CIRCLE
CENTRE FOR
INNOVATION RESEARCH

When Global Pipelines Generate Novelty: Evidence from AstraZeneca's International Academic Collaborations

Claudio Fassio, Pauline Mattsson, Aldo Geuna, Ioana Igna

Paper in Innovation Studies no. 2026/07

When Global Pipelines Generate Novelty: Evidence from AstraZeneca's International Academic Collaborations

Claudio Fassio

Department of Political Sciences, University of Pisa, Italy
GLINNOD (Global Innovation Dynamics) research lab, University of Pisa, Italy
& CIRCLE, University of Lund

Pauline Mattsson

LUSEM, School of Business and Economics, University of Lund,,Sweden
& CIRCLE, University of Lund

Aldo Geuna

Department of Cultures, Politics and Society, University of Torino

& Collegio Carlo Alberto

Ioana Igna

Copenhagen Business School, Denmark
& CIRCLE, University of Lund

JEL codes: L24; L65; O32; D83; F23; I23.

Keywords: International university-industry collaborations; Novelty; Social proximity; Geographic proximity; R&D sites

Acknowledgments

Thanks to Sotaro Shibayama for providing the novelty indicators. Thanks for the valuable suggestions from participants of the seminars in SPRU, University of Sussex; The Economic Geography and Spatial Economics seminar series, LSE; and DRUID.

Abstract

International university-industry collaboration expands access to heterogeneous knowledge environments but simultaneously raises coordination costs that may impede the deep, exploratory exchange needed to produce genuinely novel science. This paper examines the relational conditions under which geographically dispersed firm-academia collaborations generate knowledge novelty. We argue that social proximity, operationalized as prior shared institutional affiliation between AstraZeneca researchers and their academic collaborators, serves as a critical enabling mechanism, particularly under geographic distance, where institutional and cultural frictions are highest. Using a longitudinal dataset of 17,522 co-authored publications by AstraZeneca scientists from 2000 to 2020, we measure novelty through word-embedding indicators capturing both recombination novelty and element novelty. Exploiting the within-firm variation across AstraZeneca's globally distributed R&D network, we test whether the novelty-enhancing effect of social ties is stronger in international than in domestic academic collaborations. Results support an asymmetric substitution mechanism: prior social ties are positively associated with novelty specifically in international collaborations, where they compensate for the absence of spatial and institutional proximity, but not in domestic ones. These findings refine the proximity literature's substitution hypothesis and contribute to the understanding of how multinational firms organize knowledge recombination across geographically dispersed innovation networks.

1. Introduction

The geography of knowledge production has long occupied a central place in economic geography. From Marshall's (1920) industrial districts to contemporary literature on clusters, buzz, and global pipelines, scholars have argued that where knowledge is created, and who creates it, is inseparable from how innovation unfolds (Bathelt, Malmberg, and Maskell, 2004; Boschma, 2005; Fassio et al., 2023). Geographic proximity is not merely a spatial convenience but a mechanism since it enables repeated, informal, and often unplanned interactions through which tacit knowledge flows between actors -who might not otherwise- exchange it. Yet as innovation networks have become increasingly international, the assumption that proximity is primarily spatial has come under pressure. Firms routinely collaborate across national borders, connecting to distant scientific communities whose knowledge is inaccessible from within local environments. This raises a question that economic geography has not fully resolved: in the absence of spatial proximity, what enables actors to move beyond information exchange and engage in genuinely novel collaborative knowledge creation?

This paper addresses that question in the context of the pharmaceutical industry, where the tension between proximity, distance, and knowledge novelty is particularly important. Declining R&D productivity, the exhaustion of established molecular platforms, and successive waves of patent expirations dismantled the closed, internally driven model of drug discovery that defined the industry through the 20th century (Pisano, 2006; Schuhmacher et al., 2016). The pipeline crises of the 2000s and 2010s were not temporary disruptions but signaled a structural failure of the blockbuster model, forcing firms to fundamentally reconstitute their knowledge-sourcing strategies. In response, major pharmaceutical corporations turned outward toward universities and public research organizations (Owen-Smith and Powell, 2004; Bignami & Mattsson, 2019). The knowledge at stake in these collaborations is precisely the kind that geographic proximity was theorized to transmit: namely tacit, context-dependent, and rooted in ongoing experimental practice, rather than codified (Gertler, 2003; Balland et al., 2015; Broekel, 2015).

Economic geography has long argued that spatial proximity enables tacit knowledge transfer by reducing the costs of face-to-face interaction, supporting shared frameworks, and creating conditions for serendipitous encounters that underpin knowledge recombination (Storper and Venables, 2004; Boschma, 2005). For pharmaceutical firms, proximity to leading academic

science, through co-location near research universities or recruitment of star scientists, provides access to knowledge that cannot be reliably sourced at a distance (Zucker and Darby, 1996; Abramovsky et al., 2011; Melnychuk et al., 2021). Yet, sustained geographic proximity carries costs: repeated reliance on local knowledge environments can lead to familiar, overlapping, and redundant knowledge, limiting the production of genuinely new scientific combinations. This is what Boschma (2005) defined as the proximity trap, where co-location that facilitates exchange can, over time, generate cognitive lock-in as actors converge on shared frameworks and assumptions. Local buzz is rich but not diverse, and diversity, the combination of previously distant knowledge elements, lies at the core of novelty (Fleming, 2001; Uzzi et al., 2013).

This tension has motivated pharmaceutical firms to increasingly pursue international research collaborations, connecting to academic communities with different scientific trajectories, therapeutic traditions, and access to unique patient populations and disease environments (Bathelt et al., 2004; Bignami et al., 2020). While international collaboration expands the knowledge landscape for novel combinations, it removes the spatial infrastructure essential for tacit knowledge exchange. Distance introduces institutional friction, cultural differences, and practical barriers to coordinating complex scientific work, creating a central tension: the geographic configurations that maximize knowledge diversity are precisely those that complicate the deep, exploratory exchange needed to make diverse knowledge combinable. Empirical evidence from Wagner et al. (2019) suggests that international collaborations (not limited to academia-industry collaboration), despite their potential for diversity, tend to produce less novel and more conventional knowledge combinations due to transaction costs and communication barriers. This counterintuitive finding highlights the need to examine not just the presence of international collaboration but the relational mechanisms that may mitigate these barriers.

The proximity literature addresses this tension with the substitution hypothesis, where non-geographic forms of proximity, cognitive, organizational, institutional, and social, can partially compensate for geographic distance by providing alternative coordination mechanisms (Boschma, 2005; Huber, 2012; Balland, Boschma, and Frenken, 2015). However, empirical tests of this hypothesis have largely relied on outcomes that cannot distinguish between knowledge transfer and knowledge creation. Measures such as collaboration tie formation, co-

publications, citation counts, or patent output capture whether proximity enabled exchange, but not what kind of knowledge it produced. A co-authored paper evidences collaboration, but not necessarily the generation of genuinely novel knowledge. This difference matters because the conditions that bring actors together are not the same as those that generate novelty: proximity may bring actors together while constraining them to familiar knowledge spaces, and collaboration may be productive in volume but incremental in knowledge terms. Thus, the question of whether proximity shapes the novelty of collaboration outputs has remained largely unasked in economic geography.

This study addresses this gap by examining how social proximity, operationalized as prior shared institutional affiliation between firm-based and academic researchers, conditions the novelty of knowledge produced in co-publications, and whether that effect differs between domestic and international collaborations. Social proximity, in Boschma's (2005) framework, refers to the degree to which actors are embedded in relationships of trust, reciprocity, and shared relational history at the interpersonal level. While cognitive and institutional proximity shape the capacity for mutual understanding, social proximity shapes the willingness to move beyond familiar knowledge, to share ideas before they are fully formed, and to embrace the uncertainty that genuine intellectual exploration entails.

In international collaboration, where geographic and institutional distances are large and the potential for knowledge diversity is greatest, we argue that social proximity -i.e. social ties- functions as a relational enabling condition. It does not substitute for geography, by replicating co-location's effects, but rather provides the trust infrastructure needed to transform geographic diversity into novel knowledge combinations. Recent work supports this argument, Tu (2024) shows that the interaction between international openness and tie strength is critical for knowledge creation, while Tzabbar and Vestal (2015) demonstrate that relational strength can bridge the social distance in geographically distributed R&D teams. In contrast, in domestic collaboration, where institutional, cognitive, and cultural proximities already provide coordination capacity, the marginal contribution of social ties to novelty production is substantially lower.

We test this argument using a longitudinal dataset of co-authored publications between AstraZeneca and academic collaborators from 2000 to 2020. Novelty is measured through

word-embedding-based indicators that capture both recombination novelty (the extent to which a paper draws on unusual combinations of prior knowledge) and element novelty (the extent to which it introduces knowledge elements not previously present in the literature) (Shibayama et al., 2021; Yin et al., 2023). To measure social proximity, we use publication data and check whether in past papers AstraZeneca authors used to have the same university affiliation of their current academic coauthors. The single-firm design is deliberate, since by restricting the sample to one multinational corporation, we hold firm-level strategy, absorptive capacity, organizational culture, and resource endowments constant, isolating proximity effects from the firm-level heterogeneity.

The paper makes three contributions to economic geography. First, it shifts the dependent variable in proximity research from collaborative activity to knowledge novelty, arguing that the field's dominant outcome measures cannot distinguish between knowledge transfer and knowledge creation and thereby distort the geographic conditions under which innovation produces genuinely new scientific combinations. Second, it provides empirical evidence that the effect of social proximity on novelty is contingent on geographic distance. Social ties enhance novelty specifically in international collaboration, where they bridge actors separated by large institutional and cultural distances, but not in national collaboration, where background proximities already provide the necessary coordination capacity. This asymmetry refines the substitution hypothesis by showing that non-geographic proximity does not operate uniformly but is activated by geographic distance. Third, by situating this analysis within a single globally active pharmaceutical firm, the paper contributes to a growing literature on how multinational corporations organize knowledge recombination across geographically dispersed innovation networks, extending that literature's attention beyond network formation to the nature of knowledge produced within those networks.

The paper proceeds as follows. Section 2 develops the theoretical framework, positioning the argument within the proximity literature and elaborating the mechanisms linking geographic distance, social proximity, and knowledge novelty. Section 3 describes the case study, AstraZeneca. Section 4 describes the main sources of data for the empirical analysis. Section 5 illustrates the main strategies adopted by AstraZeneca to source external knowledge. Section 6 presents the main variables and the estimation strategy. Section 7 presents the empirical results. Section 8 discusses the findings and their implications for economic geography's

understanding of proximity, knowledge creation, and the geography of innovation in multinational firms.

2. Literature Review

2.1. The Strategic Turn Toward Novelty in Multinational Pharmaceutical Firms

Novelty is a cornerstone of scientific progress and a critical driver of long-term competitive advantage in knowledge-intensive industries (Nelson and Winter, 1982; Teece et al., 1997; Uzzi et al., 2013). In pharmaceuticals, this is particularly important since first-in-class drugs that introduce new mechanisms of action can redefine treatment paradigms, create temporary monopolies, and generate substantial economic benefits. In contrast, incremental innovations in crowded therapeutic areas face intensifying generic competition, pricing pressure, and regulatory scrutiny.

Historically, large pharmaceutical firms relied on vertically integrated R&D models, where innovation was cumulative and trajectory-bound, building on proprietary chemical platforms and established therapeutic domains. Competitive advantage stemmed from scale, internal capabilities, and strong intellectual property protection. However, beginning in the 1980s and accelerating through the 1990s, this model was destabilized by advances in molecular biology, genomics, and biotechnology, which transformed the scientific foundations of drug discovery and increased dependence on frontier academic research. Simultaneously, R&D expenditures soared while productivity declined, a dynamic widely described as the pharmaceutical productivity crisis (Pisano, 2006; Schuhmacher et al., 2016). The sustainability of blockbuster pipelines weakened, regulatory requirements intensified, and therapeutic innovation increasingly required the identification of entirely new biological targets rather than the refinement of existing compounds. As a result of these structural shifts, novelty moved from an epistemic ideal to a core strategic concern.

Novelty in pharmaceuticals is often recombinant, emerging from the combination of previously unconnected or cognitively distant knowledge elements (Fleming, 2001; Azoulay et al., 2011). Empirical research demonstrates that atypical knowledge combinations are more likely to generate high-impact outcomes (Uzzi et al., 2013; Wang et al., 2017; Foster et al., 2015). In science-based sectors like pharmaceuticals, innovation frequently depends on integrating

insights across biological, chemical, computational, and clinical domains. Novelty thus commonly takes the form of new knowledge combinations rather than entirely isolated discoveries, a form of creativity measurable through word-embedding approaches that capture the unusualness of knowledge element combinations in published research (Shibayama et al., 2021; Yin et al., 2023).

Yet, established firms often struggle to generate exploratory recombination internally. Research on incumbent adaptation highlights how accumulated routines, commitments, and capability structures can bias firms toward competence-deepening trajectories, limiting their responsiveness to technological shifts (Eggers, 2018). Incumbents tend to rely on familiar knowledge neighborhoods and established problem-solving techniques, making sustained broad or distant search difficult to maintain alongside exploitation demands (Audretsch et al., 2024; Li, 2023). Chen et al. (2025) provides further evidence for this dynamic, finding that repeated collaboration tends to reduce novelty, disruptiveness, and interdisciplinarity, while new collaborations facilitate the integration of unconventional knowledge. In pharmaceuticals, case studies demonstrate that established competencies can simultaneously facilitate and limit firms' efforts to explore radically new technological directions (Phillips and Pandza, 2023).

To overcome these constraints, pharmaceutical firms have increasingly turned to externally networked and geographically dispersed collaboration, particularly with universities. Cross-national R&D partnerships provide access to heterogeneous scientific communities, institutional settings, and region-specific competencies (Cantwell and Mudambi, 2005). Universities and public research organizations play a distinctive role as generators of foundational scientific knowledge, early-stage discoveries, and specialized expertise that complement firms' applied R&D capabilities (Zucker et al. 2002; Perkmann et al., 2013). In pharmaceuticals, where drug discovery builds on advances in molecular biology, genomics, and translational medicine, academic partnerships are often central to identifying new therapeutic targets and mechanisms of action (Owen-Smith and Powell, 2004; Bignami & Mattsson, 2019). Moreover, differences in regulatory regimes, clinical infrastructures, and patient populations across countries make international academic collaboration strategically important for testing and validating new therapies across institutional contexts (Bignami et al., 2020).

However, the act of forming external collaborations does not guarantee novelty (Wagner et al., 2019). A fundamental and underexplored question is therefore under which relational and geographic conditions firm-academia collaborations produce genuinely novel scientific outputs. This question motivates the analysis that follows, as it requires engaging with both the geography of collaboration and the relational mechanisms that shape what geographically dispersed partnerships can produce.

2.2. The Geography of Collaboration and Knowledge Diversity

Research in economic geography has long demonstrated that innovation is spatially structured. Early empirical studies showed that knowledge spillovers decay with geographic distance, highlighting the localized character of innovative activity (Jaffe et al., 1993; Audretsch and Feldman, 1996). The underlying mechanism is that spatial proximity facilitates the transfer of tacit knowledge, knowledge that is context-dependent, experiential, and most effectively transmitted through repeated face-to-face interaction (Polanyi, 1966; Gertler, 2003; Storper and Venables, 2004). This is particularly consequential in pharmaceutical R&D, where basic biomedical knowledge is often embedded in laboratory practices and unpublished experimental routines rather than codifiable results (Owen-Smith and Powell, 2004; Belderos et al., 2021).

Yet, geographic proximity is not without its limits. Excessive embeddedness in local knowledge environments may actively constrain novelty. Firms embedded in dense local networks may rely on familiar partners and established trajectories, reducing exposure to diverse perspectives (Grabher, 1993; Garcia Martinez et al., 2025). Boschma (2005) coined this the “proximity trap,” as the co-location that facilitates tacit knowledge transfer can, over time, produce cognitive lock-in. Actors may converge on shared frameworks, methods, and assumptions, limiting their ability to generate novel knowledge. Since novelty in science-based industries most commonly arises from the combination of previously distant knowledge elements (Fleming, 2001; Azoulay et al., 2011; Shibayama et al., 2021), geographic proximity that fosters cognitive convergence risks pushing collaborations toward competent but incremental outcomes. As Nooteboom et al. (2007) demonstrate, there is an optimal cognitive distance for learning and innovation where too little distance stifles novelty, because actors share the same knowledge base, and too much distance impedes effective communication and integration. Sustained geographic proximity tends to erode cognitive distance over time, pushing collaboration toward the former condition.

In response to these limitations, later work extended the perspective by conceptualizing innovation as shaped by multiple proximity dimensions. Boschma (2005) distinguishes geographic, cognitive, organizational, social, and institutional proximity, arguing that innovation outcomes depend on their configuration rather than any single dimension. Later contributions emphasize that proximities are dynamic and relational, evolving through collaboration processes (Balland et al., 2015; Huber, 2012; Hansen, 2012). Recent empirical research confirms that intersecting proximity dimensions jointly shape innovation performance (Shkolnykova, 2023; Garcia Martinez et al., 2025), and that proximity configurations are particularly relevant for recombinant innovation at the dyadic level (Nan et al., 2024).

Despite this theoretical richness, existing empirical work in the proximity literature do not distinguish between knowledge transfer and knowledge creation**.** By relying on co-publications, patents, or collaboration ties as outcome measures, studies demonstrate that proximity enables exchange without addressing what kind of knowledge that exchange produces (Knoben and Oerlemans, 2006; Hansen, 2015). A co-publication is evidence that collaboration occurred, but it is not evidence that the collaboration generated novel knowledge combinations. A paper can emerge from a geographically proximate, institutionally aligned, and cognitively close partnership and still draw on entirely familiar knowledge elements. Conversely, a collaboration bridging distant knowledge environments may produce novel recombinations even if its collaboration volume is modest. The proximity literature, by focusing on the quantity and probability of collaborative output, has therefore left unasked the question of whether proximity shapes the novelty of what collaboration produces.

To overcome these constraints, firms need access to distant knowledge, not just local knowledge. Multinational corporations increasingly participate in geographically dispersed innovation networks. Bathelt et al. (2004) conceptualize this as the combination of "local buzz" and "global pipelines." While local networks provide intensive interaction and trust-based collaboration, global pipelines connect firms to distant knowledge environments, introducing novel ideas and specialized expertise unavailable within local milieus. Recent studies emphasize that multi-scalar collaboration networks have become central to contemporary knowledge-sourcing strategies (Audretsch et al., 2024; Frigon, 2024). Empirical studies further show that innovation outcomes depend on how firms balance regional embeddedness, through,

for example, subsidiaries, with cross-border connections (Galaso et al., 2020; Belderbos et al., 2021). This dynamic aligns with the broader substitution logic in the proximity literature, which argues that one form of proximity can compensate for the absence of another (Huber, 2012; Hansen, 2015).

However, geographic dispersion introduces a tension between diversity and integration. Collaborating across borders expands access to diverse knowledge bases, specialist expertise, and research communities operating on different scientific trajectories, all of which increases the potential for novel recombination. Yet spatial and institutional distance increases coordination costs, creates interpretative challenges, and complicates the transfer of tacit knowledge, challenges that co-location infrastructure typically mitigates (Gertler, 2003; Knoben and Oerlemans, 2006). Contrary to the assumption that international collaboration inherently enhances novelty, Wagner et al. (2019) find that such collaborations often produce less novel and more conventional knowledge combinations due to transaction costs and communication barriers. Diversity broadens the opportunity space for recombination, but whether it translates into genuinely novel outcomes depends on relational and institutional conditions that facilitate effective integration across spatial boundaries. This paper explores this central tension, as the same geographic configurations that expand knowledge diversity also hinder the deep, tacit exchanges required to combine that knowledge effectively. Resolving this tension requires shifting focus from geography to the relational mechanisms that determine the outcomes of internationally dispersed collaboration.

2.3. Social Proximity and Relational Mechanisms Under Distance

If geographic dispersion expands access to diverse knowledge while simultaneously increasing integration challenges, relational mechanisms become critical for enabling effective collaboration across distance. While the proximity literature identifies social proximity as one such mechanism, its role in shaping the novelty of knowledge produced, rather than merely the probability of collaboration, has received limited theoretical and empirical attention.

Social proximity, as defined by Boschma (2005), refers to interpersonal relationships embedded in trust, shared experience, and relational history. Grounded in Granovetter's (1985) insight that economic action is embedded in social relations, social proximity facilitates open, exploratory communication, the sharing of unpublished ideas, and tolerance for the uncertainty

inherent in novel knowledge combinations. This embeddedness fosters trust, shared interpretive frames, and reputational monitoring, all of which reduce the coordination costs that are particularly high when knowledge is tacit, complex, and not yet codifiable (Reagans and McEvily, 2003; Ooms et al., 2018).

However, the mechanism through which social proximity shapes novelty is distinct from its role in enabling collaboration formation. While collaboration can occur between actors with low social proximity, particularly when mandated by institutional or organizational arrangements, the depth and exploratory character of knowledge exchange may be limited. Social proximity enables actors to venture beyond familiar knowledge by providing relational safety to share speculative, uncertain, and not-yet-validated ideas. As Malmberg and Maskell (2006) note, trust often exists in local milieus as an inherited feature of co-location, but prior social ties can recreate this relational infrastructure across geographic distance, enabling the kind of deep, exploratory exchange that "local buzz" typically provides.

Proximity dimensions are not static but co-evolve through collaboration processes (Balland et al., 2015; Broekel, 2015). Prior social ties, built through shared institutional experiences earlier in collaborators' careers, represent accumulated relational capital that can be reactivated in new collaborative contexts. This perspective is supported by Dahlander and McFarland (2013), who demonstrate how tie persistence in research collaborations depends on the strength and history of relationships, and Rost (2011), who highlights the critical role of strong ties, not just weak ones, in fostering innovation, particularly in contexts requiring deep trust and coordination. Further evidence for this mechanism comes from Roth and Mattes (2023), who show that relational infrastructures sustain interaction across spatial separation, and Fassio, Geuna, and Rossi (2023), who demonstrate that firms' engagement with foreign universities is mediated by inventors' relational trajectories, underscoring the individual-level networks enabling cross-border knowledge exchange in multinational R&D.

Yet, social proximity is not uniformly beneficial for novelty. Chen et al. (2025) emphasizes that prolonged collaboration can reduce novelty by reinforcing familiar knowledge pathways, while new or weaker ties may be more effective in integrating unconventional knowledge. This suggests that the novelty-enhancing effects of social proximity may depend on the balance between repeated interaction and exposure to new perspectives. Uzzi (1997) warns that over-

embeddedness, where social ties become too dense and inward-looking, can generate knowledge redundancy and network myopia, reinforcing homophily and limiting exposure to diverse perspectives (Inkpen and Tsang, 2005). When social ties connect actors who already share similar knowledge bases and institutional frameworks, they may deepen cognitive convergence rather than bridge cognitive distance. Boschma (2005) himself notes that social embeddedness in industrial districts can decrease knowledge heterogeneity as firms sharing social proximity increasingly adopt common assumptions and cognitive frameworks. This represents the social version of the geographic proximity trap, too much social proximity, like too much geographic proximity, can produce redundancy rather than novelty.

2.4. Hypotheses

The preceding discussion highlights a structural tension in the geography of innovation. International collaboration expands access to heterogeneous scientific environments and enlarges the opportunity space for novel knowledge recombination. At the same time, spatial and institutional dispersion introduce coordination and interpretative frictions. Geographic diversity therefore does not automatically translate into novel outcomes, its effects are likely to depend on the relational conditions under which collaboration unfolds. Existing research often emphasizes the diversity-enhancing role of international collaboration. However, this perspective remains incomplete if it neglects the relational mechanisms that shape how geographically dispersed actors interact. Rather than asking whether international collaboration increases novelty per se, this study examines whether relational embeddedness conditions the association between cross-border collaboration and novel outcomes.

Social proximity constitutes one such relational condition. Prior ties between collaborators, such as shared institutional affiliation at earlier career stages, can foster trust, reduce uncertainty, and support sustained interaction across organizational boundaries. These mechanisms are especially salient when collaborators operate in different national and institutional contexts, where shared norms and opportunities for face-to-face interaction are more limited. We therefore expect prior social ties to be positively associated with novelty in collaborative research. However, we do not anticipate this association to be uniform across geographic configurations. In nationally bounded collaborations, shared institutional environments and regulatory frameworks may already reduce uncertainty and facilitate interaction. In international collaborations, by contrast, spatial and institutional distance heighten coordination challenges and increase the relevance of relational embeddedness. In

international collaboration contexts, prior work experience at a partner institution by a member of the collaborating team may help mitigate the challenges associated with spatial and institutional distance, thereby facilitating coordination and interaction across organizational boundaries

Accordingly, we propose:

H1: Collaborations involving prior social ties between partners are associated with higher levels of novelty.

H2: Social ties are positively associated with novelty in international collaborations, but not in domestic ones where other proximity dimensions already provide sufficient coordination capacity.

3. The case study: AstraZeneca

This study focuses on AstraZeneca to investigate how multinational pharmaceutical firms establish collaborations with international academics and how such collaborations relate to scientific novelty. The pharmaceutical sector provides a particularly relevant empirical setting due to its reliance on frontier science, the complexity and uncertainty of drug discovery, and the increasing globalization of R&D activities (Belderbos et al., 2010; Narula & Santangelo, 2012).

AstraZeneca constitutes a suitable case for several reasons. First, it is one of the world's largest research-intensive pharmaceutical firms, employing approximately 90,000 people globally and ranking among the leading companies by market capitalization. Its scale and scientific breadth ensure that it operates across multiple therapeutic areas and institutional environments, making it representative of large multinational pharmaceutical corporations. Second, the firm exhibits substantial internal heterogeneity in its global R&D network, with research centers differing in size, specialization, and degree of academic engagement. This variation enables the examination of collaboration patterns and novelty within a shared corporate context. Third, the period under study coincides with significant strategic and organizational transformation, creating a dynamic setting in which shifts in knowledge sourcing can be observed.

A particularly important turning point occurred in the early 2010s. A wave of patent expirations weakened AstraZeneca's core revenue base, intensifying pressure to renew its drug portfolio. In 2012, Pascal Soriot was appointed Chief Executive Officer, succeeding David Brennan and initiated a strategic reorientation centered on rebuilding the pipeline and strengthening science-driven innovation. In 2014, Pfizer launched an attempted takeover bid for the company. Although the acquisition ultimately failed, the episode heightened organizational uncertainty and accelerated strategic reassessment. Together, these developments marked a period of turbulence that placed exploratory research and external collaboration at the center of the firm's renewal strategy.

Around the same time AstraZeneca reorganized its R&D structure around major strategic hubs while closing several sites, including Lund and Södertälje in Sweden in 2013, where more than 1,000 research positions were affected. These closures reflect broader industry trends toward concentrating R&D in fewer, larger and more strategically aligned centers (Danzon et al., 2005). The current R&D organization centers on strategic hubs in Cambridge (UK) Mölndal (Sweden) and Gaithersburg (US), complemented by additional global centers located in major scientific regions such as Boston, Shanghai and Beijing. This distributed structure embeds the firm in multiple regional innovation systems and heterogeneous scientific communities.

Table 1 provides a detailed overview of AstraZeneca's R&D sites from 2000 to 2020, illustrating the dynamic evolution of its global research footprint and highlighting the years during which each center was active. Out of the 18 centers that have been active in the period considered, only Boston/Waltham (US), Macclesfield/Alderley Park (UK) and Mölndal (Sweden) were always in operation. The two strategic R&D centers of Cambridge (UK) and Gaithersburg, (US) were created in 2007 with a major restructuring of all R&D activities. At the end of the period there were 10 active R&D sites, the three main R&D centers of Molndal in Sweden, Cambridge and Macclsefield in the UK, four new centers in the US (Boston, Gaithersburg, Mountain View and New York), and three other small international centers in Japan (Osaka), China (Shanghai) and Poland (Warsaw).

INSERT TABLES 1 ABOUT HERE

Beyond geographic restructuring, AstraZeneca also shifted its therapeutic orientation. While earlier portfolios were diversified across primary-care areas, the 2010s saw a pronounced reorientation toward science-intensive domains, particularly oncology, immunology and

molecularly targeted therapies. These fields are characterized by rapidly evolving biological knowledge, high uncertainty and strong dependence on academic research in molecular biology, genomics, and translational medicine. Entering and expanding within such domains increases reliance on frontier science and heightens the importance of accessing geographically dispersed and cognitively diverse knowledge communities. In this sense. AstraZeneca's therapeutic shift reinforces the relevance of international academic collaboration for exploratory and potentially novel research.

AstraZeneca's global R&D footprint exhibits substantial heterogeneity across locations in terms of publication intensity and engagement with academic partners. As described in the data section, sites differ in their degree of international academic collaboration and internal research orientation. This differentiated R&D geography provides a structured yet internally diverse empirical setting in which to examine how international collaboration and relational configurations are associated with the production of novelty in multinational pharmaceutical research.

4. Sources of data

Annual reports

Spanning 2000 to 2020. For each year, we identified the therapeutic areas in which the company was active by categorizing pipeline components across pre-clinical and clinical phases, tracking products already on the market, and recording R&D expenditure by area. All therapeutic areas in which AstraZeneca held any degree of activity were included as controls.

The annual reports also served as the basis for identifying institutional agreements signed with academic partners in each given year. This was supplemented by deal-level data from IQVIA Pharma Deals (formerly IMS Health), which captures agreements between pharmaceutical companies and a range of external organizations, including deal type and counterparty. Finally, the reports were used to determine the locations of AstraZeneca's R&D centers across the study period.

Interviews

Prior to launching our empirical study, we conducted four interviews with directors and vice presidents for science relations and policy to deepen our understanding of AstraZeneca's

historical and current pipeline. Our questions focused on strategic decision-making around therapeutic areas and R&D site selection.

Publication data

We downloaded data from Scopus about all the papers that were published between 2000 and 2020, in which at least one of the authors had an Astra-Zeneca affiliation or an acquired company[1]. Through this procedure we were able to identify 17,522 papers. We recovered information about all the 59,021 authors included in such papers. For the 12,789 authors with an AstraZeneca affiliation, using their Scopus ID, we downloaded information in Scopus about all the papers that they published throughout their career (hence also before 2000), regardless of their affiliation. In such a way we could retrieve information also about the publishing activity of AstraZeneca authors before the chosen period of 2000-2020 to track their professional history. We can see their history of affiliations and the location of the affiliation. We can also know if the coauthors with whom they publish in the period 2000-2020 have coauthored with them in the past or not. For further details, including how we managed that some AZ authors had more than one affiliation, see Appendix A.

Affiliations and geographic location

Using both the affiliation name listed in the papers and the affiliation ID provided by Scopus we recovered the institution, city and country of the affiliation of each author (both Astra Zeneca authors and their coauthors). In the case of AstraZeneca, we realized that Scopus wrongly assigned most of AstraZeneca authors to the sites of Cambridge in UK, or to Sodertalje in Sweden, because it often automatically assigns AZ author to the Headquarters of the company. Therefore, we had to manually check for cases in which the place of the AstraZeneca affiliation listed in the papers was different from the one that Scopus attributed through the Scopus affiliation ID. Whenever the two affiliations differed, we attributed the location listed in the original printed papers. Out of 12,789 AZ authors we had to correct the city of affiliation for 9,134 of them, that is for more than 70% of them. This allowed us to identify the full network of R&D subsidiaries by Astra Zeneca, where their scientists were working. As a further double check, we used information from the annual reports of AstraZeneca to identify the R&D centers of the company and their evolution throughout the period. This is because, as

[1] We checked which companies were acquired by AstraZeneca during the years 2000-2020. This was the case for example of MedImmune in 2007. In the Appendix A we explain in detail how we managed the inclusion of papers in which the affiliation was of one of the acquired companies.

explained above, some of the R&D centers were newly created during the period 2000-2020 and some others were closed.

For the 46,232 non-AstraZeneca authors included in the 17,522 papers, we classified their affiliations as either academic researchers, researchers from public research organizations, researchers from other companies or researchers working in hospitals. Using the name of the city we were also able to find the coordinates of the cities of affiliation of both AstraZeneca authors and their co-authors and calculate the geographical distance between each AstraZeneca R&D sites and the academic coauthors affiliation.

National and international collaborations with academic researchers

We are interested in the papers that involve academic coauthors, distinguishing between the collaborations that involve academics located in the same country as the AZ researchers (*national academic collaborations*) and those that engage academics located in universities in other countries than the AZ researchers (*international academic collaborations*). To classify the academic collaborations between national and international ones, we take advantage of the location of the institution listed in the affiliation of each author. Out of the 17,522 papers, the number of papers with academic coauthors is 8,568. Among those, 5,466 papers are with international academic coauthors. To simplify the classification of international and national collaborations we only consider papers in which all AstraZeneca authors have affiliations in the same country.[2] We also make sure that the publications are actual papers and not research notes or commentaries. After cleaning for this, our final sample is made of 5,363 papers: of which 2,806 are international collaborations and 2,557 national collaborations.

On average AstraZeneca researchers publish 25 papers each throughout their careers, although the average does not provide a clear picture, since it is a very skewed distribution. Indeed, the median is 6 papers per AstraZeneca author and then there are some extremely productive authors with more than 100 publications (approximately 5% of the sample). In Appendix A in table A3 we also show that the share of academic publications has increased between 2000 to 2020 from less than 40% of all 17,522 papers in the first three years to more than 50% in the

[2] Without this restriction, classifying a collaboration as domestic or foreign would become ambiguous. Consider, for example, a paper co-authored by AstraZeneca researchers based in both the UK and the US, an academic collaborator at a US university would be classified as foreign relative to the UK-based AZ authors, yet domestic relative to their US-based colleagues. Applying the restriction has limited impact on the overall dataset, as in more than 80% of the 17,522 papers examined, all AstraZeneca authors are in the same country.

last 5 years. Among academic papers the share of international ones, with at least one author located in a different country than the rest of the coauthors team, increased from 40%-50% in the first years of the period to more than 70% in the last years (Table A4). The academic co-authors of AstraZeneca authors are located mostly in UK, US and Sweden: these three countries account for more than 60% of the total. However, as shown in table A5, also Germany, Canada, China, Japan and Australia account each for at least 2%, showing a true global nature of the academic collaborations of AstraZeneca researchers.

5. AZ R&D geography and external sourcing strategy

5.1 Sourcing knowledge from other companies: acquisition and partnerships

AstraZeneca has adopted a diversified approach to knowledge sourcing strategies. First, as shown in Table 2 the company realized a few important acquisitions of biotech companies such as Cambridge Antibody Technology that laid the foundation for the takeover of MedImmune in 2007 which was important for strengthening the biologics arm which was central to the oncology and immunology pipelines.

Second, over time, AstraZeneca has increasingly emphasized international collaboration and open innovation to complement its internal capabilities. The company's strategic shift toward external partnerships is driven by the recognition that breakthrough discoveries often originate outside firm boundaries. One example of such a collaboration is the acquisition of Amylin Pharmaceuticals or the key alliance with Ionis Pharmaceuticals to strengthen antisense therapies (see Table 3).

INSERT TABLES 2 AND 3 ABOUT HERE

5.2 AZ R&D geography and collaborations with universities

An important source of external knowledge comes from collaboration with universities. As often stated in the AstraZeneca's Annual reports by leveraging global scientific networks and collaborating with universities and research institutes, the company aims to enhance its absorptive capacity, accelerate innovation cycles, and manage the balance between exploration

of novel knowledge and exploitation of existing competences, core themes that are central to this study.

Table 4 provides the publishing activity of the different R&D centers in the period 2000-2020, using Scopus data. Most publications by AZ researchers originated from three primary centers: Cambridge (UK), Macclesfield (UK), and Mölndal (Sweden). These three sites accounted for approximately 60% of non-academic and 66% of academic publications, with at least one author affiliated with them. Another notable center, Södertälje (near Stockholm), was active until its closure in 2012. Additionally, several other centers demonstrated significant publishing activity, including Loughborough (UK) and three US-based locations: Boston, Wilmington (Delaware), and Gaithersburg (Maryland). In Sweden, the Lund R&D center was also a key hub for publications until 2012.

