A., & Staats, B. R. (2020). Maintaining Continuity in Service: An Empirical Examination of Primary Care Physicians. *Manufacturing & Service Operations Management*, *22*(5), 1088–1106. https://doi.org/10.1287/msom.2019.0808

Akan, M., Ata, B., & Lariviere, M. A. (2011). Asymmetric Information and Economies of Scale in Service Contracting. *Manufacturing & Service Operations Management*, *13*(1), 58–72. https://doi.org/10.1287/msom.1100.0303

Allison, P. (2014). Cox Regression. In *Event History and Survival Analysis* (pp. 33–52). SAGE Publications, Inc. https://doi.org/10.4135/9781452270029

Ambulkar, S., Ramaswami, S., Blackhurst, J., & Rungtusanatham, J. M. (2022). Supply chain disruption risk: An unintended consequence of product innovation. *International Journal of Production Research*, *60*(24), 7194–7213. https://doi.org/10.1080/00207543.2022.2027038

Andersen, P. K., & Gill, R. D. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, *10*(4), 1100–1120. https://doi.org/10.1214/aos/1176345976

Andritsos, D. A., & Tang, C. S. (2018). Incentive Programs for Reducing Readmissions when Patient Care is Co-Produced. *Production and Operations Management*, *27*(6), 999–1020. https://doi.org/10.1111/poms.12847

Angelopoulos, S., Bendoly, E., Fransoo, J., Hoberg, K., Ou, C., & Tenhiälä, A. (2023). Digital transformation in operations management: Fundamental change through agency reversal. *Journal of Operations Management*, *69*(6), 876–889. https://doi.org/10.1002/joom.1271

Anzanello, M. J., & Fogliatto, F. S. (2011). Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics*, *41*(5), 573–583. https://doi.org/10.1016/j.ergon.2011.05.001

Argent, R., Daly, A., & Caulfield, B. (2018). Patient Involvement With Home-Based Exercise Programs: Can Connected Health Interventions Influence Adherence? *JMIR mHealth and uHealth*, *6*(3), e8518. https://doi.org/10.2196/mhealth.8518

Asgari, N., Lévesque, M., Soh, P.-H., & Subramanian, A. M. (2024). It's Time to Break Up: Dynamics Surrounding Young-Established Firm Alliance Duration. *Production and Operations Management*, 10591478241279551. https://doi.org/10.1177/10591478241279551

Asgari, N., Tandon, V., Singh, K., & Mitchell, W. (2018). Creating and taming discord: How firms manage embedded competition in alliance portfolios to limit alliance termination. *Strategic Management Journal*, *39*(12), 3273–3299. https://doi.org/10.1002/smj.2784

Bagno, R. B., Salerno, M. S., & da Silva, D. O. (2017). Models with graphical representation for innovation management: A literature review. *R&D Management*, *47*(4), 637–653. https://doi.org/10.1111/radm.12254

Bakker, R. M. (2016). Stepping in and stepping out: Strategic alliance partner reconfiguration and the unplanned termination of complex projects. *Strategic Management Journal*, *37*(9), 1919–1941. https://doi.org/10.1002/smj.2429

Bakker, R. M., & Shepherd, D. A. (2017). Pull the Plug or Take the Plunge: Multiple Opportunities and the Speed of Venturing Decisions in the Australian Mining Industry. *Academy of Management Journal*, *60*(1), 130–155. https://doi.org/10.5465/amj.2013.1165

Banner, M., Alosert, H., Spencer, C., Cheeks, M., Farid, S. S., Thomas, M., & Goldrick, S. (2021). A decade in review: Use of data analytics within the biopharmaceutical sector. *Current Opinion in Chemical Engineering*, *34*, 100758. https://doi.org/10.1016/j.coche.2021.100758

Bassett, S. (2003). The assessment of patient adherence to physiotherapy rehabilitation. *New Zealand Journal of Physiotherapy*, *31*(2), 60–67.

Baum, J. A. C., Calabrese, T., & Silverman, B. S. (2000). Don't go it alone: Alliance network composition and startups' performance in Canadian biotechnology. *Strategic Management Journal*, *21*(3), 267–294. https://doi.org/10.1002/(SICI)1097-0266(200003)21:3%253C267::AID-SMJ89%253E3.0.CO;2-8

Bavafa, H., Savin, S., & Terwiesch, C. (2021). Customizing Primary Care Delivery Using E-Visits. *Production and Operations Management*, *30*(11), 4306–4327. https://doi.org/10.1111/poms.13528

Baxter, D., Trott, P., & Ellwood, P. (2023). Reconceptualising innovation failure. *Research Policy*, *52*(7), 104811. https://doi.org/10.1016/j.respol.2023.104811

Bellemare, M. F., & Wichman, C. J. (2020). Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin of Economics and Statistics*, *82*(1), 50–61. https://doi.org/10.1111/obes.12325

Benner, M. J., & Tushman, M. L. (2003). Exploitation, Exploration, and Process Management: The Productivity Dilemma Revisited. *The Academy of Management Review*, *28*(2), 238–256. https://doi.org/10.2307/30040711

Bercovitz, J., & Feldman, M. (2011). The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. *Research Policy, Special Section on Heterogeneity and University-Industry Relations*, *40*(1), 81–93. https://doi.org/10.1016/j.respol.2010.09.008

Bessant, J., Lamming, R., Noke, H., & Phillips, W. (2005). Managing innovation beyond the steady state. *Technovation*, *25*(12), 1366–1376. https://doi.org/10.1016/j.technovation.2005.04.007

Bianchi, M., & Brière, M. (2024). Human-Robot Interactions in Investment Decisions. *Management Science*, mnsc.2022.03886. https://doi.org/10.1287/mnsc.2022.03886

Bierly, P. E., & Coombs, J. E. (2004). Equity alliances, stages of product development, and alliance instability. *Journal of Engineering and Technology Management*, *21*(3), 191–214. https://doi.org/10.1016/j.jengtecman.2004.05.001

Bleich, J., Kapelner, A., George, E. I., & Jensen, S. T. (2014). Variable selection for BART: An application to gene regulation. *The Annals of Applied Statistics*, *8*(3). https://doi.org/10.1214/14-AOAS755

Blut, M., & Wang, C. (2020). Technology readiness: A meta-analysis of conceptualizations of the construct and its impact on technology usage. *Journal of the Academy of Marketing Science*, *48*(4), 649–669. https://doi.org/10.1007/s11747-019-00680-8

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, *49*(4), 632–658. https://doi.org/10.1007/s11747-020-00762-y

Boeing, P., Mueller, E., & Sandner, P. (2016). China's R&D explosion—Analyzing productivity effects across ownership types and over time. *Research Policy*, *45*(1), 159–176. https://doi.org/10.1016/j.respol.2015.07.008

Bonnechre, B., Jansen, B., Omelina, L., & Jan, S. (2016). Do Patients Perform Their Exercises at Home and why (not)? A Survey on Patients' Habits during Rehabilitation Exercises. *Ulutas Medical Journal*, *2*(1), 41. https://doi.org/10.5455/umj.20160210060312

Boutilier, J. J., Jónasson, J. O., & Yoeli, E. (2022). Improving Tuberculosis Treatment Adherence Support: The Case for Targeted Behavioral Interventions. *Manufacturing & Service Operations Management*, *24*(6), 2925–2943. https://doi.org/10.1287/msom.2021.1046

Boyer, P., Burns, D., Razmjou, H., Renteria, C., Sheth, U., Richards, R., & Whyne, C. (2025). Quantifying the Relationship Between At-Home Shoulder Physiotherapy Participation and Outcome: What can a Watch Tell Us? *Journal of the American Academy of Orthopaedic Surgeons*. https://doi.org/10.5435/JAAOS-D-24-00499

Boyer, P., Burns, D., & Whyne, C. (2021). Out-of-Distribution Detection of Human Activity Recognition with Smartwatch Inertial Sensors. *Sensors*, *21*(5), Article 5. https://doi.org/10.3390/s21051669

Boyer, P., Burns, D., & Whyne, C. (2023). Evaluation of at-home physiotherapy. *Bone & Joint Research*, *12*(3), 165–177. https://doi.org/10.1302/2046-3758.123.BJR-2022-0126.R1

Boyer, P., Donia, J., Whyne, C., Burns, D., & Shaw, J. (2022). Regulatory regimes and procedural values for health-related motion data in the United States and Canada. *Health Policy and Technology*, *11*(3), 100648. https://doi.org/10.1016/j.hlpt.2022.100648

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at Work*. *The Quarterly Journal of Economics*, *140*(2), 889–942. https://doi.org/10.1093/qje/qjae044

Burns, D., Boyer, P., Arrowsmith, C., & Whyne, C. (2022). Personalized Activity Recognition with Deep Triplet Embeddings. *Sensors*, *22*(14), Article 14. https://doi.org/10.3390/s22145222

Burns, D., Boyer, P., Razmjou, H., Richards, R., & Whyne, C. (2021). Adherence Patterns and Dose Response of Physiotherapy for Rotator Cuff Pathology: Longitudinal Cohort Study. *JMIR Rehabilitation and Assistive Technologies*, *8*(1), e21374. https://doi.org/10.2196/21374

Burns, D., Leung, N., Hardisty, M., Whyne, C. M., Henry, P., & McLachlin, S. (2018). Shoulder physiotherapy exercise recognition: Machine learning the inertial signals from a smartwatch. *Physiological Measurement*, *39*(7), 075007. https://doi.org/10.1088/1361-6579/aacfd9

Burns, D., Razmjou, H., Shaw, J., Richards, R., McLachlin, S., Hardisty, M., Henry, P., & Whyne, C. (2020). Adherence Tracking With Smart Watches for Shoulder Physiotherapy in Rotator Cuff Pathology: Protocol for a Longitudinal Cohort Study. *JMIR Research Protocols*, *9*(7), e17841. https://doi.org/10.2196/17841

Carnabuci, G., & Operti, E. (2013). Where do firms' recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination. *Strategic Management Journal*, *34*(13), 1591–1613. https://doi.org/10.1002/smj.2084

Carver, C. S., & Scheier, M. F. (1981). *Attention and Self-Regulation*. Springer. https://doi.org/10.1007/978-1-4612-5887-2

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin*, *92*(1), 111–135. https://doi.org/10.1037/0033-2909.92.1.111

Caton, S., & Haas, C. (2024). Fairness in Machine Learning: A Survey. *ACM Computing Surveys*, *56*(7), 1–38. https://doi.org/10.1145/3616865

Cazavan-Jeny, A., Jeanjean, T., & Joos, P. (2011). Accounting choice and future performance: The case of R&D accounting in France. *Journal of Accounting and Public Policy*, *30*(2), 145–165. https://doi.org/10.1016/j.jaccpubpol.2010.09.016

Chaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of Explainable AI Techniques in Healthcare. *Sensors*, *23*(2), Article 2. https://doi.org/10.3390/s23020634

Chan, T. C. Y., Mahmood, R., O'Connor, D. L., Stone, D., Unger, S., Wong, R. K., & Zhu, I. Y. (2023). Got (Optimal) Milk? Pooling Donations in Human Milk Banks with Machine Learning and Optimization. *Manufacturing & Service Operations Management*. https://doi.org/10.1287/msom.2022.0455

Chen, Z., Chen, J., Feng, Y., Liu, S., Zhang, T., Zhang, K., & Xiao, W. (2022). Imbalance fault diagnosis under long-tailed distribution: Challenges, solutions and prospects. *Knowledge-Based Systems*, *258*, 110008. https://doi.org/10.1016/j.knosys.2022.110008

Chester, R., Jerosch-Herold, C., Lewis, J., & Shepstone, L. (2018). Psychological factors are associated with the outcome of physiotherapy for people with shoulder pain: A multicentre longitudinal cohort study. *British Journal of Sports Medicine*, *52*(4), 269–275. https://doi.org/10.1136/bjsports-2016-096084

Chester, R., Shepstone, L., Daniell, H., Sweeting, D., Lewis, J., & Jerosch-Herold, C. (2013). Predicting response to physiotherapy treatment for musculoskeletal shoulder pain: A systematic review. *BMC Musculoskeletal Disorders*, *14*, 203. https://doi.org/10.1186/1471-2474-14-203

Chipman, H., George, E., & McCulloch, R. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, *4*(1), 266–298. https://doi.org/10.1214/09-AOAS285

Choi, T., Kumar, S., Yue, X., & Chan, H. (2022). Disruptive Technologies and Operations Management in the Industry 4.0 Era and Beyond. *Production and Operations Management*, *31*(1), 9–31. https://doi.org/10.1111/poms.13622

Chou, Y.-C., Chuang, H. H.-C., Chou, P., & Oliva, R. (2023). Supervised machine learning for theory building and testing: Opportunities in operations management. *Journal of Operations Management*, *1*(33). https://doi.org/10.1002/joom.1228

Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. *Strategic Management Journal*, *42*(1), 30–57. https://doi.org/10.1002/smj.3215

Cohen, M., Dai, T., & Li, B. (2024). Call for Papers: *Service Science* Special Issue on the Impact of AI on Service Design and Delivery. *Service Science*, serv.2024.cfp.v16.n2. https://doi.org/10.1287/serv.2024.cfp.v16.n2

Cooper, R. G., Edgett, S. J., & Kleinschmidt, E. J. (2004). New Product Portfolio Management: Practices and Performance. *Journal of Product Innovation Management*, *16*(4), 333–351. https://doi.org/10.1111/1540-5885.1640333

Croson, R., Schultz, K., Siemsen, E., & Yeo, M. L. (2013). Behavioral operations: The state of the field. *Journal of Operations Management*, *31*(1–2), 1–5. https://doi.org/10.1016/j.jom.2012.12.001

Cui, A. S. (2013). Portfolio Dynamics and Alliance Termination: The Contingent Role of Resource Dissimilarity. *Journal of Marketing*, *77*(3), 15–32. https://doi.org/10.1509/jm.11.0164

Cui, A. S., Calantone, R. J., & Griffith, D. A. (2011). Strategic Change and Termination of Interfirm Partnerships. *Strategic Management Journal*, *32*(4), 402–423.

Cui, A. S., & O'Connor, G. (2012). Alliance Portfolio Resource Diversity and Firm Innovation. *Journal of Marketing*, *76*(4), 24–43. https://doi.org/10.1509/jm.11.0130

Cui, V., Yang, H., & Vertinsky, I. (2018). Attacking your partners: Strategic alliances and competition between partners in product markets. *Strategic Management Journal*, *39*(12), 3116–3139. https://doi.org/10.1002/smj.2746

Cyert, R. M., & March, J. G. (1963). A Behavioral Theory of the Firm. *Prentice Hall/Pearson Education*.