Tables 4 and 5 use the number of publications and the share of publications with academic coauthors (national or international) in the period 2000-2020 to provide a first illustration of how each AZ R&D center collaborates with academic scholars. We distinguish between large R&D centers, as measured by the total number of publications (more than 3,000 publications between 2000 and 2020), such as Cambridge UK or Mölndal; middle-sized ones (from 600 to 3,000 publications), such as Wilmington in Delaware, or Loughborough in the UK; and small centers (less than 600 publications), such as the ones in Japan (Osaka) or in China (Shanghai).[3]

Among the large R&D sites we find two different typologies. The first typology, which we label as "*Global Academic Hubs*", are highly engaged with academics (at least 35% of total publications produced by AZ researchers in the center) and the share of publications with at least one academic that located in a different country (international academics) is also relatively high (at least 50% of academic publications). Cambridge (UK), Mölndal and Södertälje (Sweden) exemplify this type, pointing to an international network of academics collaborating with AstraZeneca researchers at these locations. The second type, which we label "*Corporate R&D Hub*", is illustrated by Macclesfield (UK). Here, most publications do not involve academic co-authors, and international academic collaborations are particularly rare. The site remains a significant knowledge producer in terms of publication output but is oriented towards intra-firm knowledge flows or collaborations with other types of partners such as hospitals and

[3] The size classification does not change even if we account for the number of years that the R&D centers have been active (using the info from Table 1), hence measuring the number of publications per year of activity.

other firms. This more limited engagement with the academic community is also reflected in the geographic scope of its collaborative network: academic partnerships involving Macclesfield tend to be predominantly with UK-based universities rather than international institutions.

INSERT TABLES 4 AND 5 ABOUT HERE

Among mid-sized R&D sites, we identify a first type, "*Global Academic Centre*", which is similar to the characteristics of a Global Academic Hub in terms of strong engagement with academics, particularly international ones, but with a somewhat lower publication output. Gaithersburg (US) and Lund (Sweden) are representative examples.

Two further typologies emerge among mid-sized sites. The first, which we label "*International R&D Centre*", includes sites such as Boston and Waltham, characterized by relatively low overall engagement with academics but a disproportionately high share of international academic collaborators. This pattern may reflect a limited embeddedness in the local academic community. When these sites do draw on academic knowledge, they tend to source it from universities abroad, where AstraZeneca has more established relationships. The second, which we term "*Corporate R&D Centre*", is exemplified by Loughborough (UK) and Wilmington (US). These sites show both low levels of academic co-authorship and a limited international dimension to what collaborative activity does exist, they engage rarely with universities in general, and even more rarely with foreign ones.

When it comes to the small R&D centers, we notice some relevant differences: some centers have little focus on academic collaborations: this is the case of Bangalore in India, Rheims in France or Montreal in Canada: we label these "*Local R&D outpost*": these centers rely mostly on intra-firm knowledge and have little international academic collaborations. The other typology of small R&D site is what we define "*Outpost for local academic search*": this is the case for Osaka and Shanghai, where there is a high level of engagement with academics, but most collaborations are with local/national academics. This hints at a strategy of local sourcing for academic knowledge.

The distinction between the 7 typologies of R&D sites (Global Academic Hubs, Corporate R&D Hubs, International Academic Centers, International R&D centers, Corporate R&D centers, Local R&D Outposts and Outpost for Local Academic Search) reveal the considerable heterogeneity within AstraZeneca's R&D network in terms of engagement with the academic community, and with international academics in particular. Not all sites interact with universities to the same degree, and only a subset maintain a substantial share of international collaborations. This heterogeneity suggests that different nodes in the company's R&D network serve distinct purposes when it comes to accessing and sourcing academic knowledge from abroad.

By the end of the observed period, AstraZeneca's R&D network had taken on a clearly defined structure. The company maintained three world-class strategic R&D centers: the Discovery Centre (DISC) in Cambridge (UK), Gaithersburg (Maryland, US), and Gothenburg (Sweden) as its primary hubs of scientific activity. These were complemented by a set of medium-sized centers in the US and elsewhere, including Boston, which served both as nodes of internal knowledge production and as points of connection to local and international academic communities. Macclesfield remained active in the UK, though in a pharmaceutical technology and manufacturing capacity rather than as a core research site. Notably, China emerged as a significant and growing part of the R&D network in the post-2020 period. Shanghai was designated AstraZeneca's fifth strategic R&D center in 2021. A sixth strategic center was subsequently announced in Beijing in 2025.

6 Econometric methodology

Our empirical strategy is composed of a two steps strategy, First, we develop an event study to check whether, after the difficult years of the early 2010’s, AstraZeneca increased its reliance on new international academic collaborations with respect to other types of collaborations. Secondly, we check whether the existence of social ties by the AZ authors is correlated to higher novelty for the papers with international academic coauthors.

6.1 Increase in International Academic Collaboration

As discussed in the previous sections, the early 2010s marked a period of strategic disruption and organizational reorientation at AstraZeneca. The combination of patent expirations, leadership change, and the attempted takeover in 2013 prompted a renewed emphasis on science-driven innovation and external knowledge sourcing. If this strategic shift involved a

stronger orientation toward exploratory search, one observable implication would be an increased reliance on new international collaborators, particularly in academic partnerships, where frontier scientific knowledge is generated.

To examine whether collaboration patterns changed during this period, we analyzed the share of first-time international coauthors in AstraZeneca publications before and after 2013. Our unit of analysis is the individual scientific publication authored between 2000 and 2020. For each paper, we identified all coauthors, their institutional affiliation, and their prior co-publication history with AstraZeneca researchers. This allowed us to determine whether a collaboration represented a first-time co-authorship and whether the coauthor was located in a foreign country.

We distinguished between two types of collaborative papers:

1. Papers with at least one academic coauthor and
2. Papers with at least one coauthor from a firm other than AstraZeneca.

To facilitate interpretation, we excluded publications that included both academic and non-AZ company coauthors (1,398 papers out of 10,333 papers with academics and/or non-AZ company coauthor, i.e. 13% of that sample).

To identify how much a paper relied on new international coauthors (whether academic coauthors or individuals working for other companies) we adapted the procedure implemented by Liu et al. (2022). For each paper i, we constructed a dependent variable capturing the share of new international coauthors relative to all possible AstraZeneca–external coauthor pairs[4]. For academic papers, the measure is:

$$ShareNewIntlAcademic_i = \frac{\sum_j \text{NewIntlAcademic}_{ji}}{\text{AZAuthors}_i \times \text{AcademicCoauthors}_i}$$

[4] For example, in a paper with 2 AstraZeneca authors and one foreign academic coauthor if the academic coauthor is new only for one of the two AstraZeneca authors our variable will count 1 on the numerator and 2 (the two possible pairs between AstraZeneca authors and the foreign coauthor) on the denominator, hence its value will be 0,5.

where the numerator sums first-time foreign academic coauthors for each AstraZeneca author $j$in paper i, and the denominator equals the total number of potential AZ–academic coauthor pairs in that paper[5].

For company papers, we constructed an analogous measure:

$$ShareNewIntlCompany_i = \frac{\sum_j \text{NewIntlCompany}_{ji}}{\text{AZAuthors}_i \times \text{CompanyCoauthors}_i}$$

Because our data consists of repeated cross-sections of publications rather than a panel of organizational units, we implemented a grouped difference-in-differences design. Papers with academic coauthors constituted the treatment group ($g = 1$), while papers with company coauthors served as the comparison group ($g = 0$). The post-2013 period defines the time dimension.

We estimated the following specification:

$$ShareNewIntl_{igt} = \alpha + \beta_1 \text{Post2013}_t + \beta_2 \text{Academic}_g + \delta\left(\text{Post2013}_t \times \text{Academic}_g\right) + \gamma X_i + \lambda_t + \varepsilon_{igt}$$

Where $Post2013_t$equals 1 for publications after 2013, $Academic_g$ equals 1 for academic papers and 0 for company papers, X_i includes controls for team composition (number of AZ and external authors), therapeutic area, and R&D site, λ_t denotes year fixed effects, and ε_{igt} is the error term.

The coefficient of interest, δ, captures whether academic papers experienced a differential shift toward new international collaborators after 2013, relative to company papers. Under the identifying assumption of parallel trends in the absence of the strategic disruption, δ measures the relative change in collaborative orientation associated with the post-2013 period.

6.2 Social Ties and Novelty in Academic Collaboration

We next examined how relational configurations were correlated to the novelty of collaborative research outcomes.

[5] In these papers we can also have other types of co-authors: with PRO, hospital or unknown affiliation

Our dependent variable is the novelty of publication i, measured using the two word-embedding–based indicators developed by Shibayama et al. (2021) and Yin et al. (2023). These measures capture both recombinant novelty (unusual combinations of existing knowledge elements) and element-level novelty (the introduction of new semantic components).

We estimated the following baseline model for academic papers:

$$Novelty_i = \alpha + \beta_1 PersonalTies_i + \beta_2 InstitutionalAgreement_i + \theta Z_i + \lambda_t + \varepsilon_i$$

Where PersonalTie_i indicates whether at least one AstraZeneca author previously shared an institutional affiliation with a collaborating academic, $\text{InstitutionalAgreement}_i$ captures the presence of formal collaboration agreements, Z_i includes controls for team composition, R&D site, therapeutic domain, geographic distance, and other collaboration characteristics, and $\lambda_t$denotes year fixed effects.

To reflect our theoretical argument that relational mechanisms may operate differently across spatial contexts, we estimated the model separately for international and domestic academic collaborations. All models were estimated using ordinary least squares with heteroskedasticity-robust standard errors.

6.2.1 The dependent variable: Novelty

Novelty is widely regarded as a core value of scientific research, yet its empirical measurement remains conceptually and methodologically challenging. As emphasized in recent work, novelty is a multifaceted construct that can refer to new elements, new combinations of existing knowledge, or new positioning within scholarly discourse (Shibayama et al., 2021; Yin et al., 2023). Bibliometric operationalizations therefore differ substantially in what aspect of novelty they capture. Following this literature, this study adopted two complementary, text-based indicators that reflect distinct dimensions of novelty: *recombinant novelty* and *element novelty*.

Recombinant novelty captures the extent to which a publication recombines previously unconnected knowledge components. Rooted in the recombinant view of innovation (Fleming, 2001), this approach assumes that new knowledge emerges from atypical or cognitively distant

combinations of existing elements. Bibliometric implementations typically assess the novelty of reference combinations, journal pairings, or semantic distances between cited works (Uzzi et al., 2013; Wang et al., 2017). In this study, recombinant novelty follows the operationalization proposed by Shibayama et al. (2021), which uses document-level semantic representations derived from word embeddings. To account for the fact that novelty measures depend on the threshold used to define what constitutes an atypical combination or a new semantic element, we construct multiple variants of both indicators. The indicator r100 corresponds to the most stringent definition, identifying publications whose recombination lies at the extreme tail of the distance distribution. The measures r99, r95 and r90 progressively relax this threshold, classifying as novel those publications that fall within the top 1%, 5%, and 10% of the distribution, respectively. Lower thresholds therefore capture increasingly broader forms of atypical recombination. Specifically, the indicator measures the extent to which the cited knowledge components of a focal publication represent unusual or distant combinations relative to the prior knowledge space. Higher values indicate more atypical recombination of existing knowledge.

Element-level novelty captures a different dimension of novelty: the introduction of new semantic elements into the scientific corpus. Rather than focusing on combinations, this approach identifies whether a publication contains linguistic or conceptual elements that were previously absent from the knowledge space. Following Yin et al. (2023), element-level novelty is derived using contextual word-embedding models that map scientific texts (title and abstract) into high-dimensional semantic space. The indicator evaluates the semantic distance between a focal publication and the existing body of literature, identifying documents that introduce conceptually new elements rather than merely recombining known ones. Similarly, for element-level novelty, we implement alternative cutoffs. The indicator g0 reflects the most stringent criterion, identifying publications that introduce semantic elements absent from the prior corpus. The measures g1, g5, and g10 relax this requirement by allowing increasing degrees of semantic proximity to existing elements, thereby capturing progressively less radical but still substantively new contributions. As highlighted by Shibayama et al. (2025), element-based indicators may capture forms of novelty that recombinant measures overlook, particularly when novelty lies in the articulation of new concepts, phenomena or theoretical framings rather than in atypical citation combinations.

In pharmaceutical research, these two dimensions of novelty correspond to distinct but complementary innovation processes. Recombinant novelty reflects the integration of previously unconnected scientific domains, for example, combining insights from molecular biology, chemistry and clinical science to identify new therapeutic mechanisms. Such atypical knowledge combinations are central to first-in-class drug discovery, where breakthroughs often arise from bridging cognitively distant fields. Element novelty, by contrast, captures the introduction of new concepts, targets, molecular entities or mechanistic framings that expand the scientific knowledge base underlying drug development. In an industry characterized by high scientific uncertainty and strong reliance on frontier research, both recombinant configurations and genuinely new knowledge elements are critical for advancing beyond incremental innovation. Employing both indicators therefore enables a nuanced assessment of how multinational pharmaceutical collaborations relate to different forms of scientific newness.

Recent validation studies demonstrate that no single bibliometric indicator captures all forms of novelty (Shibayama et al., 2025). Recombinant and element-based measures reflect partially overlapping but distinct dimensions of newness. Using both indicators therefore allows for a more comprehensive assessment of novelty in multinational pharmaceutical R&D. Importantly, both indicators are ex-ante, text-based and independent of forward citations, ensuring that the dependent variable reflects knowledge novelty rather than ex-post impact or disruption.

6.2.2 *Independent variable: measure of social ties*

Based on the history of affiliations (listed in the papers that the AstraZeneca researcher coauthored until paper *i*) of each AstraZeneca author, we checked whether any of the academic coauthors in paper *i* had an affiliation that also the AstraZeneca author had in other papers in the past. This implies that the AstraZeneca researcher must have published some papers in the past in which her affiliation was not AstraZeneca, but rather an academic institution (the researcher worked for a period in an academic institution). We built a dummy variable (*Academic social ties*) at the paper level that is equal to 1 if at least one of the AstraZeneca authors in the paper had been listed in a publication in the past with the same affiliation as one of the academic coauthors of the paper. So, for example, the *Academic social ties* dummy would be equal to 1 if we find that in paper *i* published in 2018 one of the authors of the

Goteborg-based AstraZeneca team has published a paper in the past with an "Imperial College" affiliation in London, and in paper *i* there is at least one academic coauthor with an Imperial College affiliation. This implies that one of the AstraZeneca authors has some familiarity with and/or ties to the Imperial College environment, possibly helping her to establish such a collaboration.

Since the notion of having a publication *in the past* with the same academic institution as one of the academic coauthors is still quite vague, we include the additional condition that at least some time must have passed between the year the paper *i* is published and the last time the AZ author was affiliated with the academic institution. We set this to be at least 3 years. To assess the robustness of our findings, we conducted sensitivity analyses extending the observation period to four, five, and six years.

6.2.3 *Controls*

Institutional agreements

To identify *institutional agreements* with universities we used both annual reports and data from PharmaDeals an online database from IQVIA that provides historical information about deals and alliances in the pharmaceutical and biotech sector (see Appendix B for details). For each academic author we checked if their academic institution has signed an institutional agreement with AstraZeneca and in which year this was done. We create a variable *Institutional Agreement* that is equal to one if at least one of the academic authors involved is affiliated with a university with which AZ had an ongoing institutional agreement, and zero otherwise. The variable is time varying, meaning that it is equal to 1 only from the year of establishment of the institutional agreement.[6]

Share of new contacts

The novelty of the paper can also be influenced by the fact that new collaborations are established by authors that have never published together before we therefore adapt the procedure implemented by Liu et al. (2022). Therefore, we control for the effect of new contacts on the overall novelty of papers adapting the procedure already used in the previous section. For each paper *i*, we constructed a variable capturing the share of new academic

[6] Unfortunately, we cannot ascertain whether these collaborations were interrupted at some point in time, therefore our institutional agreement dummy variable is equal to 1 for all the years after the establishment of the agreement.

coauthors relative to all possible AstraZeneca–academic coauthor pairs. For each academic paper the measure is:

$$ShareNewAcademic_i = \frac{\sum_j \text{NewAcademic}_{ji}}{\text{AZAuthors}_i \times \text{AcademicCoauthors}_i}$$

where the numerator sums first-time academic coauthors for each AstraZeneca author *j* in paper i, and the denominator equals the total number of potential AZ–academic coauthor pairs in that paper.

Geographic variables

We also checked for some geographical characteristics of the team of coauthors which may also influence the establishment of collaborations with foreign academics. First, we included the geographical distance (*Distance to R&D sites*) between the foreign academic coauthors and the AstraZeneca R&D centers to understand how far the academics are with respect to the overall R&D network of the company. This variable is computed as follows, for each academic coauthor, we calculated the distance (in kilometers) between their institution and the nearest AstraZeneca R&D center (see Appendix C for details). We then averaged these distances across all coauthors for each paper and took the natural logarithm of the result. Secondly, we counted the number of different countries of the foreign academics (*Num. countries academics*): if the foreign academics are in more than one foreign country the variable takes a value greater than 1.

Company areas of expertise

AstraZeneca operates across multiple therapeutic domains that differ in their scientific maturity, technological trajectories, and depth of internal expertise. Variation across therapeutic areas may influence both collaboration patterns and novelty outcomes. In domains where AstraZeneca has accumulated substantial scientific and organizational experience, researchers may rely more heavily on established collaboration networks and existing academic partners. By contrast, in therapeutic areas that are relatively new or strategically expanding, the firm may engage in broader or more exploratory knowledge sourcing, including collaborations with new and potentially international academic partners.

To account for this heterogeneity, we included controls for the therapeutic area of each publication. Therapeutic areas are identified using Medical Subject Headings (MeSH) terms associated with each publication. MeSH is a standardized controlled vocabulary developed by the U.S. National Library of Medicine (NLM) and used to index articles in PubMed. Each publication is assigned a set of MeSH descriptors that summarize its biomedical content.
We mapped MeSH terms to eight therapeutic areas in which AstraZeneca was active during the study period, namely: Oncology, Immunology, Cardiovascular, Infection, Endocrinology, Respiratory, Muscoloskeletal, and Central Nervous System.

For each publication, we identified how many of its MeSH terms fell within the predefined set of descriptors corresponding to each therapeutic area. We then created a dummy variable that was equal to one if the paper had at least one term that could be assigned to a specific therapeutical area, and zero otherwise. Because articles may span multiple areas, the dummy variables are not mutually exclusive and can jointly describe multidisciplinary research content.

Clinical research

To control for differences in the stage of scientific development, we classified publications according to whether they contribute primarily to basic or clinical knowledge. This classification was constructed using MeSH terms following the methodology developed by Weber (2013). Each publication was first assigned to one or more of three broad topic categories based on its MeSH descriptors including cells and molecules (C), animals and other complex organisms (A) and humans (H). These categories reflect the biological level at which the research is conducted. We then aggregated these topic areas into two knowledge domains: basic science and clinical science. Publications categorized as A and/or C but not H were classified as contributing to basic science, as they predominantly involve cellular, molecular, or pre-clinical animal research. Such research typically corresponds to early-stage discovery and pre-clinical experimentation. Publications categorized as including at least one H descriptor were classified as contributing to clinical science, reflecting research involving human subjects and corresponding to clinical development phases. Articles that could not be assigned to any of the three topic groups (A, C, or H) were excluded from the dataset. A binary indicator for clinical research was included in all regression models to account for systematic differences in collaboration patterns and novelty between basic and clinically oriented research.

Team controls

We also controlled for author-level characteristics that may influence the likelihood of a paper exhibiting higher levels of recombinant or element novelty. The seniority, experience, and productivity of AstraZeneca authors on a given paper may all be correlated with its degree of novelty. Besides counting the number of AstraZeneca coauthors in each paper *i* (*Num of AZ authors*), we also counted the average of the past number of publications among the AstraZeneca researchers in each paper i (*Average past num papers).* We also included a control for whether at least one of the AstraZeneca authors on a given paper had accumulated more than 15 publications at the time of writing, serving as a proxy for the presence of an experienced researcher on the team.[7] This indicator is particularly useful in papers with relatively large AstraZeneca author teams, where an average publication count may obscure the presence of a highly experienced individual contributor.

In order to account for the possible heterogeneity in the team of coauthors we also controlled for the total number of authors in each paper (*Num total authors*), and for the number of coauthors that were affiliated to other companies than AstraZeneca (*Num non-AZ company authors*), the number of coauthors affiliated to a public research organization *(Num PRO authors)*, the number of authors with a hospital affiliation (*Num hospital authors*), and the number of authors with an academic affiliation (*Num academic authors*).

R&D Centers and International Academic Collaboration

As outlined above, we distinguished R&D sites into seven functional categories reflecting different knowledge-sourcing strategies (Global Academic Hubs, Corporate R&D Hubs, International Academic Centers, International R&D centers, Corporate R&D centers, Local R&D Outposts and Outpost for Local Academic Search). From the perspective of knowledge novelty, this heterogeneity is analytically important. Sites with strong international academic engagement are better positioned to access diverse scientific communities and cognitively distant knowledge, thereby broadening the opportunity space for both recombinant and element-level novelty. By contrast, sites with limited academic or international interaction are more likely to draw on established internal trajectories, potentially favoring incremental over novel contributions. AstraZeneca's differentiated R&D geography thus provides a structured

[7] 15 publications represent the 4th quartile of the distribution of the total number of papers by AZ authors.

yet internally varied empirical setting in which to examine how international collaboration and different relational configurations relate to the production of novelty in multinational pharmaceutical research.

INSERT TABLES 6 AND 7 AND FIGURES 1 AND 2 ABOUT HERE

6.3 Descriptive statistics

Table 6 lists each of the variables described with a detailed explanation of what each measures. Table 7 reports descriptive statistics for our sample, that is, all AstraZeneca papers with at least one academic coauthor, broken down into papers involving only domestic academic co-authors and those involving international ones. As noted in Section 4, the sample comprises 2,806 internationally coauthored papers and 2,557 nationally co-authored ones, for all of which a corresponding element novelty score is available. Coverage is somewhat more restricted for the recombinant novelty variable, which reduces the sample to 2,544 international and 2,018 national.

It is worth noticing that regardless of how we measure novelty (Recombinant-type or Element-type novelty) we always find a slightly higher level of novelty among papers with national academic coauthors. This is well illustrated by Figures 1 and 2 where we report the average novelty among international and national academic collaborations. Novelty is always slightly higher among national collaborations, which is in accordance with Wagner et al. 2019.

Secondly, the presence of a social tie is more frequent among national collaborations (12% of the cases) than among international ones (3%). This is not surprising, as it is much more common to have AZ authors working in an R&D center in a specific country who were in the past affiliated to a university in the same country. It speaks to the fact that most of AZ researchers have a relatively nationally bounded career and continue to collaborate with researchers from their alma mater in the same country.

As would be expected, internationally co-authored papers display a greater average geographic distance from AstraZeneca's R&D centers. They also tend to involve author teams, counting

AstraZeneca authors, academic co-authors, and other affiliated contributors, with higher average publication counts than those found in national collaborations. Papers with international academic co-authors are more likely to include authors with hospital or public research organization (PRO) affiliations and have a slightly higher average number of academic co-authors (4.4 versus 3.5 for domestic papers).

Most international collaborations originate from AstraZeneca authors based at the Global Academic Hubs of Cambridge, Mölndal, and Södertälje, which account for 70% of international papers compared to 46% of national ones. Conversely, international collaborations are less likely to involve Local R&D Centres (4% vs. 12%) or Corporate R&D Hubs (9% vs. 23%). Among therapeutic areas, the most notable difference concerns cardiology, which is proportionally more common in international papers than national (23% vs. 11%); for the remaining therapeutic areas, no substantial differences emerge.

7 Results

7.1 The increase in academic international collaborations

Table 8 reports the results of the difference-in-differences estimations. The dependent variable is the share of new international coauthors in each paper. Across all specifications, the coefficient on *Academic paper* is negative and statistically significant. This indicates that, prior to the treatment period, papers involving academic coauthors exhibit a lower share of new international collaborators compared to papers coauthored with researchers from other companies. In other words, academic collaborations appear more stable and recurrent on average, whereas company collaborations involve a higher proportion of first-time foreign partners.

The main coefficient of interest is the interaction term between Academic paper and the post-treatment period. When the treatment year is set to 2013 (Column 1), the interaction coefficient is positive and statistically significant ($\beta = 0.064$, $p < 0.01$). This implies that after 2013, academic papers experienced a differential increase in the share of new international collaborators relative to company papers. The magnitude of the coefficient suggests that the post-2013 period is associated with an approximately 6–8 percentage point increase in the share of new foreign academic coauthors, depending on the specification. This result is robust to alternative definitions of the treatment year. When the post period is shifted to 2014 or 2015

(Columns 2 and 3), the interaction term remains positive and statistically significant, indicating that the shift in collaboration patterns was not confined to a single year but persisted in the subsequent period. Importantly, placebo tests using earlier years as the treatment threshold do not produce significant interaction effects (results available upon request), providing support for the identifying assumption that the observed shift is specific to the post-2013 period.

INSERT TABLE 8 AND FIGURES 3 ABOUT HERE

Figure 3 complements these results by plotting adjusted predictions of the share of new international collaborators over time for academic and non-academic papers. Prior to 2013, the two trends are relatively stable and move in parallel. After 2013, however, the trajectory of academic papers diverges upward, while company papers display no comparable increase. The figure therefore visually corroborates the regression results, illustrating a post-2013 acceleration in the propensity of AstraZeneca publications to involve new foreign academic collaborators.

The coefficients on the R&D center controls further indicate substantial heterogeneity across AstraZeneca's research network. Relative to *Global Academic Hubs* (the reference category), most other site types exhibit significantly lower shares of new international collaborators. This pattern is consistent with the descriptive evidence that internationally oriented and academically embedded hubs play a central role in connecting the firm to foreign knowledge networks. Importantly, the positive and significant interaction effect remains robust after controlling for this spatial heterogeneity, suggesting that the post-2013 shift is not solely driven by changes in the geographic composition of publishing activity.

Taken together, the evidence suggests that the period following the strategic disruption in the early 2010s was associated with a reorientation toward new international academic partnerships, beyond general trends in external collaboration.

7.2 Social ties and novelty

Table 9 reports the OLS estimates for Recombinant-type novelty (R-novelty), distinguishing between papers co-authored with international academics and those with national academic collaborators. Across all specifications, social ties exhibit a clear and systematic pattern. For international collaborations, the coefficient on *Academic social tie* is positive and statistically significant across all R-novelty thresholds (r100–r90). The magnitude declines as the novelty definition becomes less stringent, but the effect remains robust. Substantively, this indicates that when at least one AstraZeneca researcher has previously worked at the foreign institution of the academic collaborator, the resulting publication exhibits higher recombinant novelty. In contrast, no statistically significant association emerges for national collaborations. Across all R-novelty specifications, prior institutional ties within the same country are not associated with higher novelty levels. The absence of an effect in the national sample suggests that the value of social ties is not universal, but conditional on geographic configuration.

The control variables further support this interpretation. First, we find that the share of new academic contacts is positive and significant only among international academic papers: this finding resonates the previous results of Table 8. The surge in new international academic contacts observed after 2013 is also linked to a higher degree of novelty of the papers that engage new international academics. This is not the case for papers that only engage national academics.

Institutional agreements between AZ and the universities of the academic coauthors are not associated with higher levels of recombinant novelty. Interestingly, the coefficient for national collaboration is negative and significant. This indicates that institutional agreements with local universities rather produce conventional science. The distance to R&D sites is negatively associated with novelty, and clinical research projects exhibit lower recombinant novelty, consistent with the incremental and regulatory-driven nature of later-stage development. The number of countries represented among academic collaborators is negatively related to novelty in the international sample, suggesting that excessive geographic dispersion may introduce coordination frictions that limit effective recombination. Importantly, the social-tie effect persists after controlling for team composition, therapeutic area, and R&D center typology.

INSERT TABLES 9 AND 10 ABOUT HERE

Table 10 presents parallel estimations using element-type novelty (G-novelty). The results mirror the recombinant findings but are slightly more nuanced. For international collaborations, social ties are positively associated with G-novelty across the less stringent thresholds (g1, g5, g10). The effect is not statistically significant under the most restrictive definition (g0), which captures only the most extreme element-level novelty. This suggests that prior institutional ties facilitate the integration of relatively new knowledge elements but may be insufficient to generate the most radical breakthroughs. Again, no significant effect is detected in the national sample. The divergence between international and national collaborations is consistent across both novelty measures, reinforcing the conditional interpretation.

Taken together, the results indicate that social proximity operates as a compensatory mechanism under geographic distance. Prior social ties appear to enable more effective recombination when collaborations span national borders, where interpretative and coordination barriers are higher. In national collaborations, where institutional and contextual alignment is already stronger, such ties do not yield additional novelty benefits.

The findings therefore support a conditional view of proximity: social ties matter not universally, but specifically when knowledge integration occurs across spatial and institutional divides.

7.2.1 Robustness checks: definition of social ties and total number of authors

In Appendix D we run a series of robustness checks. First, we examine whether our results are sensitive to how we measure social ties. Our baseline definition uses a three-year lag to identify a social tie; that is, at least three years must have passed between the last year in which an AZ researcher was affiliated with a given academic institution and the year of publication of the paper. We then relax this assumption to explore what happens when shorter or longer lags are allowed.

We do this by plotting the estimated marginal effects across novelty thresholds for both recombinant novelty and element novelty, changing in each figure the lag used for the definition of social tie. In each figure we report the marginal coefficients for both international academic papers (figure on the left) and national papers (figure on the right).

Figures A2 and A6 (based on a minimum lag of three years) serve as the benchmark, reproducing the marginal coefficients from the main specifications in Tables 9 and 10 and confirming that the results are consistent with those reported in the main regression tables. Figures A1 and A5 apply a two-year minimum lag.[8] and show that the positive association with novelty remains strong for recombinant novelty, while becoming slightly less statistically significant for element-based novelty. Increasing the minimum lag to four or five years (in the remaining figures) continues to confirm the positive effect of social ties, with one exception: for recombinant novelty, the coefficient loses some statistical significance when the lag is extended to five years. For national collaborations, by contrast, the relationship with novelty is never statistically distinguishable from zero regardless of the lag chosen, for either recombinant or element-based novelty. Taken together, these results confirm that the main findings are not sensitive to the choice of lag, and that the relevance of prior social ties for novelty in international collaborations remains robust across specifications.

A second robustness check concerns the number of authors in a paper. Although we already exclude papers with more than 100 authors from our sample (and control for the total number of authors in our regressions), we further verify that our results are not driven by outliers with an unusually large number of authors. In other words, we want to ensure that we focus on papers in which the contribution of each AZ researcher and academic co-author is meaningful. We therefore restrict the sample to papers with no more than 20 authors.

In the Appendix, tables A.1 and A.2 report the results of the model estimation focusing only on international collaborations. The coefficient on the social tie dummy remains positive and statistically significant, consistent with the main specification, indicating that our results are not driven by papers with a very high number of authors. The positive coefficient of social ties is robust also to the restriction of the analysis to papers with no more than 10 authors.

7.3 Heterogeneity analysis

7.3.1 Distinguishing between R&D center typology

[8] This means that we also consider cases in which only two years have passed between the last affiliation of the AZ researcher with the academic institution and the publication year of the paper in which the researcher is affiliated with AZ.

Finally, in Tables 13 and 14 we examine whether our main results regarding the importance of social ties in international collaborations vary depending on the type of AZ R&D center in which researchers are located. We distinguish between AZ researchers working in the main R&D centers, those characterized by a high volume of publications and a strong level of academic collaboration (the Global Academic Hubs, located in Cambridge, Mölndal/Göteborg, and Stockholm/Södertälje), and those working in other R&D centers.

We estimate the same model as before, focusing only on international collaborations, but now splitting the sample between papers authored by AZ researchers located in Global Academic Hubs and papers authored by AZ researchers located in other R&D centers. It should be noted that approximately 70% of international papers in our sample involve AZ researchers based in Global Academic Hubs.

INSERT TABLES 11 AND 12 ABOUT HERE

The results for recombinant novelty (Table 11) and element-based novelty (Table 12) show that the positive effect of social ties is mainly driven by AZ researchers located in the Global Academic Hubs, as the coefficient is positive and statistically significant only in this subsample. In other R&D centers, the coefficient of the social tie variable is slightly smaller and less precisely estimated, resulting in a lack of statistical significance (also due to the smaller number of observations).

These findings suggest that our baseline results are largely driven by the social ties of AZ researchers working in the main R&D centers, where publication output and the propensity to collaborate with international academic partners are particularly high.

7.3.2 Social ties and the role of distance to R&D sites

In Tables 13 and 14 we test whether the correlation between social ties and novelty changes according to how far international academic researchers are from AZ R&D sites. The variable that we interact social tie with is the average distance (in log) of academic coauthors to the closest R&D site.

INSERT TABLES 13 AND 14 AND FIGURES 4 AND 5 ABOUT HERE

The results in Tables 13 and 14 reveal a nuanced picture that varies depending on the type of novelty considered. For recombinant novelty (Table 13), the coefficient on social ties is negative but not statistically significant, while the interaction term is positive and significant. This implies that the effect of social ties on recombinant novelty is positive when the average distance between academic co-authors and AstraZeneca R&D sites is sufficiently large. As shown in figure 4, the social tie coefficient becomes positive and significant for values of log average distance above 6, corresponding to approximately 400 km.

The pattern is reversed for element-based novelty (Table 14), where the social tie coefficient is positive and significant, and the interaction term is negative and significant. Here, the beneficial effect of social ties on novelty holds even at very short distances from AstraZeneca R&D sites, but diminishes as distance increases, with the coefficient becoming indistinguishable from zero at a log average distance of approximately 6 (around 400 km), as illustrated in Figure 5. These results underscore the importance of considering different proxies for novelty as the various indicators are capturing different aspects of new knowledge. The results of previous studies that have used only one simple indicator might have been overstated.

8 Discussion and conclusions

The objective of this study was to clarify whether and under which conditions international industry–academic collaboration is associated with higher levels of scientific novelty within a multinational pharmaceutical firm. The results contribute to ongoing debates in economic geography and innovation studies concerning the relationship between geographic dispersion, proximity configurations, and knowledge creation. While early work has demonstrated that innovation is spatially structured, more recent research emphasizes that spatial proximity alone cannot fully account for innovation outcomes, and that multiple, intersecting proximity dimensions, including organizational, cognitive, and social proximities, jointly shape creative recombination across space (Boschma, 2005; Balland et al., 2015; Shkolnykova, 2023). At the

same time, studies of international research networks show that cross-border collaboration is often associated with higher impact research, but conventional rather than novel research (Wagner et al., 2019). However, comparatively less attention has been devoted to the relational conditions under which geographic diversity translates into recombinant outcomes, particularly in firm-based industry–academic collaborations within multinational corporations. By combining a grouped difference-in-differences design with publication-level novelty measures, this study provides evidence on both the strategic reorientation of collaboration and the conditional role of social proximity.

The difference-in-differences results indicate that following the strategic turbulence of the early 2010s, AstraZeneca publications with academic coauthors experienced a significant relative increase in the share of new international collaborators compared to publications coauthored with other companies. This pattern is consistent with accounts of the growing centrality of academic science in pharmaceutical innovation (Cockburn & Henderson, 1998; Owen-Smith & Powell, 2004) and with arguments that periods of strategic pressure intensify firms' reliance on external, frontier-oriented knowledge sources (Pisano, 2006; Schuhmacher et al., 2016). The evidence suggests an acceleration of "global pipelines" specifically directed toward academia (Bathelt et al., 2004), rather than a general increase in international partnering. This finding contributes micro-level evidence to the literature on multinational R&D networks by showing that the internationalization of collaboration is not static but can shift in response to organizational disruption and strategic reorientation.