Dahlander, L., O'Mahony, S., & Gann, D. M. (2016). One foot in, one foot out: How does individuals' external search breadth affect innovation outcomes? *Strategic Management Journal*, *37*(2), 280–302. https://doi.org/10.1002/smj.2342

Dai, T., & Tayur, S. (2022). Designing AI-augmented healthcare delivery systems for physician buy-in and patient acceptance. *Production and Operations Management*, *31*(12), 4443–4451. https://doi.org/10.1111/poms.13850

Dan, S. M., & Zondag, M. M. (2016). Drivers of alliance terminations: An empirical examination of the bio-pharmaceutical industry. *Industrial Marketing Management*, *54*, 107–115. https://doi.org/10.1016/j.indmarman.2015.06.006

D'Arville, C. (2004). Partnering: Key to Early-Stage Biotech Survival, But at What Cost? *Biotechnology Healthcare*, *1*(6), 26–34.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319. https://doi.org/10.2307/249008

de Groot, V., Beckerman, H., Lankhorst, G. J., & Bouter, L. M. (2003). How to measure comorbidity. A critical review of available methods. *Journal of Clinical Epidemiology*, *56*(3), 221–229. https://doi.org/10.1016/s0895-4356(02)00585-1

de Treville, S., Browning, T. R., & Oliva, R. (2023). Empirically grounding analytics (EGA) research in the Journal of Operations Management. *Journal of Operations Management*, *69*(2), 337–348. https://doi.org/10.1002/joom.1242

Diestre, L., & Rajagopalan, N. (2012). Are all 'sharks' dangerous? New biotechnology ventures and partner selection in R&D alliances. *Strategic Management Journal*, *33*(10), 1115–1134. https://doi.org/10.1002/smj.1978

DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, *47*, 20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012

Donia, J., Oyefeso, L., Embuldeniya, G., Whyne, C., Burns, D., Boyer, P., Razmjou, H., & Shaw, J. A. (2024). Lifecycles, pipelines, and value chains: Toward a focus on events in responsible artificial intelligence for health. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00594-4

Donohue, K., Özer, Ö., & Zheng, Y. (2020). Behavioral Operations: Past, Present, and Future. *Manufacturing & Service Operations Management*, *22*(1), 191–202. https://doi.org/10.1287/msom.2019.0828

Duysters, G., & De Man, A. (2003). Transitory alliances: An instrument for surviving turbulent industries? *R&D Management*, *33*(1), 49–58. https://doi.org/10.1111/1467-9310.00281

Dziallas, M., & Blind, K. (2019). Innovation indicators throughout the innovation process: An extensive literature analysis. *Technovation*, *80–81*, 3–29. https://doi.org/10.1016/j.technovation.2018.05.005

Eggers, J. P., & Suh, J.-H. (2019). Experience and Behavior: How Negative Feedback in New Versus Experienced Domains Affects Firm Action and Subsequent Performance. *Academy of Management Journal*, *62*(2), 309–334. https://doi.org/10.5465/amj.2017.0046

Eslami, H., Eshghi, K., & Sadeh, F. (2023). Premature R&D alliance termination and shareholder returns: Evidence from the biopharmaceutical industry. *Journal of Product Innovation Management*, *40*(3), 340–357. https://doi.org/10.1111/jpim.12658

Evens, R. P. (2016). Pharma Success in Product Development—Does Biotechnology Change the Paradigm in Product Development and Attrition. *The AAPS Journal*, *18*(1), 281–285. https://doi.org/10.1208/s12248-015-9833-6

Faber, M., Andersen, M. H., Sevel, C., Thorborg, K., Bandholm, T., & Rathleff, M. (2015). The majority are not performing home-exercises correctly two weeks after their initial instruction—An assessor-blinded study. *PeerJ*, *3*, e1102. https://doi.org/10.7717/peerj.1102

Faems, D., De Visser, M., Andries, P., & Van Looy, B. (2010). Technology Alliance Portfolios and Financial Performance: Value-Enhancing and Cost-Increasing Effects of Open Innovation*. *Journal of Product Innovation Management*, *27*(6), 785–796. https://doi.org/10.1111/j.1540-5885.2010.00752.x

Farhan, W., Refae, G. A. E., & Razmak, J. (2025). The role of AI applications in production and operations management: Complementing or replacing human labour. *International Journal of Economics and Business Research*, *29*(13), 10071261. https://doi.org/10.1504/IJEBR.2025.10071261

FDA Drug approval—Application Review Times Largely Reflect Agency Goals (GAO-20-244; pp. 1–49). (2020). United States Government Accountability Office. https://www.gao.gov/assets/gao-20-244.pdf

Fisher, J. A., Cottingham, M. D., & Kalbaugh, C. A. (2014). Peering into the Pharmaceutical "Pipeline": Investigational Drugs, Clinical Trials, and Industry Priorities. *Social Science & Medicine (1982)*, *131*, 322. https://doi.org/10.1016/j.socscimed.2014.08.023

Fleming, L. (2001). Recombinant Uncertainty in Technological Search. *Management Science*, *47*(1), 117–132. https://doi.org/10.1287/mnsc.47.1.117.10671

Forkan, R., Pumper, B., Smyth, N., Wirkkala, H., Ciol, M. A., & Shumway-Cook, A. (2006). Exercise Adherence Following Physical Therapy Intervention in Older Adults With Impaired Balance. *Physical Therapy*, *86*(3), 401–410. https://doi.org/10.1093/ptj/86.3.401

Frei, F. X. (2006). Breaking the Trade-Off Between Efficiency and Service. *Harvard Business Review*, *84*(11), 92.

Froehle, C. M., & Roth, A. V. (2004). New measurement scales for evaluating perceptions of the technology-mediated customer service experience. *Journal of Operations Management*, *22*(1), 1–21. https://doi.org/10.1016/j.jom.2003.12.004

From laboratory to patient—The journey of a medicine assessed by EMA (pp. 1–26). (2024). European Medicines Agency. https://www.ema.europa.eu/en/documents/other/laboratory-patient-journey-centrally-authorised-medicine_en.pdf

Fu, N., Flood, P. C., Rousseau, D. M., Morris, T., & Johnstone, M. (2023). When are clients helpful? Capitalising on client involvement in professional service delivery. *PLOS ONE*, *18*(2), e0280738. https://doi.org/10.1371/journal.pone.0280738

Gan, D., & Korsgaard, S. (2022). When courtships fail: The antecedents of failure in strategic alliances between startups and incumbents. *The International Journal of Entrepreneurship and Innovation*, 14657503221128015. https://doi.org/10.1177/14657503221128015

Ghanem, M., Ghaith, A. K., El-Hajj, V. G., Bhandarkar, A., de Giorgio, A., Elmi-Terander, A., & Bydon, M. (2023). Limitations in Evaluating Machine Learning Models for Imbalanced Binary Outcome Classification in Spine Surgery: A Systematic Review. *Brain Sciences*, *13*(12), 1723. https://doi.org/10.3390/brainsci13121723

Ghosh, A., & Klueter, T. (2022). The Role of Frictions due to Top Management in Alliance Termination Decisions: Insights from Established Bio-Pharmaceutical Firms. *Journal of Management Studies*, *59*(5), 1315–1353. https://doi.org/10.1111/joms.12780

Ghosh, S., & Wu, A. (2023). Iterative Coordination and Innovation: Prioritizing Value over Novelty. *Organization Science*, *34*(6), 2182–2206. https://doi.org/10.1287/orsc.2021.1499

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining Explanations: An Overview of Interpretability of Machine Learning. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 80–89. https://doi.org/10.1109/DSAA.2018.00018

Gnewuch, U., Morana, S., Hinz, O., Kellner, R., & Maedche, A. (2024). More Than a Bot? The Impact of Disclosing Human Involvement on Customer Interactions with Hybrid Service Agents. *Information Systems Research*, *35*(3), 936–955. https://doi.org/10.1287/isre.2022.0152

Gomez, C., Smith, B.-L., Zayas, A., Unberath, M., & Canares, T. (2024). Explainable AI decision support improves accuracy during telehealth strep throat screening. *Communications Medicine*, *4*(1), 149. https://doi.org/10.1038/s43856-024-00568-x

Gopalakrishnan, S., Scillitoe, J. L., & Santoro, M. D. (2008). Tapping Deep Pockets: The Role of Resources and Social Capital on Financial Capital Acquisition by Biotechnology Firms in Biotech-Pharma Alliances. *Journal of Management Studies*, *45*(8), 1354–1376. https://doi.org/10.1111/j.1467-6486.2008.00777.x

GoPro. (2016, May 23). *GoPro and Red Bull Form Exclusive Global Partnership*. https://gopro.com/en/us/news/gopro-and-red-bull-form-exclusive-global-partnership

GoPro. (2020, December 17). *Master the Modes: How to Live Stream*. https://gopro.com/en/us/news/how-to-live-stream-on-gopro

Greve, H. R. (2003). A behavioral theory of R&D expenditures and innovations: Evidence from shipbuilding. *Academy of Management Journal*, *46*(6), 685–702. https://doi.org/10.2307/30040661

Greve, H. R., Baum, J. a. C., Mitsuhashi, H., & Rowley, T. J. (2010). Built to Last but Falling Apart: Cohesion, Friction, and Withdrawal from Interfirm Alliances. *Academy of Management Journal*, *53*(2), 302–322. https://doi.org/10.5465/AMJ.2010.49388955

Griliches, Z. (1984). *R&D, Patents, and Productivity*. University of Chicago Press. https://www.nber.org/books-and-chapters/rd-patents-and-productivity

Gu, Q., Zhu, L., & Cai, Z. (2009). Evaluation Measures of the Classification Performance of Imbalanced Data Sets. In Z. Cai, Z. Li, Z. Kang, & Y. Liu (Eds.), *Computational Intelligence*

and Intelligent Systems (pp. 461–471). Springer. https://doi.org/10.1007/978-3-642-04962-0_53

Hampl, N. (2020). Drivers of and barriers to partner switch in interfirm alliances: A conceptual model. *Journal of Business Economics*, *90*(4), 563–589. https://doi.org/10.1007/s11573-019-00965-y

Hansen, M. T., & Birkinshaw, J. (2007). The Innovation Value Chain. *Harvard Business Review*, *85*(6), 121–130. (24998248).

Hatfield, R. C., Hoang, K., Ricci, M. A., & Thomas, E. (2025). How Auditors' Client Service Affects Managers' Cooperation with Auditor Evidence Requests. *AUDITING: A Journal of Practice & Theory*, *44*(3), 111–132. https://doi.org/10.2308/AJPT-2023-110

Havenaar, M., & Hiscocks, P. (2012). Strategic alliances and market risk. *Drug Discovery Today*, *17*(15), 824–827. https://doi.org/10.1016/j.drudis.2012.03.008

He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. IEEE Transactions on Knowledge and Data Engineering. https://doi.org/10.1109/TKDE.2008.239

Heider, F. (1958). *The psychology of interpersonal relations*. Lawrence Erlbaum Associates.

Henry, K. D., Rosemond, C., & Eckert, L. B. (1999). Effect of Number of Home Exercises on Compliance and Performance in Adults Over 65 Years of Age. *Physical Therapy*, *79*(3), 270–277. https://doi.org/10.1093/ptj/79.3.270

Hill, J. L. (2011). Bayesian Nonparametric Modeling for Causal Inference. *Journal of Computational and Graphical Statistics*, *20*(1), 217–240. https://doi.org/10.1198/jcgs.2010.08162

Hill, R., Yin, Y., Stein, C., Wang, X., Wang, D., & Jones, B. F. (2025). The pivot penalty in research. *Nature*, *642*(8069), 999–1006. https://doi.org/10.1038/s41586-025-09048-1

Hoffmann, W. H. (2007). Strategies for managing a portfolio of alliances. *Strategic Management Journal*, *28*(8), 827–856. https://doi.org/10.1002/smj.607

Hohberger, J., Kruger, H., & Almeida, P. (2020). Does separation hurt? The impact of premature termination of R&D alliances on knowledge acquisition and innovation. *Research Policy*, *49*(6), 103944. https://doi.org/10.1016/j.respol.2020.103944

Hora, M., & Dutta, D. K. (2013). Entrepreneurial Firms and Downstream Alliance Partnerships: Impact of Portfolio Depth and Scope on Technology Innovation and Commercialization Success. *Production and Operations Management*, *22*(6), 1389–1400. https://doi.org/10.1111/j.1937-5956.2012.01410.x

Hu, T., Jain, A., & Delios, A. (2021). Centrality Asymmetry and Partner Complementarity as Influences on Alliance Dissolution. *British Journal of Management*, *32*(1), 59–79. https://doi.org/10.1111/1467-8551.12440

Jack, K., McLean, S. M., Moffett, J. K., & Gardiner, E. (2010). Barriers to treatment adherence in physiotherapy outpatient clinics: A systematic review. *Manual Therapy*, *15*(3), 220–228. https://doi.org/10.1016/j.math.2009.12.004

Jamieson, K. H., McNutt, M., Kiermer, V., & Sever, R. (2019). Signaling the trustworthiness of science. *Proceedings of the National Academy of Sciences*, *116*(39), 19231–19236. https://doi.org/10.1073/pnas.1913039116

Jin, J., Sklar, G. E., Min Sen Oh, V., & Chuen Li, S. (2008). Factors affecting therapeutic compliance: A review from the patient's perspective. *Therapeutics and Clinical Risk Management*, *4*(1), 269–286. https://doi.org/10.2147/tcrm.s1458

Jui, T. D., & Rivas, P. (2024). Fairness issues, current approaches, and challenges in machine learning models. *International Journal of Machine Learning and Cybernetics*, *15*(8), 3095–3125. https://doi.org/10.1007/s13042-023-02083-2

Kalaignanam, K., Shankar, V., & Varadarajan, R. (2007). Asymmetric New Product Development Alliances: Win-Win or Win-Lose Partnerships? *Management Science*, *53*(3), 357–374.

Katila, R., Rosenberger, J. D., & Eisenhardt, K. M. (2008). Swimming with Sharks: Technology Ventures, Defense Mechanisms and Corporate Relationships. *Administrative Science Quarterly*, *53*(2), 295–332. https://doi.org/10.2189/asqu.53.2.295

Keil, T., Syrigos, E., Kostopoulos, K. C., D. Meissner, F., & G. Audia, P. (2024). (In)Consistent Performance Feedback and the Locus of Search. *Journal of Management*, *50*(7), 2927–2954. https://doi.org/10.1177/01492063231185519

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, *28*(2), 107–128. https://doi.org/10.1037/h0034225

Kim, J.-Y., & Miner, A. S. (2007). Vicarious Learning from the Failures and Near-Failures of Others: Evidence from the U.S. Commercial Banking Industry. *Academy of Management Journal*, *50*(3), 687–714.

Kim, M. (2023). Connecting artificial intelligence to value creation in services: Mechanism and implications. *Service Business*, *17*(4), 851–878. https://doi.org/10.1007/s11628-023-00547-7

Kılıç, H., & Güneş, E. D. (2024). Patient adherence in healthcare operations: A narrative review. *Socio-Economic Planning Sciences*, *91*, 101795. https://doi.org/10.1016/j.seps.2023.101795

Klassen, R. D., & Menor, L. J. (2007). The process management triangle: An empirical investigation of process trade-offs. *Journal of Operations Management, Special Issue on Innovative Data Sources for Emperically Building and Validating Theories in Operations Management*, *25*(5), 1015–1034. https://doi.org/10.1016/j.jom.2006.10.004

Klossek, A., Meyer, K. E., & Nippa, M. (2015). Why Do Strategic Alliances Persist? A Behavioral Decision Model. *Managerial and Decision Economics*, *36*(7), 470–486.

Koehler, M., & Sauermann, H. (2025). *Automation or Augmentation? Task Characteristics, Human-Centered Goals, and the Deployment of AI* (SSRN Scholarly Paper No. 4910282). Social Science Research Network. https://papers.ssrn.com/abstract=4910282

Kok, H., Faems, D., & de Faria, P. (2019). Dusting Off the Knowledge Shelves: Recombinant Lag and the Technological Value of Inventions. *Journal of Management*, *45*(7), 2807–2836. https://doi.org/10.1177/0149206318765926

Koval, M. (2021). Whether and when do alliance terminations pay off? *Industrial Marketing Management*, *98*, 149–160. https://doi.org/10.1016/j.indmarman.2021.08.008

Krieger, J. L. (2021). Trials and Terminations: Learning from Competitors' R&D Failures. *Management Science*, *67*(9), 5525–5548. https://doi.org/10.1287/mnsc.2020.3775

Krieger, J. L., Li, X., & Thakor, R. T. (2022). Find and Replace: R&D Investment Following the Erosion of Existing Products. *Management Science*, *68*(9), 6552–6571. https://doi.org/10.1287/mnsc.2021.4243

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*(2), 573–603. https://doi.org/10.1037/a0029146

Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., Wünderlich, N. V., & De Keyser, A. (2017). "Service Encounter 2.0": An investigation into the roles of technology, employees and customers. *Journal of Business Research*, *79*, 238–246. https://doi.org/10.1016/j.jbusres.2017.03.008

Larsson, R., & Bowen, D. E. (1989). Organization and Customer: Managing Design and Coordination of Services. *The Academy of Management Review*, *14*(2), 213–233. https://doi.org/10.2307/258417

Leevy, J. L., Khoshgoftaar, T. M., Bauder, R. A., & Seliya, N. (2018). A survey on addressing high-class imbalance in big data. *Journal of Big Data*, *5*(1), 42. https://doi.org/10.1186/s40537-018-0151-6

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255–275. https://doi.org/10.1037/0033-2909.125.2.255

Lévesque, M., Subramanian, A. M., & Vrande, V. V. D. (2022). The inseparable two: Impact of prior success and failure on new product development project discontinuation. *Journal of Operations Management*, joom.1214. https://doi.org/10.1002/joom.1214

Li, M., Li, Q., & Yao, Y. (2025). AI in operations management: Literature review and future research. *Reference Module in Social Sciences*. https://doi.org/10.1016/B978-0-443-28993-4.00101-3

Li, M. M., Reinhard, P., Peters, C., Oeste-Reiss, S., & Leimeister, J. M. (2024). A Value Co-Creation Perspective on Data Labeling in Hybrid Intelligence Systems: A Design Study. *Information Systems*, *120*, 102311. https://doi.org/10.1016/j.is.2023.102311

Light, D. W., & Lexchin, J. R. (2021). Pharmaceuticals as a market for "lemons": Theory and practice. *Social Science & Medicine*, *268*, 113368. https://doi.org/10.1016/j.socscimed.2020.113368

Love, J. H., Roper, S., & Vahter, P. (2023a). The dynamics of abandoned innovation activities: Learning from failure or learning to prevent failure? *Research Policy*, *52*(8), 104842. https://doi.org/10.1016/j.respol.2023.104842

Love, J. H., Roper, S., & Vahter, P. (2023b). The dynamics of abandoned innovation activities: Learning from failure or learning to prevent failure? *Research Policy*, *52*(8), 104842. https://doi.org/10.1016/j.respol.2023.104842

Lugovoi, I., Andritsos, D. A., & Senot, C. (2022). Manufacturing Process Innovation in the Pharmaceutical Industry. *Manufacturing & Service Operations Management*, *24*(3), 1760–1778. https://doi.org/10.1287/msom.2021.1035

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, 4768–4777.

Lunnan, R., & Haugland, S. A. (2008). Predicting and measuring alliance performance: A multidimensional analysis. *Strategic Management Journal*, *29*(5), 545–556. https://doi.org/10.1002/smj.660

MacCormack, A., & Mishra, A. (2015). Managing the Performance Trade-Offs from Partner Integration: Implications of Contract Choice in R&D Projects. *Production and Operations Management*, *24*(10), 1552–1569. https://doi.org/10.1111/poms.12374

Madhok, A., Keyhani, M., & Bossink, B. (2015). Understanding alliance evolution and termination: Adjustment costs and the economics of resource value. *Strategic Organization*, *13*(2), 91–116. https://doi.org/10.1177/1476127015580309

Madsen, P. M., & Desai, V. (2010). Failing to Learn? The Effects of Failure and Success on Organizational Learning in the Global Orbital Launch Vehicle Industry. *Academy of Management Journal*, *53*(3), 451–476. https://doi.org/10.5465/amj.2010.51467631

Malik, T. H., & Zhao, Y. (2013). Cultural distance and its implication for the duration of the international alliance in a high technology sector. *International Business Review*, *22*(4), 699–712. https://doi.org/10.1016/j.ibusrev.2012.11.001

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1,), 71–87.

Margolin, M., Reimer, M., & Schaupp, D. (2025). The Effects of Real-Time Feedback on Effort and Performance: Evidence from a Natural Quasi-Experiment. *Management Science*, *71*(4), 3558–3582. https://doi.org/10.1287/mnsc.2022.02084

Mata, J., & Portugal, P. (2015). The termination of international joint ventures: Closure and acquisition by domestic and foreign partners. *International Business Review*, *24*(4), 677–689.

McCarter, M. W., & Northcraft, G. B. (2007). Happy together?: Insights and implications of viewing managed supply chains as a social dilemma. *Journal of Operations Management*, *25*(2), 498–511. https://doi.org/10.1016/j.jom.2006.05.005

McCutchen, W. W., Swamidass, P. M., & Teng, B.-S. (2008). Strategic alliance termination and performance: The role of task complexity, nationality, and experience. *The Journal of High Technology Management Research*, *18*(2), 191–202. https://doi.org/10.1016/j.hitech.2007.12.001

McCutcheon, D., & Stuart, F. I. (2000). Issues in the choice of supplier alliance partners. *Journal of Operations Management*, *18*(3), 279–301. https://doi.org/10.1016/S0272-6963(99)00026-1

McLean, S. M., Burton, M., Bradley, L., & Littlewood, C. (2010). Interventions for enhancing adherence with physiotherapy: A systematic review. *Manual Therapy*, *15*(6), 514–521. https://doi.org/10.1016/j.math.2010.05.012

Merck KGaA, Darmstadt, Germany, and Pfizer Announce Investigational Immunotherapy Avelumab Receives FDA Fast Track Designation for Metastatic Merkel Cell Carcinoma | Pfizer. (2015). Pfizer. https://www.pfizer.com/news/press-release/press-release-detail/merck_kgaa_darmstadt_germany_and_pfizer_announce_investigational_immunotherapy_avelumab_receives_fda_fast_track_designation_for_metastatic_merkel_cell_carcinoma

Meschi, P.-X., & Norheim-Hansen, A. (2020). Partner-diversity effects on alliance termination in the early stage of green alliance formation: Empirical evidence from carbon-emission reduction projects in Latin America. *Business Strategy and the Environment*, *29*(1), 250–261. https://doi.org/10.1002/bse.2362

Meschi, P.-X., & Riccio, E. L. (2008). Country risk, national cultural differences between partners and survival of international joint ventures in Brazil. *International Business Review*, *17*(3), 250–266. https://doi.org/10.1016/j.ibusrev.2007.11.001

Min, J. (2017). Sensitivity of Alliance Termination to Prealliance Conditions: Expectation Effects of Alliance Partners. *Organization Studies*, *38*(7), 917–936. https://doi.org/10.1177/0170840616663244

Mishra, A., Chandrasekaran, A., & MacCormack, A. (2015). Collaboration in Multi-Partner R&D Projects: The Impact of Partnering Scale and Scope. *Journal of Operations Management*, *33–34*, 1.

Mitchell, P. H., Powell, L., Blumenthal, J., Norten, J., Ironson, G., Pitula, C. R., Froelicher, E. S., Czajkowski, S., Youngblood, M., Huber, M., & Berkman, L. F. (2003). A short social support measure for patients recovering from myocardial infarction: The ENRICHD Social Support Inventory. *Journal of Cardiopulmonary Rehabilitation*, *23*(6), 398–403. https://doi.org/10.1097/00008483-200311000-00001

Mohr, A., Wang, C., & Goerzen, A. (2016). The impact of partner diversity within multiparty international joint ventures. *International Business Review*, *25*(4), 883–894. https://doi.org/10.1016/j.ibusrev.2015.11.001

Mothilal, R. K., Sharma, A., & Tan, C. (2020). Explaining machine learning classifiers through diverse counterfactual explanations. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20*, 607–617. https://doi.org/10.1145/3351095.3372850

Nawrat, A. (2019, July 2). Pfizer acquires rare disease biotech Therachon for $340m. *Pharmaceutical Technology*. https://www.pharmaceutical-technology.com/news/pfizer-acquires-therachon-340m/

Neuberger, A., Oraiopoulos, N., & Drakeman, D. L. (2017). Lemons, or Squeezed for Resources? Information Symmetry and Asymmetric Resources in Biotechnology. *Frontiers in Pharmacology*, *8*. https://doi.org/10.3389/fphar.2017.00338

Ng, J. Y. Y., Ntoumanis, N., Thøgersen-Ntoumani, C., Deci, E. L., Ryan, R. M., Duda, J. L., & Williams, G. C. (2012). Self-Determination Theory Applied to Health Contexts: A Meta-Analysis. *Perspectives on Psychological Science*, *7*(4), 325–340. https://doi.org/10.1177/1745691612447309

Nguyen, T., & Elbanna, A. (2025). Understanding Human-AI Augmentation in the Workplace: A Review and a Future Research Agenda. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-025-10591-5

Nicholas, M. K., McGuire, B. E., & Asghari, A. (2015). A 2-item short form of the Pain Self-efficacy Questionnaire: Development and psychometric evaluation of PSEQ-2. *The Journal of Pain*, *16*(2), 153–163. https://doi.org/10.1016/j.jpain.2014.11.002

Palmarozzo, A., Toffel, M. W., & Ouellet, M. (2025). *Managing Remote Work Quality: Evidence from Auditing Management Systems Standards* (SSRN Scholarly Paper No. 4520770). Social Science Research Network. https://doi.org/10.2139/ssrn.4520770

Pammolli, F., Magazzini, L., & Riccaboni, M. (2011). The productivity crisis in pharmaceutical R&D. *Nature Reviews. Drug Discovery*, *10*(6), 428–438. https://doi.org/10.1038/nrd3405

Pammolli, F., Righetto, L., Abrignani, S., Pani, L., Pelicci, P. G., & Rabosio, E. (2020). The endless frontier? The recent increase of R&D productivity in pharmaceuticals. *Journal of Translational Medicine*, *18*(1), 162. https://doi.org/10.1186/s12967-020-02313-z

Pangarkar, N. (2003). Determinants of Alliance Duration in Uncertain Environments: The Case of the Biotechnology Sector. *Long Range Planning*, *36*(3), 269–284. https://doi.org/10.1016/S0024-6301(03)00041-4

Pangarkar, N. (2009). Do Firms Learn from Alliance Terminations? An Empirical Examination. *Journal of Management Studies*, *46*(6), 982–1004. https://doi.org/10.1111/j.1467-6486.2009.00844.x

Pangarkar, N., Yuan, L., & Hussain, S. (2017). Too much of a good thing? Alliance portfolio size and alliance expansion. *European Management Journal*, *35*(4), 477–485. https://doi.org/10.1016/j.emj.2016.09.006

Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., & Schacht, A. L. (2010). How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery*, *9*(3), Article 3. https://doi.org/10.1038/nrd3078

Peck, R. W., Lendrem, D. W., Grant, I., Lendrem, B. C., & Isaacs, J. D. (2015). Why is it hard to terminate failing projects in pharmaceutical R&D? *Nature Reviews. Drug Discovery*, *14*(10), 663–664. https://doi.org/10.1038/nrd4725

Perry, M. L., Sengupta, S., & Krapfel, R. (2004). Effectiveness of horizontal strategic alliances in technologically uncertain environments: Are trust and commitment enough? *Journal of Business Research, Managing the Future of Health Care Delivery*, *57*(9), 951–956. https://doi.org/10.1016/S0148-2963(02)00501-5

Pfeffer, J., & Salancik, G. (1978). *External Control of Organizations—Resource Dependence Perspective:* Harper & Row.

Pisano, G. P. (2006). *Science business: The promise, the reality, and the future of biotech*. Boston, Mass. : Harvard Business School Press.

Polidoro, F., Ahuja, G., & Mitchell, W. (2011). When the Social Structure Overshadows Competitive Incentives: The Effects of Network Embeddedness on Joint Venture Dissolution. *Academy of Management Journal*, *54*(1), 203–223. https://doi.org/10.5465/amj.2011.59215088

Polidoro, F., & Yang, W. (2021). Corporate Investment Relationships and the Search for Innovations: An Examination of Startups' Search Shift Toward Incumbents. *Organization Science*, *32*(4), 909–939. https://doi.org/10.1287/orsc.2020.1421

Posen, H. E., Keil, T., Kim, S., & Meissner, F. D. (2018). Renewing Research on Problemistic Search—A Review and Research Agenda. *Academy of Management Annals*, *12*(1), 208–251. (127283150). https://doi.org/10.5465/annals.2016.0018

Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, *41*(1), 116.

Rajan, R., Dhir, S., & Sushil. (2020). Alliance termination research: A bibliometric review and research agenda. *Journal of Strategy and Management*, *13*(3), 351–375. https://doi.org/10.1108/JSMA-10-2019-0184

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring Fairness in Machine Learning to Advance Health Equity. *Annals of Internal Medicine*, *169*(12), 866–872. https://doi.org/10.7326/M18-1990

Raman, J., Walton, D., MacDermid, J. C., & Athwal, G. S. (2017). Predictors of outcomes after rotator cuff repair-A meta-analysis. *Journal of Hand Therapy: Official Journal of the American Society of Hand Therapists*, *30*(3), 276–292. https://doi.org/10.1016/j.jht.2016.11.002

Ramasubbu, N., Shang, J., May, J. H., Tjader, Y., & Vargas, L. (2019). Task Interdependence and Firm Performance in Outsourced Service Operations. *Manufacturing & Service Operations Management*, *21*(3), 658–673. https://doi.org/10.1287/msom.2018.0716

Rathee, C., Malik, S., & Salandra, R. (2025). Timing, modifications, and tenor of firms' mandatory R&D disclosures: The role of competition. *Research Policy*, *54*(1), 105112. https://doi.org/10.1016/j.respol.2024.105112

Razmjou, H., Boljanovic, D., Lincoln, S., Holtby, R., Gallay, S., Henry, P., Macritchie, I., WCP Consortium, Borthwick, C., Mayer, L., Roknic, C., Shore, D., Kamino, A., Grossman, J., Hill, J., Singh, G., Travers, N., Yanofsky, L., Wilson, M., … Richards, R. R. (2017). Outcome of Expedited Rotator Cuff Surgery in Injured Workers: Determinants of Successful Recovery. *Orthopaedic Journal of Sports Medicine*, *5*(5), 2325967117705319. https://doi.org/10.1177/2325967117705319

Razmjou, H., Lincoln, S., Axelrod, T., & Holtby, R. (2008). Factors contributing to failure of rotator cuff surgery in persons with work-related injuries. *Physiotherapy Canada. Physiotherapie Canada*, *60*(2), 125–133. https://doi.org/10.3138/physio.60.2.125

Reagans, R., & McEvily, B. (2003). Network Structure and Knowledge Transfer: The Effects of Cohesion and Range. *Administrative Science Quarterly*, *48*(2), 240–267. https://doi.org/10.2307/3556658

Reuer, J. J., & Zollo, M. (2005). Termination outcomes of research alliances. *Research Policy*, *34*(1), 101–115. https://doi.org/10.1016/j.respol.2004.11.003

Rhee, M., & Haunschild, P. R. (2006). The Liability of Good Reputation: A Study of Product Recalls in the U.S. Automobile Industry. *Organization Science*, *17*(1), 101–117. https://doi.org/10.1287/orsc.1050.0175

Roels, G. (2014). Optimal Design of Coproductive Services: Interaction and Work Allocation. *Manufacturing & Service Operations Management*, *16*(4), 578–594. https://doi.org/10.1287/msom.2014.0495

Rosenkopf, L., & Nerkar, A. (2001). Beyond local search: Boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal*, *22*(4), 287–306. https://doi.org/10.1002/smj.160

Rothaermel, F. T. (2001). Incumbent's advantage through exploiting complementary assets via interfirm cooperation. *Strategic Management Journal*, *22*(Supp.), 687–700.

Rothaermel, F. T., & Deeds, D. L. (2004). Exploration and exploitation alliances in biotechnology: A system of new product development. *Strategic Management Journal*, *25*(3), 201–221. https://doi.org/10.1002/smj.376

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Sadowski, B., & Duysters, G. (2008). Strategic technology alliance termination: An empirical investigation. *Journal of Engineering and Technology Management*, *25*(4), 305–320. https://doi.org/10.1016/j.jengtecman.2008.10.002

Salerno, M. S., Gomes, L. A. de V., Silva, D. O. da, Bagno, R. B., & Freitas, S. L. T. U. (2015). Innovation processes: Which process for which project? *Technovation*, *35*, 59–70. https://doi.org/10.1016/j.technovation.2014.07.012

Sampson, S. (2000). *Understanding Service Businesses: Applying Principles of Unified Services Theory*. John Wiley & Sons, Ltd.