The central contribution of the paper lies in the novelty analysis. Across multiple specifications and novelty thresholds, prior social ties, operationalized as shared institutional affiliation in earlier career stages between AZ researchers and their academic co-authors, are positively associated with novelty in international industry–academic collaborations, while no such association is observed in national collaborations. This pattern aligns with proximity theory's core proposition that innovation outcomes depend on the configuration of proximity dimensions rather than on geographic distance alone (Boschma, 2005; Huber, 2012; Balland, Boschma, & Frenken, 2015). Social proximity appears to function as a conditioning mechanism that is particularly consequential when collaborations span national and institutional boundaries. In national collaborations, where institutional environments and opportunities for interaction are more aligned, the marginal contribution of prior ties to novelty is limited. In international settings, by contrast, prior affiliation-based ties are associated with higher novelty

outcomes, suggesting that relational embeddedness becomes more salient under conditions of greater spatial and institutional dispersion.

This conditional pattern addresses an important gap in the literature. Existing research has shown that international collaboration is often associated with higher citation rate (Glänzel & Schubert, 2001), more interdisciplinarity (van Raan, 2003), but less novel outputs (Wagner et al., 2019), while relational embeddedness has been theorized to facilitate knowledge transfer (Reagans & McEvily, 2003; Ooms et al., 2018). However, there has been limited systematic evidence demonstrating that the novelty associated with cross-border collaboration depends on pre-existing social ties within multinational firm contexts. By distinguishing between international and national industry–academic collaborations, this study moves beyond estimating an average "distance effect" and instead identifies a configurational relationship between geographic dispersion and social proximity, consistent with the relational turn in economic geography (Boschma, 2005; Crescenzi et al., 2016; Roth & Mattes, 2023).

The use of both recombinant (R-type) and element-level (G-type) novelty measures further refines the interpretation. Social ties are robustly associated with recombinant novelty across thresholds in international collaborations, consistent with theories emphasizing the importance of bridging cognitively distant knowledge domains for innovative recombination (Uzzi et al., 2013; Wang et al., 2017). For element-level novelty, the effect is present for less stringent thresholds but not for the most restrictive definition, which captures only the most radical semantic departures (Shibayama et al., 2021; Yin et al., 2023). This suggests that prior affiliation-based ties may be particularly helpful for facilitating meaningful recombination and moderately strong semantic newness, while the most extreme forms of element novelty may depend on additional factors beyond relational familiarity. In doing so, the paper contributes to the growing literature emphasizing the multidimensional nature of novelty and the importance of employing complementary indicators (Shibayama et al., 2025). The divergent moderating role of geographic distance across the two novelty dimensions is theoretically interesting. For recombinant novelty, social ties matter most at greater distances, while for element-level novelty the pattern reverses. That the two novelty dimensions respond in opposite ways to the same moderator underscores that they capture qualitatively distinct knowledge creation processes, and further supports the case for multi-indicator approaches to measuring novelty in innovation research.

Several control results reinforce the view that geographic dispersion introduces barriers to knowledge integration. The negative association between geographic distance and novelty, as well as the negative relationship between the number of academic countries represented and recombinant novelty in the international sample, indicate that dispersion can introduce coordination and interpretative burdens (Storper & Venables, 2004; Boschma, 2005). These findings resonate with arguments that diversity expands the opportunity space for recombination but does not automatically generate novel outcomes (Wagner et al., 2019). Instead, novelty appears to depend on relational mechanisms that help translate geographic diversity into coherent research outputs.

The absence of a significant effect for international collaborations, and the negative coefficient for national ones, suggests that institutional agreements between AstraZeneca and academic institutions are not primarily oriented towards the generation of novel knowledge. Such agreements may instead serve other strategic purposes, for instance, facilitating access to more applied or development-stage research, signaling commitment to the academic community, or enabling the sharing of infrastructure and research facilities rather than driving genuine knowledge recombination.

Taken together, the findings advance three interrelated strands of literature. First, they contribute to economic geography by demonstrating that the relationship between international collaboration and novelty is conditional on relational embeddedness, thereby operationalizing the configurational logic of proximity theory in a multinational firm setting. Second, they enrich research on industry–academic collaboration in science-based sectors by identifying a micro-level relational mechanism, prior shared institutional affiliation, through which cross-border academic partnerships are associated with higher novelty (Cockburn & Henderson, 1998; Perkmann et al., 2013; Fassio et al., 2023). Third, they add to the novelty-measurement literature by showing that the conditioning role of social proximity varies across novelty dimensions and thresholds, underscoring the value of multi-indicator approaches.

At the same time, the results should be interpreted within the study's scope. The operationalization of social proximity captures one observable form of relational embeddedness and likely underestimates broader informal or professional networks. Moreover, the analysis identifies conditional associations at the publication level and does not directly observe within-team coordination processes. Nevertheless, the consistent asymmetry between

international and national collaborations across novelty measures provides robust empirical support for a relational conditioning interpretation.

Overall, the study suggests that in multinational pharmaceutical research, international industry–academic collaboration is not inherently more novel. Rather, novelty is more strongly associated with cross-border collaboration when it is anchored in prior social ties. This finding refines existing understandings of global knowledge sourcing by showing that the innovative benefits of international collaboration depend not only on access to diverse scientific environments but also on the relational infrastructures that connect actors across spatial and institutional divides.

The null or negative association between formal institutional agreements and novelty suggests that such partnerships may not be the most effective for stimulating frontier knowledge creation. For managers, this implies that broad institutional agreements should not be treated as a substitute for the kind of researcher-level relational embeddedness that is shown here to drive novelty. Agreements may rather serve purposes in terms of infrastructure sharing, IP management. For policymakers, these results are a cautionary note for instruments that incentivise industry–academic collaboration primarily by counting formal partnerships. The quality of knowledge exchange matters more than its institutional form, and support mechanisms that foster deeper researcher-level interactions, such as joint doctoral programmes, visiting scholar schemes, co-location initiatives, are likely to be more effective at stimulating the kind of relational ties that translate geographic and institutional diversity into genuine scientific novelty.

Several limitations of this study should be acknowledged. First, the novelty analysis does not rest on a comparable causal identification strategy. The association between prior social ties and novelty is estimated with extensive controls and is robust across specifications, but unobserved factors, such as the selective matching of experienced researchers to more ambitious projects, could contribute to the observed pattern. Second, the generalisability of the findings is inherently constrained by the single-firm design. AstraZeneca is a large, science-intensive multinational with a distinctive R&D history and a particularly strong tradition of academic engagement. Whether the relational mechanisms identified here operate similarly in other pharmaceutical firms, in different industries, or in firms with less internationalized R&D networks remains an open question. The single-firm setting is also, however, a strength in that

it allows for fine-grained analysis of collaboration patterns and organizational context that would be lost in cross-firm comparisons. Third, the operationalization of social proximity relies on a single, observable form of relational embeddedness, prior shared institutional affiliation between AstraZeneca researchers and their academic co-authors. While this is a theoretically grounded and systematically measurable proxy, it inevitably underestimates the broader landscape of informal ties, repeated interactions, and professional networks that may also shape collaboration quality and outcomes. Similarly, the novelty measures, while capturing meaningful dimensions of scientific contribution, do not exhaust all relevant aspects of knowledge creation, and results may vary with alternative operationalization. Finally, the analysis operates at the level of publications and does not directly observe the within-team coordination processes through which prior ties may facilitate or constrain knowledge integration. Future research combining bibliometric analysis with qualitative or survey-based approaches could shed further light on the mechanisms underlying the patterns documented here.

References

Ao, J., Johnson, D., & Esposito, S. E. (2026). Where open innovation partnership matters: Phase-specific patterns in R&D project counts for large pharmaceutical firms. *Technovation*, *150*, 103459.

Araújo, C., & Teixeira, A. (2014). Determinants of international technology transfer: An empirical analysis of the Enterprise Europe Network. *Journal of technology management & innovation*, *9*(3), 120-134.

Asheim, B., Coenen, L., Moodysson, J., & Vang, J. (2007). Constructing knowledge-based regional advantage: implications for regional innovation policy. *International Journal of Entrepreneurship and Innovation Management*, *7*(2-5), 140-155.

Audretsch, D. B., & Belitski, M. (2024). Geography of knowledge collaboration and innovation in Schumpeterian firms. *Regional Studies*, *58*(4), 821-840.

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American economic review*, *86*(3), 630-640.

Autant-Bernard, C., Mairesse, J., & Massard, N. (2007). Spatial knowledge diffusion through collaborative networks. *Papers in Regional science*, *86*(3), 341-350.

Balland, P. A., Boschma, R., & Frenken, K. (2015). Proximity and innovation: From statics to dynamics. *Regional studies*, *49*(6), 907-920.

Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in human geography*, *28*(1), 31-56.

Belderbos, R., Faems, D., Leten, B., & Looy, B. V. (2010). Technological activities and their impact on the financial performance of the firm: Exploitation and exploration within and between firms. *Journal of Product Innovation Management*, *27*(6), 869-882.

Belderbos, R., Grabowska, M., Kelchtermans, S., Leten, B., Jacob, J., & Riccaboni, M. (2021). Whither geographic proximity? Bypassing local R&D units in foreign university collaboration. *Journal of international business studies*, *52*(7), 1302.

Bignami, F., Mattsson, P., & Hoekman, J. (2020). The importance of geographical distance to different types of R&D collaboration in the pharmaceutical industry. *Industry and Innovation*, *27*(5), 513-537

Bignami, F., & Mattsson, P. (2019). Potential effects of increased openness in pharma: the original knowledge behind new drugs. *Drug discovery today*, *24*(10), 1957-1962.

Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional studies*, *39*(1), 61-74.

Boschma, R. A., & Lambooy, J. G. (1999). Evolutionary economics and economic geography. *Journal of evolutionary economics*, *9*(4), 411-429.

Cantwell, J., & Mudambi, R. (2005). MNE competence-creating subsidiary mandates. *Strategic management journal*, *26*(12), 1109-1128.

Cockburn, I. M., & Henderson, R. M. (2001). Scale and scope in drug development: unpacking the advantages of size in pharmaceutical research. *Journal of health economics*, *20*(6), 1033-1057.

Crescenzi, R., Nathan, M., & Rodríguez-Pose, A. (2016). Do inventors talk to strangers? On proximity and collaborative knowledge creation. *Research policy*, *45*(1), 177-194.

D'Este, P., & Patel, P. (2007). University–industry linkages in the UK: What are the factors underlying the variety of interactions with industry?. *Research policy*, *36*(9), 1295-1313.

Fassio, C., Geuna, A., & Rossi, F. (2023). How do firms reach out to foreign universities? Inventors' personal characteristics and the multinational structure of firms. *Journal of World Business*, *58*(3), 101431.

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research policy*, *30*(7), 1019-1039.

Frigon, A., & Rigby, D. L. (2024). Geographies of knowledge sourcing and the complexity of knowledge in multilocational firms. *Economic Geography*, *100*(4), 329-350.

Galaso, P., & Kovářík, J. (2021). Collaboration networks, geography and innovation: Local and national embeddedness. *Papers in Regional Science*, *100*(2), 349-378.

Garcia Martinez, M., Zouaghi, F., & Sanchez Garcia, M. (2025). Geographical and cognitive proximity effects on innovation performance: Which types of proximity for which types of innovation?. *European Management Review*, *22*(2), 291-317.

Gertler, M. S. (2003). Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of economic geography*, *3*(1), 75-99.

Grabher, G. (1993). The weakness of strong ties; the lock-in of regional development in Ruhr area. *The embedded firm; on the socioeconomics of industrial networks*, 255-277.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American journal of sociology*, *91*(3), 481-510.

Gupta, A. K., & Govindarajan, V. (2000). Knowledge flows within multinational corporations. *Strategic management journal*, *21*(4), 473-496.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, *108*(3), 577-598.

de Jong, J. P., & Freel, M. (2010). Absorptive capacity and the reach of collaboration in high technology small firms. *Research policy*, *39*(1), 47-54.

Liu M, Bu Y, Chen C, Xu J, Li D, Leng Y, Freeman RB, Meyer ET, Yoon W, Sung M, Jeong M, Lee J, Kang J, Min C, Song M, Zhai Y, Ding Y. (2022) Pandemics are catalysts of scientific novelty: Evidence from COVID-19. *Journal of the Association for Information Science and Technology*. 73(8):1065-1078.

Nan, D. (2024). The influence of proximity and knowledge base on recombination innovation in R&D collaboration. *Plos one*, *19*(2), e0298735.

Narula, R., & Santangelo, G. D. (2012). Location and collocation advantages in international innovation. *Multinational Business Review*, *20*(1), 6-25.

Ooms, W., Werker, C., & Caniëls, M. (2018). Personal and social proximity empowering collaborations: the glue of knowledge networks. *Industry and Innovation*, *25*(9), 833-840.

Owen-Smith, J., & Powell, W. W. (2004). Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community. *Organization science*, *15*(1), 5-21.

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'este, P., ... & Sobrero, M. (2013). Academic engagement and commercialisation: A review of the literature on university–industry relations. *Research policy*, *42*(2), 423-442.

Perri, A., Scalera, V. G., & Mudambi, R. (2017). What are the most promising conduits for foreign knowledge inflows? innovation networks in the Chinese pharmaceutical industry. *Industrial and Corporate Change*, *26*(2), 333–355.

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*, 116-145.

Reagans, R., & McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative science quarterly*, *48*(2), 240-267.

Plotnikova, T., & Rake, B. (2014). Collaboration in pharmaceutical research: Exploration of country-level determinants. *Scientometrics*, *98*(2), 1173-1202.

Roth, P., & Mattes, J. (2023). Distance creates proximity: Unraveling the influence of geographical distance on social proximity in interorganizational collaborations. *Environment and Planning A: Economy and Space*, *55*(6), 1372-1391.

Schuhmacher, A., Gassmann, O., & Hinder, M. (2016). Changing R&D models in research-based pharmaceutical companies. *Journal of translational medicine*, *14*(1), 105.

Shibayama, S., Wu, Z., Yin, D., & Yokota, K. (2025). State of the Art of Novelty Indicators. *Available at SSRN 5379973*.

Shibayama, S., Yin, D., & Matsumoto, K. (2021). Measuring novelty in science with word embedding. *PloS one*, *16*(7), e0254034.

Shkolnykova, M. (2023). Assessing the importance of proximity dimensions for the diffusion of radical innovations in German biotechnology. *European Planning Studies*, *31*(7), 1510-1531.

Storper, M., & Venables, A. J. (2004). Buzz: face-to-face contact and the urban economy. *Journal of economic geography*, *4*(4), 351-370.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468-472.

Visser, E. J., & Boschma, R. (2004). Learning in districts: Novelty and lock-in in a regional context. *European Planning Studies*, *12*(6), 793-808.

Wagner, C. S., Whetsell, T. A., & Mukherjee, S. (2019). International research collaboration: Novelty, conventionality, and atypicality in knowledge recombination. *Research policy*, *48*(5), 1260-1270.

Wang, J., Veugelers, R. & P.Stephan, (2017) Bias against novelty in science: a cautionary tale for users of bibliometric indicators, *Research Policy*, 46 (8) (2017), pp. 1416-1436

Werker, C., Ooms, W., & Caniëls, M. C. (2016). Personal and related kinds of proximity driving collaborations: a multi-case study of Dutch nanotechnology researchers. *SpringerPlus*, *5*(1), 1751.

Yin, D., Wu, Z., Yokota, K., Matsumoto, K., & Shibayama, S. (2023). Identify novel elements of knowledge with word embedding. *Plos one*, *18*(6), e0284567.

Zucker, L. G., Darby, M. R., & Armstrong, J. S. (2002). Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology. *Management science*, *48*(1), 138-153.

TABLES AND FIGURES

Table 1. The list of R&D centers of AstraZeneca during the period 2000-2020.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Bangalore (IND)	1	1	1	1	1	1	1	1	1	1	1	1	1	1							
Boston/Waltham (US)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cambridge (UK)								1	1	1	1	1	1	1	1	1	1	1	1	1	1
Gaithersburg (US)								1	1	1	1	1	1	1	1	1	1	1	1	1	1
Loughbourhood (UK)	1	1	1	1	1	1	1	1	1	1	1	1									
Lund (SWE)	1	1	1	1	1	1	1	1	1	1	1	1									
Macclesfield/Alderley park (UK)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Montreal (CAN)	1	1	1	1	1	1	1	1	1	1	1	1									
Mountain View (US)								1	1	1	1	1	1	1	1	1	1	1	1	1	1
Mölndal (SWE)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
New York (US)																				1	1
Osaka (JAP)						1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Rheims (FRA)			1	1	1	1	1	1	1	1	1	1									
Shanghai (CHI)							1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Södertälje (SWE)	1	1	1	1	1	1	1	1	1	1	1	1									
Tokyo (JAP)						1	1	1	1	1											
Warsaw (POL)																1	1	1	1	1	1
Wilmington (US)	1	1	1	1	1	1	1	1	1	1	1	1									

Source: authors' calculations using company annual reports

Table 2. Major Acquisitions by AstraZeneca (2000–2020)

year	Acquired company	Cost	Significance
2005	KuDos	approx. £120 million	Smaller deal of a DNA repair technology that could be used to treat cancer. One of the products was undergoing clinical trials.
2006	Cambridge Antibody Technology (UK)	approx. £702 million	Brought in leading expertise in monoclonal antibody technologies. This acquisition laid the foundation for AZ's biologics strategy before MedImmune
2007	MedImmune (USA)	approx. $15.6 billion	Transformed AstraZeneca into a major player in biologics and vaccines. MedImmune became AZ's dedicated biologics arm, central to its oncology and immunology pipeline
2007	Spinal Therapeutics / Arrow Therapeutics (UK)	-	Smaller deals, focused on anti-infectives and niche therapeutic areas
2012	Amylin Pharmaceuticals	approx. $3.4 billion	A joint acquisition together with Bristol-Myers Squibb. This was an expansion of AZ's existing diabetes alliance.
2013	Amplimmune (USA)	approx. $225 million.	Strengthened AZ's immuno-oncology capabilities, particularly in T-cell modulation
2013	Pearl Therapeutics	approx. $1.15 billion	Access to Pearl's promising pipeline of inhaled bronchodilator products for the treatment of asthma and chronic obstructive pulmonary disease.
2010's	Zoladex rights in certain markets	-	Regional acquisitions of rights to specific drugs to consolidate oncology portfolio.
2015	ZS Pharma	approx. $2.7 billion	Primarily driven by AZ's goal to strengthen its cardiovascular and metabolic disease portfolio.
2016	Acerta Pharma (Netherlands/USA)	$4 billion (initial + milestones)	Gave AZ access to acalabrutinib, a next-generation BTK inhibitor for blood cancers

Table 3. Key alliances by AstraZeneca

Year	Name of partner	Therapeutic area and significance
2004	FibroGen (USA)	focused on anemia drug Roxadustat
2010	Rigel Pharmaceuticals (USA)	immunology/oncology
2015	Ionis Pharmaceuticals (USA)	Antisense therapies. Multiple collaborations
2018	Innate Pharma (France)	immuno-oncology
2020	Silence Therapeutics (UK)	RNAi partnership

Table 4. Number of academic papers and number of international academic papers in each R&D site[9]

R&D Site	R&D Employees estimates[10]	Non-Academic Pubs.	Academic Pubs.	National Academic Pubs.	International Academic Pubs.	Share Academic Pubs.	Share International Academic Pubs.
Bangalore (IND)	300-700	781	178	36	142	19%	80%
Boston (US)[11]	800-1,500	1,871	404	186	218	18%	54%
Cambridge (UK)	2,500-3,500	5,466	4,209	1,386	2,823	44%	67%
Gaithersburg (US)	2,500-4,000	1,052	660	257	403	39%	61%
Loughborough (UK)	1,200	1,221	517	392	125	30%	24%
Lund (SWE)	900	684	426	134	292	38%	69%
Macclesfield (UK)	300-800	5,835	2,645	1,718	927	31%	35%
Molndal (SWE)	2,000-3,000	5,190	3,795	1,244	2,551	42%	67%
Montreal (CAN)	100-250	390	105	81	24	21%	23%
Osaka (JAP)	300-600	79	117	94	23	60%	20%
Reims (FRA)	50-200	296	16	8	8	5%	50%
Shanghai (CHI)	100-1800	222	311	253	58	58%	19%
Sodertalje (SWE)	1,400-1,800	3,041	1,952	655	1,297	39%	66%
Waltham (US)	700-1200	1,032	361	173	188	26%	52%

[9] If a focal author publishes two papers she is counted twice, if a paper includes two AZ focal authors the paper is counted twice.

[10] The number of R&D employees in each site is based on an AI-assisted online search on websites, company annual reports and other online sources to reach an approximate understanding of the size of the R&D center. The number of employees refers to the last available year, so for R&D centers that were closed it refers to the period before the closing. For some still existing sites it refers to the last available year.

[11] Differently from the previous table, here we separate the more recent R&D center of Boston from the older one based in Waltham Massachusetts, although they are less than 25 km distant. This allows us to look at the specific publication dynamics of each of the two research centers.

Wilmington (US)	400-900	1,616	714	449	265	31%	37%
Small R&D sites							
Mountain View (US)	150-350	4	0	0	0	0%	0%
New York (US)	50-200	0	2	1	1	100%	50%
Warsaw (POL)	700-1,500	5	6	3	3	55%	50%
Tokyo (JAP)	100-300	4	7	6	1	64%	14%
Total		**28,789**	**16,425**	**7,076**	**9,349**	**36%**	**57%**

Table 5. Typology of R&D centers, based on academic collaborations[12]

R&D center	Size center	High academic engagement (>35%)	International academic networks (>50%)	Typology	Center open at the end of the period
Cambridge	LARGE	YES	YES	Global Academic Hub	YES
Molndal	LARGE	YES	YES	Global Academic Hub	YES
Sodertalje	LARGE	YES	YES	Global Academic Hub	NO
Macclesfield	LARGE	NO	NO	Corporate R&D Hub	YES
Gaithersburg	MEDIUM	YES	YES	International Academic center	YES
Lund	MEDIUM	YES	YES	International Academic center	NO
Boston	MEDIUM	NO	YES	International R&D center	YES
Waltham	MEDIUM	NO	YES	International R&D center	YES
Loughborough	MEDIUM	NO	NO	Local R&D center	NO
Wilmington	MEDIUM	NO	NO	Local R&D center	NO
Bangalore	SMALL	NO	YES	Local R&D Outpost	NO
Reims	SMALL	NO	YES	Local R&D Outpost	NO
Osaka	SMALL	YES	NO	Outpost for local academic search	YES
Shanghai	SMALL	YES	NO	Outpost for local academic search	YES
Montreal	SMALL	NO	NO	Local R&D center	NO

Table 6. Description of the variables used in the regressions

Dependent variable	Description of variables
Recombinant-Novelty of the paper (R)	Recombinant novelty measured using document-level semantic representations derived from word embeddings, following the procedure by Shibayama, Yin and Matsumoto (2021)
Element-Novelty of the paper (G)	Element novelty measured using contextual word-embedding models to identify documents that introduce conceptually new elements, rather than merely recombining known ones (Yin et al., 2023)
Share of new international contacts	The number of coauthors of the AZ authors that are new for them and that are affiliated to other institutions (universities or companies), divided by all possible pair combinations between the AZ authors and their coauthors (based on Liu et al. 2022).
Independent variables	
Academic social tie	Dummy equal to 1 if in the past (up to 3 years before) at least one of the AZ authors has published a paper with the same affiliation of one (or more) of the academic coauthors.
Share of new academic contacts	Number of AZ authors that coauthor for the first time with each of the academic coauthors in a specific paper, divided by all the possible combinations between AZ and academic authors present in the paper (based on Liu et al. 2022)

[12] We did not include the smallest R&D sites in terms of publications (Mountain View, New York, Warsaw and Tokyo), since there are not enough publications to draw robust conclusions about their academic engagement.

Institutional agreement	Dummy equal to 1 if at least one of the academic co-authors are affiliated to a university that has an ongoing (at the time of the publication) institutional agreement with AZ.
Average distance to R&D sites	For each academic coauthor in a paper we calculate the km distance between his/her institution and the closest AstraZeneca R&D site, then take the average of such distance for all the academic coauthors in the paper and do a log transformation of 1 plus the value of such average distance.
Average past number of papers	Average of the number of papers written by the AZ authors of the paper in the previous years
Productive AZ researcher	Dummy equal to 1 if there is at least one AZ author in the paper, with a relatively high number of past publications (more than 15)
Num total authors	Total number of authors in the paper
Num AZ authors	Number of AZ authors in the paper
Num academic authors	Number of academic authors in the paper
Num non-AZ company authors	Number of authors in the paper with a *company* affiliation that is not AstraZeneca
Num PRO authors	Number of authors in the paper with a Public Research Organization affiliation
Num hospital authors	Number of authors in the paper with an affiliation to a hospital
Num countries academics	Num of different countries of affiliation of the academic authors in the paper
Clinical research	Dummy equal to 1 if the paper deals with clinical research (as opposed to basic research)
R&D typology	Dummies equal to 1 if the AZ researchers are based in one of the 7 R&D centres typologies introduced in Table 5.

Table 7. Descriptive statistics

	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
		International collaborations					***National collaborations***			
Dependent variables										
Recombinant novelty (r100)	2544	1.032	0.172	0.388	1.400	2018	1.055	0.159	0.364	1.460
Recombinant novelty (r99)	2544	0.935	0.151	0.387	1.300	2018	0.960	0.141	0.341	1.309
Recombinant novelty (r95)	2544	0.818	0.140	0.352	1.169	2018	0.844	0.132	0.311	1.203
Recombinant novelty (r90)	2544	0.743	0.135	0.324	1.131	2018	0.772	0.129	0.272	1.109
Element novelty (g0)	2806	0.054	0.027	0.008	0.301	2557	0.053	0.028	0.006	0.274
Element novelty (g1)	2806	0.237	0.064	0.098	0.532	2557	0.247	0.067	0.079	0.575
Element novelty (g5)	2806	0.321	0.074	0.143	0.645	2557	0.330	0.079	0.117	0.653
Element novelty (g10)	2806	0.393	0.081	0.183	0.709	2557	0.393	0.088	0.157	0.691
Independent variables										
Academic social tie	2806	0.034	0.181	0	1	2557	0.127	0.333	0	1
Share of new contacts	2806	0.616	0.395	0	1	2557	0.569	0.401	0	1
Institutional agreement	2806	0.202	0.402	0	1	2557	0.330	0.470	0	1
Team characteristics										
Productive AZ researcher	2806	0.837	0.865	0	11	2557	0.768	0.861	0	8
Distance to R&D sites (km log)	2806	6.060	1.539	0	9.164	2557	4.322	1.768	0	8.961
Average past num papers	2806	49.857	96.537	0	798	2557	35.445	60.105	0	988
Num AZ authors	2806	2.025	1.881	1	26	2557	2.087	1.883	1	34
Num academic authors	2806	4.407	3.945	1	42	2557	3.521	2.799	1	29
Num non-AZ company authors	2806	0.522	1.856	0	27	2557	0.381	1.553	0	28
Num PRO authors	2806	1.521	3.263	0	46	2557	0.684	2.345	0	45
Num. hospital authors	2806	2.016	3.436	0	34	2557	0.763	2.382	0	48
Num total authors	2806	11.141	8.061	2	99	2557	7.732	5.656	2	90
Num countries academics	2806	1.648	0.997	1	11	2557	1	0	1	1

Type of R&D center										
Corporate R&D hub	2806	0.095	0.293	0	1	2557	0.236	0.425	0	1
Global academic hub	2806	0.703	0.457	0	1		0.463	0.499	0	1
International Academic center	2806	0.067	0.251	0	1	2557	0.039	0.194	0	1
International R&D center	2806	0.027	0.161	0	1	2557	0.038	0.192	0	1
Local R&D centre	2806	0.043	0.202	0	1	2557	0.120	0.325	0	1
Local R&D outpost	2806	0.007	0.086	0	1	2557	0.015	0.123	0	1
Local search	2806	0.003	0.053	0	1	2557	0.021	0.142	0	1
Other R&D centers	2806	0.001	0.027	0	1	2557	0.002	0.040	0	1
Other AZ facility	2806	0.054	0.226	0	1	2557	0.066	0.248	0	1
Type of medical research										
Clinical research	2806	0.753	0.431	0	1	2557	0.587	0.493	0	1
Oncology	2806	0.170	0.375	0	1	2557	0.160	0.367	0	1
Immunology	2806	0.141	0.348	0	1	2557	0.135	0.342	0	1
Cardio	2806	0.230	0.421	0	1	2557	0.110	0.312	0	1
Infection	2806	0.066	0.248	0	1	2557	0.043	0.204	0	1
Endocrinology	2806	0.117	0.322	0	1	2557	0.089	0.284	0	1
Respiratory	2806	0.130	0.337	0	1	2557	0.087	0.282	0	1
Musculoskeletal	2806	0.068	0.252	0	1	2557	0.051	0.221	0	1
Central Nervous system	2806	0.131	0.337	0	1	2557	0.097	0.296	0	1

Figure 1. Average *Recombinant-type novelty* among international and domestic academic papers

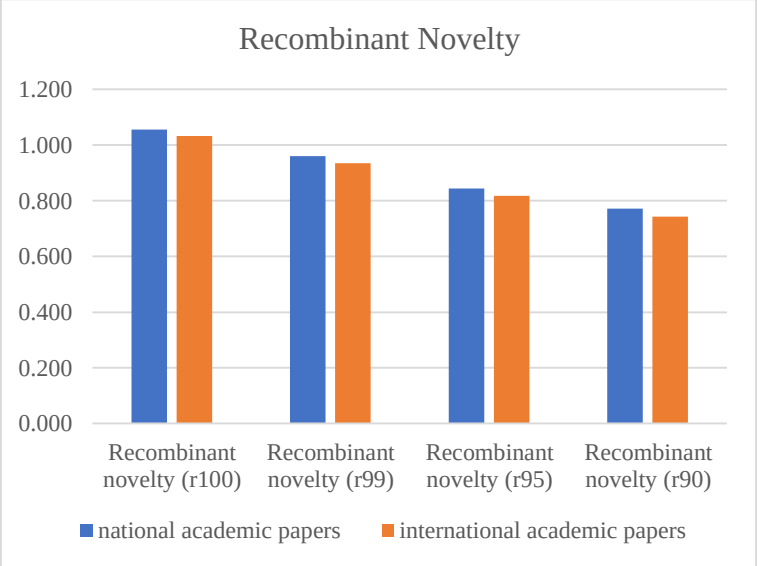


Figure 2. Average *Element-type novelty* among international and domestic academic papers

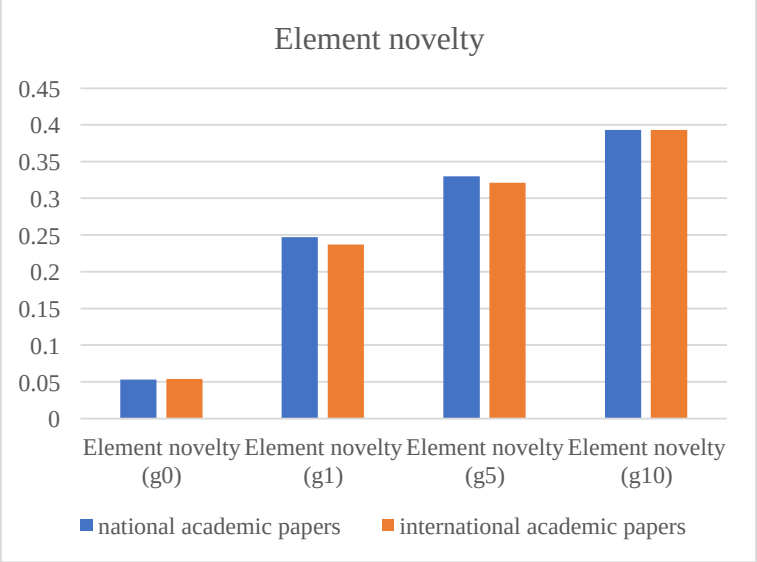


Table 8. Difference in difference estimation. The dependent variable is the *share of new international contacts*.

	(1) t=2013	(2) t=2014	(3) t=2015
Academic paper (0/1)	-0.212***	-0.210***	-0.211***
	(0.016)	(0.015)	(0.014)
From year t	0.013	0.011	0.002
	(0.034)	(0.034)	(0.034)
Academic paper * from year t	0.064***	0.066***	0.077***
	(0.020)	(0.020)	(0.020)
Controls			
Num AZ authors	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)
R&D center typology			
Baseline category: Global academic hub			
Corporate R&D hub	-0.131***	-0.131***	-0.130***
	(0.011)	(0.011)	(0.011)
International academic centre	-0.057***	-0.057***	-0.056***
	(0.018)	(0.018)	(0.018)
International R&D centre	-0.151***	-0.152***	-0.152***
	(0.021)	(0.021)	(0.021)
Local R&D center	-0.184***	-0.185***	-0.185***
	(0.015)	(0.015)	(0.015)
Local R&D outpost	-0.082**	-0.082**	-0.081**
	(0.037)	(0.037)	(0.037)
Local search	-0.182***	-0.182***	-0.182***
	(0.035)	(0.035)	(0.035)
Other R&D centre	-0.127	-0.127	-0.122
	(0.168)	(0.168)	(0.167)
Outside AZ R&D main centers	-0.109***	-0.109***	-0.109***
	(0.019)	(0.019)	(0.019)
Time dummies	YES	YES	YES
Constant	0.455***	0.453***	0.454***
	(0.029)	(0.029)	(0.029)
Observations	8939	8939	8939
R^2	0.0757	0.0758	0.0762

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3. Parallel trends assumption before 2013

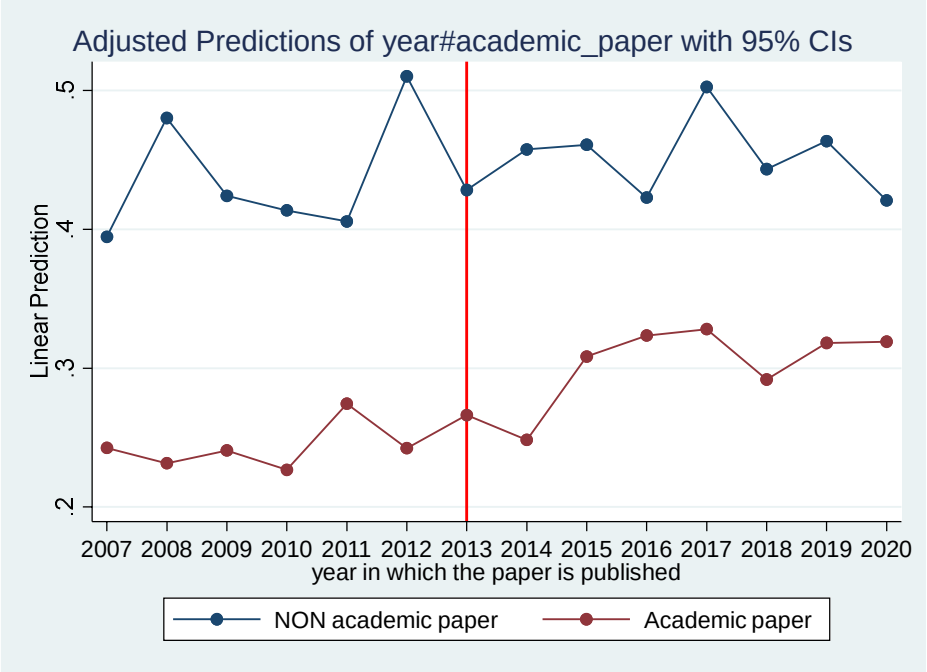


Table 9. The role of social ties on the novelty of the papers (Recombinant-type novelty).

	(1) Internat. **r100**	(2) National **r100**	(3) Internat. **r99**	(4) National **r99**	(5) Internat. **r95**	(6) National **r95**	(7) Internat. **r90**	(8) National **r90**
Academic social tie	0.040***	0.011	0.038***	0.008	0.031***	0.008	0.027**	0.011
	(0.014)	(0.010)	(0.013)	(0.009)	(0.012)	(0.008)	(0.012)	(0.008)
Share of new contacts	0.052***	-0.000	0.040***	-0.000	0.038***	-0.001	0.039***	0.000
	(0.009)	(0.009)	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)
Institutional agreement	0.004	-0.014*	0.005	-0.015**	0.001	-0.014**	0.001	-0.013**
	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Team level controls								
Productive AZ researcher	0.008*	0.007	0.005	0.007*	0.003	0.008**	0.003	0.007*
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Distance to R&D sites	-0.008***	-0.007***	-0.007***	-0.006***	-0.007***	-0.006***	-0.006***	-0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Average past num papers	0.000**	-0.000*	0.000*	-0.000	0.000*	-0.000	0.000**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Num AZ authors	0.001	0.008**	0.003	0.008**	0.004*	0.008***	0.004*	0.008***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Num non-AZ company authors	0.000	0.005	0.002	0.005	0.002	0.005*	0.003	0.005*
	(0.003)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Num PRO authors	0.004*	0.005	0.004**	0.004	0.004**	0.004	0.004**	0.004
	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Num hospital authors	-0.007***	-0.001	-0.005**	0.000	-0.005**	0.001	-0.004**	0.001
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Num academic authors	0.005**	0.009***	0.004**	0.008***	0.004**	0.008***	0.004**	0.007***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Num total authors	0.002	-0.002	0.001	-0.003	0.001	-0.003	0.001	-0.003
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Num countries academics	-0.016***	-	-0.013***	-	-0.012***	-	-0.012***	-
	(0.004)		(0.003)		(0.003)		(0.003)	
Company areas of expertise								
Clinical research	-0.073***	-0.035***	-0.060***	-0.032***	-0.054***	-0.030***	-0.056***	-0.032***
	(0.007)	(0.008)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Clinical research	-0.073***	-0.035***	-0.060***	-0.032***	-0.054***	-0.030***	-0.056***	-0.032***
	(0.007)	(0.008)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Oncology	-0.017*	-0.008	-0.013*	-0.008	-0.021***	-0.009	-0.026***	-0.009
	(0.009)	(0.010)	(0.008)	(0.009)	(0.007)	(0.008)	(0.007)	(0.008)
Immunology	0.013	0.012	0.008	0.002	0.001	-0.001	-0.001	-0.004
	(0.009)	(0.010)	(0.008)	(0.009)	(0.008)	(0.008)	(0.007)	(0.008)
Cardiovascular	-0.031***	-0.042***	-0.021***	-0.035***	-0.022***	-0.040***	-0.026***	-0.042***
	(0.008)	(0.011)	(0.007)	(0.010)	(0.006)	(0.009)	(0.006)	(0.009)

Infection	0.011	0.000	0.008	-0.000	0.006	0.004	0.007	0.004
	(0.013)	(0.017)	(0.011)	(0.015)	(0.010)	(0.014)	(0.010)	(0.014)
Endocrinology	-0.030***	-0.037***	-0.030***	-0.039***	-0.035***	-0.046***	-0.038***	-0.048***
	(0.011)	(0.012)	(0.010)	(0.010)	(0.009)	(0.009)	(0.008)	(0.009)
Respiratory	-0.036***	-0.030**	-0.027***	-0.016	-0.022**	-0.017	-0.025***	-0.018*
	(0.011)	(0.012)	(0.009)	(0.011)	(0.009)	(0.010)	(0.008)	(0.010)
Muscoloskeletal	0.041***	0.010	0.041***	0.010	0.042***	0.007	0.044***	0.005
	(0.012)	(0.012)	(0.011)	(0.011)	(0.010)	(0.011)	(0.010)	(0.010)
Central Nervous System	-0.028***	-0.005	-0.035***	-0.009	-0.038***	-0.007	-0.035***	-0.005
	-0.017*	-0.008	-0.013*	-0.008	-0.021***	-0.009	-0.026***	-0.009
R&D center typology								
Baseline: Global academic hub								
Corporate R&D hub	0.017	0.004	0.016*	0.006	0.016*	-0.000	0.016*	-0.003
	(0.011)	(0.009)	(0.010)	(0.008)	(0.009)	(0.007)	(0.009)	(0.007)
International academic center	-0.010	-0.019	-0.010	-0.022	-0.009	-0.026*	-0.010	-0.026*
	(0.013)	(0.020)	(0.012)	(0.017)	(0.011)	(0.015)	(0.010)	(0.015)
International R&D center	0.012	0.007	0.015	-0.007	0.009	-0.013	0.009	-0.016
	(0.016)	(0.017)	(0.013)	(0.015)	(0.012)	(0.014)	(0.012)	(0.013)
Local R&D center	0.026*	-0.037***	0.022*	-0.035***	0.019*	-0.035***	0.019*	-0.038***
	(0.013)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)
Local R&D outpost	-0.022	-0.016	-0.024	-0.014	-0.029	-0.039*	-0.019	-0.054***
	(0.027)	(0.024)	(0.023)	(0.022)	(0.020)	(0.020)	(0.019)	(0.020)
Local search	0.051	-0.112***	0.046	-0.100***	0.057	-0.108***	0.062	-0.113***
	(0.050)	(0.025)	(0.042)	(0.023)	(0.042)	(0.021)	(0.039)	(0.019)
Other R&D centers	0.012	0.072*	0.003	0.068	-0.038	0.061	-0.047	0.046
	(0.070)	(0.039)	(0.044)	(0.045)	(0.075)	(0.042)	(0.085)	(0.028)
Outside of AZ R&D centers	0.004	-0.016	0.005	-0.024*	0.004	-0.022*	0.001	-0.020*
	(0.015)	(0.014)	(0.013)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
TIME DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	1.139***	1.090***	1.054***	1.008***	0.941***	0.904***	0.861***	0.834***
	(0.031)	(0.026)	(0.026)	(0.025)	(0.024)	(0.023)	(0.024)	(0.023)
Num of obs	2544	2018	2544	2018	2544	2018	2544	2018
R^2	0.182	0.123	0.164	0.118	0.173	0.131	0.192	0.144

Notes: OLS Regression. Social tie must have occurred at least 3 years before the year of the paper. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10. The role of social ties on the novelty of the papers (Element-type novelty).

	(1) Internat. g0	(2) National g0	(3) Internat. g1	(4) National g1	(5) Internat. g5	(6) National g5	(7) Internat. g10	(8) National g10
Academic social tie	0.002	0.001	0.019**	-0.001	0.022**	-0.004	0.021**	-0.005
	(0.003)	(0.002)	(0.008)	(0.004)	(0.009)	(0.005)	(0.010)	(0.005)
Share of new contacts	-0.001	0.001	0.003	0.001	0.008**	0.007*	0.013***	0.015***
	(0.001)	(0.001)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Institutional agreement	-0.000	0.002	-0.001	0.000	-0.001	0.005	0.000	0.009**
	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Team level controls								
Productive AZ researcher	0.001	0.001**	0.005***	0.009***	0.005***	0.009***	0.005***	0.009***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Distance to R&D sites	-0.000	0.000	-0.002*	0.001	-0.001	0.001	-0.001	0.002
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Average past num papers	0.000*	-0.000*	0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Num AZ authors	0.000	-0.000	0.002**	0.003**	0.001	0.004**	0.001	0.004**
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Num non-AZ company authors	0.000	-0.000	0.001	0.003*	0.000	0.003	-0.001	0.002
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Num PRO authors	0.000	-0.000	0.002**	0.002	0.002**	0.003**	0.002**	0.004**
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Num hospital authors	-0.000	0.000	-0.000	0.002	-0.000	0.002*	-0.000	0.003*
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Num academic authors	0.001*	-0.000	0.004***	0.003***	0.004***	0.005***	0.005***	0.005***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Num total authors	-0.000	0.000	-0.001*	-0.001	-0.001	-0.002	-0.001	-0.002
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Num countries academics	-0.002***	0.000	-0.010***	0.000	-0.012***	0.000	-0.012***	0.000
	(0.001)	(.)	(0.001)	(.)	(0.001)	(.)	(0.001)	(.)
Company areas of expertise								
Clinical research	-0.001	-0.004***	-0.030***	-0.023***	-0.032***	-0.027***	-0.027***	-0.023***
	(0.001)	(0.001)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Oncology	-0.000	-0.002	0.006*	0.001	0.020***	0.015***	0.032***	0.029***
	(0.001)	(0.002)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Immunology	-0.001	-0.000	0.001	-0.000	0.004	0.002	0.008*	0.006
	(0.001)	(0.002)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Cardiovascular	-0.001	0.001	-0.004	-0.015***	0.004	-0.009*	0.016***	0.004
	(0.001)	(0.001)	(0.003)	(0.004)	(0.003)	(0.005)	(0.003)	(0.005)

Infection	0.001	0.007***	0.023***	0.031***	0.026***	0.036***	0.027***	0.039***
	(0.002)	(0.003)	(0.005)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
Endocrinology	-0.002	-0.002	-0.005	-0.015***	0.001	-0.010**	0.009**	-0.003
	(0.001)	(0.002)	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.006)
Respiratory	0.001	0.003	-0.003	-0.001	-0.003	-0.005	0.000	-0.006
	(0.001)	(0.002)	(0.003)	(0.005)	(0.004)	(0.005)	(0.004)	(0.006)
Muscoloskeletal	0.000	0.007***	0.008*	0.007	0.016***	0.008	0.027***	0.016**
	(0.002)	(0.002)	(0.004)	(0.005)	(0.005)	(0.006)	(0.005)	(0.007)
Central Nervous System	-0.003**	-0.004***	-0.000	-0.000	0.004	-0.000	0.009**	0.003
	(0.001)	(0.001)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
R&D center typology *Baseline: Global academic hub*								
Corporate R&D hub	-0.002	0.000	0.012***	0.005	0.015***	0.011***	0.015***	0.014***
	(0.002)	(0.001)	(0.004)	(0.003)	(0.005)	(0.004)	(0.005)	(0.005)
International academic center	-0.000	0.001	-0.006	-0.005	-0.010**	-0.008	-0.014***	-0.009
	(0.002)	(0.003)	(0.004)	(0.007)	(0.005)	(0.008)	(0.005)	(0.008)
International R&D center	-0.004*	-0.003	0.019**	-0.003	0.028***	0.003	0.035***	0.010
	(0.002)	(0.002)	(0.008)	(0.006)	(0.008)	(0.007)	(0.009)	(0.008)
Local R&D center	0.004*	0.000	0.007	-0.010**	0.008	-0.010*	0.006	-0.005
	(0.002)	(0.002)	(0.006)	(0.005)	(0.006)	(0.005)	(0.007)	(0.006)
Local R&D outpost	-0.006*	-0.001	0.037***	-0.014	0.044***	-0.004	0.050***	0.010
	(0.004)	(0.004)	(0.011)	(0.011)	(0.013)	(0.014)	(0.016)	(0.015)
Local search	0.010	-0.003	-0.014	-0.003	-0.013	0.002	-0.006	0.006
	(0.008)	(0.003)	(0.013)	(0.009)	(0.015)	(0.010)	(0.018)	(0.011)
Other R&D centers	0.010*	0.001	0.077***	0.015	0.086***	0.009	0.072***	0.008
	(0.005)	(0.009)	(0.005)	(0.030)	(0.006)	(0.034)	(0.008)	(0.029)
Outside of AZ R&D centers	0.003	-0.000	0.007	-0.009*	0.012*	-0.010*	0.017**	-0.007
	(0.002)	(0.002)	(0.006)	(0.005)	(0.006)	(0.006)	(0.007)	(0.006)
TIME DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	0.057***	0.057***	0.266***	0.267***	0.342***	0.341***	0.393***	0.389***
	(0.004)	(0.004)	(0.012)	(0.009)	(0.015)	(0.011)	(0.016)	(0.013)
N	2806	2557	2806	2557	2806	2557	2806	2557
R^2	0.177	0.172	0.160	0.121	0.166	0.116	0.163	0.107

Notes: OLS Regression. Social tie must have occurred at least 3 years before the year of the paper. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11. Distinguishing between typology of R&D centers (Recombinant novelty)

	(1) r100	(2) r99	(3) r95	(4) r90	(5) r100	(6) r99	(7) r95	(8) r90
	Only international academic papers				Only international academic papers			
	Global Academic Hubs				**Other R&D centers**			
Academic social tie	0.035**	0.039**	0.033**	0.030**	0.055***	0.040**	0.029*	0.020
	(0.017)	(0.016)	(0.015)	(0.015)	(0.021)	(0.019)	(0.017)	(0.017)
Paper-level controls	YES	YES	YES	YES	YES	YES	YES	YES
R&D center typology	YES	YES	YES	YES	YES	YES	YES	YES
Company areas of exp. dummies	YES	YES	YES	YES	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES	YES	YES	YES	YES
Constant	1.176***	1.082***	0.966***	0.885***	1.087***	1.021***	0.928***	0.854***
	(0.036)	(0.029)	(0.027)	(0.026)	(0.080)	(0.068)	(0.065)	(0.063)
N	1784	1784	1784	1784	760	760	760	760
R^2	0.192	0.175	0.183	0.202	0.198	0.178	0.188	0.210

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12. Distinguishing between typology of R&D centers (Element novelty).

	(1) g0	(2) g1	(3) g5	(4) g10	(5) g0	(6) g1	(7) g5	(8) g10
	Only international academic papers				*Only international academic papers*			
	Global Academic Hubs				**Other R&D centers**			
Academic social tie	0.003	0.020**	0.021**	0.019*	0.003	0.013	0.024	0.031
	(0.004)	(0.008)	(0.010)	(0.011)	(0.005)	(0.016)	(0.022)	(0.025)
Paper-level controls	YES	YES	YES	YES	YES	YES	YES	YES
R&D center typology	YES	YES	YES	YES	YES	YES	YES	YES
Company areas of exp. dummies	YES	YES	YES	YES	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES	YES	YES	YES	YES
Constant	0.054***	0.268***	0.346***	0.398***	0.073***	0.250***	0.323***	0.380***
	(0.005)	(0.014)	(0.017)	(0.019)	(0.013)	(0.027)	(0.030)	(0.032)
N	1973	1973	1973	1973	833	833	833	833
R^2	0.171	0.142	0.141	0.136	0.227	0.244	0.260	0.258

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13. Interaction Social ties and distance to AZ R&D centers (Recombinant-novelty).

	(1) r100 International	(2) r99 International	(3) r95 International	(4) r90 International
Academic social tie	-0.059	-0.053	-0.052	-0.034
	(0.045)	(0.037)	(0.036)	(0.034)
Distance to R&D sites	-0.009***	-0.008***	-0.008***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Academic social tie * Distance to R&D centers	0.016**	0.015**	0.013**	0.010*
	(0.007)	(0.006)	(0.006)	(0.005)
Other controls	YES	YES	YES	YES
R&D center typology	YES	YES	YES	YES
Company areas of exp. dummies	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES
Constant	1.142***	1.057***	0.944***	0.863***
	(0.032)	(0.026)	(0.024)	(0.024)
N	2544	2544	2544	2544
R^2	0.183	0.165	0.174	0.192

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14. Interaction Social ties and distance to AZ R&D centers (Element-novelty).

	(1) g0 International	(2) g1 International	(3) g5 International	(4) g10 International
Academic social tie	0.004	0.074***	0.081***	0.080***
	(0.010)	(0.022)	(0.028)	(0.031)
Distance to R&D sites	-0.000	-0.001	-0.001	-0.001
	(0.000)	(0.001)	(0.001)	(0.001)
Academic social tie * Distance to R&D sites	-0.000	-0.009***	-0.010**	-0.010**
	(0.001)	(0.004)	(0.004)	(0.005)
Other controls	YES	YES	YES	YES
R&D center typology	YES	YES	YES	YES
Company areas of exp. dummies	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES
Constant	0.057***	0.263***	0.339***	0.390***
	(0.004)	(0.012)	(0.015)	(0.016)
N	2806	2806	2806	2806
R^2	0.177	0.162	0.168	0.165

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 4. Marginal effects: interaction between Academic social tie and Distance to R&D centers (Recombinant novelty)

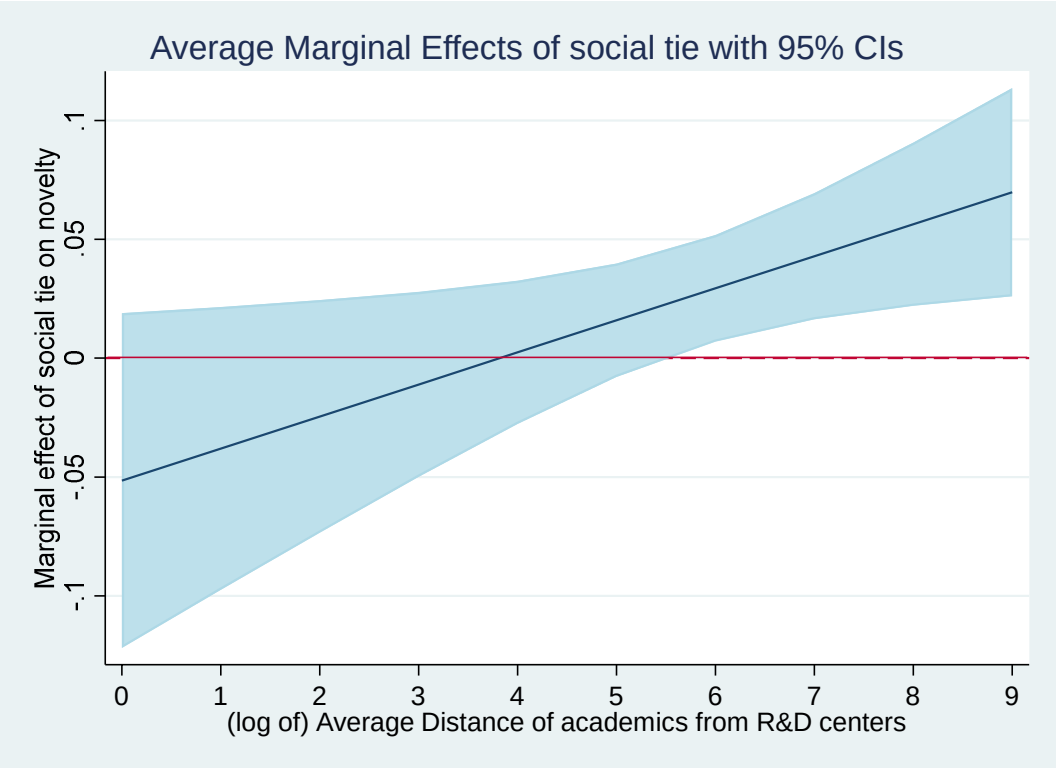
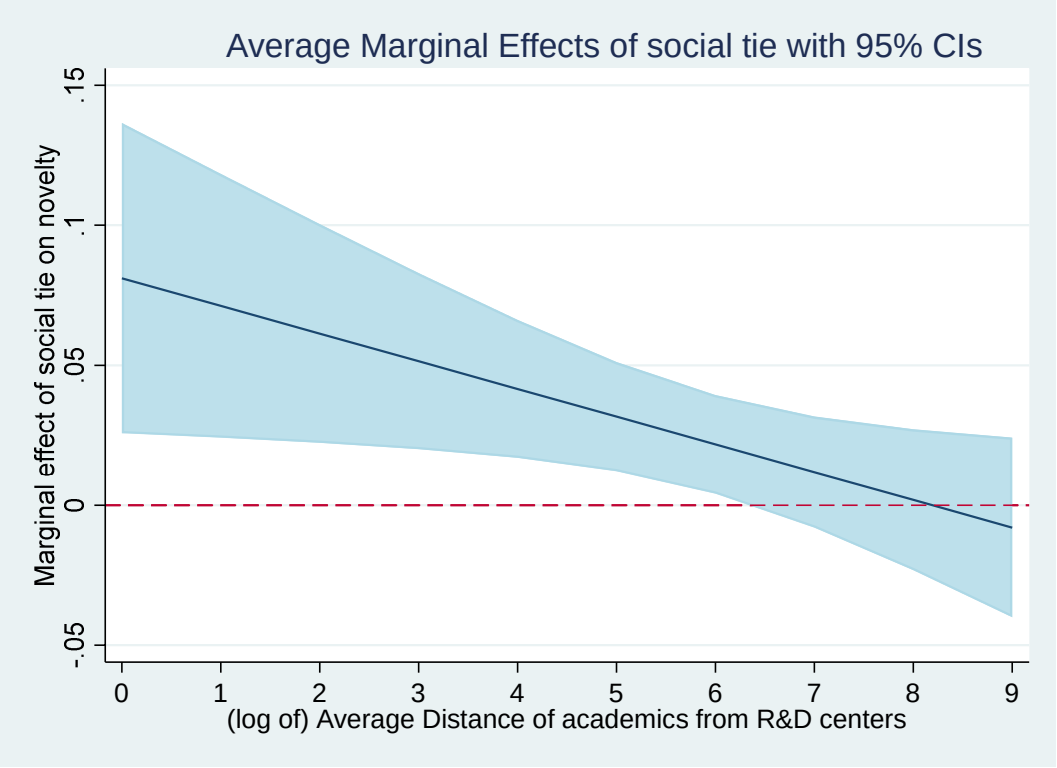


Figure 5. Marginal effects: interaction between Academic social tie and Distance to R&D centers (Element novelty)



APPENDICES

Appendix A: Dataset construction

Publication data

We downloaded data from Scopus about all the papers that were published between 2000 and 2020, in which at least one of the authors had an Astra-Zeneca affiliation. We used the search terms like Astra*, AstraZeneca, Zeneca*. We manually checked the articles to make sure that they were AstraZeneca publications since there also exist other companies with similar names such as Astra Rocket and Astra International. We also went through AstraZeneca's acquisition history to make sure we included publications that were written under the name of acquired companies but only after the companies were formally acquired by AstraZeneca. These include for example KuDOs (2005), MedImmune Biologics incl. Cambridge Antibody Technology (2006) and MedImmune (2007), Arrow Therapeutics (2007), Amylin Pharmaceuticals (2012), Spirogen (2013), Pearl Therapeutics (2013), Omthera Pharmaceuticals (2013), ZS Pharma (2015), and some additional small companies.

Through this procedure we were able to identify 17,522 papers. We did not include papers with more than 100 authors (only a few papers had so many authors), because we considered that authorship in these cases did not imply any real collaboration with all the other coauthors of the paper. We recovered information about all the 59,021 authors included in such papers. For the 12,789 authors with an AstraZeneca affiliation, using their Scopus ID, we downloaded information in Scopus about all the papers that they published throughout their career (hence also before 2000), regardless of their affiliation. In such a way we could retrieve information also about the publishing activity of AstraZeneca authors before the chosen period of 2000-2020, to track their professional history. We can see their history of affiliations and the location of the affiliation. We can also know if the coauthors with whom they publish in the period 2000-2020 have coauthored with them in the past or not.

Recovering the publication history of AstraZeneca authors

We first downloaded data from Scopus about all the papers that were published between 2000 and 2020 (henceforth *focal papers*), in which at least one author had an Astra-Zeneca affiliation. We define this set of authors the "*focal authors*". Among these "focal authors", which include also authors that have never worked for AZ, we are interested in AZ authors, i.e. those authors that at least at some point in their career (between 2000 and 2020) published a paper with an AZ affiliation. This strategy allows us to identify all the papers of Astra Zeneca in the period 2000-2020, all the authors that in that period were affiliated with AZ, and their co-authors.

Using the Scopus author id we downloaded from Scopus all the papers for each of the focal authors that had an AZ affiliation between 2000 and 2020. For each AZ focal author we downloaded all the papers available in Scopus, regardless of the year in which it was published. This means for example that we may have cases in which a focal author with an AZ affiliation in the period 2000-2020 may also have papers without an AZ affiliation, before (or after) that

period.[13] We retrieved all the information related to the articles of the AZ focal authors. This included affiliation name and affiliation id, city and country of the affiliation for each of the co-authors of the paper, the id of the paper, typology of paper (article, note, book, etc.) and other variables. The most important variables are a) the affiliation name, which allows us to identify the institution of affiliation of the authors of the paper and whether it is a university, company, hospital, public research organization, etc., and b) the location of the institution, in order to identify collaboration within the same country or international ones. The identification of the institution was done using different search words like univ., university, faculty, hospital, hospital, sjukhus, clinic, ltd., inc., ab.. For institutions where none of these search words were found we manually assigned the type by searching for the institution online or asking around in our academic and professional network.

In such a way we could retrieve information also about the publishing activity of focal authors before (and after) the chosen period of 2000-2020, to track their educational and professional history. Most importantly we were able to see their history of affiliations and the location of the affiliation. We can also know if the coauthors with whom they publish in the period 2000-2020 have coauthored with them in the past.

Authors with double affiliations: AstraZeneca and academic

In many cases the authors listed among the papers that we identified as AZ papers had more than one affiliation. In these cases, Scopus data could not handle the complexity of the data. Many authors have an academic affiliation and AZ affiliation. In these cases, Scopus usually attributes only the first affiliation, be it a company or an academic institution. Hence for many of the papers that we had originally selected because at least one of the authors had an AZ affiliation, when we downloaded data about authors' affiliations we could not find any author with an AZ affiliation, because AZ was not the first affiliation listed in the paper and hence Scopus dropped the AZ affiliation for that specific author. For example, an author that has two affiliations in a paper -Department of Biology of Cambridge University and the R&D center of AZ in Cambridge- would only have one affiliation id attributed by Scopus and that is the one of Cambridge University. This means that, if this author is the only author in that paper with an AZ affiliation, among the whole list of co-authors in the paper we would not see any AZ affiliation. In all these cases then we had to implement a cleaning procedure to make sure that we do not miss AZ affiliations. Scopus also provides a string that includes all the affiliations listed in the paper for each author. This data is not processed by Scopus, i.e. there is not a simple way to derive the affiliation id of the institutions listed. For all the papers that we selected and for which we could not find any author with an AZ affiliation we used the original string provided by Scopus with the affiliations' name, isolate the AstraZeneca affiliation and created a new variable that listed AZ as a second affiliation. Besides adding the affiliation name of the institution, we also had to identify the city and country of the second affiliation. This was done manually by checking the city and country of the second affiliation listed in the original string

[13] We also have information about additional co-authors that are not focal authors, i.e. they did not publish a paper in 2000-2020 where there was at least an author with an AZ affiliation (see the cases of author M, N and O in the Figure). For those authors we do not have complete information about their publication history.

of affiliations. In total we found that AstraZeneca was listed as the second affiliation for 2,076 authors in the period 2000-2020. For each of these authors this second affiliation issue often occurred in more than one paper so in the end we recovered AstraZeneca as the second affiliation in 11,098 cases.

Cleaning the location of AZ affiliations

Among the authors with AZ affiliation, Scopus attributes an affiliation id, the city and the country of the affiliation. After a first check we realized however that in more than 92% of the cases all the authors with an AZ affiliation had either an affiliation in the AZ subsidiary of Cambridge or in the AZ subsidiary of Molndal, close to Goteborg in Sweden. Therefore, we implemented a series of routine checks to see if in the original string of the affiliation names that we recovered from Scopus the name of the AZ location listed was different from the one that Scopus routinely attributed. By looking at the original string containing information about the affiliation names we were able to implement in Stata some changes to attribute the correct affiliation to the AZ authors. Our strategy was to check if the terms "Cambridge" and "Molndal" were included in the original string of the affiliation. When this was not the case, we codified the real location of the AZ subsidiary and attributed this new location to that specific AZ author. We did this for 9,134 AZ authors. After our cleaning procedure the share of AZ authors with an affiliation in either Cambridge or Molndal dropped from 92% to 39%, as shown in Tables A1 and A2.

Table A1. Number of times we observe an authorship from an *AstraZeneca author* with an affiliation in a specific city. Most common cities BEFORE correcting for the AZ affiliations.

City	number	%	Cum. %
Cambridge	34,216	63.48	63.48
Sodertalje	15,746	29.21	92.69
Stockholm	565	1.05	93.74
Gothenburg	393	0.73	94.46
Uppsala	229	0.42	94.89
London	182	0.34	95.23
New York	159	0.29	95.52
Gaithersburg (US)	144	0.27	95.79
Lund	87	0.16	95.95
Boston	83	0.15	96.10
Manchester	82	0.15	96.26
Montreal	74	0.14	96.39
Madrid	51	0.09	96.49
Brisbane	48	0.09	96.58
Aurora	39	0.07	96.65
Bethesda (Maryland, US)	37	0.07	96.72
Brentford (UK)	36	0.07	96.79
Philadelphia	35	0.06	96.85
San Diego	34	0.06	96.91
Barcelona	32	0.06	96.97

Table A2. Number of times we observe an authorship from an *AstraZeneca author* with an affiliation in a specific city. Most common cities AFTER correcting for the AZ affiliations.

City	number	%	Cum. %
Cambridge	10,867	20.32	20.32
Molndal	10,219	19.11	39.42
Macclesfield	9,868	18.45	57.88
Sodertalje	5,758	10.77	68.64
Wilmington	2,901	5.42	74.07
Boston	2,699	5.05	79.11
Loughborough	2,186	4.09	83.20
Gaithersburg	1,683	3.15	86.35
Waltham	1,534	2.87	89.21
Lund	1,297	2.43	91.64
Bangalore	1,089	2.04	93.67
Montreal	613	1.15	94.82
Shanghai	555	1.04	95.86
Reims	321	0.60	96.46
Brixham	292	0.55	97.00
Gothenburg	208	0.39	97.39
Osaka	205	0.38	97.78
Madrid	192	0.36	98.14
Stockholm	165	0.31	98.44
Barcelona	106	0.20	98.64

Table A3. Number of academic (and non-academic) papers per year

year	Academic papers	Non academic papers	share academic papers	total num of papers
2000	198	353	35.93	551
2001	189	265	41.63	454
2002	186	289	39.16	475
2003	278	322	46.33	600
2004	316	356	47.02	672
2005	372	428	46.50	800
2006	376	418	47.36	794
2007	347	422	45.12	769
2008	333	389	46.12	722
2009	353	424	45.43	777
2010	352	396	47.06	748
2011	364	441	45.22	805
2012	420	553	43.17	973
2013	388	466	45.43	854
2014	457	511	47.21	968
2015	484	476	50.42	960
2016	570	485	54.03	1,055

2017	547	487	52.90	1,034
2018	615	496	55.36	1,111
2019	662	497	57.12	1,159
2020	761	480	61.32	1,241
total	8,568	8,954	48.90	17,522

Table A4. International and non-international academic papers by year.

year	international papers	non international papers	share of international	Total num of academic papers
2000	79	119	40%	198
2001	99	90	52%	189
2002	91	95	49%	186
2003	143	135	51%	278
2004	172	143	54%	316
2005	203	169	55%	372
2006	215	161	57%	376
2007	200	147	58%	347
2008	182	150	55%	333
2009	205	148	58%	353
2010	204	148	58%	352
2011	241	123	66%	364
2012	269	151	64%	420
2013	241	147	62%	388
2014	282	175	62%	457
2015	334	150	69%	484
2016	407	163	71%	570
2017	400	147	73%	547
2018	446	169	73%	615
2019	484	178	73%	662
2020	569	191	75%	761
Total	5,466	2984	64%	8,568

Table A5. Number of times we observe an authorship from an ***academic*** author with an affiliation in a specific country (the affiliation of an author is counted multiple times, according to the number of papers included in the sample in which she/he is listed as an author).

Country	num	%	Cum %
United Kingdom	11,353	36.66	36.66
United States	5,024	16.22	52.88
Sweden	4,512	14.57	67.45
Germany	1,315	4.25	71.69
Canada	1,161	3.75	75.44
China	792	2.56	78.00
Japan	633	2.04	80.04

Australia	632	2.04	82.08
Italy	573	1.85	83.93
Denmark	548	1.77	85.70
France	484	1.56	87.27
Netherlands	403	1.30	88.57
Spain	383	1.24	89.80
Finland	345	1.11	90.92
South Korea	320	1.03	91.95
Belgium	286	0.92	92.87
Switzerland	273	0.88	93.76
Norway	257	0.83	94.59
Brazil	186	0.60	95.19
Austria	156	0.50	95.69
Greece	144	0.46	96.15
Other countries	1191	3.84	100.00
Total	30,971	100.00	

Appendix B: Institutional partnerships

Based on data from PharmaDeals (IMS Health now IQVIA) and the analysis of AstraZeneca annual reports, we were able to identify the formal partnerships that AstraZeneca has signed with academic institutions over the period 2000 to 2020. Indeed, AstraZeneca often establishes collaborations lasting for several years with some universities. These collaborations or partnerships usually have some broad targets focused on specific areas of research, they may involve staff exchanges between AZ and the university personnel, it may involve research funding, sharing of infrastructure, access to samples and data, and sharing costs of clinical trials. Table 3 presents these deals. We only have the year when agreements were signed but no information regarding when the partnership ended, therefore this variable remains a 1 throughout the study period even though in reality they may have ended.

Table B1. The list of formal partnerships with academic institutions.

Year	University partner	City	Country	Therapeutic area
2000	University of California Irvine	Irvine	US	Gastro
2000	Karolinska Institutet	Stockholm	SE	All
2000	McMaster University	Hamilton	CA	Gastro
2000	Washington University	St. Louis	US	Gastro
2000	Conaris/Kiel University	Kiel	DE	Gastro/Tech
2000	The University of Manchester	Manchester	UK	Oncology/Cardiovascular
2000	Shanghai University	Shanghai	CN	CNS
2000	Rockefeller University	New York	US	CNS
2000	University of Southampton	Southampton	UK	Respiratory/Inflammation
2000	Oxford University	Oxford	UK	Respiratory/Inflammation/Tech
2000	Baylor College of Medicine	Houston	US	Tech platforms
2000	Griffith University	Nathan	AU	Tech platforms
2000	University of Pennsylvania	Philadelphia	US	Cardiovascular

2000	University of Liverpool	Liverpool	UK	Cardiovascular
2000	McGill University	Montreal	CA	Pain
2000	University of Gothenburg	Gothenburg	SE	Infection
2001	Jiao Tong University	Shanghai	CN	CNS
2003	University of Dundee	Dundee	UK	
2003	University of Gratz	Gratz	AT	
2006	University of Pennsylvania	Philadelphia	US	Cardiovascular
2006	The University of Manchester	Manchester	UK	Oncology
2006	The University of Manchester	Manchester	UK	Cardiovascular
2007	The University of Manchester	Manchester	UK	Inflammation
2007	Keio University School of Medicine	Tokyo	JP	
2007	University of Texas	Austin	US	CNS
2007	Yale University	New Haven	US	Immunology
2007	Peking University	Peking	SCN	Pharmacology
2008	Jagiellonian University	Krakow	PL	
2008	Washington University	St. Louis	US	CNS
2008	Columbia University	New York	US	Cardiovascular/Diabetes
2008	Newcastle University	Newcastle	UK	
2009	University of Virginia	Charlottesville	US	Cardiovascular
2009	Duke University	Durham	US	CNS
2009	University of Heidelberg	Heidelberg	DE	Pain
2009	Bar-Ilan University	Ramat Gan	IL	
2010	University of Pennsylvania	Philadelphia	US	CNS
2010	University of Liverpool	Liverpool	UK	Cardiovascular
2010	McGill University	Montreal	CA	Pain
2010	University College London	London	UK	Diabetes/Opthamology
2010	Peking University	Peking	CN	Pharmacology
2011	University of Michigan	Ann Arbor	US	Diabetes
2011	The University of Manchester	Manchester	UK	Inflammation
2011	Kuwait University	Kuwait	KW	Cardiovascular
2011	Indiana University	Bloomington	US	
2012	Texas A&M University	Collega Station	US	Immunology
2012	University of Dundee	Dundee	UK	Oncology
2012	Johann Wolfgang Goethe University of Frankfurt, Germany	Frankfurt	DE	
2012	Case Western Reserve University School of Medicine	Cleveland	US	Infection
2012	The University of Manchester	Manchester	UK	Oncology
2012	University of Bristol	Bristol	UK	CNS/Oncology
2012	Fudan University	Shanghai	CN	Cardiovascular
2012	University of California San Francisco	San Francisco	US	Tech
2013	Vanderbilt University	Nashville	US	CNS
2013	University of Maryland	Baltimore	US	Oncology/Respiratory/Immunology
2013	Tufts University	Medfort	US	CNS

2013	Leiden University	Leiden	NL	
2013	University of Cambridge	Cambridge	UK	Oncology/CNS
2013	University of Cambridge	Cambridge	UK	Tech
2013	University of Birmingham	Birmingham	UK	Musculoskeletal
2013	Aston University	Birmingham	UK	
2013	Johns Hopkins University	Baltimore	US	Cardiovascular/CNS/Inflammatory/Oncology/Infection
2013	Harvard University	Boston	US	Tech
2013	Uppsala University	Uppsala	SE	Cardiovascular/CNS/Inflammatory/Oncology
2013	University of Colorado	Boulder	US	Cardiovascular
2014	University of Texas	Austin	US	Immunology
2014	University of Michigan	Ann Arbor	US	Diabetes
2016	University of Leeds	Leeds	UK	Oncology
2016	University of Sheffield	Sheffield	UK	Oncology/Vaccine
2016	Harvard University	Boston	US	Oncology
2017	Washington University	St. Louis	US	Oncology
2018	University of Bern	Bern	CH	Tech
2018	University of Bonn	Bonn	DE	Tech
2019	University of Colorado	Colorado	US	Oncology
2019	University of Pittsburgh	Pittsburgh	US	Cardiovascular
2020	University of Oxford	Oxford	UK	Vaccine

Source: authors' own calculations based on the Annual reports of AstraZeneca and deal data from PharmaDeals (IMS Health).