Sampson, S., & Chase, R. B. (2022). Optimizing Customer Involvement: How Close Should You Be to Your Customers? *California Management Review*, *65*(1), 119–146. https://doi.org/10.1177/00081256221118117

Sampson, S., & Froehle, C. M. (2006). Foundations and Implications of a Proposed Unified Services Theory. *Production and Operations Management*, *15*(2), 329–343. https://doi.org/10.1111/j.1937-5956.2006.tb00248.x

Sassi, M., Villa Corta, M., Pisani, M. G., Nicodemi, G., Schena, E., Pecchia, L., & Longo, U. G. (2024). Advanced Home-Based Shoulder Rehabilitation: A Systematic Review of Remote Monitoring Devices and Their Therapeutic Efficacy. *Sensors*, *24*(9), 2936. https://doi.org/10.3390/s24092936

Savva, N., & Scholtes, S. (2014). Opt-Out Options in New Product Co-development Partnerships. *Production and Operations Management*, *23*(8), 1370–1386. https://doi.org/10.1111/poms.12059

Shaik, A. S., Alshibani, S. M., Mishra, S., Papa, A., & Cuomo, M. T. (2023). Does learning from innovation failure enhance innovation performance? A quantitative investigation of small businesses. *Technovation*, *127*, 102818. https://doi.org/10.1016/j.technovation.2023.102818

Sharaf, T., Williams, T., Chehade, A., & Pokhrel, K. (2020). BLNN: An R package for training neural networks using Bayesian inference. *SoftwareX*, *11*, 100432. https://doi.org/10.1016/j.softx.2020.100432

Shavinina, L. (2003). *The International Handbook on Innovation* (1st ed.). Elsevier. https://doi.org/10.1016/B978-0-08-044198-6.X5000-1

Sheetal, A., Jiang, Z., & Di Milia, L. (2023). Using machine learning to analyze longitudinal data: A tutorial guide and best-practice recommendations for social science researchers. *Applied Psychology*, *72*(3), 1339–1364. https://doi.org/10.1111/apps.12435

Shenhar, A. (2001). One Size Does Not Fit All Projects: Exploring Classical Contingency Domains. *Management Science*, *47*(3), 394–414. https://doi.org/mnsc.47.3.394.9772

Shepherd, D. A., & Haynie, J. M. (2011). Venture failure, stigma, and impression management: A self-verification, self-determination view. *Strategic Entrepreneurship Journal*, *5*(2), 178–197. https://doi.org/10.1002/sej.113

Shepherd, D. A., Patzelt, H., & Wolfe, M. (2011). Moving Forward from Project Failure: Negative Emotions, Affective Commitment, and Learning from the Experience. *Academy of Management Journal*, *54*(6), 1229–1259.

Shi, Y. (1998). The effects of tasks on semiconductor start-up alliance dissolution: A structuration perspective. *The Journal of High Technology Management Research*, *9*(1), 87–113. https://doi.org/10.1016/1047-8310(88)90007-7

Shimao, H., Kim, S. J., Khern-Am-Nuai, W., & Cohen, M. C. (2025). Revisiting the CEO Effect Through a Machine Learning Lens. *Management Science*, *71*(6), 5396–5408. https://doi.org/10.1287/mnsc.2023.03625

Shrestha, Y. R., He, V. F., Puranam, P., & von Krogh, G. (2021). Algorithm Supported Induction for Building Theory: How Can We Use Prediction Models to Theorize? *Organization Science*, *32*(3), 856–880. https://doi.org/10.1287/orsc.2020.1382

Silverberg, G., & Verspagen, B. (2007). The size distribution of innovations revisited: An application of extreme value statistics to citation and value measures of patent significance. *Journal of Econometrics, The Econometrics of Intellectual Property*, *139*(2), 318–339. https://doi.org/10.1016/j.jeconom.2006.10.017

Singh, J., & Fleming, L. (2010). Lone Inventors as Sources of Breakthroughs: Myth or Reality? *Management Science*, *56*(1), 41–56. https://doi.org/10.1287/mnsc.1090.1072

Soulami, M., Benchekroun, S., & Galiulina, A. (2024). Exploring how AI adoption in the workplace affects employees: A bibliometric and systematic review. *Frontiers in Artificial Intelligence*, *7*. https://doi.org/10.3389/frai.2024.1473872

Spring, M., Faulconbridge, J., & Sarwar, A. (2022). How information technology automates and augments processes: Insights from Artificial-Intelligence-based systems in professional service operations. *Journal of Operations Management*, *68*(6–7), 592–618. https://doi.org/10.1002/joom.1215

Stoltzfus, J. C. (2011). Logistic Regression: A Brief Primer. *Academic Emergency Medicine*, *18*(10), 1099–1104. https://doi.org/10.1111/j.1553-2712.2011.01185.x

Stuart, T. E., Ozdemir, S. Z., & Ding, W. W. (2007). Vertical Alliance Networks: The Case of University-Biotechnology-Pharmaceutical Alliance Chains. *Research Policy*, *36*(4), 477–498. https://doi.org/10.1016/j.respol.2007.02.016

Stynes, T. (2010). *Pfizer Agrees to Acquire Drug Developer FoldRx*. WSJ. https://www.wsj.com/articles/SB10001424052748703882304575465442494683032

Subramanian, A. M., Lévesque, M., & van de Vrande, V. (2020). "Pulling the Plug:" Time Allocation between Drug Discovery and Development Projects. *Production and Operations Management*, *29*(12), 2851–2876. https://doi.org/10.1111/poms.13255

Subramanian, A. M., Lim, K., & Soh, P.-H. (2013). When birds of a feather don't flock together: Different scientists and the roles they play in biotech R&D alliances. *Research Policy*, *42*(3), 595–612. https://doi.org/10.1016/j.respol.2012.12.002

Sukhatme, L. (2007). Deterring Fraud: Mandatory Disclosure and the Fda Drug Approval Process. *New York University Law Review*, *82*(3). https://nyulawreview.org/wp-content/uploads/2018/08/NYULawReview-82-4-Sukhatme.pdf

Tandon, V., Asgari, N., & Ranganathan, R. (2022). Divestment of relational assets following acquisitions: Evidence from the biopharmaceutical industry. *Strategic Management Journal*, smj.3456. https://doi.org/10.1002/smj.3456

Taneri, N., & De Meyer, A. (2017). Contract Theory: Impact on Biopharmaceutical Alliance Structure and Performance. *Manufacturing & Service Operations Management*, *19*(3), 453–471. https://doi.org/10.1287/msom.2017.0617

Taylor, N. P. (2023, January 6). *Pfizer pivots from early-stage rare disease R&D, shifting to external innovation and putting assets up for sale*. Fierce Biotech. https://www.fiercebiotech.com/biotech/pfizer-pivots-early-stage-rare-disease-rd-shifting-external-innovation-and-putting-assets

Tian, Z., Gurnani, H., & Xu, Y. (2021). Collaboration in Development of New Drugs. *Production and Operations Management*, *30*(11), 3943–3966. https://doi.org/10.1111/poms.13377

Tsinopoulos, C., Yan, J., & Sousa, C. M. P. (2019). Abandoning innovation activities and performance: The moderating role of openness. *Research Policy*, *48*(6), 1399–1411. https://doi.org/10.1016/j.respol.2019.02.005

Tyagi, H., Hermosilla, M., & Shah, R. (2025). Does Transparency Hinder Technological Novelty? Evidence from Large Pharmaceutical Firms. *Manufacturing & Service Operations Management*, *27*(5), 1415–1432. https://doi.org/10.1287/msom.2023.0527

Tyebjee, T., & Hardin, J. (2004). Biotech–pharma alliances: Strategies, structures and financing. *Journal of Commercial Biotechnology*, *10*(4), 329–339. https://doi.org/10.1057/palgrave.jcb.3040091

Ucbasaran, D., Shepherd, D. A., Lockett, A., & Lyon, S. J. (2013). Life After Business Failure: The Process and Consequences of Business Failure for Entrepreneurs. *Journal of Management*, *39*(1), 163–202. https://doi.org/10.1177/0149206312457823

Uotila, J., Maula, M., Keil, T., & Zahra, S. A. (2009). Exploration, exploitation, and financial performance: Analysis of S&P 500 corporations. *Strategic Management Journal*, *30*(2), 221–231. https://doi.org/10.1002/smj.738

Utterback, J. M. (1971). The Process of Technological Innovation within the Firm. *The Academy of Management Journal*, *14*(1), 75–88. https://doi.org/10.2307/254712

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, *8*(12), 2293–2303. https://doi.org/10.1038/s41562-024-02024-1

Vassolo, R. S., Anand, J., & Folta, T. B. (2004). Non-additivity in portfolios of exploration activities: A real options-based analysis of equity alliances in biotechnology. *Strategic Management Journal*, *25*(11), 1045–1061. https://doi.org/10.1002/smj.414

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Verbyla, A. P. (2019). A note on model selection using information criteria for general linear models estimated using REML. *Australian & New Zealand Journal of Statistics*, *61*(1), 39–50. https://doi.org/10.1111/anzs.12254

Villena, V. H., Revilla, E., & Choi, T. Y. (2011). The dark side of buyer–supplier relationships: A social capital perspective. *Journal of Operations Management*, *29*(6), 561–576. https://doi.org/10.1016/j.jom.2010.09.001

von Zahn, M., Bauer, K., Mihale-Wilson, C., Jagow, J., Speicher, M., & Hinz, O. (2024). Smart Green Nudging: Reducing Product Returns Through Digital Footprints and Causal Machine Learning. *Marketing Science*. https://doi.org/10.1287/mksc.2022.0393

Walsh, J. P., Lee, Y.-N., & Nagaoka, S. (2016). Openness and innovation in the US: Collaboration form, idea generation and implementation. *Research Policy*, *45*(8), 1660–1671. https://doi.org/10.1016/j.respol.2016.04.013

Wang, G. (2022). The Effect of Medicaid Expansion on Wait Time in the Emergency Department. *Management Science*, *68*(9), 6648–6665. https://doi.org/10.1287/mnsc.2021.4239

Wang, L., Huang, N., Hong, Y., Liu, L., Guo, X., & Chen, G. (2023). Voice-based AI in call center customer service: A natural field experiment. *Production and Operations Management*, *32*(4), 1002–1018. https://doi.org/10.1111/poms.13953

Wang, T. (2023). Toward an understanding of innovation failure: The timing of failure experience. *Technovation*, *125*, 102787. https://doi.org/10.1016/j.technovation.2023.102787

Wang, X.-X., & Jiao, H.-Y. (2022). The impact of network positions in scientific collaboration on pharmaceutical firms' technological innovation performance: Moderating roles of scientific collaboration strength and patent stock. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.980845

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548–573. https://doi.org/10.1037/0033-295X.92.4.548

Weng, D. H., & Tang, Y. (2023). How do Status Differentials Affect the Unplanned Dissolution of Alliances? *Journal of Management Studies*. https://doi.org/10.1111/joms.12938

Whittaker, E., & Bower, D. J. (1994). A shift to external alliances for product development in the pharmaceutical industry. *R&D Management*, *24*(3), 249–260. https://doi.org/10.1111/j.1467-9310.1994.tb00877.x

Williamson, J., Spicer, A. J., & Colcomb, P.-A. (2024). The Effect of the RMAT Designation on Biotechnology Stock Prices, Drug Development Timelines and Outcomes: An Empirical Analysis. *Journal of Pharmaceutical Research International*, *36*(9), 1–15. https://doi.org/10.9734/jpri/2024/v36i97572

Wright, B. J., Galtieri, N. J., & Fell, M. (2014). Non-adherence to prescribed home rehabilitation exercises for musculoskeletal injuries: The role of the patient-practitioner relationship. *Journal of Rehabilitation Medicine*, *46*(2), 153–158. https://doi.org/10.2340/16501977-1241

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, *316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099

Xia, J. (2011). Mutual dependence, partner substitutability, and repeated partnership: The survival of cross-border alliances. *Strategic Management Journal*, *32*(3), 229–253. https://doi.org/10.1002/smj.873

Xiong, J., Wang, T., & Yang, H. (2024). Innovation search after failure: The role of peers. *Journal of Business Research*, *183*, 114873. https://doi.org/10.1016/j.jbusres.2024.114873

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., & Zhu, J. (2019). Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. In J. Tang, M.-Y. Kan, D. Zhao, S. Li, & H. Zan (Eds.), *Natural Language Processing and Chinese Computing* (pp. 563–574). Springer International Publishing. https://doi.org/10.1007/978-3-030-32236-6_51

Yadav, S., Singh, A., Singhal, R., & Yadav, J. P. (2024). Revolutionizing drug discovery: The impact of artificial intelligence on advancements in pharmacology and the pharmaceutical industry. *Intelligent Pharmacy*, *2*(3), 367–380. https://doi.org/10.1016/j.ipha.2024.02.009

Yamanoi, J., & Cao, Q. (2014). Competition and termination of the alliances between asymmetric partners: The case of Japanese department stores. *Asia Pacific Journal of Management*, *31*(4), 949–971. https://doi.org/10.1007/s10490-014-9383-y

Yashiro, K., Lim, Y., Sengoku, S., & Kodama, K. (2023). Recent trends in interorganizational deal networks in pharmaceutical and biotechnology industries. *Drug Discovery Today*, *28*(3), 103483. https://doi.org/10.1016/j.drudis.2022.103483

Zarin, D. A., Tse, T., Williams, R. J., Califf, R. M., & Ide, N. C. (2011). The ClinicalTrials.gov Results Database—Update and Key Issues. *New England Journal of Medicine*, *364*(9), 852–860. https://doi.org/10.1056/NEJMsa1012065

Zeijen, A., Romagnoli, M., & Marengo, L. (2025). Signposts for Problemistic Search: Reference Points and Adaptation in Rugged Landscapes. *Strategy Science*, *10*(3), 263–279. https://doi.org/10.1287/stsc.2023.0072

Zhang, Z., Gu, X., & Yang, X. (2024). Performance Feedback and Innovative Search Strategies: An Integrative Perspective of Motivation and Capacity for Risk Taking. *British Journal of Management*, *35*(4), 1867–1885. https://doi.org/10.1111/1467-8551.12794

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x

Appendices
Supplementary Materials for Chapter 2

Appendix A. Inclusion and Exclusion Criteria

Inclusion criteria	Exclusion criteria
1. Age ≥18 years;	1. Upper body neurological conditions;
2. Diagnosis of unilateral rotator cuff tendinosis, shoulder impingement syndrome, or rotator cuff tear;	2. Should injuries present in both shoulders; or
3. Planned physiotherapy treatment; and	3. Failed surgical shoulder intervention
4. Capacity to participate in home shoulder physiotherapy	

Table A1. Inclusion and Exclusion Criteria for Participant Selection

Appendix B. Selection of Control Variables for Analysis

To ensure the robustness of our analyses, we include a set of control variables with established relevance to physiotherapy (Bassett, 2003; Chester et al., 2013, 2018; Jack et al., 2010; Jin et al., 2008). *Age* is associated with physical capacity and may influence both participation and movement quality. *Body Mass Index (BMI)* serves as a proxy for body composition, which can affect movement accuracy and fatigue. *Sex* accounts for potential physiological and behavioral differences in exercise performance across male and female participants. Finally, the *Pain Self-Efficacy Questionnaire* (*PSEQ2*) score captures patients' confidence in managing pain (Nicholas et al., 2015), a factor shown to influence motivation and engagement in rehabilitation. Additional demographic information was collected through an intake questionnaire and is reported in Table B1. While not all variables are used as controls, they confirm that the intervention and control cohorts were well balanced.

	Control Group (n = 100)		Intervention Group (n = 24)	
Age, mean (SD)	*n* = 100	62.8 (12.9)	*n* = 23	57.1 (21.6)
Baseline physical activity (hours/week), mean (SD)	93*	2.1 (4.0)	21	1.8 (3.2)
Body Mass Index (BMI), mean (SD)	96	25.4 (6.4)	20	30.0 (7.5)
Dominant side injured	100	24% (24/100)	22	36% (8/22)
Duration of symptoms (months), mean (SD)	87	39.5 (55.3)	18	14.4 (14.0)
Sex (female)	99	59% (58/99)	23	78% (18/23)
Highest Level of Education, *n* (%)	**89**		**20**	
No high school diploma		3% (3/89)		5% (1/20)
High school diploma		14% (12/89)		20% (4/20)
Some college/university		11% (10/89)		15% (3/20)
College diploma / professional degree		34% (30/89)		40% (8/20)
University degree		38% (34/89)		20% (4/20)
Income (per annum), *n* (%)	**81**		**13**	
Less than $40,000		35% (28/81)		33% (4/12)
$40,000 to less than $80,000		28% (23/81)		33% (4/12)
$80,000 or more		37% (30/81)		41% (5/12)
Married/common law, *n* (%)	96	59% (57/96)	19	84% (16/19)
Severity of Injury (%)	**100**		**24**	
No tear		39% (39/100)		33% (8/24)
Partial tear		23% (23/100)		21% (5/24)
Full thickness tear		38% (38/100)		46% (11/24)
Smoking status, *n* (%)	**50**		**14**	
Non-smoker		80% (40/50)		71% (10/14)
Ex-smoker		14% (7/50)		21% (3/14)
Current smoker		6% (3/50)		7% (1/14)
Traumatic injury, *n* (%)	99	45% (45/99)	21	29% (6/21)
Work status (working), *n* (%)	91	58% (53/91)	18	33% (6/18)
Questionnaires, mean (SD)				
Cumulative Illness Rating Scale (CIRS), mean (SD)	85	3.31 (2.67)	22	3.86 (2.73)
ENRICHED Social Support Inventory (ESSI)	100	26.2 (6.4)	24	28.4 (5.9)
2-Item Pain Self Efficacy Questionnaire (PSEQ2)	99	7.8 (3.2)	24	8.6 (3.1)
Patient Expectation Questionnaire	99	17.3 (3.5)	24	16.9 (3.9)
Hospital Anxiety and Depression Scale (Anxiety)	100	7.0 (4.4)	24	6.2 (3.9)
Hospital Anxiety and Depression Scale (Depression)	100	5.4 (3.9)	24	4.8 (3.1)

* The baseline physical activity data from one participant was removed due to an extreme outlier (99 hours per week), which reduced the response from 94 to 93 participants.
**Control participants had significantly longer symptom duration before treatment, explained in part by trial-related COVID-19 delays and potentially by patient hesitancy with seeking treatment during COVID-19 restrictions. Injury severity was comparable across groups.
Note: *n* represents the number of responses for each variable. The control group contains a total of 100 participants while the intervention group has 24.

Table B1. Demographics Data

PSEQ2 was selected from a broader set of psychometric measures, including the *Cumulative Illness Rating Scale* (*CIRS*) for overall health burden (de Groot et al., 2003), the *ENRICHED Social Support Inventory* (*ESSI*) for perceived social support (Mitchell et al., 2003), the *Patient Expectation*

Questionnaire for treatment expectations (Razmjou et al., 2008), and the *Hospital Anxiety and Depression Scales* (*HADS*) (Zigmond & Snaith, 1983). Expectedly, these psychological measures are highly correlated (Table B2). As well, severity of injury is moderately correlated with *HADS(D)* (r = -0.47), suggesting more severe injuries were associated with higher depression scores. To avoid multicollinearity and confounding, we retained only *PSEQ2*, which showed the weakest correlation with *severity* (r = -0.22). Additionally, *PSEQ2* is conceptually aligned with exercise participation, as this measure of self efficacy in managing pain is a known predictor of participating in physical rehabilitation activities (Nicholas et al., 2015). Thus, *age*, *BMI*, *sex*, and *PSEQ2* were included as control variables in all models.

	Expectation	**HADS(A)**	**HADS(D)**	**CIRS**	**PSEQ2**	**Severity**	**Age**	**BMI**	**Sex**
Expectation	1.00	0.42	0.51	0.03	-0.54	-0.26	-0.32	0.38	0.22
HADS(A)		1.00	0.60	-0.12	-0.49	-0.34	-0.17	0.12	0.43
HADS(D)			1.00	-0.05	-0.56	-0.47	-0.34	0.20	0.36
CIRS				1.00	0.33	0.01	0.12	0.21	-0.15
PSEQ2					1.00	0.22	0.50	-0.37	-0.41
Severity						1.00	-0.01	0.17	-0.20
Age							1.00	-0.14	-0.48
BMI								1.00	0.23
Sex									1.00

Table B2. Correlation Matrix Between Relevant Control Variables

We calculated the Variance Inflation Factor (VIF) for each predictor variable in input quantity (Table 2.2, main text) and input quality (Table 2.3, main text) models to assess for problematic multicollinearity. Generally, VIF values below 5 indicate low multicollinearity. We used the Generalized Variance Inflation Factor (GVIF) for the input quantity model to account for the multi-dimensional fixed effects (month and year fixed effects). The Adjusted GVIF allows for a direct interpretation similar to the traditional VIF used in the input quality model. The results, presented in Table B3, show that all VIF and Adjusted GVIF values are well below the threshold (of 5), confirming that multicollinearity is not a concern.

Input Quantity				Input Quality	
Variable	**GVIF**	**Df**	**Adjusted GVIF**	**Variable**	**VIF**
Intervention	2.247	1	1.499	log(Repetition)	1.03
Days since last clinical visit	1.384	1	1.176	Injury severity	1.26
Number of physio visits	1.047	1	1.023	Age	1.63
Age	1.220	1	1.104	BMI	1.31
BMI	1.137	1	1.066	Sex	1.40
Sex	1.168	1	1.081	PSEQ2	1.79
PSEQ2	1.182	1	1.087		
Month	2.008	11	1.032		
Year	4.216	5	1.155		

Table B3. Variance Inflation Factor (VIF)

Appendix C. Qualitative Analysis Protocol

We conducted qualitative interviews to examine the mechanisms underlying the intervention's effects. The resulting insights provide context for interpreting the quantitative results by offering an in-depth view of participant perspectives. This approach is also valuable to understand unexpected patterns in the data and clarify how the digital technology influenced participant behavior.

Interview participants were selected from the intervention cohort to ensure they had exposure to the SPARS technology. The sampling strategy employed was purposive by targeting participants who had completed most of the trial and therefore had a comprehensive experience with the technology. Of the five participants who completed the full 12-week treatment, three consented to be interviewed. The analysis of these three interviews formed the initial basis of the qualitative analysis. To corroborate these initial findings, the interview with a fourth participant, who exited the study after eight weeks of treatment, was then analyzed. An interview was also conducted with the physiotherapist who managed the weekly in-clinic visits for hundreds of patients in the study. This interview ensures that the experiences and perspectives shared by the four patient interviewees are consistent with the physiotherapist's overall perception of the participant population. All interviews were conducted in person during the final weeks of their treatment, with each session lasting approximately 60 minutes.

C1. Interview Protocol

Semi-structured interview guides were used to facilitate conversations with the patients and the physiotherapist. The guide provided a list of topics to ensure consistency across interviews while allowing flexibility to explore emergent themes and individual experiences. Sample questions for patients are provided in Tables C1 and for the physiotherapist in Table C2. To ensure a natural flow of conversation, questions were not asked verbatim but were instead introduced conversationally. All interviews were conducted in person by a member of the research team with formal training in qualitative research methods. Each interview prompted participants to discuss their experiences with the SPARS/PT+ technology, its perceived utility, and any concerns regarding privacy, data use, and the ethics of AI-enabled health monitoring. The interviews were audio-recorded and transcribed verbatim.

Table C1. Patient Interview Guide

Background and Context

- Can you start by telling me what brought you to physiotherapy at the Holland Centre and what your experience there has been like?
- Was your injury work-related? Had you done physiotherapy before this, or was this your first time?
- Have any family members or friends been involved in your rehabilitation, now or in the past? Do they help you with exercises, or do you do them on your own?

Enrollment in the Study:

- How did you decide to enroll in this study? Was there something specific that interested you about the technology?
- Do you generally tend to like new technologies, or do you feel indifferent about them?
- What were your expectations about how this technology might affect your recovery?

Experiences with Physiotherapy:

- Could you describe your past experience with physiotherapy?
- Comparing this experience with your past physiotherapy experiences, did anything stand out?
- Has your experience at the Holland Centre been mostly positive or negative? Did anything stand out as particularly notable, or was it fairly routine?
- Would you say the balance of in-person versus at-home therapy was right for you? Would you prefer more in-person sessions, or more at-home work?

Experiences with the Technology:

- Could you describe the technology in your own words, as if explaining it to someone unfamiliar? What does it look like, how do you use it, and what does it do?
- Walk me through your experience using the technology, step by step, from the first meeting with your physiotherapist to your current stage.
- Is there anything about the technology that stood out to you? positive, negative, or unique?
- How do you feel about the technique ratings and feedback provided by the system?
- Seeing your progress, does it make you feel motivated, stressed, or something else?
- Did the technology affect your motivation or diligence? If so, how?

- Is there anything you would change about the technology?
- Do you use any other technologies to manage your health? How does this compare?

Technology and Service Relationship:
- If you had the choice to continue physiotherapy with or without this technology, which would you prefer, and why?
- How do you see the ideal balance between in-person sessions and at-home technology use (e.g., 50/50, 75/25)?
- How do you think using the tablet and watch changes or affects your relationship with your physiotherapist?
- Does it affect the work you do with your physiotherapist in-session?

Concerns and Reflections
- Is there anything that worries you about the technology (e.g., the tablet, the watch)?
- Hypothetically, how would you feel if your exercise data were shared with someone outside the hospital or outside your care team?
- Overall, besides motivation, are there any other ways the technology makes you feel?

Table C2. Physiotherapist Interview Guide

Background
- Could you start by telling me about your educational and professional background? When did you begin practicing as a physiotherapist, and where did you study?
- What does an average day look like for you? How much of your time is spent interacting with patients versus administrative or other responsibilities?
- How would you describe the typical patients who come to the Holland Centre? How are they usually referred?
- What is the usual course of care at the centre, and how does discharge to the community typically work?

Perspectives on Physiotherapy Practice
- From your experience, what would you say are some essential elements of good physiotherapy?
- Do you think patients define "good physiotherapy" in the same way? Have you noticed differences between your professional view and patients' expectations or preferences?

Experiences with SPARS/PT+
- Can you recall your first impressions when you heard about SPARS/PT+? Has your opinion changed over time?
- Could you walk me through what happens when you start working with a patient using SPARS? How do you introduce or onboard them to the technology?
- What role do you see SPARS playing in physiotherapy?
- Are there patients who found it difficult to use SPARS? What kinds of challenges did they face?
- Have you noticed differences in how effective SPARS is depending on the quality of the physiotherapist–patient relationship? For example, if rapport is weak, does SPARS still support adherence and improvement?
- For patients who did not improve, did you ever discuss SPARS with them? Did they attribute lack of improvement to the technology, or was it viewed as neutral?

Reflections on Future Models of Care
- Thinking about how SPARS is currently designed, what would be your preference in terms of the balance between in-person visits and technology-supported home exercise?
- Do you see a possible model of care where SPARS might replace some in-person contact altogether, with physiotherapists mainly monitoring data? If so, what would that ideal model

look like in your opinion?

- Are there contexts where you think SPARS should not be used?
- Do you have any concerns about SPARS (ethical, practical, or clinical) that have come up in your time working with it?

C2. Analytical Approach

The thematic analysis was performed by a research team member who had no prior relationship with the participants and was not involved in the technology's development nor the trial's implementation. The analysis was guided by a grounded theory approach. The process began with open coding of the verbatim transcripts to identify recurrent concepts and patterns in the data. These initial codes were then refined into more focused codes, which were subsequently consolidated through an iterative process to form emergent themes.

To ensure the trustworthiness of the findings, we implemented several methodological safeguards. First, we conducted peer debriefing through regular meetings with the research team to review coding decisions and interpretations, which helped validate and finalize the thematic structure. Second, we engaged in iterative refinement of codes, revisiting and adjusting both codes and themes throughout the analysis to ensure they accurately captured participants' experiences. Finally, we employed triangulation with quantitative results by comparing qualitative insights with the observed exercise patterns from the smartwatch data. Figure C1 provides a summary coding tree that demonstrates the development of codes into emergent themes.

Figure C1. Coding Tree

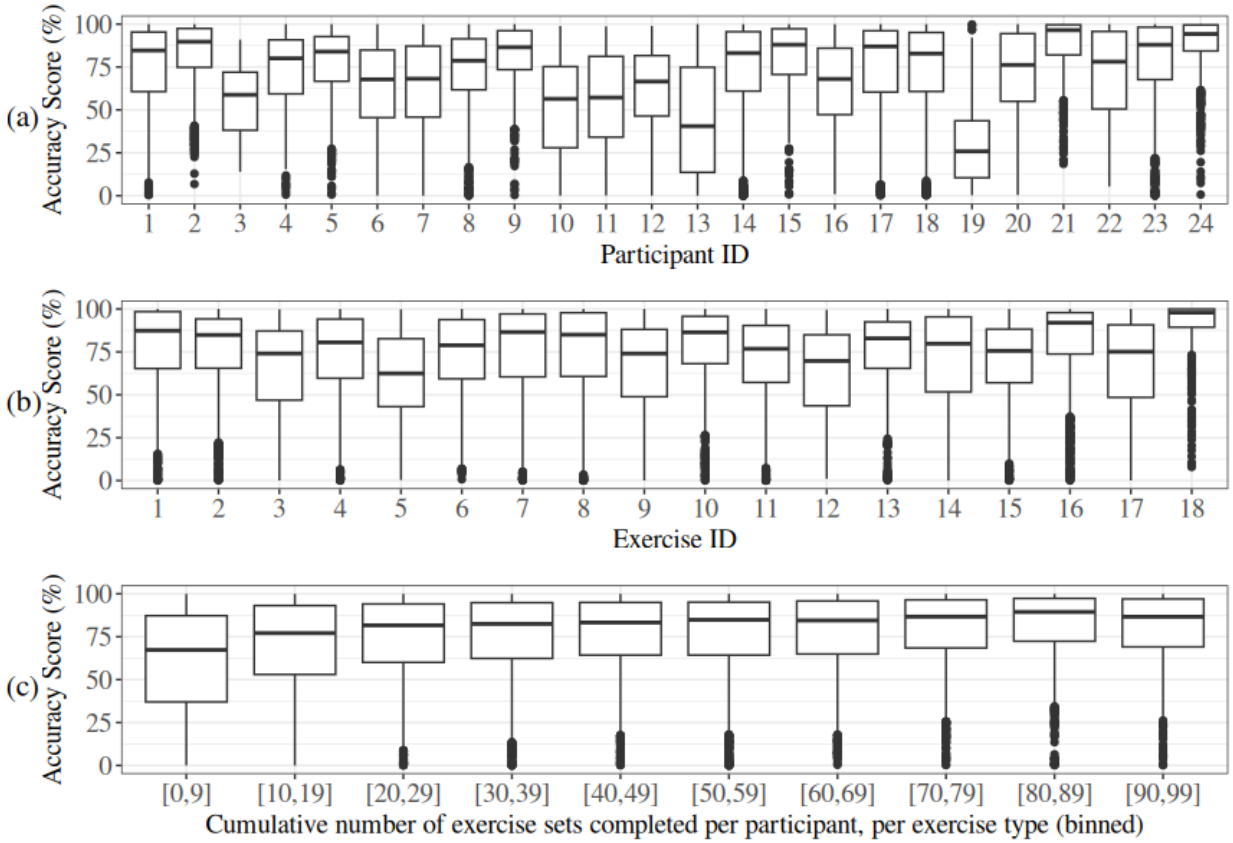
First Order Codes	Focused Codes	Emergent Themes
"The tablet shows me the movement. So, I get to see [...] the exact movement that I'm trying to replicate."	Technology reduces doubt about technique correctness	Uncertainty reduction
"Yeah, sort of the fact that I know I've done the exercise correctly, makes me want to continue doing it."		
"The watch would be kind of measuring or calculating how good the patient was doing the exercise." (Physiotherapist)		
"I like that aspect, it's giving you immediate feedback. Where before […] are you doing it quite correctly? Think so, but don't know."	Feedback provides reassurance	
"So, the way the program is set up, it gives you time to learn each step and so you then are more likely to be doing it correctly."		
"With the tablet, it's kind of giving me a rating each time […] at the end, each individual exercise, it says sort of are you in the green zone or are you in the red zone."		
"Exercises will never make you worse […] if aggravated, there is a physiological reason and we have to treat you." (Physiotherapist)		
"It's nice to see, okay, I'm in a green zone on this one and then I'm in a red zone on this one, which means I haven't done it correctly or something, so I've got to think about it to prevent that from happening next time."	Builds routine/ improvement	
"By the time you got to number ten, your form had improved … when you move right into the second set, your brain and your memory for how you're doing the exercise is still with you."		
"My exercise improves, my methodology of doing that exercise improves from one week to the next, because of the way that I'm monitored"		
"It's nice to see, okay, I'm in a green zone on this one and then I'm in a red zone on this one, which means I haven't done it correctly or something, so I've got to think about it to prevent that from happening next time"		
"They start feeling stronger, less pain… they are more motivated to keep doing it." (Physiotherapist)		
"The tablet is sitting there, it's ready to go … instead of saying, oh screw it, I'll do it tomorrow, it's like, no, I've got to do this."	Device creates a commitment	Intrinsic accountability
"It keeps you honest. Sort of, hey, did it record you doing all of the exercises?"		
"I'm one of these personalities that might, I need to be monitored a bit, because it increases my competitiveness and increases my desire to achieve."		
"What is very important is… the patient has that commitment that I have the watch." (Physiotherapist)		
"I like that […] I can see my progress."	Tracking progress over time motivates	
"Each week I can see a progression in how much I can do"		
"Maybe I'm competitive, I like to see the green in the 80% completion range. And that's motivation."		
"There's at least something tangible that I can look at and see what my progress is like."		