Appendix C: Computation of the distances to AZ R&D centers

Using the Stata software geocode we also attributed to each affiliation of the authors in our sample the geographic coordinates of their institution. This required to manually check the name and location of their affiliation. Also in this case, we realized that the data provided by Scopus had some limitations, so we double checked manually the location of each affiliation to make sure that the city/country was correct. Once this task was completed, we were able to attribute to each institution a geographic coordinate and then we computed the distance to each of the R&D centers of AZ. In this way for each non-AZ author, we can compute the distance between his institutions and the closest R&D center.

Appendix D. Robustness checks

Sensitivity analysis on the lag structure of the Academic social tie

Recombinant-type novelty

Figure D1. Academic social tie. International papers vs national papers (*degrees of Recombinant-novelty*) at least 2 years before

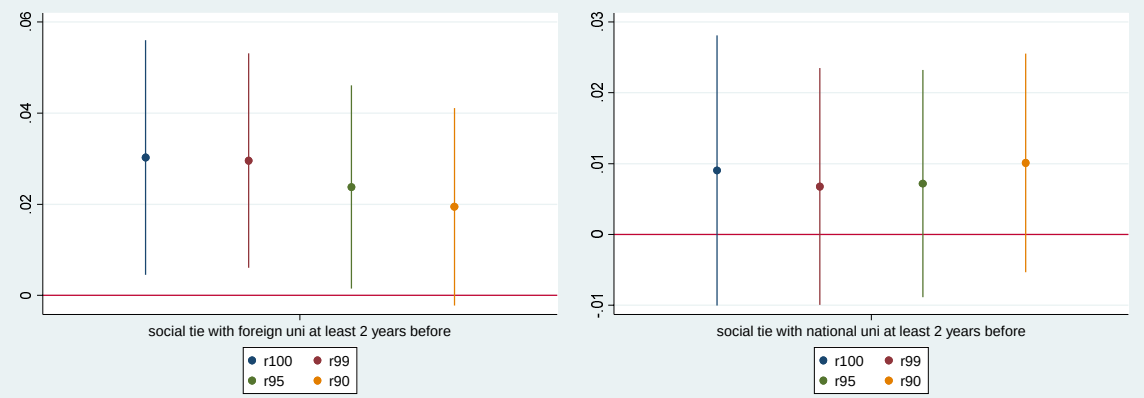


Figure D2. Academic social tie. International papers vs national papers (*degrees of Recombinant -novelty*) at least 3 years before

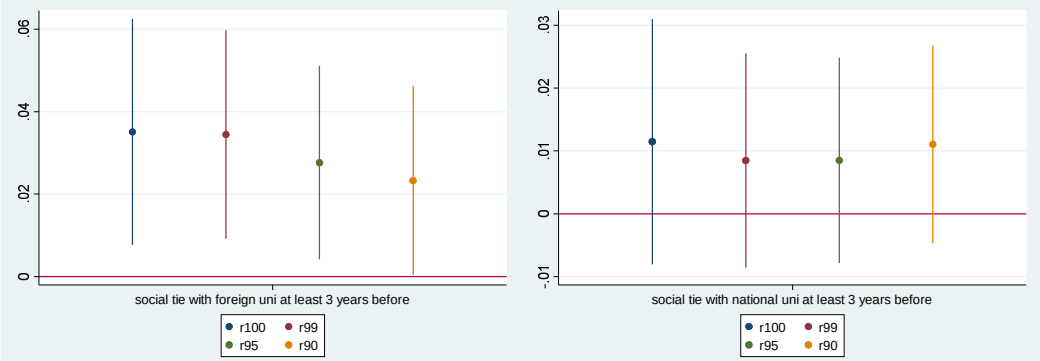


Figure D3. Academic social tie. International papers vs national papers (*degrees of Recombinant-novelty*) at least 4 years before

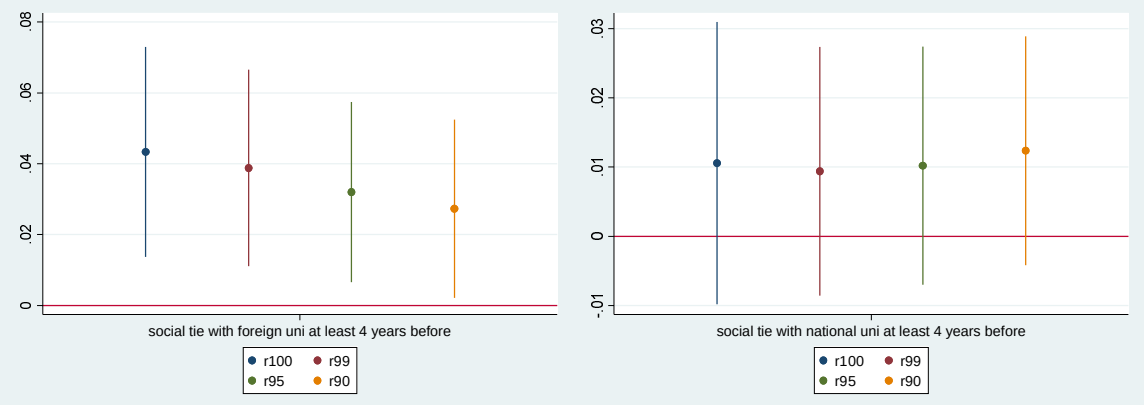
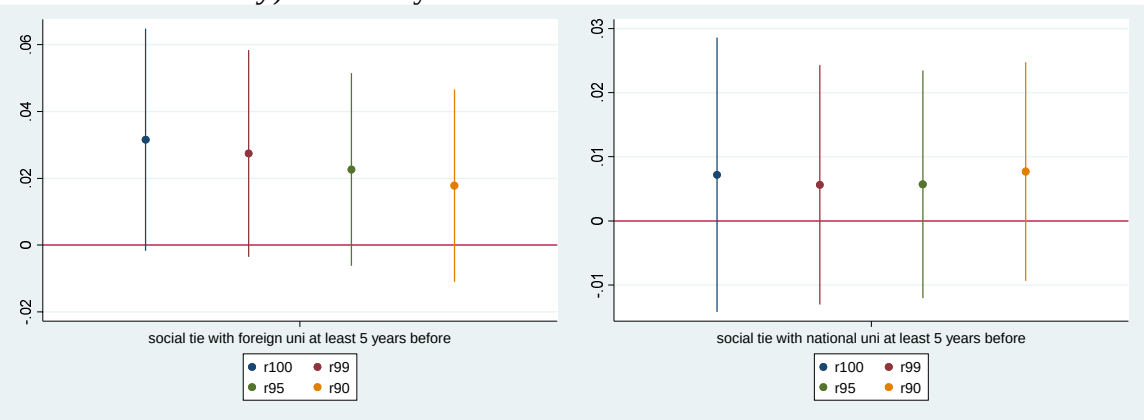


Figure D4. Academic social tie. International papers vs national papers (*degrees of Recombinant-novelty*) at least 5 years before



Element-type novelty

Figure D5. Academic social tie. International papers vs national papers (*degrees of Element-novelty*) at least 2 years before

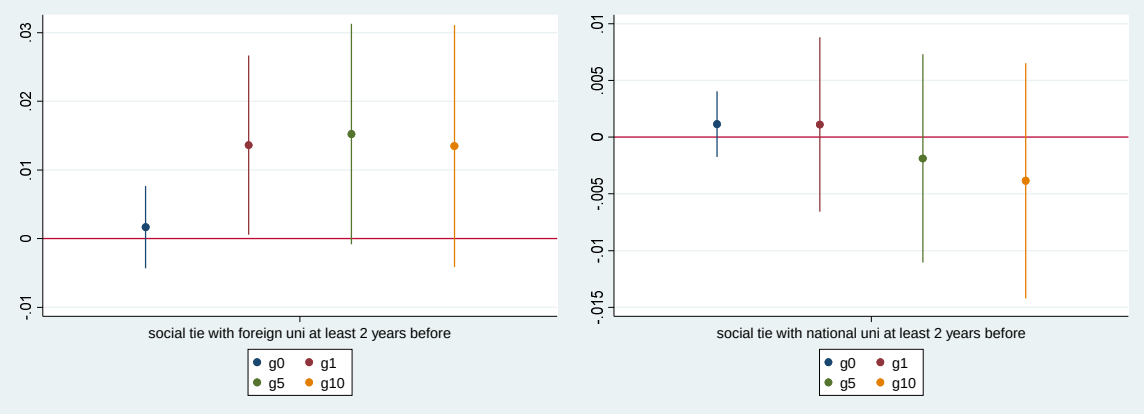


Figure D6. Academic social tie. International papers vs national papers (*degrees of Element-novelty*) at least 3 years before

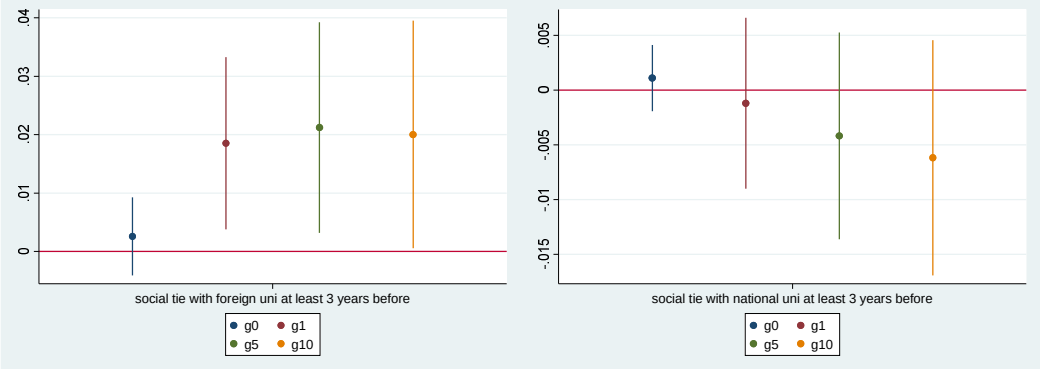


Figure D7. Academic social tie. International papers vs national papers (*degrees of Element-novelty*) at least 4 years before

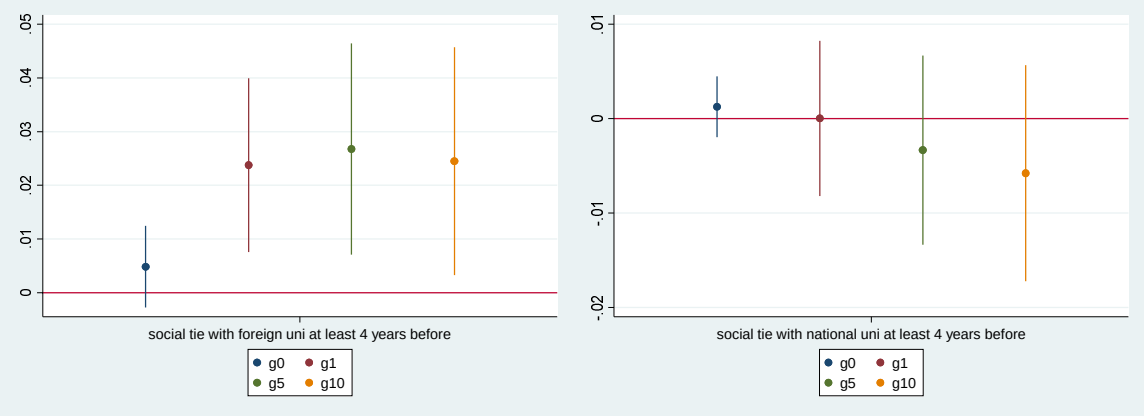


Figure D8. Academic social tie. International papers vs national papers (*degrees of Element-novelty*) at least 4 years before

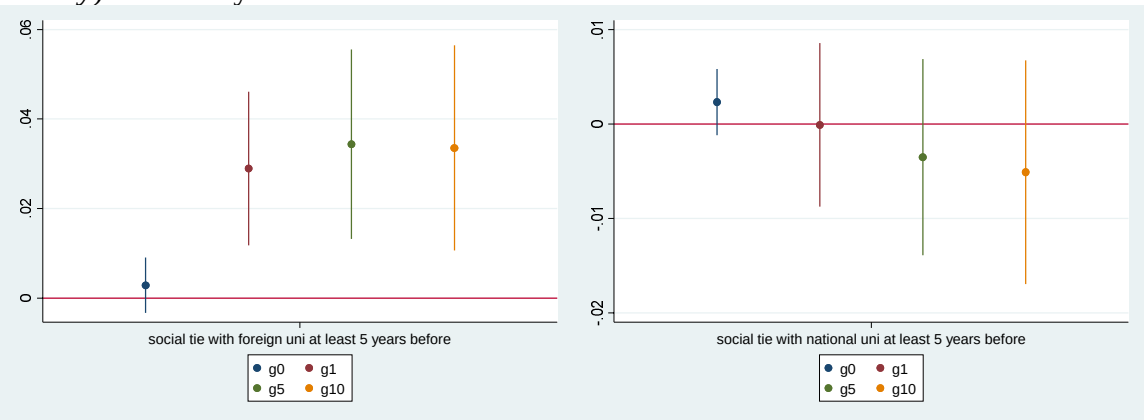


Table D1. Only papers with up to 20 authors in total (Recombinant-novelty).

	(1) r100 International	(2) r99 International	(3) r95 International	(4) r90 International
Academic social tie	0.043***	0.041***	0.035***	0.032***
	(0.014)	(0.013)	(0.012)	(0.012)
Other controls	YES	YES	YES	YES
R&D center typology	YES	YES	YES	YES
Company areas of exp. dummies	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES
Constant	1.132***	1.049***	0.937***	0.856***
	(0.032)	(0.026)	(0.024)	(0.024)
N	2271	2271	2271	2271
R^2	0.190	0.171	0.182	0.202

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D2. Only papers with up to 20 authors in total (Element-novelty).

	(1) g0 International	(2) g1 International	(3) g5 International	(4) g10 International
Academic social tie	0.004	0.021***	0.025***	0.024**
	(0.004)	(0.008)	(0.009)	(0.010)
Other controls	YES	YES	YES	YES
R&D center typology	YES	YES	YES	YES
Company areas of exp. dummies	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES
constant	0.056***	0.259***	0.334***	0.384***
	(0.004)	(0.012)	(0.015)	(0.016)
N	2519	2519	2519	2519
R^2	0.186	0.180	0.185	0.181

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

When Global Pipelines Generate Novelty: Evidence from AstraZeneca's International Academic Collaborations

Claudio Fassio

Department of Political Sciences, University of Pisa, Italy
GLINNOD (Global Innovation Dynamics) research lab, University of Pisa, Italy
& CIRCLE, University of Lund

Pauline Mattsson

LUSEM, School of Business and Economics, University of Lund,,Sweden
& CIRCLE, University of Lund

Aldo Geuna

Department of Cultures, Politics and Society, University of Torino

& Collegio Carlo Alberto

Ioana Igna

Copenhagen Business School, Denmark
& CIRCLE, University of Lund

JEL codes: L24; L65; O32; D83; F23; I23.

Keywords: International university-industry collaborations; Novelty; Social proximity; Geographic proximity; R&D sites

Acknowledgments

Thanks to Sotaro Shibayama for providing the novelty indicators. Thanks for the valuable suggestions from participants of the seminars in SPRU, University of Sussex; The Economic Geography and Spatial Economics seminar series, LSE; and DRUID.

Abstract

International university-industry collaboration expands access to heterogeneous knowledge environments but simultaneously raises coordination costs that may impede the deep, exploratory exchange needed to produce genuinely novel science. This paper examines the relational conditions under which geographically dispersed firm-academia collaborations generate knowledge novelty. We argue that social proximity, operationalized as prior shared institutional affiliation between AstraZeneca researchers and their academic collaborators, serves as a critical enabling mechanism, particularly under geographic distance, where institutional and cultural frictions are highest. Using a longitudinal dataset of 17,522 co-authored publications by AstraZeneca scientists from 2000 to 2020, we measure novelty through word-embedding indicators capturing both recombination novelty and element novelty. Exploiting the within-firm variation across AstraZeneca's globally distributed R&D network, we test whether the novelty-enhancing effect of social ties is stronger in international than in domestic academic collaborations. Results support an asymmetric substitution mechanism: prior social ties are positively associated with novelty specifically in international collaborations, where they compensate for the absence of spatial and institutional proximity, but not in domestic ones. These findings refine the proximity literature's substitution hypothesis and contribute to the understanding of how multinational firms organize knowledge recombination across geographically dispersed innovation networks.

1. Introduction

The geography of knowledge production has long occupied a central place in economic geography. From Marshall's (1920) industrial districts to contemporary literature on clusters, buzz, and global pipelines, scholars have argued that where knowledge is created, and who creates it, is inseparable from how innovation unfolds (Bathelt, Malmberg, and Maskell, 2004; Boschma, 2005; Fassio et al., 2023). Geographic proximity is not merely a spatial convenience but a mechanism since it enables repeated, informal, and often unplanned interactions through which tacit knowledge flows between actors -who might not otherwise- exchange it. Yet as innovation networks have become increasingly international, the assumption that proximity is primarily spatial has come under pressure. Firms routinely collaborate across national borders, connecting to distant scientific communities whose knowledge is inaccessible from within local environments. This raises a question that economic geography has not fully resolved: in the absence of spatial proximity, what enables actors to move beyond information exchange and engage in genuinely novel collaborative knowledge creation?

This paper addresses that question in the context of the pharmaceutical industry, where the tension between proximity, distance, and knowledge novelty is particularly important. Declining R&D productivity, the exhaustion of established molecular platforms, and successive waves of patent expirations dismantled the closed, internally driven model of drug discovery that defined the industry through the 20th century (Pisano, 2006; Schuhmacher et al., 2016). The pipeline crises of the 2000s and 2010s were not temporary disruptions but signaled a structural failure of the blockbuster model, forcing firms to fundamentally reconstitute their knowledge-sourcing strategies. In response, major pharmaceutical corporations turned outward toward universities and public research organizations (Owen-Smith and Powell, 2004; Bignami & Mattsson, 2019). The knowledge at stake in these collaborations is precisely the kind that geographic proximity was theorized to transmit: namely tacit, context-dependent, and rooted in ongoing experimental practice, rather than codified (Gertler, 2003; Balland et al., 2015; Broekel, 2015).

Economic geography has long argued that spatial proximity enables tacit knowledge transfer by reducing the costs of face-to-face interaction, supporting shared frameworks, and creating conditions for serendipitous encounters that underpin knowledge recombination (Storper and Venables, 2004; Boschma, 2005). For pharmaceutical firms, proximity to leading academic

science, through co-location near research universities or recruitment of star scientists, provides access to knowledge that cannot be reliably sourced at a distance (Zucker and Darby, 1996; Abramovsky et al., 2011; Melnychuk et al., 2021). Yet, sustained geographic proximity carries costs: repeated reliance on local knowledge environments can lead to familiar, overlapping, and redundant knowledge, limiting the production of genuinely new scientific combinations. This is what Boschma (2005) defined as the proximity trap, where co-location that facilitates exchange can, over time, generate cognitive lock-in as actors converge on shared frameworks and assumptions. Local buzz is rich but not diverse, and diversity, the combination of previously distant knowledge elements, lies at the core of novelty (Fleming, 2001; Uzzi et al., 2013).

This tension has motivated pharmaceutical firms to increasingly pursue international research collaborations, connecting to academic communities with different scientific trajectories, therapeutic traditions, and access to unique patient populations and disease environments (Bathelt et al., 2004; Bignami et al., 2020). While international collaboration expands the knowledge landscape for novel combinations, it removes the spatial infrastructure essential for tacit knowledge exchange. Distance introduces institutional friction, cultural differences, and practical barriers to coordinating complex scientific work, creating a central tension: the geographic configurations that maximize knowledge diversity are precisely those that complicate the deep, exploratory exchange needed to make diverse knowledge combinable. Empirical evidence from Wagner et al. (2019) suggests that international collaborations (not limited to academia-industry collaboration), despite their potential for diversity, tend to produce less novel and more conventional knowledge combinations due to transaction costs and communication barriers. This counterintuitive finding highlights the need to examine not just the presence of international collaboration but the relational mechanisms that may mitigate these barriers.

The proximity literature addresses this tension with the substitution hypothesis, where non-geographic forms of proximity, cognitive, organizational, institutional, and social, can partially compensate for geographic distance by providing alternative coordination mechanisms (Boschma, 2005; Huber, 2012; Balland, Boschma, and Frenken, 2015). However, empirical tests of this hypothesis have largely relied on outcomes that cannot distinguish between knowledge transfer and knowledge creation. Measures such as collaboration tie formation, co-

publications, citation counts, or patent output capture whether proximity enabled exchange, but not what kind of knowledge it produced. A co-authored paper evidences collaboration, but not necessarily the generation of genuinely novel knowledge. This difference matters because the conditions that bring actors together are not the same as those that generate novelty: proximity may bring actors together while constraining them to familiar knowledge spaces, and collaboration may be productive in volume but incremental in knowledge terms. Thus, the question of whether proximity shapes the novelty of collaboration outputs has remained largely unasked in economic geography.

This study addresses this gap by examining how social proximity, operationalized as prior shared institutional affiliation between firm-based and academic researchers, conditions the novelty of knowledge produced in co-publications, and whether that effect differs between domestic and international collaborations. Social proximity, in Boschma's (2005) framework, refers to the degree to which actors are embedded in relationships of trust, reciprocity, and shared relational history at the interpersonal level. While cognitive and institutional proximity shape the capacity for mutual understanding, social proximity shapes the willingness to move beyond familiar knowledge, to share ideas before they are fully formed, and to embrace the uncertainty that genuine intellectual exploration entails.

In international collaboration, where geographic and institutional distances are large and the potential for knowledge diversity is greatest, we argue that social proximity -i.e. social ties- functions as a relational enabling condition. It does not substitute for geography, by replicating co-location's effects, but rather provides the trust infrastructure needed to transform geographic diversity into novel knowledge combinations. Recent work supports this argument, Tu (2024) shows that the interaction between international openness and tie strength is critical for knowledge creation, while Tzabbar and Vestal (2015) demonstrate that relational strength can bridge the social distance in geographically distributed R&D teams. In contrast, in domestic collaboration, where institutional, cognitive, and cultural proximities already provide coordination capacity, the marginal contribution of social ties to novelty production is substantially lower.

We test this argument using a longitudinal dataset of co-authored publications between AstraZeneca and academic collaborators from 2000 to 2020. Novelty is measured through

word-embedding-based indicators that capture both recombination novelty (the extent to which a paper draws on unusual combinations of prior knowledge) and element novelty (the extent to which it introduces knowledge elements not previously present in the literature) (Shibayama et al., 2021; Yin et al., 2023). To measure social proximity, we use publication data and check whether in past papers AstraZeneca authors used to have the same university affiliation of their current academic coauthors. The single-firm design is deliberate, since by restricting the sample to one multinational corporation, we hold firm-level strategy, absorptive capacity, organizational culture, and resource endowments constant, isolating proximity effects from the firm-level heterogeneity.

The paper makes three contributions to economic geography. First, it shifts the dependent variable in proximity research from collaborative activity to knowledge novelty, arguing that the field's dominant outcome measures cannot distinguish between knowledge transfer and knowledge creation and thereby distort the geographic conditions under which innovation produces genuinely new scientific combinations. Second, it provides empirical evidence that the effect of social proximity on novelty is contingent on geographic distance. Social ties enhance novelty specifically in international collaboration, where they bridge actors separated by large institutional and cultural distances, but not in national collaboration, where background proximities already provide the necessary coordination capacity. This asymmetry refines the substitution hypothesis by showing that non-geographic proximity does not operate uniformly but is activated by geographic distance. Third, by situating this analysis within a single globally active pharmaceutical firm, the paper contributes to a growing literature on how multinational corporations organize knowledge recombination across geographically dispersed innovation networks, extending that literature's attention beyond network formation to the nature of knowledge produced within those networks.

The paper proceeds as follows. Section 2 develops the theoretical framework, positioning the argument within the proximity literature and elaborating the mechanisms linking geographic distance, social proximity, and knowledge novelty. Section 3 describes the case study, AstraZeneca. Section 4 describes the main sources of data for the empirical analysis. Section 5 illustrates the main strategies adopted by AstraZeneca to source external knowledge. Section 6 presents the main variables and the estimation strategy. Section 7 presents the empirical results. Section 8 discusses the findings and their implications for economic geography's

understanding of proximity, knowledge creation, and the geography of innovation in multinational firms.

2. Literature Review

2.1. The Strategic Turn Toward Novelty in Multinational Pharmaceutical Firms

Novelty is a cornerstone of scientific progress and a critical driver of long-term competitive advantage in knowledge-intensive industries (Nelson and Winter, 1982; Teece et al., 1997; Uzzi et al., 2013). In pharmaceuticals, this is particularly important since first-in-class drugs that introduce new mechanisms of action can redefine treatment paradigms, create temporary monopolies, and generate substantial economic benefits. In contrast, incremental innovations in crowded therapeutic areas face intensifying generic competition, pricing pressure, and regulatory scrutiny.

Historically, large pharmaceutical firms relied on vertically integrated R&D models, where innovation was cumulative and trajectory-bound, building on proprietary chemical platforms and established therapeutic domains. Competitive advantage stemmed from scale, internal capabilities, and strong intellectual property protection. However, beginning in the 1980s and accelerating through the 1990s, this model was destabilized by advances in molecular biology, genomics, and biotechnology, which transformed the scientific foundations of drug discovery and increased dependence on frontier academic research. Simultaneously, R&D expenditures soared while productivity declined, a dynamic widely described as the pharmaceutical productivity crisis (Pisano, 2006; Schuhmacher et al., 2016). The sustainability of blockbuster pipelines weakened, regulatory requirements intensified, and therapeutic innovation increasingly required the identification of entirely new biological targets rather than the refinement of existing compounds. As a result of these structural shifts, novelty moved from an epistemic ideal to a core strategic concern.

Novelty in pharmaceuticals is often recombinant, emerging from the combination of previously unconnected or cognitively distant knowledge elements (Fleming, 2001; Azoulay et al., 2011). Empirical research demonstrates that atypical knowledge combinations are more likely to generate high-impact outcomes (Uzzi et al., 2013; Wang et al., 2017; Foster et al., 2015). In science-based sectors like pharmaceuticals, innovation frequently depends on integrating

insights across biological, chemical, computational, and clinical domains. Novelty thus commonly takes the form of new knowledge combinations rather than entirely isolated discoveries, a form of creativity measurable through word-embedding approaches that capture the unusualness of knowledge element combinations in published research (Shibayama et al., 2021; Yin et al., 2023).

Yet, established firms often struggle to generate exploratory recombination internally. Research on incumbent adaptation highlights how accumulated routines, commitments, and capability structures can bias firms toward competence-deepening trajectories, limiting their responsiveness to technological shifts (Eggers, 2018). Incumbents tend to rely on familiar knowledge neighborhoods and established problem-solving techniques, making sustained broad or distant search difficult to maintain alongside exploitation demands (Audretsch et al., 2024; Li, 2023). Chen et al. (2025) provides further evidence for this dynamic, finding that repeated collaboration tends to reduce novelty, disruptiveness, and interdisciplinarity, while new collaborations facilitate the integration of unconventional knowledge. In pharmaceuticals, case studies demonstrate that established competencies can simultaneously facilitate and limit firms' efforts to explore radically new technological directions (Phillips and Pandza, 2023).

To overcome these constraints, pharmaceutical firms have increasingly turned to externally networked and geographically dispersed collaboration, particularly with universities. Cross-national R&D partnerships provide access to heterogeneous scientific communities, institutional settings, and region-specific competencies (Cantwell and Mudambi, 2005). Universities and public research organizations play a distinctive role as generators of foundational scientific knowledge, early-stage discoveries, and specialized expertise that complement firms' applied R&D capabilities (Zucker et al. 2002; Perkmann et al., 2013). In pharmaceuticals, where drug discovery builds on advances in molecular biology, genomics, and translational medicine, academic partnerships are often central to identifying new therapeutic targets and mechanisms of action (Owen-Smith and Powell, 2004; Bignami & Mattsson, 2019). Moreover, differences in regulatory regimes, clinical infrastructures, and patient populations across countries make international academic collaboration strategically important for testing and validating new therapies across institutional contexts (Bignami et al., 2020).

However, the act of forming external collaborations does not guarantee novelty (Wagner et al., 2019). A fundamental and underexplored question is therefore under which relational and geographic conditions firm-academia collaborations produce genuinely novel scientific outputs. This question motivates the analysis that follows, as it requires engaging with both the geography of collaboration and the relational mechanisms that shape what geographically dispersed partnerships can produce.

2.2. The Geography of Collaboration and Knowledge Diversity

Research in economic geography has long demonstrated that innovation is spatially structured. Early empirical studies showed that knowledge spillovers decay with geographic distance, highlighting the localized character of innovative activity (Jaffe et al., 1993; Audretsch and Feldman, 1996). The underlying mechanism is that spatial proximity facilitates the transfer of tacit knowledge, knowledge that is context-dependent, experiential, and most effectively transmitted through repeated face-to-face interaction (Polanyi, 1966; Gertler, 2003; Storper and Venables, 2004). This is particularly consequential in pharmaceutical R&D, where basic biomedical knowledge is often embedded in laboratory practices and unpublished experimental routines rather than codifiable results (Owen-Smith and Powell, 2004; Belderos et al., 2021).

Yet, geographic proximity is not without its limits. Excessive embeddedness in local knowledge environments may actively constrain novelty. Firms embedded in dense local networks may rely on familiar partners and established trajectories, reducing exposure to diverse perspectives (Grabher, 1993; Garcia Martinez et al., 2025). Boschma (2005) coined this the “proximity trap,” as the co-location that facilitates tacit knowledge transfer can, over time, produce cognitive lock-in. Actors may converge on shared frameworks, methods, and assumptions, limiting their ability to generate novel knowledge. Since novelty in science-based industries most commonly arises from the combination of previously distant knowledge elements (Fleming, 2001; Azoulay et al., 2011; Shibayama et al., 2021), geographic proximity that fosters cognitive convergence risks pushing collaborations toward competent but incremental outcomes. As Nooteboom et al. (2007) demonstrate, there is an optimal cognitive distance for learning and innovation where too little distance stifles novelty, because actors share the same knowledge base, and too much distance impedes effective communication and integration. Sustained geographic proximity tends to erode cognitive distance over time, pushing collaboration toward the former condition.

In response to these limitations, later work extended the perspective by conceptualizing innovation as shaped by multiple proximity dimensions. Boschma (2005) distinguishes geographic, cognitive, organizational, social, and institutional proximity, arguing that innovation outcomes depend on their configuration rather than any single dimension. Later contributions emphasize that proximities are dynamic and relational, evolving through collaboration processes (Balland et al., 2015; Huber, 2012; Hansen, 2012). Recent empirical research confirms that intersecting proximity dimensions jointly shape innovation performance (Shkolnykova, 2023; Garcia Martinez et al., 2025), and that proximity configurations are particularly relevant for recombinant innovation at the dyadic level (Nan et al., 2024).

Despite this theoretical richness, existing empirical work in the proximity literature do not distinguish between knowledge transfer and knowledge creation**.** By relying on co-publications, patents, or collaboration ties as outcome measures, studies demonstrate that proximity enables exchange without addressing what kind of knowledge that exchange produces (Knoben and Oerlemans, 2006; Hansen, 2015). A co-publication is evidence that collaboration occurred, but it is not evidence that the collaboration generated novel knowledge combinations. A paper can emerge from a geographically proximate, institutionally aligned, and cognitively close partnership and still draw on entirely familiar knowledge elements. Conversely, a collaboration bridging distant knowledge environments may produce novel recombinations even if its collaboration volume is modest. The proximity literature, by focusing on the quantity and probability of collaborative output, has therefore left unasked the question of whether proximity shapes the novelty of what collaboration produces.

To overcome these constraints, firms need access to distant knowledge, not just local knowledge. Multinational corporations increasingly participate in geographically dispersed innovation networks. Bathelt et al. (2004) conceptualize this as the combination of "local buzz" and "global pipelines." While local networks provide intensive interaction and trust-based collaboration, global pipelines connect firms to distant knowledge environments, introducing novel ideas and specialized expertise unavailable within local milieus. Recent studies emphasize that multi-scalar collaboration networks have become central to contemporary knowledge-sourcing strategies (Audretsch et al., 2024; Frigon, 2024). Empirical studies further show that innovation outcomes depend on how firms balance regional embeddedness, through,

for example, subsidiaries, with cross-border connections (Galaso et al., 2020; Belderbos et al., 2021). This dynamic aligns with the broader substitution logic in the proximity literature, which argues that one form of proximity can compensate for the absence of another (Huber, 2012; Hansen, 2015).

However, geographic dispersion introduces a tension between diversity and integration. Collaborating across borders expands access to diverse knowledge bases, specialist expertise, and research communities operating on different scientific trajectories, all of which increases the potential for novel recombination. Yet spatial and institutional distance increases coordination costs, creates interpretative challenges, and complicates the transfer of tacit knowledge, challenges that co-location infrastructure typically mitigates (Gertler, 2003; Knoben and Oerlemans, 2006). Contrary to the assumption that international collaboration inherently enhances novelty, Wagner et al. (2019) find that such collaborations often produce less novel and more conventional knowledge combinations due to transaction costs and communication barriers. Diversity broadens the opportunity space for recombination, but whether it translates into genuinely novel outcomes depends on relational and institutional conditions that facilitate effective integration across spatial boundaries. This paper explores this central tension, as the same geographic configurations that expand knowledge diversity also hinder the deep, tacit exchanges required to combine that knowledge effectively. Resolving this tension requires shifting focus from geography to the relational mechanisms that determine the outcomes of internationally dispersed collaboration.

2.3. Social Proximity and Relational Mechanisms Under Distance

If geographic dispersion expands access to diverse knowledge while simultaneously increasing integration challenges, relational mechanisms become critical for enabling effective collaboration across distance. While the proximity literature identifies social proximity as one such mechanism, its role in shaping the novelty of knowledge produced, rather than merely the probability of collaboration, has received limited theoretical and empirical attention.

Social proximity, as defined by Boschma (2005), refers to interpersonal relationships embedded in trust, shared experience, and relational history. Grounded in Granovetter's (1985) insight that economic action is embedded in social relations, social proximity facilitates open, exploratory communication, the sharing of unpublished ideas, and tolerance for the uncertainty

inherent in novel knowledge combinations. This embeddedness fosters trust, shared interpretive frames, and reputational monitoring, all of which reduce the coordination costs that are particularly high when knowledge is tacit, complex, and not yet codifiable (Reagans and McEvily, 2003; Ooms et al., 2018).

However, the mechanism through which social proximity shapes novelty is distinct from its role in enabling collaboration formation. While collaboration can occur between actors with low social proximity, particularly when mandated by institutional or organizational arrangements, the depth and exploratory character of knowledge exchange may be limited. Social proximity enables actors to venture beyond familiar knowledge by providing relational safety to share speculative, uncertain, and not-yet-validated ideas. As Malmberg and Maskell (2006) note, trust often exists in local milieus as an inherited feature of co-location, but prior social ties can recreate this relational infrastructure across geographic distance, enabling the kind of deep, exploratory exchange that "local buzz" typically provides.