"The importance is the patient will have a goal… They can see are these goals being met or not" (Physiotherapist)		
"I actually look forward to getting it done versus oh geez, here's another chore to do."	acts as a habit-reinforcing tool	
"If it wasn't for this, this treatment would have followed the same path as what I've done in the past."		
"I need to make sure that the strength that I've now achieved, I continue."		

"With this, they're able to see how correctly I'm doing the exercise."	Feeling of being monitored by physiotherapist	Extrinsic accountability
"It's that idea that someone is following, someone is knowing if I do the exercises or not, and then the real motivation is basically to get better."		
"They feel that, with the watch, someone will know if they are doing or not doing their exercises." (Physiotherapist)		
"If he's helping me, I don't want to let him down either."	Data visibility creates external pressure not to disappoint provider	
"If I skip it, I have to tell [my physiotherapist], 'Sorry … I did not do it for how many days."		
"When I go to the physiotherapist, like, he tells me, hey, am I doing it right or am I doing it wrong."		
"That sense of [the physiotherapist] watching is a very good motivation"		

Appendix D. Input Quality Exploratory Data Analysis

We validate the inclusion of random effects in our input quality analysis (Table 2.3, main text) and examine the relationship between exercise technique accuracy and the number of exercise repetitions completed by a participant for a given exercise, by performing an exploratory data analysis presented in Figure D1. The results show distinct variation in accuracy scores, both between participants and exercise types, which supports the addition of random effects by participant and exercise type. As well, the improvement in accuracy scores follows a well-established learning curve pattern: an initial period with rapid improvement followed by a slowing rate of improvement, eventually reaching an almost stable level. Learning curve effects are commonly modeled as log-linear relationships (Anzanello & Fogliatto, 2011).



Notes: (a) Variation by participant.
(b) Variation by exercise type.
(c) Variation by number of sets performed for a given exercise type

Figure D1. Exploratory Analysis of Input Quality Accuracy Scores

Appendix E. Robustness for Input Quantity Analysis

We assess the robustness of our main findings (Table 2.2, main text) by estimating alternative model specifications with and without control variables. For ease of comparison, Models 1 and 4 in Table E1 reproduce the main analysis. Models 2 and 5 exclude all control variables and continue to support our findings. Models 3 and 6 include month and year fixed effects but omit demographic controls, and the results remain consistent with the main specification.

Dependent Variable	Exercise frequency			Exercise duration		
	(1) **All controls**	(2) **No controls**	(3) **Month & Year FE Only**	(4) **All controls**	(5) **No controls**	(6) **Month & Year FE Only**
Intervention	-0.369 (0.473)	0.034 (0.324)	-0.604 (0.430)	-0.701 (1.206)	-0.134 (0.855)	-1.442 (1.118)
Days since last physio visit	-0.055*** (0.004)	-0.055*** (0.003)	-0.053*** (0.004)	-0.030*** (0.003)	-0.032*** (0.003)	-0.029*** (0.004)
Number of physio visits	-0.195*** (0.012)	-0.185*** (0.010)	-0.199*** (0.012)	-0.188*** (0.026)	-0.175*** (0.026)	-0.200*** (0.025)
Intervention × Number of visits	0.071** (0.026)	0.069** (0.023)	0.084*** (0.025)	0.687*** (0.058)	0.679*** (0.054)	0.653*** (0.056)
Age	0.016+ (0.009)			0.043+ (0.023)		
Sex (Female)	-0.249 (0.267)			-0.949 (0.702)		
BMI	0.002 (0.020)			0.031 (0.053)		
PSEQ2	0.032 (0.041)			0.141 (0.108)		
Year fixed effects	Yes	No	Yes	Yes	No	Yes
Month fixed effects	Yes	No	Yes	Yes	No	Yes
Random Effects	**Variance**					
Participant	1.637 (1.279)	1.613 (1.270)	1.650 (1.285)	11.62 (3.408)	11.94 (3.456)	12.32 (3.510)
Residual				28.94 (5.380)	29.73 (5.453)	29.53 (5.434)
Num.Obs	9,310	9,310	9,310	9,310	9,310	9,310
AIC	9,357	10,124	10,097	58,201	62,091	62,047
BIC	9,542	10,167	10,254	58,394	62,141	62,213

$+ p < 0.1$, $* p < 0.05$, $** p < 0.01$, $*** p < 0.001$

Table E1. Regression Coefficient Estimates for Input Quantity With and Without Controls

Second, we re-estimated our models while restricting the sample to the 78 participants enrolled during the overlap period between cohorts to address potential time related confounding. The univariate results were consistent: intervention participants exercised for 10.07 minutes per day, on average, and control participants performed a daily average of 4.11 minutes of exercises. This difference is statistically significant (t(7.8064) = 3.16, $p = 0.014$) and represents a large effect size (Cohen's d = 1.65, [95% CI: 0.86, 2.43]). Similarly, Table E2 shows that the mixed-effects model remains consistent for this subset of participants. Results are reported in Table E2 (Model 2), with the full sample model

(Model 1) from the main text reproduced for ease of comparison. Because of the limited sample size (only 8 intervention participants), the binary outcome model for exercise frequency did not converge reliably, whereas the continuous measure of exercise duration provided sufficient information for estimation. Therefore, only exercise duration is reported.

Dependent Variable	**Exercise duration**	
	(1) **Full Sample**	(2) **Overlap Sample**
Intervention	-0.701 (1.206)	-0.469 (1.873)
Days since last physio visit	-0.030*** (0.003)	-0.039*** (0.005)
Number of physio visits	-0.188*** (0.026)	-0.134*** (0.032)
Intervention × Number of visits	0.687*** (0.058)	0.889*** (0.094)
Age	0.043+ (0.023)	0.017 (0.041)
Sex (Female)	-0.949 (0.702)	0.430 (1.001)
BMI	0.031 (0.053)	0.050 (0.096)
PSEQ2	0.141 (0.108)	0.051 (0.150)
Year fixed effects	Yes	Yes
Month fixed effects	Yes	Yes
Random Effects	**Variance**	
Participant	11.62 (3.408)	14.47 (3.804)
Residual	28.94 (5.380)	27.32 (5.227)
Num.Obs	9,310	5,839
AIC	58,201	36,185
BIC	58,394	36,357

$+\ p < 0.1$, $*\ p < 0.05$, $**\ p < 0.01$, $***\ p < 0.001$

Note: Columns 1 includes all participants as reported in the main analysis. Columns 2 only includes participants which were enrolled between January 2022 and April 2023.

Table E2. Regression Coefficient Estimates for Input Quantity With and Without Overlap Sample

Third, we re-estimated our models with a subsample constructed through pairwise matching of intervention and control participants. Nearest-neighbor matching without replacement was implemented at a 1:1 ratio, using a logistic propensity score estimated from all available demographic

information (Table B1). Missing values were imputed using the median for continuous variables and the mode for categorical variables, with indicator variables included to capture missingness.

Propensity scores were estimated via logit regression, and nearest-neighbor matching was performed based on the logit score. Each intervention participant was paired with the closest available control participant; unmatched control participants were discarded. The resulting matched dataset was then used to re-estimate the mixed-effects regression models. Results are reported in Table E3 (Models 2 and 4), with the unmatched estimates (Models 1 and 3) from the main text reproduced for ease of comparison. The results remain consistent with our main findings.

Dependent Variable	**Exercise frequency**		**Exercise duration**	
	(1) **Without Matching**	(2) **With Matching**	(3) **Without Matching**	(4) **With Matching**
Intervention	-0.369 (0.473)	-0.471 (0.701)	-0.701 (1.206)	-2.886 (2.652)
Days since last physio visit	-0.055*** (0.004)	-0.053*** (0.007)	-0.030*** (0.003)	-0.021* (0.009)
Number of physio visits	-0.195*** (0.012)	-0.275*** (0.029)	-0.188*** (0.026)	-0.390*** (0.072)
Intervention × Number of visits	0.071** (0.026)	0.138*** (0.036)	0.687*** (0.058)	0.912*** (0.090)
Age	0.016+ (0.009)	0.005 (0.011)	0.043+ (0.023)	0.079+ (0.039)
Sex (Female)	-0.249 (0.267)	-0.827+ (0.461)	-0.949 (0.702)	0.415 (1.670)
BMI	0.002 (0.020)	0.013 (0.024)	0.031 (0.053)	0.181 (0.093)
PSEQ2	0.032 (0.041)	0.024 (0.068)	0.141 (0.108)	0.167 (0.247)
Year fixed effects	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes
Random Effects			**Variance**	
Participant	1.637 (1.279)	1.086 (1.042)	11.62 (3.408)	14.52 (3.810)
Residual			28.94 (5.380)	34.26 (5.853)
Num.Obs	9,310	3,142	9,310	3,142
AIC	9,357	3,323	58,201	5,340
BIC	9,542	3,481	58,394	5,176

$+ p < 0.1$, $* p < 0.05$, $** p < 0.01$, $*** p < 0.001$

Table E3. Regression Coefficient Estimates for Input Quantity With and Without Matching

Note: This table summarizes the regression coefficients for exercise frequency (columns 1–2) and exercise duration (columns 3–4). Columns 1 and 3 include all participants as reported in the main analysis. Columns 2 and 4 only include pairwise matched participants.

Appendix F. Robustness for Input Quality Analysis

We trained and compared mixed-effects regression models with and without a log transformation of the repetition variable to validate the use of a log-linear model. The results, presented in Table F1, indicate that the logarithmic model provides a better fit, as demonstrated by lower values for the REML, AIC, and BIC criteria (Verbyla, 2019). In our analysis, the REML (143,195 vs. 143,459), AIC (143,215 vs. 143,479), and BIC (143,304 vs. 143,568) values favor the logarithmic model. Despite the relative differences appearing small compared to the absolute values, such differences are considered substantial within the model selection framework (Burnham & Anderson, 2004).

Model	Fixed Effects	Estimate	Std. Error	t-value	Random Effects (Variance)	Std. Dev.	AIC	BIC	REML
Linear	Intercept	69.68***	9.76	7.14	Participant ID (Intercept)	14.16	143,479	143,568	143,459
	Repetition	0.079***	0.0058	13.59	Exercise ID (Intercept)	14.57			
	Severity	-2.237	3.95	-0.57	Exercise ID × Severity	5.13			
	Residual				Residual	22.51			
Logarithmic	Intercept	59.54***	9.48	6.28	Participant ID (Intercept)	13.64	143,215	143,304	143,195
	log(Repetition)	3.97***	0.19	21.15	Exercise ID (Intercept)	14.7			
	Severity	-2.008	3.83	-0.52	Exercise ID × Severity	5.31			
	Residual				Residual	22.33			

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table F1. Mixed-effects Regression Model Fit Comparison (Linear vs. Logarithmic)

References

Anzanello, M. J., & Fogliatto, F. S. (2011). Learning curve models and applications: Literature review and research directions. International Journal of Industrial Ergonomics, 41(5), 573–583. https://doi.org/10.1016/j.ergon.2011.05.001

Bassett, S. (2003). The assessment of patient adherence to physiotherapy rehabilitation. New Zealand Journal of Physiotherapy, 31, 60–66.

Chester, R., Jerosch-Herold, C., Lewis, J., & Shepstone, L. (2018). Psychological factors are associated with the outcome of physiotherapy for people with shoulder pain: A multicentre

longitudinal cohort study. British Journal of Sports Medicine, 52(4), 269–275. https://doi.org/10.1136/bjsports-2016-096084

Chester, R., Shepstone, L., Daniell, H., Sweeting, D., Lewis, J., & Jerosch-Herold, C. (2013). Predicting response to physiotherapy treatment for musculoskeletal shoulder pain: A systematic review. BMC Musculoskeletal Disorders, 14, 203. https://doi.org/10.1186/1471-2474-14-203

de Groot, V., Beckerman, H., Lankhorst, G. J., & Bouter, L. M. (2003). How to measure comorbidity. A critical review of available methods. Journal of Clinical Epidemiology, 56(3), 221–229. https://doi.org/10.1016/s0895-4356(02)00585-1

Jack, K., McLean, S. M., Moffett, J. K., & Gardiner, E. (2010). Barriers to treatment adherence in physiotherapy outpatient clinics: A systematic review. Manual Therapy, 15(3), 220–228. https://doi.org/10.1016/j.math.2009.12.004

Jin, J., Sklar, G. E., Min Sen Oh, V., & Chuen Li, S. (2008). Factors affecting therapeutic compliance: A review from the patient's perspective. Therapeutics and Clinical Risk Management, 4(1), 269–286. https://doi.org/10.2147/tcrm.s1458

Mitchell, P. H., Powell, L., Blumenthal, J., Norten, J., Ironson, G., Pitula, C. R., Froelicher, E. S., Czajkowski, S., Youngblood, M., Huber, M., & Berkman, L. F. (2003). A short social support measure for patients recovering from myocardial infarction: The ENRICHD Social Support Inventory. Journal of Cardiopulmonary Rehabilitation, 23(6), 398–403. https://doi.org/10.1097/00008483-200311000-00001

Nicholas, M. K., McGuire, B. E., & Asghari, A. (2015). A 2-item short form of the Pain Self-efficacy Questionnaire: Development and psychometric evaluation of PSEQ-2. The Journal of Pain, 16(2), 153–163. https://doi.org/10.1016/j.jpain.2014.11.002

Razmjou, H., Lincoln, S., Axelrod, T., & Holtby, R. (2008). Factors contributing to failure of rotator cuff surgery in persons with work-related injuries. Physiotherapy Canada. Physiotherapie Canada, 60(2), 125–133. https://doi.org/10.3138/physio.60.2.125

Verbyla, A. P. (2019). A note on model selection using information criteria for general linear models estimated using REML. Australian & New Zealand Journal of Statistics, 61(1), 39–50. https://doi.org/10.1111/anzs.12254

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. Acta Psychiatrica Scandinavica, 67(6), 361–370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x

Supplementary Materials for Chapter 3

Appendix G. Descriptive Statistics for Dataset of Terminated Alliances

Variable name	Missingness rate	Mean	Std. Dev.	Min	Max
Biotech firm size	0%	3.58	7.70	0	49
Pharma firm size	0%	43.91	35.62	0	150
Biotech failure experience	0%	4.38	16.52	0	201
Pharma failure experience	0%	86.68	94.25	0	375
Biotech technology experience	0%	0.01	0.11	0	1
Pharma technology experience	0%	0.05	0.67	0	13
Alliance experience	0%	1.09	2.93	0	25
Biotech alliance portfolio size	0%	10.89	21.51	0	250
Pharma alliance portfolio size	0%	210.73	172.32	1	678
Biotech R&D portfolio size	0%	8.38	19.90	0	307
Pharma R&D portfolio size	0%	142.70	158.40	0	713
Principal investigator (biotech =1)	0%	0.73	0.48	0	1
Technology novelty	0%	0.02	0.14	0	1
Indication novelty	0%	0.02	0.15	0	1

Table G1. Descriptive Statistics for Numerical Variables

Variable name	Missingness rate	Ordinal variable	Number of categories
Development stage at the time of alliance formation	0%	Yes	5
Therapy area	0%	No	18

Table G2. Descriptive Statistics for Categorical Variables

Appendix H. Datasets for CPH Models

The Cox proportional hazards (CPH) models in Section 3.3.3 (main text) consider alliances that did not end in termination. Some of these alliances concluded as planned based on contractual terms and are recorded as *completed*, while others remained *active* through the end of the study period. We prepare two datasets for comparison: one that includes active alliances ($n = 34{,}208$) and one that excludes them ($n = 3{,}496$). Descriptive statistics for both are presented in Tables H1 and H2.

Variable name	Alliances with Known Outcomes (Terminated of completed)					All Alliances				
	Missingness rate	Mean	Std. Dev.	Min	Max	Missingness rate	Mean	Std. Dev.	Min	Max
Biotech firm size	2.5%	3.56	7.45	0	62	4.2%	2.57	6.59	0	62
Pharma firm size	2.5%	14.77	25.47	0	164	4.2%	9.01	20.99	0	164
Biotech failure experience	2.5%	4.41	18.06	0	201	4.2%	4.13	17.91	0	204
Pharma failure experience	2.5%	27.04	61.34	0	440	4.2%	17.74	48.67	0	440
Biotech technology experience	2.5%	0.01	0.09	0	1	4.2%	0.01	.015	0	7
Pharma technology experience	2.5%	0.02	0.39	0	13	4.2%	0.03	0.57	0	42
Alliance experience	0%	0.42	1.65	0	25	0%	0.27	1.10	0	25
Biotech alliance portfolio size	2.5%	10.82	23.66	0	271	4.2%	9.57	23.25	0	271
Pharma alliance portfolio size	2.5%	70.06	126.83	1	756	4.2%	48.48	110.00	1	1052
Biotech R&D portfolio size	2.5%	7.47	18.27	0	307	4.2%	5.71	16.76	0	307
Pharma R&D portfolio size	2.5%	48.19	100.25	0	713	4.2%	29.95	82.52	0	715
Principal investigator (biotech =1)	2.5%	0.58	0.49	0	1	4.2%	0.40	0.49	0	1
Technology uncertainty	0%	0.02	0.15	0	1	0%	0.03	0.18	0	1
Indication uncertainty	0%	0.02	0.15	0	1	0%	0.07	0.26	0	1

Table H1. Descriptive Statistics for Numerical Variables

Variable name	Alliances with Known Outcomes (Terminated of completed)			All Alliances		
	Missingness rate	Ordinal variable	Number of categories	Missingness rate	Ordinal variable	Number of categories
Development stage at the time of alliance formation	32.1%	Yes	5	47%	Yes	5
Therapy area	0%	No	19	0%	No	19

Table H2. Descriptive Statistics for Categorical Variables

Compared to the characteristics of terminated alliances (see Tables G1 and G2), pharma firm attributes, including firm size, failure experience, R&D portfolio size, and alliance portfolio size, are substantially smaller in the two datasets described in Tables H1 and H2. This likely reflects the fact that alliances marked as *terminated* are disproportionately associated with large pharmaceutical firms, for which information is more publicly available and likely more closely tracked by analysts. While this showcases a bias in the termination dataset toward larger firms, it remains appropriate for our study, as large pharmaceutical firms account for the majority of alliances with biotech firms and are of great interest in our analysis. We nonetheless acknowledge that the set of terminated alliances is biased against smaller pharma firms.

The differences between the *known outcomes* dataset (i.e., terminated or completed alliances) and the *all alliances* dataset are also pronounced. The higher proportion of small pharma firms in the *all alliances* data suggests that many alliances involving smaller pharmaceutical firms are recorded in the database without follow-up information and thus remain "active". These firms are also less likely to experience a drug launch, which can inflate the estimated effect of drug launches in our CPH models. The main challenge is that firms with no drug launches are less likely to have termination events recorded, even if a termination occurred, violating the missing-at-random assumption. While not a perfect match, the *known outcomes* dataset more closely resembles the *termination* dataset than the *full alliance* dataset does.

In contrast, the attributes of biotech firms are more consistent across all three datasets. This suggests that potential misclassification of active alliances is more likely to involve smaller pharmaceutical firms than smaller biotech firms. This is plausible, as biotech firms are typically small thus less likely to be the source of public disclosures than their pharma partner. Termination information is thus more likely reported through pharmaceutical firm communications.

Our assumption that some alliances receive less scrutiny and are misclassified as active is further supported by the missingness rates reported in Tables H1 and H2. While the *termination* dataset contains complete data for the variables used in our analysis, the datasets used for survival analysis have higher rates of missing data. In a small number of cases ($\leq$ 4.2%), firm names could not be matched to Cortellis' Company Intelligence data. In these instances, neither the biotech nor the pharma firm names matched. In addition, a substantial number of drug projects are missing information on development stage at the time of alliance formation (missing in >32% of cases). Notably, none of these projects resulted in a drug launch, supporting our view that alliances between firms without drug launches receive less scrutiny and are more likely to be misclassified as active. Due to the high missingness rate for the predictor *development stage at the time of alliance formation*, we exclude this variable from our CPH models. As later reported in Figure 3.5 of the main text, feature importance analysis show that this variable has minimal influence on model predictions.

Appendix I. Statistical Assumption Tests

We test the statistical assumptions underlying the application of inference-related approaches on the dataset of terminated alliances ($n = 1{,}649$). We show that several issues threaten the reliability of regression-based methods, making machine learning (ML) the more appropriate analytics tool. Such justification is consistent with best practices when using ML methods for social science and management research problems (e.g., Choudhury et al., 2021; Lundberg et al., 2022; Sheetal et al., 2023). Specifically, we detect and assess multicollinearity among feature variables, examine the linearity between the continuous predictors and the dependent variable, and evaluate the influence of outliers. We also test whether the presence of interaction effects between predictors can improve model fit, which may justify the use of more complex ML models.

I1. Tests for Multicollinearity Amongst Predictors

Since no single measure is considered definitive for detecting multicollinearity (Kovàcs et al., 2005), we perform a series of tests including Determinant [X'X], Farrar Chi-Square, Red Indication, Sum of Lambda Inverse, Theil's Method, and Condition Number (Imdadullah et al., 2016). Table I1 indicates that the predictors failed four out of the six multicollinearity tests. Moreover, each individual predictor failed at least one of the tests, indicating the presence of significant dependencies between the feature variables.

Statistical test	Results	Multicollinearity detected	Threshold value
Determinant |X'X|	0.0001	Yes	0.01
Farrar Chi-Square	240046.61	Yes	Conf. = 0.99
Red Indicator	0.4434	No	0.5
Sum of Lambda Inverse	60.535	Yes	30
Theil's Method	6.8411	Yes	0.5
Condition Number	21.0508	No	30

Table I1. Multicollinearity Tests of All Predictors Used
(threshold values as in Imdadullah et al., 2016)

I2. Test for Linearity

We perform a Box-Tidwell test using a significance threshold of 0.05 to assess the assumption of linearity between the continuous predictors and the log odds of the dependent variable. Table I2 shows

that the linearity assumption is violated for *Pharma firm size* ($p = 0.0196$), *pharma R&D portfolio size* ($p = 0.0111$), and *pharma failure experience* ($p = 0.0134$). Nonetheless, from Figure 3.5 in our main analysis these three variables are found to be important predictors.

Variable	Estimate	Std. error	Statistic	*p*-value
Biotech: Alliance portfolio size	-0.000213	0.000421	-0.504856	0.61366
Pharma: Alliance portfolio size	-0.000439	0.000461	-0.952174	0.34100
Biotech: Firm size	0.0011913	0.000654	1.821379	0.068549
Pharma: Firm size	0.0010669	0.000457	2.3348884	**0.019549***
Biotech: R&D portfolio size	-0.000229	0.000289	-0.790625	0.429163
Pharma: R&D portfolio size	0.0009021	0.000355	2.5412249	**0.011047***
Elapse time between termination & drug launch	0.0000337	0.000314	0.107506	0.914388
Pharma: Failure experience	-0.001682	0.00068	-2.473961	**0.013362***
Biotech: Failure experience	0.0011299	0.001186	0.953014	0.340583
Alliance experience	0.0248959	0.01815	1.3716523	0.170172

Table I2. Box-Tidwell Test for Linearity Between Predictor and Dependent Variable's Log Odds ($p < 0.05$ is statistically significant)

I3. Test for Outliers

We use Grubbs' test (Grubbs, 1950) to assess the presence of outliers. Table I3 shows that 5 out of 11 continuous predictors reject the null hypothesis, suggesting the presence of at least one outlier. Since outliers are only a problem if they have a strong influence on model predictions, we also assess their impact by measuring Cook's distance on a logistic regression model. Observations with Cook's distance values significantly higher than $4/n$, with n as the sample size, are considered influential outliers (Altman & Krzywinski, 2016). Figure I1 shows observations appearing as outliers with strong influence; roughly 2%.

Variable	G	*p*-value	Alternative hypothesis
Alliance length	6.049407	0.0000097	highest value 11109 is an outlier
Biotech: Alliance portfolio size	5.330877	0.0006565	highest value 1914 is an outlier
Pharma: Alliance portfolio size	2.978993	1	highest value 2018 is an outlier
Biotech: Firm size	3.653547	1	highest value 696 is an outlier
Pharma: Firm size	1.801667	1	highest value 696 is an outlier
Biotech: R&D portfolio size	5.140179	0.0018472	highest value 2642 is an outlier
Pharma: R&D portfolio size	3.107801	1	highest value 2797 is an outlier
Elapse time between termination & drug launch	1.735343	1	highest value 1094 is an outlier
Pharma: Failure experience	3.266567	1	highest value 722 is an outlier
Biotech: Failure experience	6.379223	0.0000012	highest value 664 is an outlier
Alliance experience	6.916252	0	highest value 47 is an outlier

Table I3. Grubbs' Test for Outliers ($p < 0.05$ is statistically significant)

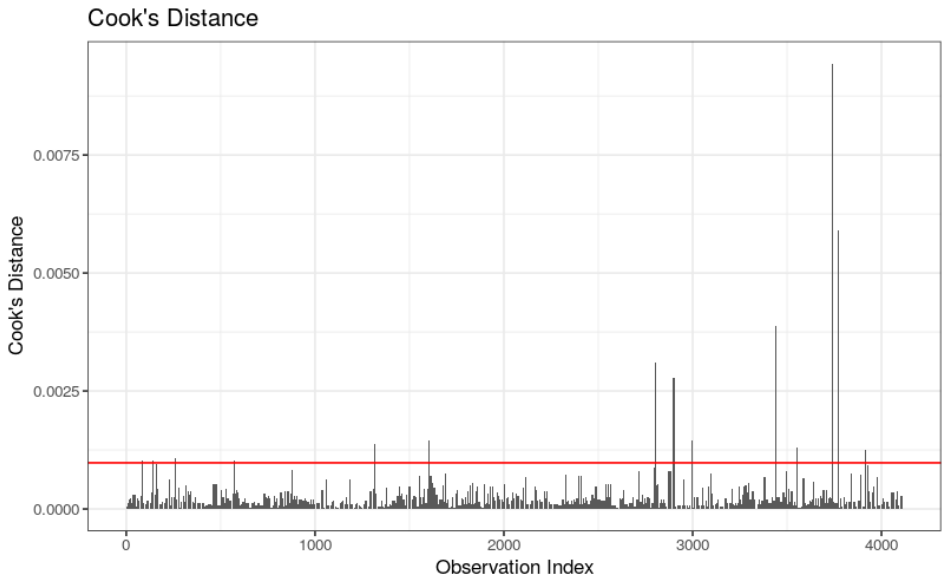
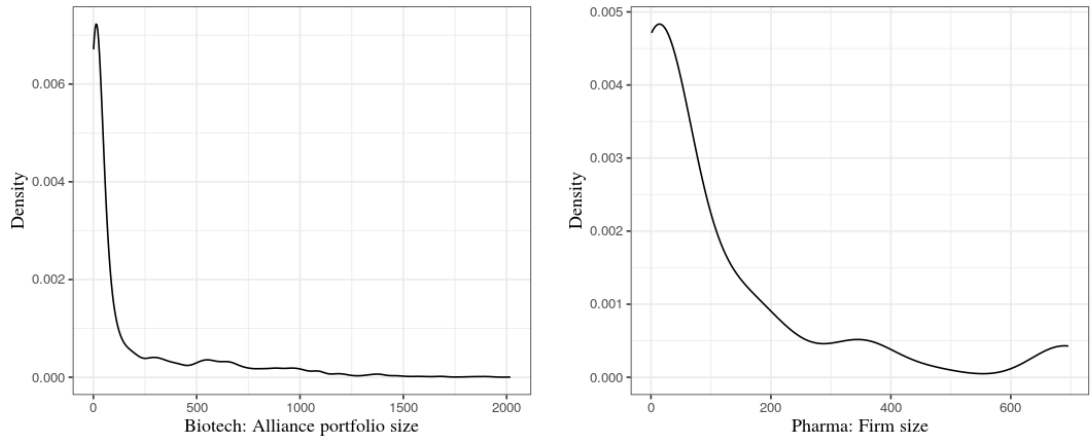


Figure I1. Cook's Distance Analysis with Influential Outliers Lying Above the Red Line

Since these outliers represent observations of large firms, and they have significant leverage (i.e., they influence the fit of a regression line), the coefficients of a regression model would change substantially depending on whether these data points were included in the analysis, or not (Sarkar et al., 2010). However, we cannot exclude such observations because they represent a feasible choice of an alliance partner for biotech firms. As is common in research involving startups, many independent variables follow a power law distribution, for which Crawford et al. (2015) emphasize the importance of incorporating data from the tails. For instance, the distribution of biotech alliance portfolio sizes shown in Figure I2a is highly skewed. Removing the largest biotech firms from the analysis, due to their classification as outliers, risks introducing reverse survivorship bias into the final dataset since these firms overwhelmingly represent cases with drug launches. In contrast, the distribution of pharma firm sizes shown in Figure I2b portrays concentrations of very small firms and very large firms, aligning with findings that the industry is an oligopolistic market (Spitz & Wickham, 2012). This precludes the identification of pharma firm outliers based on size. Thus, removing observations based on outlier detection would significantly bias any analysis by overrepresenting alliances between successful pharma firms and less successful biotech firms.



(a) (b)

Figure I2. Density Plots

I4. Test for Significance of Interactions between Predictors

We determine whether interactions between predictors improve model fit using ANOVA, to compare a logistic regression model with only main effects to models that include interaction terms. The ANOVA results (Table I4) show that adding two-way interaction terms significantly reduce the residual deviance as compared to a model with only main effects ($p < 0.05$). We observe additional improvement when the model includes all three-way interactions ($p < 0.05$). These results indicate that important higher-order interaction effects may exist and they should not all be overlooked (Lavelle-Hill et al., 2023).

	df residual	**residual deviance**	**df**	**deviance**	***p*-value**
Main effects only	7845	10256.54	NA	NA	NA
Two-way interaction effects	7800	10007.48	45	249.0613	4.68E-30
Three-way interaction effects	7680	9730.08	120	277.3961	1.74E-14

Table I4. Logistic Regression Model Fits With and Without Interactions

Overall, the tests in sections C1–C4 indicate significant violations of the prerequisite requirements for using logistic regression and call into question the reliability of interpretations from statistical inference more broadly, as well as any model that ignores interactions and nonlinear relationships between the predictors and outcome. Thus, our choice of using ML for theory building is well-grounded, as XGBoost and BART make no statistical assumptions and can automatically capture nonlinear relationships and interaction effects without the need for explicit specification (Tan & Roy, 2019). Moreover, in our main analysis (Table 3.3), we benchmark BART against a Naïve Bayes

classifier and a Bayesian Neural Network (BNN). We show that the Naïve Bayes classifier has lower out-of-sample performance as compared to BART, which, in addition to Table I4, strengthens the argument for choosing a more complex ML model. Yet, we also show that BART retains similar out-of-sample performance as compared to the BNN, indicating that relevant and interpretable knowledge can, indeed, be extracted from the ML approach. Further, using ML to obtain managerial insights from large datasets overcomes the challenge of interpreting *p*-values, which is a major issue when analyzing big data in biomedical research (Gómez-de-Mariscal et al., 2021).