Proximity dimensions are not static but co-evolve through collaboration processes (Balland et al., 2015; Broekel, 2015). Prior social ties, built through shared institutional experiences earlier in collaborators' careers, represent accumulated relational capital that can be reactivated in new collaborative contexts. This perspective is supported by Dahlander and McFarland (2013), who demonstrate how tie persistence in research collaborations depends on the strength and history of relationships, and Rost (2011), who highlights the critical role of strong ties, not just weak ones, in fostering innovation, particularly in contexts requiring deep trust and coordination. Further evidence for this mechanism comes from Roth and Mattes (2023), who show that relational infrastructures sustain interaction across spatial separation, and Fassio, Geuna, and Rossi (2023), who demonstrate that firms' engagement with foreign universities is mediated by inventors' relational trajectories, underscoring the individual-level networks enabling cross-border knowledge exchange in multinational R&D.

Yet, social proximity is not uniformly beneficial for novelty. Chen et al. (2025) emphasizes that prolonged collaboration can reduce novelty by reinforcing familiar knowledge pathways, while new or weaker ties may be more effective in integrating unconventional knowledge. This suggests that the novelty-enhancing effects of social proximity may depend on the balance between repeated interaction and exposure to new perspectives. Uzzi (1997) warns that over-

embeddedness, where social ties become too dense and inward-looking, can generate knowledge redundancy and network myopia, reinforcing homophily and limiting exposure to diverse perspectives (Inkpen and Tsang, 2005). When social ties connect actors who already share similar knowledge bases and institutional frameworks, they may deepen cognitive convergence rather than bridge cognitive distance. Boschma (2005) himself notes that social embeddedness in industrial districts can decrease knowledge heterogeneity as firms sharing social proximity increasingly adopt common assumptions and cognitive frameworks. This represents the social version of the geographic proximity trap, too much social proximity, like too much geographic proximity, can produce redundancy rather than novelty.

2.4. Hypotheses

The preceding discussion highlights a structural tension in the geography of innovation. International collaboration expands access to heterogeneous scientific environments and enlarges the opportunity space for novel knowledge recombination. At the same time, spatial and institutional dispersion introduce coordination and interpretative frictions. Geographic diversity therefore does not automatically translate into novel outcomes, its effects are likely to depend on the relational conditions under which collaboration unfolds. Existing research often emphasizes the diversity-enhancing role of international collaboration. However, this perspective remains incomplete if it neglects the relational mechanisms that shape how geographically dispersed actors interact. Rather than asking whether international collaboration increases novelty per se, this study examines whether relational embeddedness conditions the association between cross-border collaboration and novel outcomes.

Social proximity constitutes one such relational condition. Prior ties between collaborators, such as shared institutional affiliation at earlier career stages, can foster trust, reduce uncertainty, and support sustained interaction across organizational boundaries. These mechanisms are especially salient when collaborators operate in different national and institutional contexts, where shared norms and opportunities for face-to-face interaction are more limited. We therefore expect prior social ties to be positively associated with novelty in collaborative research. However, we do not anticipate this association to be uniform across geographic configurations. In nationally bounded collaborations, shared institutional environments and regulatory frameworks may already reduce uncertainty and facilitate interaction. In international collaborations, by contrast, spatial and institutional distance heighten coordination challenges and increase the relevance of relational embeddedness. In

international collaboration contexts, prior work experience at a partner institution by a member of the collaborating team may help mitigate the challenges associated with spatial and institutional distance, thereby facilitating coordination and interaction across organizational boundaries

Accordingly, we propose:

H1: Collaborations involving prior social ties between partners are associated with higher levels of novelty.

H2: Social ties are positively associated with novelty in international collaborations, but not in domestic ones where other proximity dimensions already provide sufficient coordination capacity.

3. The case study: AstraZeneca

This study focuses on AstraZeneca to investigate how multinational pharmaceutical firms establish collaborations with international academics and how such collaborations relate to scientific novelty. The pharmaceutical sector provides a particularly relevant empirical setting due to its reliance on frontier science, the complexity and uncertainty of drug discovery, and the increasing globalization of R&D activities (Belderbos et al., 2010; Narula & Santangelo, 2012).

AstraZeneca constitutes a suitable case for several reasons. First, it is one of the world's largest research-intensive pharmaceutical firms, employing approximately 90,000 people globally and ranking among the leading companies by market capitalization. Its scale and scientific breadth ensure that it operates across multiple therapeutic areas and institutional environments, making it representative of large multinational pharmaceutical corporations. Second, the firm exhibits substantial internal heterogeneity in its global R&D network, with research centers differing in size, specialization, and degree of academic engagement. This variation enables the examination of collaboration patterns and novelty within a shared corporate context. Third, the period under study coincides with significant strategic and organizational transformation, creating a dynamic setting in which shifts in knowledge sourcing can be observed.

A particularly important turning point occurred in the early 2010s. A wave of patent expirations weakened AstraZeneca's core revenue base, intensifying pressure to renew its drug portfolio. In 2012, Pascal Soriot was appointed Chief Executive Officer, succeeding David Brennan and initiated a strategic reorientation centered on rebuilding the pipeline and strengthening science-driven innovation. In 2014, Pfizer launched an attempted takeover bid for the company. Although the acquisition ultimately failed, the episode heightened organizational uncertainty and accelerated strategic reassessment. Together, these developments marked a period of turbulence that placed exploratory research and external collaboration at the center of the firm's renewal strategy.

Around the same time AstraZeneca reorganized its R&D structure around major strategic hubs while closing several sites, including Lund and Södertälje in Sweden in 2013, where more than 1,000 research positions were affected. These closures reflect broader industry trends toward concentrating R&D in fewer, larger and more strategically aligned centers (Danzon et al., 2005). The current R&D organization centers on strategic hubs in Cambridge (UK) Mölndal (Sweden) and Gaithersburg (US), complemented by additional global centers located in major scientific regions such as Boston, Shanghai and Beijing. This distributed structure embeds the firm in multiple regional innovation systems and heterogeneous scientific communities.

Table 1 provides a detailed overview of AstraZeneca's R&D sites from 2000 to 2020, illustrating the dynamic evolution of its global research footprint and highlighting the years during which each center was active. Out of the 18 centers that have been active in the period considered, only Boston/Waltham (US), Macclesfield/Alderley Park (UK) and Mölndal (Sweden) were always in operation. The two strategic R&D centers of Cambridge (UK) and Gaithersburg, (US) were created in 2007 with a major restructuring of all R&D activities. At the end of the period there were 10 active R&D sites, the three main R&D centers of Molndal in Sweden, Cambridge and Macclsefield in the UK, four new centers in the US (Boston, Gaithersburg, Mountain View and New York), and three other small international centers in Japan (Osaka), China (Shanghai) and Poland (Warsaw).

INSERT TABLES 1 ABOUT HERE

Beyond geographic restructuring, AstraZeneca also shifted its therapeutic orientation. While earlier portfolios were diversified across primary-care areas, the 2010s saw a pronounced reorientation toward science-intensive domains, particularly oncology, immunology and

molecularly targeted therapies. These fields are characterized by rapidly evolving biological knowledge, high uncertainty and strong dependence on academic research in molecular biology, genomics, and translational medicine. Entering and expanding within such domains increases reliance on frontier science and heightens the importance of accessing geographically dispersed and cognitively diverse knowledge communities. In this sense. AstraZeneca's therapeutic shift reinforces the relevance of international academic collaboration for exploratory and potentially novel research.

AstraZeneca's global R&D footprint exhibits substantial heterogeneity across locations in terms of publication intensity and engagement with academic partners. As described in the data section, sites differ in their degree of international academic collaboration and internal research orientation. This differentiated R&D geography provides a structured yet internally diverse empirical setting in which to examine how international collaboration and relational configurations are associated with the production of novelty in multinational pharmaceutical research.

4. Sources of data

Annual reports

Spanning 2000 to 2020. For each year, we identified the therapeutic areas in which the company was active by categorizing pipeline components across pre-clinical and clinical phases, tracking products already on the market, and recording R&D expenditure by area. All therapeutic areas in which AstraZeneca held any degree of activity were included as controls.

The annual reports also served as the basis for identifying institutional agreements signed with academic partners in each given year. This was supplemented by deal-level data from IQVIA Pharma Deals (formerly IMS Health), which captures agreements between pharmaceutical companies and a range of external organizations, including deal type and counterparty. Finally, the reports were used to determine the locations of AstraZeneca's R&D centers across the study period.

Interviews

Prior to launching our empirical study, we conducted four interviews with directors and vice presidents for science relations and policy to deepen our understanding of AstraZeneca's

historical and current pipeline. Our questions focused on strategic decision-making around therapeutic areas and R&D site selection.

Publication data

We downloaded data from Scopus about all the papers that were published between 2000 and 2020, in which at least one of the authors had an Astra-Zeneca affiliation or an acquired company[1]. Through this procedure we were able to identify 17,522 papers. We recovered information about all the 59,021 authors included in such papers. For the 12,789 authors with an AstraZeneca affiliation, using their Scopus ID, we downloaded information in Scopus about all the papers that they published throughout their career (hence also before 2000), regardless of their affiliation. In such a way we could retrieve information also about the publishing activity of AstraZeneca authors before the chosen period of 2000-2020 to track their professional history. We can see their history of affiliations and the location of the affiliation. We can also know if the coauthors with whom they publish in the period 2000-2020 have coauthored with them in the past or not. For further details, including how we managed that some AZ authors had more than one affiliation, see Appendix A.

Affiliations and geographic location

Using both the affiliation name listed in the papers and the affiliation ID provided by Scopus we recovered the institution, city and country of the affiliation of each author (both Astra Zeneca authors and their coauthors). In the case of AstraZeneca, we realized that Scopus wrongly assigned most of AstraZeneca authors to the sites of Cambridge in UK, or to Sodertalje in Sweden, because it often automatically assigns AZ author to the Headquarters of the company. Therefore, we had to manually check for cases in which the place of the AstraZeneca affiliation listed in the papers was different from the one that Scopus attributed through the Scopus affiliation ID. Whenever the two affiliations differed, we attributed the location listed in the original printed papers. Out of 12,789 AZ authors we had to correct the city of affiliation for 9,134 of them, that is for more than 70% of them. This allowed us to identify the full network of R&D subsidiaries by Astra Zeneca, where their scientists were working. As a further double check, we used information from the annual reports of AstraZeneca to identify the R&D centers of the company and their evolution throughout the period. This is because, as

[1] We checked which companies were acquired by AstraZeneca during the years 2000-2020. This was the case for example of MedImmune in 2007. In the Appendix A we explain in detail how we managed the inclusion of papers in which the affiliation was of one of the acquired companies.

explained above, some of the R&D centers were newly created during the period 2000-2020 and some others were closed.

For the 46,232 non-AstraZeneca authors included in the 17,522 papers, we classified their affiliations as either academic researchers, researchers from public research organizations, researchers from other companies or researchers working in hospitals. Using the name of the city we were also able to find the coordinates of the cities of affiliation of both AstraZeneca authors and their co-authors and calculate the geographical distance between each AstraZeneca R&D sites and the academic coauthors affiliation.

National and international collaborations with academic researchers

We are interested in the papers that involve academic coauthors, distinguishing between the collaborations that involve academics located in the same country as the AZ researchers (*national academic collaborations*) and those that engage academics located in universities in other countries than the AZ researchers (*international academic collaborations*). To classify the academic collaborations between national and international ones, we take advantage of the location of the institution listed in the affiliation of each author. Out of the 17,522 papers, the number of papers with academic coauthors is 8,568. Among those, 5,466 papers are with international academic coauthors. To simplify the classification of international and national collaborations we only consider papers in which all AstraZeneca authors have affiliations in the same country.[2] We also make sure that the publications are actual papers and not research notes or commentaries. After cleaning for this, our final sample is made of 5,363 papers: of which 2,806 are international collaborations and 2,557 national collaborations.

On average AstraZeneca researchers publish 25 papers each throughout their careers, although the average does not provide a clear picture, since it is a very skewed distribution. Indeed, the median is 6 papers per AstraZeneca author and then there are some extremely productive authors with more than 100 publications (approximately 5% of the sample). In Appendix A in table A3 we also show that the share of academic publications has increased between 2000 to 2020 from less than 40% of all 17,522 papers in the first three years to more than 50% in the

[2] Without this restriction, classifying a collaboration as domestic or foreign would become ambiguous. Consider, for example, a paper co-authored by AstraZeneca researchers based in both the UK and the US, an academic collaborator at a US university would be classified as foreign relative to the UK-based AZ authors, yet domestic relative to their US-based colleagues. Applying the restriction has limited impact on the overall dataset, as in more than 80% of the 17,522 papers examined, all AstraZeneca authors are in the same country.

last 5 years. Among academic papers the share of international ones, with at least one author located in a different country than the rest of the coauthors team, increased from 40%-50% in the first years of the period to more than 70% in the last years (Table A4). The academic co-authors of AstraZeneca authors are located mostly in UK, US and Sweden: these three countries account for more than 60% of the total. However, as shown in table A5, also Germany, Canada, China, Japan and Australia account each for at least 2%, showing a true global nature of the academic collaborations of AstraZeneca researchers.

5. AZ R&D geography and external sourcing strategy

5.1 Sourcing knowledge from other companies: acquisition and partnerships

AstraZeneca has adopted a diversified approach to knowledge sourcing strategies. First, as shown in Table 2 the company realized a few important acquisitions of biotech companies such as Cambridge Antibody Technology that laid the foundation for the takeover of MedImmune in 2007 which was important for strengthening the biologics arm which was central to the oncology and immunology pipelines.

Second, over time, AstraZeneca has increasingly emphasized international collaboration and open innovation to complement its internal capabilities. The company's strategic shift toward external partnerships is driven by the recognition that breakthrough discoveries often originate outside firm boundaries. One example of such a collaboration is the acquisition of Amylin Pharmaceuticals or the key alliance with Ionis Pharmaceuticals to strengthen antisense therapies (see Table 3).

INSERT TABLES 2 AND 3 ABOUT HERE

5.2 AZ R&D geography and collaborations with universities

An important source of external knowledge comes from collaboration with universities. As often stated in the AstraZeneca's Annual reports by leveraging global scientific networks and collaborating with universities and research institutes, the company aims to enhance its absorptive capacity, accelerate innovation cycles, and manage the balance between exploration

of novel knowledge and exploitation of existing competences, core themes that are central to this study.

Table 4 provides the publishing activity of the different R&D centers in the period 2000-2020, using Scopus data. Most publications by AZ researchers originated from three primary centers: Cambridge (UK), Macclesfield (UK), and Mölndal (Sweden). These three sites accounted for approximately 60% of non-academic and 66% of academic publications, with at least one author affiliated with them. Another notable center, Södertälje (near Stockholm), was active until its closure in 2012. Additionally, several other centers demonstrated significant publishing activity, including Loughborough (UK) and three US-based locations: Boston, Wilmington (Delaware), and Gaithersburg (Maryland). In Sweden, the Lund R&D center was also a key hub for publications until 2012.

Tables 4 and 5 use the number of publications and the share of publications with academic coauthors (national or international) in the period 2000-2020 to provide a first illustration of how each AZ R&D center collaborates with academic scholars. We distinguish between large R&D centers, as measured by the total number of publications (more than 3,000 publications between 2000 and 2020), such as Cambridge UK or Mölndal; middle-sized ones (from 600 to 3,000 publications), such as Wilmington in Delaware, or Loughborough in the UK; and small centers (less than 600 publications), such as the ones in Japan (Osaka) or in China (Shanghai).[3]

Among the large R&D sites we find two different typologies. The first typology, which we label as "*Global Academic Hubs*", are highly engaged with academics (at least 35% of total publications produced by AZ researchers in the center) and the share of publications with at least one academic that located in a different country (international academics) is also relatively high (at least 50% of academic publications). Cambridge (UK), Mölndal and Södertälje (Sweden) exemplify this type, pointing to an international network of academics collaborating with AstraZeneca researchers at these locations. The second type, which we label "*Corporate R&D Hub*", is illustrated by Macclesfield (UK). Here, most publications do not involve academic co-authors, and international academic collaborations are particularly rare. The site remains a significant knowledge producer in terms of publication output but is oriented towards intra-firm knowledge flows or collaborations with other types of partners such as hospitals and

[3] The size classification does not change even if we account for the number of years that the R&D centers have been active (using the info from Table 1), hence measuring the number of publications per year of activity.

other firms. This more limited engagement with the academic community is also reflected in the geographic scope of its collaborative network: academic partnerships involving Macclesfield tend to be predominantly with UK-based universities rather than international institutions.

INSERT TABLES 4 AND 5 ABOUT HERE

Among mid-sized R&D sites, we identify a first type, "*Global Academic Centre*", which is similar to the characteristics of a Global Academic Hub in terms of strong engagement with academics, particularly international ones, but with a somewhat lower publication output. Gaithersburg (US) and Lund (Sweden) are representative examples.

Two further typologies emerge among mid-sized sites. The first, which we label "*International R&D Centre*", includes sites such as Boston and Waltham, characterized by relatively low overall engagement with academics but a disproportionately high share of international academic collaborators. This pattern may reflect a limited embeddedness in the local academic community. When these sites do draw on academic knowledge, they tend to source it from universities abroad, where AstraZeneca has more established relationships. The second, which we term "*Corporate R&D Centre*", is exemplified by Loughborough (UK) and Wilmington (US). These sites show both low levels of academic co-authorship and a limited international dimension to what collaborative activity does exist, they engage rarely with universities in general, and even more rarely with foreign ones.

When it comes to the small R&D centers, we notice some relevant differences: some centers have little focus on academic collaborations: this is the case of Bangalore in India, Rheims in France or Montreal in Canada: we label these "*Local R&D outpost*": these centers rely mostly on intra-firm knowledge and have little international academic collaborations. The other typology of small R&D site is what we define "*Outpost for local academic search*": this is the case for Osaka and Shanghai, where there is a high level of engagement with academics, but most collaborations are with local/national academics. This hints at a strategy of local sourcing for academic knowledge.

The distinction between the 7 typologies of R&D sites (Global Academic Hubs, Corporate R&D Hubs, International Academic Centers, International R&D centers, Corporate R&D centers, Local R&D Outposts and Outpost for Local Academic Search) reveal the considerable heterogeneity within AstraZeneca's R&D network in terms of engagement with the academic community, and with international academics in particular. Not all sites interact with universities to the same degree, and only a subset maintain a substantial share of international collaborations. This heterogeneity suggests that different nodes in the company's R&D network serve distinct purposes when it comes to accessing and sourcing academic knowledge from abroad.

By the end of the observed period, AstraZeneca's R&D network had taken on a clearly defined structure. The company maintained three world-class strategic R&D centers: the Discovery Centre (DISC) in Cambridge (UK), Gaithersburg (Maryland, US), and Gothenburg (Sweden) as its primary hubs of scientific activity. These were complemented by a set of medium-sized centers in the US and elsewhere, including Boston, which served both as nodes of internal knowledge production and as points of connection to local and international academic communities. Macclesfield remained active in the UK, though in a pharmaceutical technology and manufacturing capacity rather than as a core research site. Notably, China emerged as a significant and growing part of the R&D network in the post-2020 period. Shanghai was designated AstraZeneca's fifth strategic R&D center in 2021. A sixth strategic center was subsequently announced in Beijing in 2025.

6 Econometric methodology

Our empirical strategy is composed of a two steps strategy, First, we develop an event study to check whether, after the difficult years of the early 2010’s, AstraZeneca increased its reliance on new international academic collaborations with respect to other types of collaborations. Secondly, we check whether the existence of social ties by the AZ authors is correlated to higher novelty for the papers with international academic coauthors.

6.1 Increase in International Academic Collaboration

As discussed in the previous sections, the early 2010s marked a period of strategic disruption and organizational reorientation at AstraZeneca. The combination of patent expirations, leadership change, and the attempted takeover in 2013 prompted a renewed emphasis on science-driven innovation and external knowledge sourcing. If this strategic shift involved a

stronger orientation toward exploratory search, one observable implication would be an increased reliance on new international collaborators, particularly in academic partnerships, where frontier scientific knowledge is generated.

To examine whether collaboration patterns changed during this period, we analyzed the share of first-time international coauthors in AstraZeneca publications before and after 2013. Our unit of analysis is the individual scientific publication authored between 2000 and 2020. For each paper, we identified all coauthors, their institutional affiliation, and their prior co-publication history with AstraZeneca researchers. This allowed us to determine whether a collaboration represented a first-time co-authorship and whether the coauthor was located in a foreign country.

We distinguished between two types of collaborative papers:

1. Papers with at least one academic coauthor and
2. Papers with at least one coauthor from a firm other than AstraZeneca.

To facilitate interpretation, we excluded publications that included both academic and non-AZ company coauthors (1,398 papers out of 10,333 papers with academics and/or non-AZ company coauthor, i.e. 13% of that sample).

To identify how much a paper relied on new international coauthors (whether academic coauthors or individuals working for other companies) we adapted the procedure implemented by Liu et al. (2022). For each paper i, we constructed a dependent variable capturing the share of new international coauthors relative to all possible AstraZeneca–external coauthor pairs[4]. For academic papers, the measure is:

$$ShareNewIntlAcademic_i = \frac{\sum_j \text{NewIntlAcademic}_{ji}}{\text{AZAuthors}_i \times \text{AcademicCoauthors}_i}$$

[4] For example, in a paper with 2 AstraZeneca authors and one foreign academic coauthor if the academic coauthor is new only for one of the two AstraZeneca authors our variable will count 1 on the numerator and 2 (the two possible pairs between AstraZeneca authors and the foreign coauthor) on the denominator, hence its value will be 0,5.

where the numerator sums first-time foreign academic coauthors for each AstraZeneca author $j$in paper i, and the denominator equals the total number of potential AZ–academic coauthor pairs in that paper[5].

For company papers, we constructed an analogous measure:

$$ShareNewIntlCompany_i = \frac{\sum_j \text{NewIntlCompany}_{ji}}{\text{AZAuthors}_i \times \text{CompanyCoauthors}_i}$$

Because our data consists of repeated cross-sections of publications rather than a panel of organizational units, we implemented a grouped difference-in-differences design. Papers with academic coauthors constituted the treatment group ($g = 1$), while papers with company coauthors served as the comparison group ($g = 0$). The post-2013 period defines the time dimension.

We estimated the following specification:

$$ShareNewIntl_{igt} = \alpha + \beta_1 \text{Post2013}_t + \beta_2 \text{Academic}_g + \delta\left(\text{Post2013}_t \times \text{Academic}_g\right) + \gamma X_i + \lambda_t + \varepsilon_{igt}$$

Where $Post2013_t$equals 1 for publications after 2013, $Academic_g$ equals 1 for academic papers and 0 for company papers, X_i includes controls for team composition (number of AZ and external authors), therapeutic area, and R&D site, λ_t denotes year fixed effects, and ε_{igt} is the error term.

The coefficient of interest, δ, captures whether academic papers experienced a differential shift toward new international collaborators after 2013, relative to company papers. Under the identifying assumption of parallel trends in the absence of the strategic disruption, δ measures the relative change in collaborative orientation associated with the post-2013 period.

6.2 Social Ties and Novelty in Academic Collaboration

We next examined how relational configurations were correlated to the novelty of collaborative research outcomes.

[5] In these papers we can also have other types of co-authors: with PRO, hospital or unknown affiliation

Our dependent variable is the novelty of publication i, measured using the two word-embedding–based indicators developed by Shibayama et al. (2021) and Yin et al. (2023). These measures capture both recombinant novelty (unusual combinations of existing knowledge elements) and element-level novelty (the introduction of new semantic components).

We estimated the following baseline model for academic papers:

$$Novelty_i = \alpha + \beta_1 PersonalTies_i + \beta_2 InstitutionalAgreement_i + \theta Z_i + \lambda_t + \varepsilon_i$$

Where PersonalTie_i indicates whether at least one AstraZeneca author previously shared an institutional affiliation with a collaborating academic, $\text{InstitutionalAgreement}_i$ captures the presence of formal collaboration agreements, Z_i includes controls for team composition, R&D site, therapeutic domain, geographic distance, and other collaboration characteristics, and $\lambda_t$denotes year fixed effects.

To reflect our theoretical argument that relational mechanisms may operate differently across spatial contexts, we estimated the model separately for international and domestic academic collaborations. All models were estimated using ordinary least squares with heteroskedasticity-robust standard errors.

6.2.1 The dependent variable: Novelty

Novelty is widely regarded as a core value of scientific research, yet its empirical measurement remains conceptually and methodologically challenging. As emphasized in recent work, novelty is a multifaceted construct that can refer to new elements, new combinations of existing knowledge, or new positioning within scholarly discourse (Shibayama et al., 2021; Yin et al., 2023). Bibliometric operationalizations therefore differ substantially in what aspect of novelty they capture. Following this literature, this study adopted two complementary, text-based indicators that reflect distinct dimensions of novelty: *recombinant novelty* and *element novelty*.

Recombinant novelty captures the extent to which a publication recombines previously unconnected knowledge components. Rooted in the recombinant view of innovation (Fleming, 2001), this approach assumes that new knowledge emerges from atypical or cognitively distant

combinations of existing elements. Bibliometric implementations typically assess the novelty of reference combinations, journal pairings, or semantic distances between cited works (Uzzi et al., 2013; Wang et al., 2017). In this study, recombinant novelty follows the operationalization proposed by Shibayama et al. (2021), which uses document-level semantic representations derived from word embeddings. To account for the fact that novelty measures depend on the threshold used to define what constitutes an atypical combination or a new semantic element, we construct multiple variants of both indicators. The indicator r100 corresponds to the most stringent definition, identifying publications whose recombination lies at the extreme tail of the distance distribution. The measures r99, r95 and r90 progressively relax this threshold, classifying as novel those publications that fall within the top 1%, 5%, and 10% of the distribution, respectively. Lower thresholds therefore capture increasingly broader forms of atypical recombination. Specifically, the indicator measures the extent to which the cited knowledge components of a focal publication represent unusual or distant combinations relative to the prior knowledge space. Higher values indicate more atypical recombination of existing knowledge.

Element-level novelty captures a different dimension of novelty: the introduction of new semantic elements into the scientific corpus. Rather than focusing on combinations, this approach identifies whether a publication contains linguistic or conceptual elements that were previously absent from the knowledge space. Following Yin et al. (2023), element-level novelty is derived using contextual word-embedding models that map scientific texts (title and abstract) into high-dimensional semantic space. The indicator evaluates the semantic distance between a focal publication and the existing body of literature, identifying documents that introduce conceptually new elements rather than merely recombining known ones. Similarly, for element-level novelty, we implement alternative cutoffs. The indicator g0 reflects the most stringent criterion, identifying publications that introduce semantic elements absent from the prior corpus. The measures g1, g5, and g10 relax this requirement by allowing increasing degrees of semantic proximity to existing elements, thereby capturing progressively less radical but still substantively new contributions. As highlighted by Shibayama et al. (2025), element-based indicators may capture forms of novelty that recombinant measures overlook, particularly when novelty lies in the articulation of new concepts, phenomena or theoretical framings rather than in atypical citation combinations.

In pharmaceutical research, these two dimensions of novelty correspond to distinct but complementary innovation processes. Recombinant novelty reflects the integration of previously unconnected scientific domains, for example, combining insights from molecular biology, chemistry and clinical science to identify new therapeutic mechanisms. Such atypical knowledge combinations are central to first-in-class drug discovery, where breakthroughs often arise from bridging cognitively distant fields. Element novelty, by contrast, captures the introduction of new concepts, targets, molecular entities or mechanistic framings that expand the scientific knowledge base underlying drug development. In an industry characterized by high scientific uncertainty and strong reliance on frontier research, both recombinant configurations and genuinely new knowledge elements are critical for advancing beyond incremental innovation. Employing both indicators therefore enables a nuanced assessment of how multinational pharmaceutical collaborations relate to different forms of scientific newness.

Recent validation studies demonstrate that no single bibliometric indicator captures all forms of novelty (Shibayama et al., 2025). Recombinant and element-based measures reflect partially overlapping but distinct dimensions of newness. Using both indicators therefore allows for a more comprehensive assessment of novelty in multinational pharmaceutical R&D. Importantly, both indicators are ex-ante, text-based and independent of forward citations, ensuring that the dependent variable reflects knowledge novelty rather than ex-post impact or disruption.

6.2.2 *Independent variable: measure of social ties*

Based on the history of affiliations (listed in the papers that the AstraZeneca researcher coauthored until paper *i*) of each AstraZeneca author, we checked whether any of the academic coauthors in paper *i* had an affiliation that also the AstraZeneca author had in other papers in the past. This implies that the AstraZeneca researcher must have published some papers in the past in which her affiliation was not AstraZeneca, but rather an academic institution (the researcher worked for a period in an academic institution). We built a dummy variable (*Academic social ties*) at the paper level that is equal to 1 if at least one of the AstraZeneca authors in the paper had been listed in a publication in the past with the same affiliation as one of the academic coauthors of the paper. So, for example, the *Academic social ties* dummy would be equal to 1 if we find that in paper *i* published in 2018 one of the authors of the

Goteborg-based AstraZeneca team has published a paper in the past with an "Imperial College" affiliation in London, and in paper *i* there is at least one academic coauthor with an Imperial College affiliation. This implies that one of the AstraZeneca authors has some familiarity with and/or ties to the Imperial College environment, possibly helping her to establish such a collaboration.

Since the notion of having a publication *in the past* with the same academic institution as one of the academic coauthors is still quite vague, we include the additional condition that at least some time must have passed between the year the paper *i* is published and the last time the AZ author was affiliated with the academic institution. We set this to be at least 3 years. To assess the robustness of our findings, we conducted sensitivity analyses extending the observation period to four, five, and six years.

6.2.3 *Controls*

Institutional agreements

To identify *institutional agreements* with universities we used both annual reports and data from PharmaDeals an online database from IQVIA that provides historical information about deals and alliances in the pharmaceutical and biotech sector (see Appendix B for details). For each academic author we checked if their academic institution has signed an institutional agreement with AstraZeneca and in which year this was done. We create a variable *Institutional Agreement* that is equal to one if at least one of the academic authors involved is affiliated with a university with which AZ had an ongoing institutional agreement, and zero otherwise. The variable is time varying, meaning that it is equal to 1 only from the year of establishment of the institutional agreement.[6]

Share of new contacts

The novelty of the paper can also be influenced by the fact that new collaborations are established by authors that have never published together before we therefore adapt the procedure implemented by Liu et al. (2022). Therefore, we control for the effect of new contacts on the overall novelty of papers adapting the procedure already used in the previous section. For each paper *i*, we constructed a variable capturing the share of new academic

[6] Unfortunately, we cannot ascertain whether these collaborations were interrupted at some point in time, therefore our institutional agreement dummy variable is equal to 1 for all the years after the establishment of the agreement.

coauthors relative to all possible AstraZeneca–academic coauthor pairs. For each academic paper the measure is:

$$ShareNewAcademic_i = \frac{\sum_j \text{NewAcademic}_{ji}}{\text{AZAuthors}_i \times \text{AcademicCoauthors}_i}$$

where the numerator sums first-time academic coauthors for each AstraZeneca author *j* in paper i, and the denominator equals the total number of potential AZ–academic coauthor pairs in that paper.

Geographic variables

We also checked for some geographical characteristics of the team of coauthors which may also influence the establishment of collaborations with foreign academics. First, we included the geographical distance (*Distance to R&D sites*) between the foreign academic coauthors and the AstraZeneca R&D centers to understand how far the academics are with respect to the overall R&D network of the company. This variable is computed as follows, for each academic coauthor, we calculated the distance (in kilometers) between their institution and the nearest AstraZeneca R&D center (see Appendix C for details). We then averaged these distances across all coauthors for each paper and took the natural logarithm of the result. Secondly, we counted the number of different countries of the foreign academics (*Num. countries academics*): if the foreign academics are in more than one foreign country the variable takes a value greater than 1.

Company areas of expertise

AstraZeneca operates across multiple therapeutic domains that differ in their scientific maturity, technological trajectories, and depth of internal expertise. Variation across therapeutic areas may influence both collaboration patterns and novelty outcomes. In domains where AstraZeneca has accumulated substantial scientific and organizational experience, researchers may rely more heavily on established collaboration networks and existing academic partners. By contrast, in therapeutic areas that are relatively new or strategically expanding, the firm may engage in broader or more exploratory knowledge sourcing, including collaborations with new and potentially international academic partners.

To account for this heterogeneity, we included controls for the therapeutic area of each publication. Therapeutic areas are identified using Medical Subject Headings (MeSH) terms associated with each publication. MeSH is a standardized controlled vocabulary developed by the U.S. National Library of Medicine (NLM) and used to index articles in PubMed. Each publication is assigned a set of MeSH descriptors that summarize its biomedical content.
We mapped MeSH terms to eight therapeutic areas in which AstraZeneca was active during the study period, namely: Oncology, Immunology, Cardiovascular, Infection, Endocrinology, Respiratory, Muscoloskeletal, and Central Nervous System.

For each publication, we identified how many of its MeSH terms fell within the predefined set of descriptors corresponding to each therapeutic area. We then created a dummy variable that was equal to one if the paper had at least one term that could be assigned to a specific therapeutical area, and zero otherwise. Because articles may span multiple areas, the dummy variables are not mutually exclusive and can jointly describe multidisciplinary research content.

Clinical research

To control for differences in the stage of scientific development, we classified publications according to whether they contribute primarily to basic or clinical knowledge. This classification was constructed using MeSH terms following the methodology developed by Weber (2013). Each publication was first assigned to one or more of three broad topic categories based on its MeSH descriptors including cells and molecules (C), animals and other complex organisms (A) and humans (H). These categories reflect the biological level at which the research is conducted. We then aggregated these topic areas into two knowledge domains: basic science and clinical science. Publications categorized as A and/or C but not H were classified as contributing to basic science, as they predominantly involve cellular, molecular, or pre-clinical animal research. Such research typically corresponds to early-stage discovery and pre-clinical experimentation. Publications categorized as including at least one H descriptor were classified as contributing to clinical science, reflecting research involving human subjects and corresponding to clinical development phases. Articles that could not be assigned to any of the three topic groups (A, C, or H) were excluded from the dataset. A binary indicator for clinical research was included in all regression models to account for systematic differences in collaboration patterns and novelty between basic and clinically oriented research.

Team controls

We also controlled for author-level characteristics that may influence the likelihood of a paper exhibiting higher levels of recombinant or element novelty. The seniority, experience, and productivity of AstraZeneca authors on a given paper may all be correlated with its degree of novelty. Besides counting the number of AstraZeneca coauthors in each paper *i* (*Num of AZ authors*), we also counted the average of the past number of publications among the AstraZeneca researchers in each paper i (*Average past num papers).* We also included a control for whether at least one of the AstraZeneca authors on a given paper had accumulated more than 15 publications at the time of writing, serving as a proxy for the presence of an experienced researcher on the team.[7] This indicator is particularly useful in papers with relatively large AstraZeneca author teams, where an average publication count may obscure the presence of a highly experienced individual contributor.

In order to account for the possible heterogeneity in the team of coauthors we also controlled for the total number of authors in each paper (*Num total authors*), and for the number of coauthors that were affiliated to other companies than AstraZeneca (*Num non-AZ company authors*), the number of coauthors affiliated to a public research organization *(Num PRO authors)*, the number of authors with a hospital affiliation (*Num hospital authors*), and the number of authors with an academic affiliation (*Num academic authors*).

R&D Centers and International Academic Collaboration

As outlined above, we distinguished R&D sites into seven functional categories reflecting different knowledge-sourcing strategies (Global Academic Hubs, Corporate R&D Hubs, International Academic Centers, International R&D centers, Corporate R&D centers, Local R&D Outposts and Outpost for Local Academic Search). From the perspective of knowledge novelty, this heterogeneity is analytically important. Sites with strong international academic engagement are better positioned to access diverse scientific communities and cognitively distant knowledge, thereby broadening the opportunity space for both recombinant and element-level novelty. By contrast, sites with limited academic or international interaction are more likely to draw on established internal trajectories, potentially favoring incremental over novel contributions. AstraZeneca's differentiated R&D geography thus provides a structured

[7] 15 publications represent the 4th quartile of the distribution of the total number of papers by AZ authors.

yet internally varied empirical setting in which to examine how international collaboration and different relational configurations relate to the production of novelty in multinational pharmaceutical research.