Appendix J. Additional Robustness Test

To validate the robustness of our analysis, we train additional XGBoost models given their high predictive performance (Table 3.3, main text) and significantly faster computing speed compared to BART. Table J1 presents the results after modifying the maximum time span between the drug launch and alliance termination to 2-year, 3-year, and 4-year intervals. The results remain consistent across all models.

	Time horizon (around launch)	2-year	3-year	4-year
Prediction on validation data	Threshold	0.901	0.929	0.940
	PPV (positive prediction value)	0.938	0.972	0.986
	NPV (negative prediction value)	0.468	0.488	0.422
	Detection rate	0.0096	0.0213	0.0258
Prediction on held-out testing data	PPV	0.900	0.974	0.985
	NPV	0.455	0.505	0.424
	Detection rate	0.0064	0.0419	0.0260

Table J1. XGBoost Performance Varying the Time Horizon Around Drug Launch

Though unlikely, we also verify whether the results could be influenced by systematic timing errors—like the possibility that alliance termination dates have been erroneously recorded with a lag, while drug launch dates were recorded accurately. We thus train two additional models that exclude data points where the drug launch and alliance termination occur within a specified time window of each other, focusing on 30-days and 60-days time windows. The same 3-year time horizon as in the main analysis is applied to select each alliance termination–drug launch pair. Table J2 shows that the results remain consistent with our main findings. These results also corroborate the survival analysis in

Section 3.3.3 (main text), which shows that the increase in alliance terminations extends for three years after a drug launch.

Moreover, to rule out data leakage between the training, validation, and testing datasets as a potential cause of our findings, we use a temporal split. Data points corresponding to the earliest 60% of alliance terminations are assigned to the training set, the next 20% to the validation set, and the most recent 20% to the testing set. Specifically, the training data covers terminations from January 2000 to October 2008, the validation data spans from October 2008 to April 2012, and the testing data covers April 2012 to December 2016. Alliance terminations occurring after December 2016 are excluded to mitigate censoring bias since drug launches are only considered until December 2019. This ensures that data points in the testing set have the same opportunity to observe terminations before and after drug launches as those in the training and validation sets. Table J3 presents the results.

	Time window (around launch)	30 days	60 days
Prediction on validation data	Threshold	0.916	0.895
	PPV (positive prediction value)	0.987	0.952
	NPV (negative prediction value)	0.480	0.504
	Detection rate	0.0433	0.0377
Prediction on held-out testing data	PPV	0.962	0.909
	NPV	0.491	0.495
	Detection rate	0.0289	0.0336

Table J2. XGBoost Performance: No Datapoints Within a Time Window Around Drug Launch

Prediction on validation data	Threshold	0.537
	PPV (positive prediction value)	0.778
	NPV (negative prediction value)	0.513
	Detection rate	0.0129
Prediction on held-out testing data	PPV	0.6111
	NPV	0.4935
	Detection rate	0.0051

Table J3. XGBoost Performance with Temporal Train-Validation-Test Data Split

While the performance of the model in Table J3 is not as strong as those achieved with XGBoost in our main analysis, it demonstrates that predicting whether a termination occurs after a drug launch remains possible with an accuracy greater than random chance. However, analyzing the data with a temporal split introduces challenges that are not fully addressed within our data. For instance, firms

considered large in the testing dataset (2012-2016) have feature values significantly larger than those observed in the training dataset (2000-2008). Consequently, XGBoost must extrapolate beyond the training data to make predictions, leading to inaccuracies in its outputs. Although time-series models could improve predictive accuracy in such case (Ahmed et al., 2010; Masini et al., 2023), their application is beyond the scope of this study. We chose to use traditional ML models to pair them with well-established interpretation methods that allow us to investigate the factors driving model predictions.

Appendix K. Feature Reduction

To reduce the number of features (predictors) so that only important ones are in the model, we first retrain each ML model from Table 3.3 (main text) using only the important predictors identified in Figure 3.5 (main text) and test their performance on the validation and held-out test sets. As shown in Table K1, XGBoost and BART continue to perform better than non-tree-based models, while Random Forest performs significantly better on the held-out dataset than its corresponding full model. This indicates that removing unnecessary predictors reduces the Random Forest model's tendency to overfit. Among the simplified models, the forest-based ones consistently perform the best. However, their structure, which aggregates multiple decision trees, is still not easily interpretable.

We assess the feasibility of approximating an interpretable decision tree (Sagi & Rokach, 2021) based on our XGBoost model to prioritize inherently interpretable models over *post-hoc* interpretations. Our goal is to eliminate unnecessary decision rules and minimize the number of trees while optimizing for the highest PPV. XGBoost was chosen because it achieved the highest out-of-sample PPV in Table 3.3 (main text). This analysis produces a list of simplified decision trees, ranked by their effectiveness in increasing the PPV on the validation set. Figure K1 shows the tradeoff between interpretability and performance as the PPV increases as more trees are included in the model. To reach a PPV of 75%, a minimum ensemble of 20 trees is required (based on limiting the tree depth—number of splitting decisions—to 5 as increasing the tree depth is at the expense of model interpretability). As identified

in Appendix C above, our model contains nonlinear effects and high-level interactions that may not be easily captured with only a few trees. We also tested methods developed for other tree-based ensembles on our BART and Random Forest models and found similar results (Banerjee et al., 2012; Laabs et al., 2024).

		Bayesian Models			Frequentist Models			
		BART	BNN	Naïve Bayes	Random Forest	XGBoost	Ridge	Lasso
	Threshold	0.965	0.420	0.993	0.997	0.922	0.742	0.708
Prediction on validation data	**PPV (Positive Prediction Value)**	0.917	0.99	0.616	0.826	0.989	0.950	0.944
	NPV (Negative Prediction Value)	0.400	0.446	0.269	0.417	0.408	0.386	0.385
	Detection Rate	0.055	0.0005	0.598	0.123	0.068	0.014	0.012
Prediction on held-out testing data	**PPV**	**0.936**	0.545	0.604	0.868	**0.953**	0.928	0.923
	NPV	**0.414**	0.442	0.214	0.435	**0.420**	0.392	0.391
	Detection Rate	**0.067**	0.0028	0.582	0.134	**0.080**	0.0085	0.0079

Legend: Gray cells represent the best performing model from the Bayesian (left) and Frequentist (right) algorithms.

Table K1. Comparisons of Predictive Performance Results for the Simplified Models

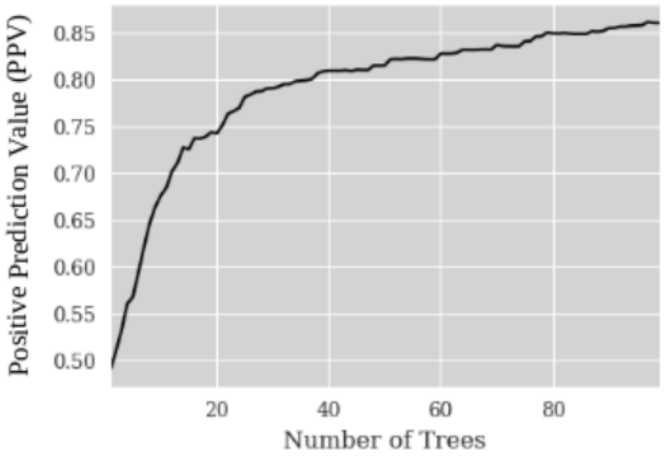


Figure K1. PPV vs. No. of Trees: Each Added Tree is the Next Most Effective in Improving Performance

Appendix L. Example of DiCE Counterfactual Generation

Predictors	Original Data Values (Risk of post-launch termination)	Counterfactual Data Values (No risk of post-launch termination) #1	#2	#3
Pharma failure experience	6	75	6	17
Pharma firm size	7	7	112	7
Pharma alliance portfolio size	33	33	33	33
Pharma R&D portfolio size	39	39	39	9
Biotech firm size	1	1	1	1
Biotech alliance portfolio size	5	5	5	5
Alliance experience	1	1	1	1
Principal investigator	1	1	1	1

Legend: Gray cells represent modified predictors.

Table L1. Three DiCE Counterfactual Scenarios Based on One Alliance from the Held-out Sample

Appendix M. Excluded Variables

Alliance duration	It measures the length of time (in days) between alliance formation and alliance termination. This variable is not used in Section 3.4 of the main text as we are interested in determining if there are *ex-ante* predictors of alliances terminated post-drug launch.
Alliance experience asymmetry	It measures whether both firms have an equal number of past alliances. We do not explicitly construct this variable since we include a separate variable for the biotech and pharma firms' alliance experience. Should experience asymmetry be relevant, our BART model would have constructed an interaction between the two *alliance experience* variables to represent this asymmetry.
International distance between partners	It was excluded as most firms have a global footprint with subsidiaries in multiple countries. Without information regarding where R&D activities take place, determining the international distance between collaborating workers is not feasible.
R&D portfolio size asymmetry	For the same reasoning as the exclusion of an alliance experience asymmetry variable, it is not measured because this information is already captured by including a portfolio size variable for both partners in the alliance.
Substitution option	It is not feasible to determine whether an immediate substitute partner is available. We assume that from the perspective of a biotech firm, a substitute pharma partner is always possible since the resources contributed by the pharma firms are less differentiated than the specialized resources of the biotech firm. From the perspective of the pharma firm, while a substitute biotech firm with identical capabilities is not likely, we assume that it is always

	possible to find alternative projects to include in their drug development pipeline.
Project expected progress	Some articles in the literature identified that the expectation that an R&D project will not progress sufficiently can lead to an alliance termination. However, measures of expectation are not observable in practice, thus we exclude this variable.
Alliance scope	It is measured as the number of project activities encompassed by the collaborative agreements. This information was not available within our dataset.
Alliance centrality	It is measured by counting the number of active alliances that each partner has in common at the time of alliance formation. As we restricted the analysis on alliances between pharma and biotech firms, there were few alliances in common.

Table M1. Variables Present in the Literature but Excluded From This Study

References

Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Econometric Reviews*, *29*(5–6), 594–621. https://doi.org/10.1080/07474938.2010.481556

Altman, N., & Krzywinski, M. (2016). Analyzing outliers: Influential or nuisance? *Nature Methods*, *13*(4), 281–282. https://doi.org/10.1038/nmeth.3812

Banerjee, M., Ding, Y., & Noone, A.-M. (2012). Identifying representative trees from ensembles. *Statistics in Medicine*, *31*(15), 1601–1616. https://doi.org/10.1002/sim.4492

Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. *Strategic Management Journal*, *42*(1), 30–57. https://doi.org/10.1002/smj.3215

Crawford, G. C., Aguinis, H., Lichtenstein, B., Davidsson, P., & McKelvey, B. (2015). Power law distributions in entrepreneurship: Implications for theory and research. *Journal of Business Venturing*, *30*(5), 696–713. https://doi.org/10.1016/j.jbusvent.2015.01.001

Gómez-de-Mariscal, E., Guerrero, V., Sneider, A., Jayatilaka, H., Phillip, J. M., Wirtz, D., & Muñoz-Barrutia, A. (2021). Use of the p-values as a size-dependent function to address practical differences when analyzing large datasets. *Scientific Reports*, *11*(1), 20942. https://doi.org/10.1038/s41598-021-00199-5

Grubbs, F. E. (1950). Sample Criteria for Testing Outlying Observations. *The Annals of Mathematical Statistics*, *21*(1), 27–58. https://doi.org/10.1214/aoms/1177729885

Imdadullah, M., Aslam, M., & Altaf, S. (2016). mctest: An R Package for Detection of Collinearity among Regressors. *The R Journal*, *8*(2), 495. https://doi.org/10.32614/RJ-2016-062

Kovàcs, P., Petres, T., & Tóth, L. (2005). A New Measure of Multicollinearity in Linear Regression Models. *International Statistical Review*, *73*(3), 405–412. https://doi.org/10.1111/j.1751-5823.2005.tb00156.x

Laabs, B.-H., Westenberger, A., & König, I. R. (2024). Identification of representative trees in random forests based on a new tree-based distance measure. *Advances in Data Analysis and Classification*, *18*(2), 363–380. https://doi.org/10.1007/s11634-023-00537-7

Lavelle-Hill, R. E., Frenzel, A. C., Goetz, T., Lichtenfeld, S., Marsh, H., Pekrun, R., Sakaki, M., Smith, G., & Murayama, K. (2023). *How the Predictors of Math Achievement Change over Time: A Longitudinal Machine Learning Approach*. https://doi.org/10.31219/osf.io/upb7f

Lundberg, I., Brand, J. E., & Jeon, N. (2022). Researcher reasoning meets computational capacity: Machine learning for social science. *Social Science Research*, *108*, 102807. https://doi.org/10.1016/j.ssresearch.2022.102807

Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of Economic Surveys*, *37*(1), 76–111. https://doi.org/10.1111/joes.12429

Rigatti, S. J. (2017). Random Forest. *Journal of Insurance Medicine*, *47*(1), 31–39. https://doi.org/10.17849/insm-47-01-31-39.1

Sagi, O., & Rokach, L. (2021). Approximating XGBoost with an interpretable decision tree. *Information Sciences*, *572*(Complete), 522–542. https://doi.org/10.1016/j.ins.2021.05.055

Sarkar, S. K., Midi, H., & Rana, S. (2010). Detection of Outliers and Influential Observations in Binary Logistic Regression: An Empirical Study. *Journal of Applied Sciences*, *11*(1), 26–35. https://doi.org/10.3923/jas.2011.26.35

Sheetal, A., Jiang, Z., & Di Milia, L. (2023). Using machine learning to analyze longitudinal data: A tutorial guide and best-practice recommendations for social science researchers. *Applied Psychology*, *72*(3), 1339–1364. https://doi.org/10.1111/apps.12435

Spitz, J., & Wickham, M. (2012). Pharmaceutical High Profits: The Value of R&D, or Oligopolistic Rents? *The American Journal of Economics and Sociology*, *71*(1), 1–36. https://doi.org/10.1111/j.1536-7150.2011.00820.x

Tan, Y. V., & Roy, J. (2019). Bayesian additive regression trees and the General BART model. *Statistics in Medicine*, *38*(25), 5048–5069. https://doi.org/10.1002/sim.8347

Supplementary Materials for Chapter 4

Appendix N. Regression model details

for event-time coefficients representing effect of discontinuations on new collaborations (mediator)

	Estimate	Std. Error	t-value	p-value
Treated × t = −5	0.036790	0.040126	0.916848	0.35925
Treated × t = −4	0.019575	0.032860	0.595699	0.55139
Treated × t = −3	−0.005635	0.028428	−0.198205	0.84289
Treated × t = −2	0.024265	0.026727	0.907888	0.36396
Treated × t = −1	−0.037049	0.022613	−1.638354	0.10139
Treated × t = 1	−0.008924	0.021875	−0.407939	0.68333
Treated × t = 2	0.039088	0.025233	1.549105	0.12139
Treated × t = 3	0.083334	0.027295	3.053106	0.00227**
Treated × t = 4	0.125745	0.030300	4.150010	0.00003 ***
Treated × t = 5	0.166035	0.034898	4.757779	0.000002 ***
Project fixed effects		Yes		
Year fixed effects		Yes		
Clustered standard errors		Yes		
Observations		12,131,493		
R^2		0.352855		
Within R^2		5.784e-4		

*** p<0.001, ** p<0.01, * p< 0.05

Table N1. Event-time Regression Model Details