INSERT TABLES 6 AND 7 AND FIGURES 1 AND 2 ABOUT HERE

6.3 Descriptive statistics

Table 6 lists each of the variables described with a detailed explanation of what each measures. Table 7 reports descriptive statistics for our sample, that is, all AstraZeneca papers with at least one academic coauthor, broken down into papers involving only domestic academic co-authors and those involving international ones. As noted in Section 4, the sample comprises 2,806 internationally coauthored papers and 2,557 nationally co-authored ones, for all of which a corresponding element novelty score is available. Coverage is somewhat more restricted for the recombinant novelty variable, which reduces the sample to 2,544 international and 2,018 national.

It is worth noticing that regardless of how we measure novelty (Recombinant-type or Element-type novelty) we always find a slightly higher level of novelty among papers with national academic coauthors. This is well illustrated by Figures 1 and 2 where we report the average novelty among international and national academic collaborations. Novelty is always slightly higher among national collaborations, which is in accordance with Wagner et al. 2019.

Secondly, the presence of a social tie is more frequent among national collaborations (12% of the cases) than among international ones (3%). This is not surprising, as it is much more common to have AZ authors working in an R&D center in a specific country who were in the past affiliated to a university in the same country. It speaks to the fact that most of AZ researchers have a relatively nationally bounded career and continue to collaborate with researchers from their alma mater in the same country.

As would be expected, internationally co-authored papers display a greater average geographic distance from AstraZeneca's R&D centers. They also tend to involve author teams, counting

AstraZeneca authors, academic co-authors, and other affiliated contributors, with higher average publication counts than those found in national collaborations. Papers with international academic co-authors are more likely to include authors with hospital or public research organization (PRO) affiliations and have a slightly higher average number of academic co-authors (4.4 versus 3.5 for domestic papers).

Most international collaborations originate from AstraZeneca authors based at the Global Academic Hubs of Cambridge, Mölndal, and Södertälje, which account for 70% of international papers compared to 46% of national ones. Conversely, international collaborations are less likely to involve Local R&D Centres (4% vs. 12%) or Corporate R&D Hubs (9% vs. 23%). Among therapeutic areas, the most notable difference concerns cardiology, which is proportionally more common in international papers than national (23% vs. 11%); for the remaining therapeutic areas, no substantial differences emerge.

7 Results

7.1 The increase in academic international collaborations

Table 8 reports the results of the difference-in-differences estimations. The dependent variable is the share of new international coauthors in each paper. Across all specifications, the coefficient on *Academic paper* is negative and statistically significant. This indicates that, prior to the treatment period, papers involving academic coauthors exhibit a lower share of new international collaborators compared to papers coauthored with researchers from other companies. In other words, academic collaborations appear more stable and recurrent on average, whereas company collaborations involve a higher proportion of first-time foreign partners.

The main coefficient of interest is the interaction term between Academic paper and the post-treatment period. When the treatment year is set to 2013 (Column 1), the interaction coefficient is positive and statistically significant ($\beta = 0.064$, $p < 0.01$). This implies that after 2013, academic papers experienced a differential increase in the share of new international collaborators relative to company papers. The magnitude of the coefficient suggests that the post-2013 period is associated with an approximately 6–8 percentage point increase in the share of new foreign academic coauthors, depending on the specification. This result is robust to alternative definitions of the treatment year. When the post period is shifted to 2014 or 2015

(Columns 2 and 3), the interaction term remains positive and statistically significant, indicating that the shift in collaboration patterns was not confined to a single year but persisted in the subsequent period. Importantly, placebo tests using earlier years as the treatment threshold do not produce significant interaction effects (results available upon request), providing support for the identifying assumption that the observed shift is specific to the post-2013 period.

INSERT TABLE 8 AND FIGURES 3 ABOUT HERE

Figure 3 complements these results by plotting adjusted predictions of the share of new international collaborators over time for academic and non-academic papers. Prior to 2013, the two trends are relatively stable and move in parallel. After 2013, however, the trajectory of academic papers diverges upward, while company papers display no comparable increase. The figure therefore visually corroborates the regression results, illustrating a post-2013 acceleration in the propensity of AstraZeneca publications to involve new foreign academic collaborators.

The coefficients on the R&D center controls further indicate substantial heterogeneity across AstraZeneca's research network. Relative to *Global Academic Hubs* (the reference category), most other site types exhibit significantly lower shares of new international collaborators. This pattern is consistent with the descriptive evidence that internationally oriented and academically embedded hubs play a central role in connecting the firm to foreign knowledge networks. Importantly, the positive and significant interaction effect remains robust after controlling for this spatial heterogeneity, suggesting that the post-2013 shift is not solely driven by changes in the geographic composition of publishing activity.

Taken together, the evidence suggests that the period following the strategic disruption in the early 2010s was associated with a reorientation toward new international academic partnerships, beyond general trends in external collaboration.

7.2 Social ties and novelty

Table 9 reports the OLS estimates for Recombinant-type novelty (R-novelty), distinguishing between papers co-authored with international academics and those with national academic collaborators. Across all specifications, social ties exhibit a clear and systematic pattern. For international collaborations, the coefficient on *Academic social tie* is positive and statistically significant across all R-novelty thresholds (r100–r90). The magnitude declines as the novelty definition becomes less stringent, but the effect remains robust. Substantively, this indicates that when at least one AstraZeneca researcher has previously worked at the foreign institution of the academic collaborator, the resulting publication exhibits higher recombinant novelty. In contrast, no statistically significant association emerges for national collaborations. Across all R-novelty specifications, prior institutional ties within the same country are not associated with higher novelty levels. The absence of an effect in the national sample suggests that the value of social ties is not universal, but conditional on geographic configuration.

The control variables further support this interpretation. First, we find that the share of new academic contacts is positive and significant only among international academic papers: this finding resonates the previous results of Table 8. The surge in new international academic contacts observed after 2013 is also linked to a higher degree of novelty of the papers that engage new international academics. This is not the case for papers that only engage national academics.

Institutional agreements between AZ and the universities of the academic coauthors are not associated with higher levels of recombinant novelty. Interestingly, the coefficient for national collaboration is negative and significant. This indicates that institutional agreements with local universities rather produce conventional science. The distance to R&D sites is negatively associated with novelty, and clinical research projects exhibit lower recombinant novelty, consistent with the incremental and regulatory-driven nature of later-stage development. The number of countries represented among academic collaborators is negatively related to novelty in the international sample, suggesting that excessive geographic dispersion may introduce coordination frictions that limit effective recombination. Importantly, the social-tie effect persists after controlling for team composition, therapeutic area, and R&D center typology.

INSERT TABLES 9 AND 10 ABOUT HERE

Table 10 presents parallel estimations using element-type novelty (G-novelty). The results mirror the recombinant findings but are slightly more nuanced. For international collaborations, social ties are positively associated with G-novelty across the less stringent thresholds (g1, g5, g10). The effect is not statistically significant under the most restrictive definition (g0), which captures only the most extreme element-level novelty. This suggests that prior institutional ties facilitate the integration of relatively new knowledge elements but may be insufficient to generate the most radical breakthroughs. Again, no significant effect is detected in the national sample. The divergence between international and national collaborations is consistent across both novelty measures, reinforcing the conditional interpretation.

Taken together, the results indicate that social proximity operates as a compensatory mechanism under geographic distance. Prior social ties appear to enable more effective recombination when collaborations span national borders, where interpretative and coordination barriers are higher. In national collaborations, where institutional and contextual alignment is already stronger, such ties do not yield additional novelty benefits.

The findings therefore support a conditional view of proximity: social ties matter not universally, but specifically when knowledge integration occurs across spatial and institutional divides.

7.2.1 Robustness checks: definition of social ties and total number of authors

In Appendix D we run a series of robustness checks. First, we examine whether our results are sensitive to how we measure social ties. Our baseline definition uses a three-year lag to identify a social tie; that is, at least three years must have passed between the last year in which an AZ researcher was affiliated with a given academic institution and the year of publication of the paper. We then relax this assumption to explore what happens when shorter or longer lags are allowed.

We do this by plotting the estimated marginal effects across novelty thresholds for both recombinant novelty and element novelty, changing in each figure the lag used for the definition of social tie. In each figure we report the marginal coefficients for both international academic papers (figure on the left) and national papers (figure on the right).

Figures A2 and A6 (based on a minimum lag of three years) serve as the benchmark, reproducing the marginal coefficients from the main specifications in Tables 9 and 10 and confirming that the results are consistent with those reported in the main regression tables. Figures A1 and A5 apply a two-year minimum lag.[8] and show that the positive association with novelty remains strong for recombinant novelty, while becoming slightly less statistically significant for element-based novelty. Increasing the minimum lag to four or five years (in the remaining figures) continues to confirm the positive effect of social ties, with one exception: for recombinant novelty, the coefficient loses some statistical significance when the lag is extended to five years. For national collaborations, by contrast, the relationship with novelty is never statistically distinguishable from zero regardless of the lag chosen, for either recombinant or element-based novelty. Taken together, these results confirm that the main findings are not sensitive to the choice of lag, and that the relevance of prior social ties for novelty in international collaborations remains robust across specifications.

A second robustness check concerns the number of authors in a paper. Although we already exclude papers with more than 100 authors from our sample (and control for the total number of authors in our regressions), we further verify that our results are not driven by outliers with an unusually large number of authors. In other words, we want to ensure that we focus on papers in which the contribution of each AZ researcher and academic co-author is meaningful. We therefore restrict the sample to papers with no more than 20 authors.

In the Appendix, tables A.1 and A.2 report the results of the model estimation focusing only on international collaborations. The coefficient on the social tie dummy remains positive and statistically significant, consistent with the main specification, indicating that our results are not driven by papers with a very high number of authors. The positive coefficient of social ties is robust also to the restriction of the analysis to papers with no more than 10 authors.

7.3 Heterogeneity analysis

7.3.1 Distinguishing between R&D center typology

[8] This means that we also consider cases in which only two years have passed between the last affiliation of the AZ researcher with the academic institution and the publication year of the paper in which the researcher is affiliated with AZ.

Finally, in Tables 13 and 14 we examine whether our main results regarding the importance of social ties in international collaborations vary depending on the type of AZ R&D center in which researchers are located. We distinguish between AZ researchers working in the main R&D centers, those characterized by a high volume of publications and a strong level of academic collaboration (the Global Academic Hubs, located in Cambridge, Mölndal/Göteborg, and Stockholm/Södertälje), and those working in other R&D centers.

We estimate the same model as before, focusing only on international collaborations, but now splitting the sample between papers authored by AZ researchers located in Global Academic Hubs and papers authored by AZ researchers located in other R&D centers. It should be noted that approximately 70% of international papers in our sample involve AZ researchers based in Global Academic Hubs.

INSERT TABLES 11 AND 12 ABOUT HERE

The results for recombinant novelty (Table 11) and element-based novelty (Table 12) show that the positive effect of social ties is mainly driven by AZ researchers located in the Global Academic Hubs, as the coefficient is positive and statistically significant only in this subsample. In other R&D centers, the coefficient of the social tie variable is slightly smaller and less precisely estimated, resulting in a lack of statistical significance (also due to the smaller number of observations).

These findings suggest that our baseline results are largely driven by the social ties of AZ researchers working in the main R&D centers, where publication output and the propensity to collaborate with international academic partners are particularly high.

7.3.2 Social ties and the role of distance to R&D sites

In Tables 13 and 14 we test whether the correlation between social ties and novelty changes according to how far international academic researchers are from AZ R&D sites. The variable that we interact social tie with is the average distance (in log) of academic coauthors to the closest R&D site.

INSERT TABLES 13 AND 14 AND FIGURES 4 AND 5 ABOUT HERE

The results in Tables 13 and 14 reveal a nuanced picture that varies depending on the type of novelty considered. For recombinant novelty (Table 13), the coefficient on social ties is negative but not statistically significant, while the interaction term is positive and significant. This implies that the effect of social ties on recombinant novelty is positive when the average distance between academic co-authors and AstraZeneca R&D sites is sufficiently large. As shown in figure 4, the social tie coefficient becomes positive and significant for values of log average distance above 6, corresponding to approximately 400 km.

The pattern is reversed for element-based novelty (Table 14), where the social tie coefficient is positive and significant, and the interaction term is negative and significant. Here, the beneficial effect of social ties on novelty holds even at very short distances from AstraZeneca R&D sites, but diminishes as distance increases, with the coefficient becoming indistinguishable from zero at a log average distance of approximately 6 (around 400 km), as illustrated in Figure 5. These results underscore the importance of considering different proxies for novelty as the various indicators are capturing different aspects of new knowledge. The results of previous studies that have used only one simple indicator might have been overstated.

8 Discussion and conclusions

The objective of this study was to clarify whether and under which conditions international industry–academic collaboration is associated with higher levels of scientific novelty within a multinational pharmaceutical firm. The results contribute to ongoing debates in economic geography and innovation studies concerning the relationship between geographic dispersion, proximity configurations, and knowledge creation. While early work has demonstrated that innovation is spatially structured, more recent research emphasizes that spatial proximity alone cannot fully account for innovation outcomes, and that multiple, intersecting proximity dimensions, including organizational, cognitive, and social proximities, jointly shape creative recombination across space (Boschma, 2005; Balland et al., 2015; Shkolnykova, 2023). At the

same time, studies of international research networks show that cross-border collaboration is often associated with higher impact research, but conventional rather than novel research (Wagner et al., 2019). However, comparatively less attention has been devoted to the relational conditions under which geographic diversity translates into recombinant outcomes, particularly in firm-based industry–academic collaborations within multinational corporations. By combining a grouped difference-in-differences design with publication-level novelty measures, this study provides evidence on both the strategic reorientation of collaboration and the conditional role of social proximity.

The difference-in-differences results indicate that following the strategic turbulence of the early 2010s, AstraZeneca publications with academic coauthors experienced a significant relative increase in the share of new international collaborators compared to publications coauthored with other companies. This pattern is consistent with accounts of the growing centrality of academic science in pharmaceutical innovation (Cockburn & Henderson, 1998; Owen-Smith & Powell, 2004) and with arguments that periods of strategic pressure intensify firms' reliance on external, frontier-oriented knowledge sources (Pisano, 2006; Schuhmacher et al., 2016). The evidence suggests an acceleration of "global pipelines" specifically directed toward academia (Bathelt et al., 2004), rather than a general increase in international partnering. This finding contributes micro-level evidence to the literature on multinational R&D networks by showing that the internationalization of collaboration is not static but can shift in response to organizational disruption and strategic reorientation.

The central contribution of the paper lies in the novelty analysis. Across multiple specifications and novelty thresholds, prior social ties, operationalized as shared institutional affiliation in earlier career stages between AZ researchers and their academic co-authors, are positively associated with novelty in international industry–academic collaborations, while no such association is observed in national collaborations. This pattern aligns with proximity theory's core proposition that innovation outcomes depend on the configuration of proximity dimensions rather than on geographic distance alone (Boschma, 2005; Huber, 2012; Balland, Boschma, & Frenken, 2015). Social proximity appears to function as a conditioning mechanism that is particularly consequential when collaborations span national and institutional boundaries. In national collaborations, where institutional environments and opportunities for interaction are more aligned, the marginal contribution of prior ties to novelty is limited. In international settings, by contrast, prior affiliation-based ties are associated with higher novelty

outcomes, suggesting that relational embeddedness becomes more salient under conditions of greater spatial and institutional dispersion.

This conditional pattern addresses an important gap in the literature. Existing research has shown that international collaboration is often associated with higher citation rate (Glänzel & Schubert, 2001), more interdisciplinarity (van Raan, 2003), but less novel outputs (Wagner et al., 2019), while relational embeddedness has been theorized to facilitate knowledge transfer (Reagans & McEvily, 2003; Ooms et al., 2018). However, there has been limited systematic evidence demonstrating that the novelty associated with cross-border collaboration depends on pre-existing social ties within multinational firm contexts. By distinguishing between international and national industry–academic collaborations, this study moves beyond estimating an average "distance effect" and instead identifies a configurational relationship between geographic dispersion and social proximity, consistent with the relational turn in economic geography (Boschma, 2005; Crescenzi et al., 2016; Roth & Mattes, 2023).

The use of both recombinant (R-type) and element-level (G-type) novelty measures further refines the interpretation. Social ties are robustly associated with recombinant novelty across thresholds in international collaborations, consistent with theories emphasizing the importance of bridging cognitively distant knowledge domains for innovative recombination (Uzzi et al., 2013; Wang et al., 2017). For element-level novelty, the effect is present for less stringent thresholds but not for the most restrictive definition, which captures only the most radical semantic departures (Shibayama et al., 2021; Yin et al., 2023). This suggests that prior affiliation-based ties may be particularly helpful for facilitating meaningful recombination and moderately strong semantic newness, while the most extreme forms of element novelty may depend on additional factors beyond relational familiarity. In doing so, the paper contributes to the growing literature emphasizing the multidimensional nature of novelty and the importance of employing complementary indicators (Shibayama et al., 2025). The divergent moderating role of geographic distance across the two novelty dimensions is theoretically interesting. For recombinant novelty, social ties matter most at greater distances, while for element-level novelty the pattern reverses. That the two novelty dimensions respond in opposite ways to the same moderator underscores that they capture qualitatively distinct knowledge creation processes, and further supports the case for multi-indicator approaches to measuring novelty in innovation research.

Several control results reinforce the view that geographic dispersion introduces barriers to knowledge integration. The negative association between geographic distance and novelty, as well as the negative relationship between the number of academic countries represented and recombinant novelty in the international sample, indicate that dispersion can introduce coordination and interpretative burdens (Storper & Venables, 2004; Boschma, 2005). These findings resonate with arguments that diversity expands the opportunity space for recombination but does not automatically generate novel outcomes (Wagner et al., 2019). Instead, novelty appears to depend on relational mechanisms that help translate geographic diversity into coherent research outputs.

The absence of a significant effect for international collaborations, and the negative coefficient for national ones, suggests that institutional agreements between AstraZeneca and academic institutions are not primarily oriented towards the generation of novel knowledge. Such agreements may instead serve other strategic purposes, for instance, facilitating access to more applied or development-stage research, signaling commitment to the academic community, or enabling the sharing of infrastructure and research facilities rather than driving genuine knowledge recombination.

Taken together, the findings advance three interrelated strands of literature. First, they contribute to economic geography by demonstrating that the relationship between international collaboration and novelty is conditional on relational embeddedness, thereby operationalizing the configurational logic of proximity theory in a multinational firm setting. Second, they enrich research on industry–academic collaboration in science-based sectors by identifying a micro-level relational mechanism, prior shared institutional affiliation, through which cross-border academic partnerships are associated with higher novelty (Cockburn & Henderson, 1998; Perkmann et al., 2013; Fassio et al., 2023). Third, they add to the novelty-measurement literature by showing that the conditioning role of social proximity varies across novelty dimensions and thresholds, underscoring the value of multi-indicator approaches.

At the same time, the results should be interpreted within the study's scope. The operationalization of social proximity captures one observable form of relational embeddedness and likely underestimates broader informal or professional networks. Moreover, the analysis identifies conditional associations at the publication level and does not directly observe within-team coordination processes. Nevertheless, the consistent asymmetry between

international and national collaborations across novelty measures provides robust empirical support for a relational conditioning interpretation.

Overall, the study suggests that in multinational pharmaceutical research, international industry–academic collaboration is not inherently more novel. Rather, novelty is more strongly associated with cross-border collaboration when it is anchored in prior social ties. This finding refines existing understandings of global knowledge sourcing by showing that the innovative benefits of international collaboration depend not only on access to diverse scientific environments but also on the relational infrastructures that connect actors across spatial and institutional divides.

The null or negative association between formal institutional agreements and novelty suggests that such partnerships may not be the most effective for stimulating frontier knowledge creation. For managers, this implies that broad institutional agreements should not be treated as a substitute for the kind of researcher-level relational embeddedness that is shown here to drive novelty. Agreements may rather serve purposes in terms of infrastructure sharing, IP management. For policymakers, these results are a cautionary note for instruments that incentivise industry–academic collaboration primarily by counting formal partnerships. The quality of knowledge exchange matters more than its institutional form, and support mechanisms that foster deeper researcher-level interactions, such as joint doctoral programmes, visiting scholar schemes, co-location initiatives, are likely to be more effective at stimulating the kind of relational ties that translate geographic and institutional diversity into genuine scientific novelty.

Several limitations of this study should be acknowledged. First, the novelty analysis does not rest on a comparable causal identification strategy. The association between prior social ties and novelty is estimated with extensive controls and is robust across specifications, but unobserved factors, such as the selective matching of experienced researchers to more ambitious projects, could contribute to the observed pattern. Second, the generalisability of the findings is inherently constrained by the single-firm design. AstraZeneca is a large, science-intensive multinational with a distinctive R&D history and a particularly strong tradition of academic engagement. Whether the relational mechanisms identified here operate similarly in other pharmaceutical firms, in different industries, or in firms with less internationalized R&D networks remains an open question. The single-firm setting is also, however, a strength in that

it allows for fine-grained analysis of collaboration patterns and organizational context that would be lost in cross-firm comparisons. Third, the operationalization of social proximity relies on a single, observable form of relational embeddedness, prior shared institutional affiliation between AstraZeneca researchers and their academic co-authors. While this is a theoretically grounded and systematically measurable proxy, it inevitably underestimates the broader landscape of informal ties, repeated interactions, and professional networks that may also shape collaboration quality and outcomes. Similarly, the novelty measures, while capturing meaningful dimensions of scientific contribution, do not exhaust all relevant aspects of knowledge creation, and results may vary with alternative operationalization. Finally, the analysis operates at the level of publications and does not directly observe the within-team coordination processes through which prior ties may facilitate or constrain knowledge integration. Future research combining bibliometric analysis with qualitative or survey-based approaches could shed further light on the mechanisms underlying the patterns documented here.

References

Ao, J., Johnson, D., & Esposito, S. E. (2026). Where open innovation partnership matters: Phase-specific patterns in R&D project counts for large pharmaceutical firms. *Technovation*, *150*, 103459.

Araújo, C., & Teixeira, A. (2014). Determinants of international technology transfer: An empirical analysis of the Enterprise Europe Network. *Journal of technology management & innovation*, *9*(3), 120-134.

Asheim, B., Coenen, L., Moodysson, J., & Vang, J. (2007). Constructing knowledge-based regional advantage: implications for regional innovation policy. *International Journal of Entrepreneurship and Innovation Management*, *7*(2-5), 140-155.

Audretsch, D. B., & Belitski, M. (2024). Geography of knowledge collaboration and innovation in Schumpeterian firms. *Regional Studies*, *58*(4), 821-840.

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American economic review*, *86*(3), 630-640.

Autant-Bernard, C., Mairesse, J., & Massard, N. (2007). Spatial knowledge diffusion through collaborative networks. *Papers in Regional science*, *86*(3), 341-350.

Balland, P. A., Boschma, R., & Frenken, K. (2015). Proximity and innovation: From statics to dynamics. *Regional studies*, *49*(6), 907-920.

Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in human geography*, *28*(1), 31-56.

Belderbos, R., Faems, D., Leten, B., & Looy, B. V. (2010). Technological activities and their impact on the financial performance of the firm: Exploitation and exploration within and between firms. *Journal of Product Innovation Management*, *27*(6), 869-882.

Belderbos, R., Grabowska, M., Kelchtermans, S., Leten, B., Jacob, J., & Riccaboni, M. (2021). Whither geographic proximity? Bypassing local R&D units in foreign university collaboration. *Journal of international business studies*, *52*(7), 1302.

Bignami, F., Mattsson, P., & Hoekman, J. (2020). The importance of geographical distance to different types of R&D collaboration in the pharmaceutical industry. *Industry and Innovation*, *27*(5), 513-537

Bignami, F., & Mattsson, P. (2019). Potential effects of increased openness in pharma: the original knowledge behind new drugs. *Drug discovery today*, *24*(10), 1957-1962.

Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional studies*, *39*(1), 61-74.

Boschma, R. A., & Lambooy, J. G. (1999). Evolutionary economics and economic geography. *Journal of evolutionary economics*, *9*(4), 411-429.

Cantwell, J., & Mudambi, R. (2005). MNE competence-creating subsidiary mandates. *Strategic management journal*, *26*(12), 1109-1128.

Cockburn, I. M., & Henderson, R. M. (2001). Scale and scope in drug development: unpacking the advantages of size in pharmaceutical research. *Journal of health economics*, *20*(6), 1033-1057.

Crescenzi, R., Nathan, M., & Rodríguez-Pose, A. (2016). Do inventors talk to strangers? On proximity and collaborative knowledge creation. *Research policy*, *45*(1), 177-194.

D'Este, P., & Patel, P. (2007). University–industry linkages in the UK: What are the factors underlying the variety of interactions with industry?. *Research policy*, *36*(9), 1295-1313.

Fassio, C., Geuna, A., & Rossi, F. (2023). How do firms reach out to foreign universities? Inventors' personal characteristics and the multinational structure of firms. *Journal of World Business*, *58*(3), 101431.

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research policy*, *30*(7), 1019-1039.

Frigon, A., & Rigby, D. L. (2024). Geographies of knowledge sourcing and the complexity of knowledge in multilocational firms. *Economic Geography*, *100*(4), 329-350.

Galaso, P., & Kovářík, J. (2021). Collaboration networks, geography and innovation: Local and national embeddedness. *Papers in Regional Science*, *100*(2), 349-378.

Garcia Martinez, M., Zouaghi, F., & Sanchez Garcia, M. (2025). Geographical and cognitive proximity effects on innovation performance: Which types of proximity for which types of innovation?. *European Management Review*, *22*(2), 291-317.

Gertler, M. S. (2003). Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of economic geography*, *3*(1), 75-99.

Grabher, G. (1993). The weakness of strong ties; the lock-in of regional development in Ruhr area. *The embedded firm; on the socioeconomics of industrial networks*, 255-277.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American journal of sociology*, *91*(3), 481-510.

Gupta, A. K., & Govindarajan, V. (2000). Knowledge flows within multinational corporations. *Strategic management journal*, *21*(4), 473-496.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, *108*(3), 577-598.

de Jong, J. P., & Freel, M. (2010). Absorptive capacity and the reach of collaboration in high technology small firms. *Research policy*, *39*(1), 47-54.

Liu M, Bu Y, Chen C, Xu J, Li D, Leng Y, Freeman RB, Meyer ET, Yoon W, Sung M, Jeong M, Lee J, Kang J, Min C, Song M, Zhai Y, Ding Y. (2022) Pandemics are catalysts of scientific novelty: Evidence from COVID-19. *Journal of the Association for Information Science and Technology*. 73(8):1065-1078.

Nan, D. (2024). The influence of proximity and knowledge base on recombination innovation in R&D collaboration. *Plos one*, *19*(2), e0298735.

Narula, R., & Santangelo, G. D. (2012). Location and collocation advantages in international innovation. *Multinational Business Review*, *20*(1), 6-25.

Ooms, W., Werker, C., & Caniëls, M. (2018). Personal and social proximity empowering collaborations: the glue of knowledge networks. *Industry and Innovation*, *25*(9), 833-840.

Owen-Smith, J., & Powell, W. W. (2004). Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community. *Organization science*, *15*(1), 5-21.

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'este, P., ... & Sobrero, M. (2013). Academic engagement and commercialisation: A review of the literature on university–industry relations. *Research policy*, *42*(2), 423-442.

Perri, A., Scalera, V. G., & Mudambi, R. (2017). What are the most promising conduits for foreign knowledge inflows? innovation networks in the Chinese pharmaceutical industry. *Industrial and Corporate Change*, *26*(2), 333–355.

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*, 116-145.

Reagans, R., & McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative science quarterly*, *48*(2), 240-267.

Plotnikova, T., & Rake, B. (2014). Collaboration in pharmaceutical research: Exploration of country-level determinants. *Scientometrics*, *98*(2), 1173-1202.

Roth, P., & Mattes, J. (2023). Distance creates proximity: Unraveling the influence of geographical distance on social proximity in interorganizational collaborations. *Environment and Planning A: Economy and Space*, *55*(6), 1372-1391.

Schuhmacher, A., Gassmann, O., & Hinder, M. (2016). Changing R&D models in research-based pharmaceutical companies. *Journal of translational medicine*, *14*(1), 105.

Shibayama, S., Wu, Z., Yin, D., & Yokota, K. (2025). State of the Art of Novelty Indicators. *Available at SSRN 5379973*.

Shibayama, S., Yin, D., & Matsumoto, K. (2021). Measuring novelty in science with word embedding. *PloS one*, *16*(7), e0254034.

Shkolnykova, M. (2023). Assessing the importance of proximity dimensions for the diffusion of radical innovations in German biotechnology. *European Planning Studies*, *31*(7), 1510-1531.

Storper, M., & Venables, A. J. (2004). Buzz: face-to-face contact and the urban economy. *Journal of economic geography*, *4*(4), 351-370.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468-472.

Visser, E. J., & Boschma, R. (2004). Learning in districts: Novelty and lock-in in a regional context. *European Planning Studies*, *12*(6), 793-808.

Wagner, C. S., Whetsell, T. A., & Mukherjee, S. (2019). International research collaboration: Novelty, conventionality, and atypicality in knowledge recombination. *Research policy*, *48*(5), 1260-1270.

Wang, J., Veugelers, R. & P.Stephan, (2017) Bias against novelty in science: a cautionary tale for users of bibliometric indicators, *Research Policy*, 46 (8) (2017), pp. 1416-1436

Werker, C., Ooms, W., & Caniëls, M. C. (2016). Personal and related kinds of proximity driving collaborations: a multi-case study of Dutch nanotechnology researchers. *SpringerPlus*, *5*(1), 1751.

Yin, D., Wu, Z., Yokota, K., Matsumoto, K., & Shibayama, S. (2023). Identify novel elements of knowledge with word embedding. *Plos one*, *18*(6), e0284567.

Zucker, L. G., Darby, M. R., & Armstrong, J. S. (2002). Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology. *Management science*, *48*(1), 138-153.

TABLES AND FIGURES

Table 1. The list of R&D centers of AstraZeneca during the period 2000-2020.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Bangalore (IND)	1	1	1	1	1	1	1	1	1	1	1	1	1	1							
Boston/Waltham (US)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cambridge (UK)								1	1	1	1	1	1	1	1	1	1	1	1	1	1
Gaithersburg (US)								1	1	1	1	1	1	1	1	1	1	1	1	1	1
Loughbourhood (UK)	1	1	1	1	1	1	1	1	1	1	1	1									
Lund (SWE)	1	1	1	1	1	1	1	1	1	1	1	1									
Macclesfield/Alderley park (UK)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Montreal (CAN)	1	1	1	1	1	1	1	1	1	1	1	1									
Mountain View (US)								1	1	1	1	1	1	1	1	1	1	1	1	1	1
Mölndal (SWE)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
New York (US)																				1	1
Osaka (JAP)						1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Rheims (FRA)			1	1	1	1	1	1	1	1	1	1									
Shanghai (CHI)							1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Södertälje (SWE)	1	1	1	1	1	1	1	1	1	1	1	1									
Tokyo (JAP)						1	1	1	1	1											
Warsaw (POL)																1	1	1	1	1	1
Wilmington (US)	1	1	1	1	1	1	1	1	1	1	1	1									

Source: authors' calculations using company annual reports

Table 2. Major Acquisitions by AstraZeneca (2000–2020)

year	Acquired company	Cost	Significance
2005	KuDos	approx. £120 million	Smaller deal of a DNA repair technology that could be used to treat cancer. One of the products was undergoing clinical trials.
2006	Cambridge Antibody Technology (UK)	approx. £702 million	Brought in leading expertise in monoclonal antibody technologies. This acquisition laid the foundation for AZ's biologics strategy before MedImmune
2007	MedImmune (USA)	approx. $15.6 billion	Transformed AstraZeneca into a major player in biologics and vaccines. MedImmune became AZ's dedicated biologics arm, central to its oncology and immunology pipeline
2007	Spinal Therapeutics / Arrow Therapeutics (UK)	-	Smaller deals, focused on anti-infectives and niche therapeutic areas
2012	Amylin Pharmaceuticals	approx. $3.4 billion	A joint acquisition together with Bristol-Myers Squibb. This was an expansion of AZ's existing diabetes alliance.
2013	Amplimmune (USA)	approx. $225 million.	Strengthened AZ's immuno-oncology capabilities, particularly in T-cell modulation
2013	Pearl Therapeutics	approx. $1.15 billion	Access to Pearl's promising pipeline of inhaled bronchodilator products for the treatment of asthma and chronic obstructive pulmonary disease.
2010's	Zoladex rights in certain markets	-	Regional acquisitions of rights to specific drugs to consolidate oncology portfolio.
2015	ZS Pharma	approx. $2.7 billion	Primarily driven by AZ's goal to strengthen its cardiovascular and metabolic disease portfolio.
2016	Acerta Pharma (Netherlands/USA)	$4 billion (initial + milestones)	Gave AZ access to acalabrutinib, a next-generation BTK inhibitor for blood cancers

Table 3. Key alliances by AstraZeneca

Year	Name of partner	Therapeutic area and significance
2004	FibroGen (USA)	focused on anemia drug Roxadustat
2010	Rigel Pharmaceuticals (USA)	immunology/oncology
2015	Ionis Pharmaceuticals (USA)	Antisense therapies. Multiple collaborations
2018	Innate Pharma (France)	immuno-oncology
2020	Silence Therapeutics (UK)	RNAi partnership

Table 4. Number of academic papers and number of international academic papers in each R&D site[9]

R&D Site	R&D Employees estimates[10]	Non-Academic Pubs.	Academic Pubs.	National Academic Pubs.	International Academic Pubs.	Share Academic Pubs.	Share International Academic Pubs.
Bangalore (IND)	300-700	781	178	36	142	19%	80%
Boston (US)[11]	800-1,500	1,871	404	186	218	18%	54%
Cambridge (UK)	2,500-3,500	5,466	4,209	1,386	2,823	44%	67%
Gaithersburg (US)	2,500-4,000	1,052	660	257	403	39%	61%
Loughborough (UK)	1,200	1,221	517	392	125	30%	24%
Lund (SWE)	900	684	426	134	292	38%	69%
Macclesfield (UK)	300-800	5,835	2,645	1,718	927	31%	35%
Molndal (SWE)	2,000-3,000	5,190	3,795	1,244	2,551	42%	67%
Montreal (CAN)	100-250	390	105	81	24	21%	23%
Osaka (JAP)	300-600	79	117	94	23	60%	20%
Reims (FRA)	50-200	296	16	8	8	5%	50%
Shanghai (CHI)	100-1800	222	311	253	58	58%	19%
Sodertalje (SWE)	1,400-1,800	3,041	1,952	655	1,297	39%	66%
Waltham (US)	700-1200	1,032	361	173	188	26%	52%

[9] If a focal author publishes two papers she is counted twice, if a paper includes two AZ focal authors the paper is counted twice.

[10] The number of R&D employees in each site is based on an AI-assisted online search on websites, company annual reports and other online sources to reach an approximate understanding of the size of the R&D center. The number of employees refers to the last available year, so for R&D centers that were closed it refers to the period before the closing. For some still existing sites it refers to the last available year.

[11] Differently from the previous table, here we separate the more recent R&D center of Boston from the older one based in Waltham Massachusetts, although they are less than 25 km distant. This allows us to look at the specific publication dynamics of each of the two research centers.

Wilmington (US)	400-900	1,616	714	449	265	31%	37%
Small R&D sites							
Mountain View (US)	150-350	4	0	0	0	0%	0%
New York (US)	50-200	0	2	1	1	100%	50%
Warsaw (POL)	700-1,500	5	6	3	3	55%	50%
Tokyo (JAP)	100-300	4	7	6	1	64%	14%
Total		**28,789**	**16,425**	**7,076**	**9,349**	**36%**	**57%**

Table 5. Typology of R&D centers, based on academic collaborations[12]

R&D center	Size center	High academic engagement (>35%)	International academic networks (>50%)	Typology	Center open at the end of the period
Cambridge	LARGE	YES	YES	Global Academic Hub	YES
Molndal	LARGE	YES	YES	Global Academic Hub	YES
Sodertalje	LARGE	YES	YES	Global Academic Hub	NO
Macclesfield	LARGE	NO	NO	Corporate R&D Hub	YES
Gaithersburg	MEDIUM	YES	YES	International Academic center	YES
Lund	MEDIUM	YES	YES	International Academic center	NO
Boston	MEDIUM	NO	YES	International R&D center	YES
Waltham	MEDIUM	NO	YES	International R&D center	YES
Loughborough	MEDIUM	NO	NO	Local R&D center	NO
Wilmington	MEDIUM	NO	NO	Local R&D center	NO
Bangalore	SMALL	NO	YES	Local R&D Outpost	NO
Reims	SMALL	NO	YES	Local R&D Outpost	NO
Osaka	SMALL	YES	NO	Outpost for local academic search	YES
Shanghai	SMALL	YES	NO	Outpost for local academic search	YES
Montreal	SMALL	NO	NO	Local R&D center	NO

Table 6. Description of the variables used in the regressions

Dependent variable	Description of variables
Recombinant-Novelty of the paper (R)	Recombinant novelty measured using document-level semantic representations derived from word embeddings, following the procedure by Shibayama, Yin and Matsumoto (2021)
Element-Novelty of the paper (G)	Element novelty measured using contextual word-embedding models to identify documents that introduce conceptually new elements, rather than merely recombining known ones (Yin et al., 2023)
Share of new international contacts	The number of coauthors of the AZ authors that are new for them and that are affiliated to other institutions (universities or companies), divided by all possible pair combinations between the AZ authors and their coauthors (based on Liu et al. 2022).
Independent variables	
Academic social tie	Dummy equal to 1 if in the past (up to 3 years before) at least one of the AZ authors has published a paper with the same affiliation of one (or more) of the academic coauthors.
Share of new academic contacts	Number of AZ authors that coauthor for the first time with each of the academic coauthors in a specific paper, divided by all the possible combinations between AZ and academic authors present in the paper (based on Liu et al. 2022)

[12] We did not include the smallest R&D sites in terms of publications (Mountain View, New York, Warsaw and Tokyo), since there are not enough publications to draw robust conclusions about their academic engagement.

Institutional agreement	Dummy equal to 1 if at least one of the academic co-authors are affiliated to a university that has an ongoing (at the time of the publication) institutional agreement with AZ.
Average distance to R&D sites	For each academic coauthor in a paper we calculate the km distance between his/her institution and the closest AstraZeneca R&D site, then take the average of such distance for all the academic coauthors in the paper and do a log transformation of 1 plus the value of such average distance.
Average past number of papers	Average of the number of papers written by the AZ authors of the paper in the previous years
Productive AZ researcher	Dummy equal to 1 if there is at least one AZ author in the paper, with a relatively high number of past publications (more than 15)
Num total authors	Total number of authors in the paper
Num AZ authors	Number of AZ authors in the paper
Num academic authors	Number of academic authors in the paper
Num non-AZ company authors	Number of authors in the paper with a *company* affiliation that is not AstraZeneca
Num PRO authors	Number of authors in the paper with a Public Research Organization affiliation
Num hospital authors	Number of authors in the paper with an affiliation to a hospital
Num countries academics	Num of different countries of affiliation of the academic authors in the paper
Clinical research	Dummy equal to 1 if the paper deals with clinical research (as opposed to basic research)
R&D typology	Dummies equal to 1 if the AZ researchers are based in one of the 7 R&D centres typologies introduced in Table 5.

Table 7. Descriptive statistics

	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
		International collaborations					***National collaborations***			
Dependent variables										
Recombinant novelty (r100)	2544	1.032	0.172	0.388	1.400	2018	1.055	0.159	0.364	1.460
Recombinant novelty (r99)	2544	0.935	0.151	0.387	1.300	2018	0.960	0.141	0.341	1.309
Recombinant novelty (r95)	2544	0.818	0.140	0.352	1.169	2018	0.844	0.132	0.311	1.203
Recombinant novelty (r90)	2544	0.743	0.135	0.324	1.131	2018	0.772	0.129	0.272	1.109
Element novelty (g0)	2806	0.054	0.027	0.008	0.301	2557	0.053	0.028	0.006	0.274
Element novelty (g1)	2806	0.237	0.064	0.098	0.532	2557	0.247	0.067	0.079	0.575
Element novelty (g5)	2806	0.321	0.074	0.143	0.645	2557	0.330	0.079	0.117	0.653
Element novelty (g10)	2806	0.393	0.081	0.183	0.709	2557	0.393	0.088	0.157	0.691
Independent variables										
Academic social tie	2806	0.034	0.181	0	1	2557	0.127	0.333	0	1
Share of new contacts	2806	0.616	0.395	0	1	2557	0.569	0.401	0	1
Institutional agreement	2806	0.202	0.402	0	1	2557	0.330	0.470	0	1
Team characteristics										
Productive AZ researcher	2806	0.837	0.865	0	11	2557	0.768	0.861	0	8
Distance to R&D sites (km log)	2806	6.060	1.539	0	9.164	2557	4.322	1.768	0	8.961
Average past num papers	2806	49.857	96.537	0	798	2557	35.445	60.105	0	988
Num AZ authors	2806	2.025	1.881	1	26	2557	2.087	1.883	1	34
Num academic authors	2806	4.407	3.945	1	42	2557	3.521	2.799	1	29
Num non-AZ company authors	2806	0.522	1.856	0	27	2557	0.381	1.553	0	28
Num PRO authors	2806	1.521	3.263	0	46	2557	0.684	2.345	0	45
Num. hospital authors	2806	2.016	3.436	0	34	2557	0.763	2.382	0	48
Num total authors	2806	11.141	8.061	2	99	2557	7.732	5.656	2	90
Num countries academics	2806	1.648	0.997	1	11	2557	1	0	1	1

Type of R&D center										
Corporate R&D hub	2806	0.095	0.293	0	1	2557	0.236	0.425	0	1
Global academic hub	2806	0.703	0.457	0	1		0.463	0.499	0	1
International Academic center	2806	0.067	0.251	0	1	2557	0.039	0.194	0	1
International R&D center	2806	0.027	0.161	0	1	2557	0.038	0.192	0	1
Local R&D centre	2806	0.043	0.202	0	1	2557	0.120	0.325	0	1
Local R&D outpost	2806	0.007	0.086	0	1	2557	0.015	0.123	0	1
Local search	2806	0.003	0.053	0	1	2557	0.021	0.142	0	1
Other R&D centers	2806	0.001	0.027	0	1	2557	0.002	0.040	0	1
Other AZ facility	2806	0.054	0.226	0	1	2557	0.066	0.248	0	1
Type of medical research										
Clinical research	2806	0.753	0.431	0	1	2557	0.587	0.493	0	1
Oncology	2806	0.170	0.375	0	1	2557	0.160	0.367	0	1
Immunology	2806	0.141	0.348	0	1	2557	0.135	0.342	0	1
Cardio	2806	0.230	0.421	0	1	2557	0.110	0.312	0	1
Infection	2806	0.066	0.248	0	1	2557	0.043	0.204	0	1
Endocrinology	2806	0.117	0.322	0	1	2557	0.089	0.284	0	1
Respiratory	2806	0.130	0.337	0	1	2557	0.087	0.282	0	1
Musculoskeletal	2806	0.068	0.252	0	1	2557	0.051	0.221	0	1
Central Nervous system	2806	0.131	0.337	0	1	2557	0.097	0.296	0	1

Figure 1. Average *Recombinant-type novelty* among international and domestic academic papers

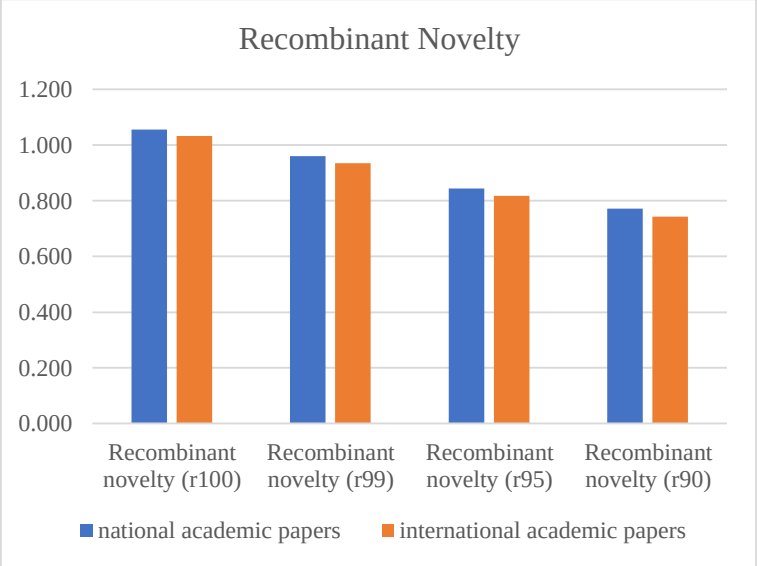


Figure 2. Average *Element-type novelty* among international and domestic academic papers

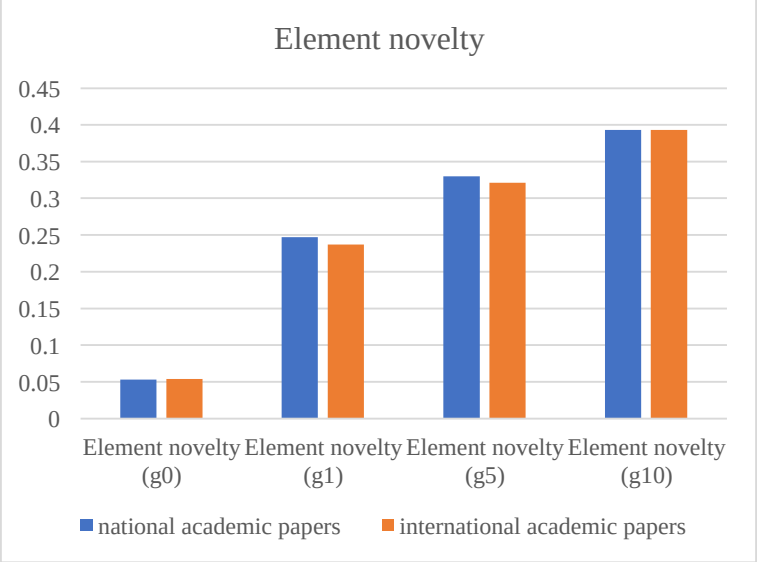


Table 8. Difference in difference estimation. The dependent variable is the *share of new international contacts*.

	(1) t=2013	(2) t=2014	(3) t=2015
Academic paper (0/1)	-0.212***	-0.210***	-0.211***
	(0.016)	(0.015)	(0.014)
From year t	0.013	0.011	0.002
	(0.034)	(0.034)	(0.034)
Academic paper * from year t	0.064***	0.066***	0.077***
	(0.020)	(0.020)	(0.020)
Controls			
Num AZ authors	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)
R&D center typology			
Baseline category: Global academic hub			
Corporate R&D hub	-0.131***	-0.131***	-0.130***
	(0.011)	(0.011)	(0.011)
International academic centre	-0.057***	-0.057***	-0.056***
	(0.018)	(0.018)	(0.018)
International R&D centre	-0.151***	-0.152***	-0.152***
	(0.021)	(0.021)	(0.021)
Local R&D center	-0.184***	-0.185***	-0.185***
	(0.015)	(0.015)	(0.015)
Local R&D outpost	-0.082**	-0.082**	-0.081**
	(0.037)	(0.037)	(0.037)
Local search	-0.182***	-0.182***	-0.182***
	(0.035)	(0.035)	(0.035)
Other R&D centre	-0.127	-0.127	-0.122
	(0.168)	(0.168)	(0.167)
Outside AZ R&D main centers	-0.109***	-0.109***	-0.109***
	(0.019)	(0.019)	(0.019)
Time dummies	YES	YES	YES
Constant	0.455***	0.453***	0.454***
	(0.029)	(0.029)	(0.029)
Observations	8939	8939	8939
R^2	0.0757	0.0758	0.0762

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3. Parallel trends assumption before 2013

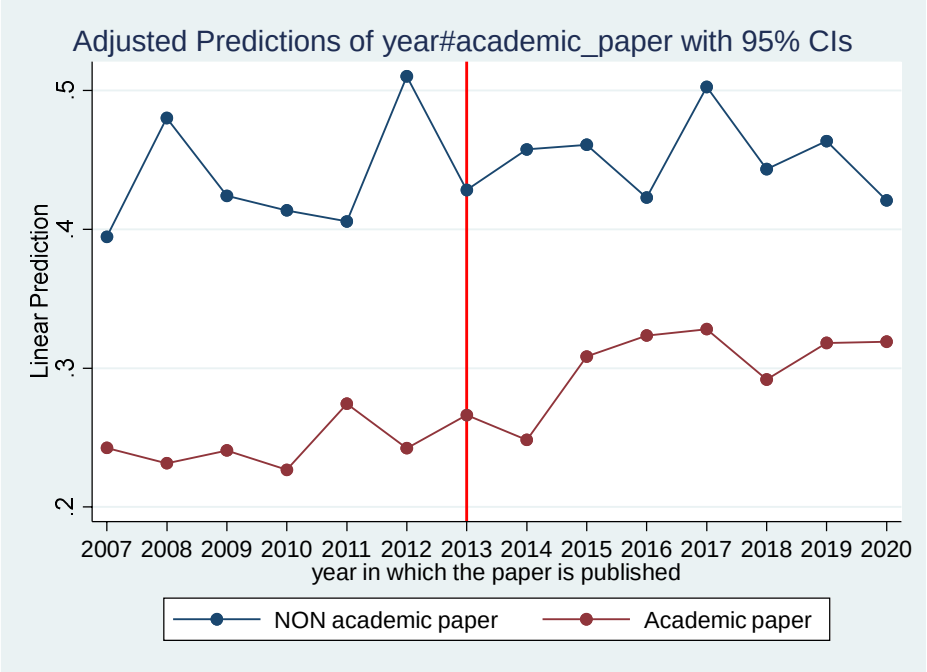


Table 9. The role of social ties on the novelty of the papers (Recombinant-type novelty).

	(1) Internat. **r100**	(2) National **r100**	(3) Internat. **r99**	(4) National **r99**	(5) Internat. **r95**	(6) National **r95**	(7) Internat. **r90**	(8) National **r90**
Academic social tie	0.040***	0.011	0.038***	0.008	0.031***	0.008	0.027**	0.011
	(0.014)	(0.010)	(0.013)	(0.009)	(0.012)	(0.008)	(0.012)	(0.008)
Share of new contacts	0.052***	-0.000	0.040***	-0.000	0.038***	-0.001	0.039***	0.000
	(0.009)	(0.009)	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)
Institutional agreement	0.004	-0.014*	0.005	-0.015**	0.001	-0.014**	0.001	-0.013**
	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Team level controls								
Productive AZ researcher	0.008*	0.007	0.005	0.007*	0.003	0.008**	0.003	0.007*
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Distance to R&D sites	-0.008***	-0.007***	-0.007***	-0.006***	-0.007***	-0.006***	-0.006***	-0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Average past num papers	0.000**	-0.000*	0.000*	-0.000	0.000*	-0.000	0.000**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Num AZ authors	0.001	0.008**	0.003	0.008**	0.004*	0.008***	0.004*	0.008***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Num non-AZ company authors	0.000	0.005	0.002	0.005	0.002	0.005*	0.003	0.005*
	(0.003)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Num PRO authors	0.004*	0.005	0.004**	0.004	0.004**	0.004	0.004**	0.004
	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Num hospital authors	-0.007***	-0.001	-0.005**	0.000	-0.005**	0.001	-0.004**	0.001
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Num academic authors	0.005**	0.009***	0.004**	0.008***	0.004**	0.008***	0.004**	0.007***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Num total authors	0.002	-0.002	0.001	-0.003	0.001	-0.003	0.001	-0.003
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Num countries academics	-0.016***	-	-0.013***	-	-0.012***	-	-0.012***	-
	(0.004)		(0.003)		(0.003)		(0.003)	
Company areas of expertise								
Clinical research	-0.073***	-0.035***	-0.060***	-0.032***	-0.054***	-0.030***	-0.056***	-0.032***
	(0.007)	(0.008)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Clinical research	-0.073***	-0.035***	-0.060***	-0.032***	-0.054***	-0.030***	-0.056***	-0.032***
	(0.007)	(0.008)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Oncology	-0.017*	-0.008	-0.013*	-0.008	-0.021***	-0.009	-0.026***	-0.009
	(0.009)	(0.010)	(0.008)	(0.009)	(0.007)	(0.008)	(0.007)	(0.008)
Immunology	0.013	0.012	0.008	0.002	0.001	-0.001	-0.001	-0.004
	(0.009)	(0.010)	(0.008)	(0.009)	(0.008)	(0.008)	(0.007)	(0.008)
Cardiovascular	-0.031***	-0.042***	-0.021***	-0.035***	-0.022***	-0.040***	-0.026***	-0.042***
	(0.008)	(0.011)	(0.007)	(0.010)	(0.006)	(0.009)	(0.006)	(0.009)

Infection	0.011	0.000	0.008	-0.000	0.006	0.004	0.007	0.004
	(0.013)	(0.017)	(0.011)	(0.015)	(0.010)	(0.014)	(0.010)	(0.014)
Endocrinology	-0.030***	-0.037***	-0.030***	-0.039***	-0.035***	-0.046***	-0.038***	-0.048***
	(0.011)	(0.012)	(0.010)	(0.010)	(0.009)	(0.009)	(0.008)	(0.009)
Respiratory	-0.036***	-0.030**	-0.027***	-0.016	-0.022**	-0.017	-0.025***	-0.018*
	(0.011)	(0.012)	(0.009)	(0.011)	(0.009)	(0.010)	(0.008)	(0.010)
Muscoloskeletal	0.041***	0.010	0.041***	0.010	0.042***	0.007	0.044***	0.005
	(0.012)	(0.012)	(0.011)	(0.011)	(0.010)	(0.011)	(0.010)	(0.010)
Central Nervous System	-0.028***	-0.005	-0.035***	-0.009	-0.038***	-0.007	-0.035***	-0.005
	-0.017*	-0.008	-0.013*	-0.008	-0.021***	-0.009	-0.026***	-0.009
R&D center typology								
Baseline: Global academic hub								
Corporate R&D hub	0.017	0.004	0.016*	0.006	0.016*	-0.000	0.016*	-0.003
	(0.011)	(0.009)	(0.010)	(0.008)	(0.009)	(0.007)	(0.009)	(0.007)
International academic center	-0.010	-0.019	-0.010	-0.022	-0.009	-0.026*	-0.010	-0.026*
	(0.013)	(0.020)	(0.012)	(0.017)	(0.011)	(0.015)	(0.010)	(0.015)
International R&D center	0.012	0.007	0.015	-0.007	0.009	-0.013	0.009	-0.016
	(0.016)	(0.017)	(0.013)	(0.015)	(0.012)	(0.014)	(0.012)	(0.013)
Local R&D center	0.026*	-0.037***	0.022*	-0.035***	0.019*	-0.035***	0.019*	-0.038***
	(0.013)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)
Local R&D outpost	-0.022	-0.016	-0.024	-0.014	-0.029	-0.039*	-0.019	-0.054***
	(0.027)	(0.024)	(0.023)	(0.022)	(0.020)	(0.020)	(0.019)	(0.020)
Local search	0.051	-0.112***	0.046	-0.100***	0.057	-0.108***	0.062	-0.113***
	(0.050)	(0.025)	(0.042)	(0.023)	(0.042)	(0.021)	(0.039)	(0.019)
Other R&D centers	0.012	0.072*	0.003	0.068	-0.038	0.061	-0.047	0.046
	(0.070)	(0.039)	(0.044)	(0.045)	(0.075)	(0.042)	(0.085)	(0.028)
Outside of AZ R&D centers	0.004	-0.016	0.005	-0.024*	0.004	-0.022*	0.001	-0.020*
	(0.015)	(0.014)	(0.013)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
TIME DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	1.139***	1.090***	1.054***	1.008***	0.941***	0.904***	0.861***	0.834***
	(0.031)	(0.026)	(0.026)	(0.025)	(0.024)	(0.023)	(0.024)	(0.023)
Num of obs	2544	2018	2544	2018	2544	2018	2544	2018
R^2	0.182	0.123	0.164	0.118	0.173	0.131	0.192	0.144

Notes: OLS Regression. Social tie must have occurred at least 3 years before the year of the paper. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10. The role of social ties on the novelty of the papers (Element-type novelty).

	(1) Internat. g0	(2) National g0	(3) Internat. g1	(4) National g1	(5) Internat. g5	(6) National g5	(7) Internat. g10	(8) National g10
Academic social tie	0.002	0.001	0.019**	-0.001	0.022**	-0.004	0.021**	-0.005
	(0.003)	(0.002)	(0.008)	(0.004)	(0.009)	(0.005)	(0.010)	(0.005)
Share of new contacts	-0.001	0.001	0.003	0.001	0.008**	0.007*	0.013***	0.015***
	(0.001)	(0.001)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Institutional agreement	-0.000	0.002	-0.001	0.000	-0.001	0.005	0.000	0.009**
	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Team level controls								
Productive AZ researcher	0.001	0.001**	0.005***	0.009***	0.005***	0.009***	0.005***	0.009***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Distance to R&D sites	-0.000	0.000	-0.002*	0.001	-0.001	0.001	-0.001	0.002
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Average past num papers	0.000*	-0.000*	0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Num AZ authors	0.000	-0.000	0.002**	0.003**	0.001	0.004**	0.001	0.004**
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Num non-AZ company authors	0.000	-0.000	0.001	0.003*	0.000	0.003	-0.001	0.002
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Num PRO authors	0.000	-0.000	0.002**	0.002	0.002**	0.003**	0.002**	0.004**
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Num hospital authors	-0.000	0.000	-0.000	0.002	-0.000	0.002*	-0.000	0.003*
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Num academic authors	0.001*	-0.000	0.004***	0.003***	0.004***	0.005***	0.005***	0.005***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Num total authors	-0.000	0.000	-0.001*	-0.001	-0.001	-0.002	-0.001	-0.002
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Num countries academics	-0.002***	0.000	-0.010***	0.000	-0.012***	0.000	-0.012***	0.000
	(0.001)	(.)	(0.001)	(.)	(0.001)	(.)	(0.001)	(.)
Company areas of expertise								
Clinical research	-0.001	-0.004***	-0.030***	-0.023***	-0.032***	-0.027***	-0.027***	-0.023***
	(0.001)	(0.001)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Oncology	-0.000	-0.002	0.006*	0.001	0.020***	0.015***	0.032***	0.029***
	(0.001)	(0.002)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Immunology	-0.001	-0.000	0.001	-0.000	0.004	0.002	0.008*	0.006
	(0.001)	(0.002)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Cardiovascular	-0.001	0.001	-0.004	-0.015***	0.004	-0.009*	0.016***	0.004
	(0.001)	(0.001)	(0.003)	(0.004)	(0.003)	(0.005)	(0.003)	(0.005)

Infection	0.001	0.007***	0.023***	0.031***	0.026***	0.036***	0.027***	0.039***
	(0.002)	(0.003)	(0.005)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
Endocrinology	-0.002	-0.002	-0.005	-0.015***	0.001	-0.010**	0.009**	-0.003
	(0.001)	(0.002)	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.006)
Respiratory	0.001	0.003	-0.003	-0.001	-0.003	-0.005	0.000	-0.006
	(0.001)	(0.002)	(0.003)	(0.005)	(0.004)	(0.005)	(0.004)	(0.006)
Muscoloskeletal	0.000	0.007***	0.008*	0.007	0.016***	0.008	0.027***	0.016**
	(0.002)	(0.002)	(0.004)	(0.005)	(0.005)	(0.006)	(0.005)	(0.007)
Central Nervous System	-0.003**	-0.004***	-0.000	-0.000	0.004	-0.000	0.009**	0.003
	(0.001)	(0.001)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
R&D center typology *Baseline: Global academic hub*								
Corporate R&D hub	-0.002	0.000	0.012***	0.005	0.015***	0.011***	0.015***	0.014***
	(0.002)	(0.001)	(0.004)	(0.003)	(0.005)	(0.004)	(0.005)	(0.005)
International academic center	-0.000	0.001	-0.006	-0.005	-0.010**	-0.008	-0.014***	-0.009
	(0.002)	(0.003)	(0.004)	(0.007)	(0.005)	(0.008)	(0.005)	(0.008)
International R&D center	-0.004*	-0.003	0.019**	-0.003	0.028***	0.003	0.035***	0.010
	(0.002)	(0.002)	(0.008)	(0.006)	(0.008)	(0.007)	(0.009)	(0.008)
Local R&D center	0.004*	0.000	0.007	-0.010**	0.008	-0.010*	0.006	-0.005
	(0.002)	(0.002)	(0.006)	(0.005)	(0.006)	(0.005)	(0.007)	(0.006)
Local R&D outpost	-0.006*	-0.001	0.037***	-0.014	0.044***	-0.004	0.050***	0.010
	(0.004)	(0.004)	(0.011)	(0.011)	(0.013)	(0.014)	(0.016)	(0.015)
Local search	0.010	-0.003	-0.014	-0.003	-0.013	0.002	-0.006	0.006
	(0.008)	(0.003)	(0.013)	(0.009)	(0.015)	(0.010)	(0.018)	(0.011)
Other R&D centers	0.010*	0.001	0.077***	0.015	0.086***	0.009	0.072***	0.008
	(0.005)	(0.009)	(0.005)	(0.030)	(0.006)	(0.034)	(0.008)	(0.029)
Outside of AZ R&D centers	0.003	-0.000	0.007	-0.009*	0.012*	-0.010*	0.017**	-0.007
	(0.002)	(0.002)	(0.006)	(0.005)	(0.006)	(0.006)	(0.007)	(0.006)
TIME DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	0.057***	0.057***	0.266***	0.267***	0.342***	0.341***	0.393***	0.389***
	(0.004)	(0.004)	(0.012)	(0.009)	(0.015)	(0.011)	(0.016)	(0.013)
N	2806	2557	2806	2557	2806	2557	2806	2557
R^2	0.177	0.172	0.160	0.121	0.166	0.116	0.163	0.107

Notes: OLS Regression. Social tie must have occurred at least 3 years before the year of the paper. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11. Distinguishing between typology of R&D centers (Recombinant novelty)

	(1) r100	(2) r99	(3) r95	(4) r90	(5) r100	(6) r99	(7) r95	(8) r90
	Only international academic papers				Only international academic papers			
	Global Academic Hubs				**Other R&D centers**			
Academic social tie	0.035**	0.039**	0.033**	0.030**	0.055***	0.040**	0.029*	0.020
	(0.017)	(0.016)	(0.015)	(0.015)	(0.021)	(0.019)	(0.017)	(0.017)
Paper-level controls	YES	YES	YES	YES	YES	YES	YES	YES
R&D center typology	YES	YES	YES	YES	YES	YES	YES	YES
Company areas of exp. dummies	YES	YES	YES	YES	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES	YES	YES	YES	YES
Constant	1.176***	1.082***	0.966***	0.885***	1.087***	1.021***	0.928***	0.854***
	(0.036)	(0.029)	(0.027)	(0.026)	(0.080)	(0.068)	(0.065)	(0.063)
N	1784	1784	1784	1784	760	760	760	760
R^2	0.192	0.175	0.183	0.202	0.198	0.178	0.188	0.210

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12. Distinguishing between typology of R&D centers (Element novelty).

	(1) g0	(2) g1	(3) g5	(4) g10	(5) g0	(6) g1	(7) g5	(8) g10
	Only international academic papers				*Only international academic papers*			
	Global Academic Hubs				**Other R&D centers**			
Academic social tie	0.003	0.020**	0.021**	0.019*	0.003	0.013	0.024	0.031
	(0.004)	(0.008)	(0.010)	(0.011)	(0.005)	(0.016)	(0.022)	(0.025)
Paper-level controls	YES	YES	YES	YES	YES	YES	YES	YES
R&D center typology	YES	YES	YES	YES	YES	YES	YES	YES
Company areas of exp. dummies	YES	YES	YES	YES	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES	YES	YES	YES	YES
Constant	0.054***	0.268***	0.346***	0.398***	0.073***	0.250***	0.323***	0.380***
	(0.005)	(0.014)	(0.017)	(0.019)	(0.013)	(0.027)	(0.030)	(0.032)
N	1973	1973	1973	1973	833	833	833	833
R^2	0.171	0.142	0.141	0.136	0.227	0.244	0.260	0.258

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13. Interaction Social ties and distance to AZ R&D centers (Recombinant-novelty).

	(1) r100 International	(2) r99 International	(3) r95 International	(4) r90 International
Academic social tie	-0.059	-0.053	-0.052	-0.034
	(0.045)	(0.037)	(0.036)	(0.034)
Distance to R&D sites	-0.009***	-0.008***	-0.008***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Academic social tie * Distance to R&D centers	0.016**	0.015**	0.013**	0.010*
	(0.007)	(0.006)	(0.006)	(0.005)
Other controls	YES	YES	YES	YES
R&D center typology	YES	YES	YES	YES
Company areas of exp. dummies	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES
Constant	1.142***	1.057***	0.944***	0.863***
	(0.032)	(0.026)	(0.024)	(0.024)
N	2544	2544	2544	2544
R^2	0.183	0.165	0.174	0.192

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14. Interaction Social ties and distance to AZ R&D centers (Element-novelty).

	(1) g0 International	(2) g1 International	(3) g5 International	(4) g10 International
Academic social tie	0.004	0.074***	0.081***	0.080***
	(0.010)	(0.022)	(0.028)	(0.031)
Distance to R&D sites	-0.000	-0.001	-0.001	-0.001
	(0.000)	(0.001)	(0.001)	(0.001)
Academic social tie * Distance to R&D sites	-0.000	-0.009***	-0.010**	-0.010**
	(0.001)	(0.004)	(0.004)	(0.005)
Other controls	YES	YES	YES	YES
R&D center typology	YES	YES	YES	YES
Company areas of exp. dummies	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES
Constant	0.057***	0.263***	0.339***	0.390***
	(0.004)	(0.012)	(0.015)	(0.016)
N	2806	2806	2806	2806
R^2	0.177	0.162	0.168	0.165

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 4. Marginal effects: interaction between Academic social tie and Distance to R&D centers (Recombinant novelty)

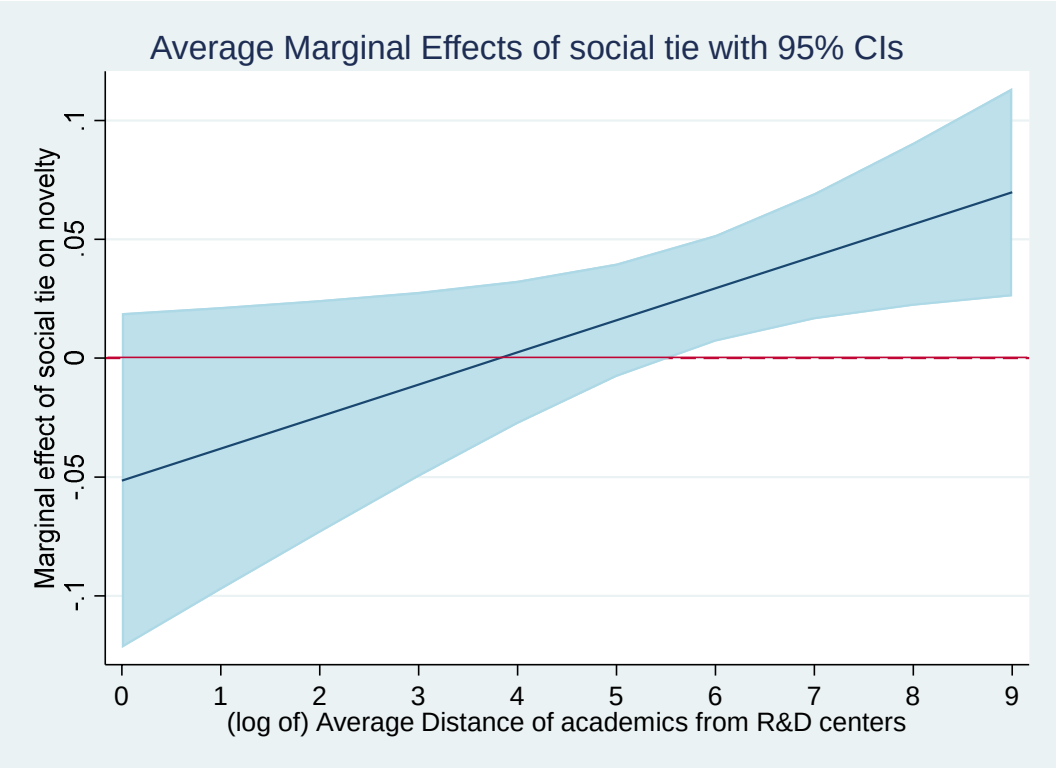
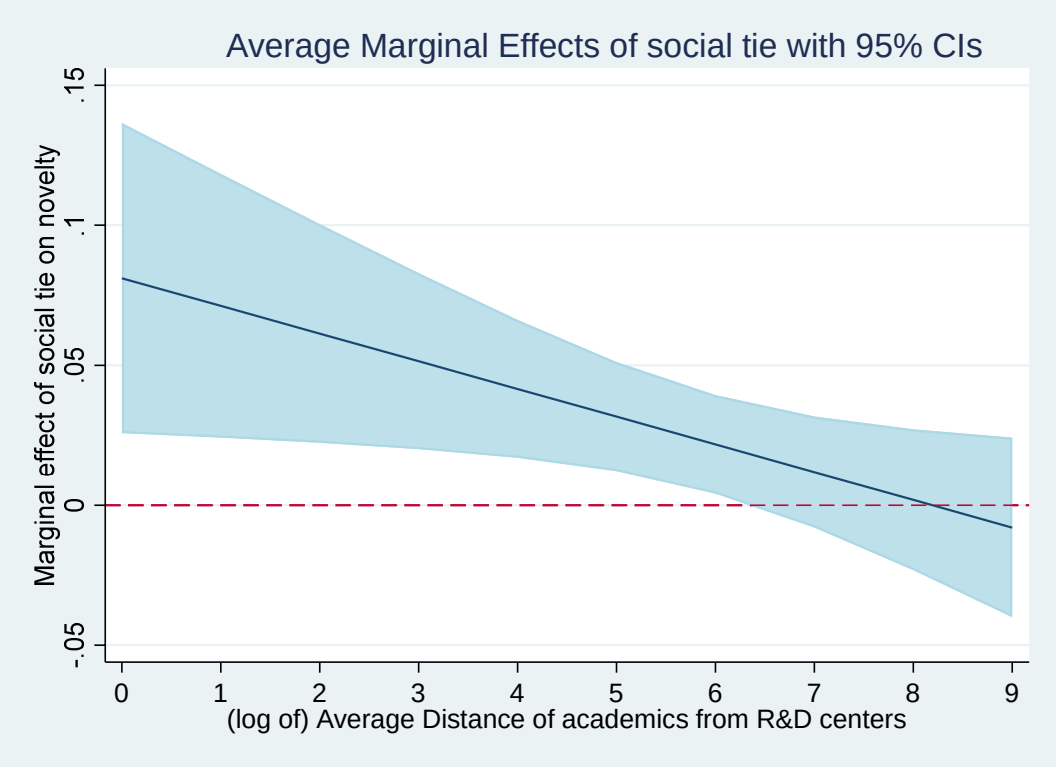


Figure 5. Marginal effects: interaction between Academic social tie and Distance to R&D centers (Element novelty)



APPENDICES

Appendix A: Dataset construction

Publication data

We downloaded data from Scopus about all the papers that were published between 2000 and 2020, in which at least one of the authors had an Astra-Zeneca affiliation. We used the search terms like Astra*, AstraZeneca, Zeneca*. We manually checked the articles to make sure that they were AstraZeneca publications since there also exist other companies with similar names such as Astra Rocket and Astra International. We also went through AstraZeneca's acquisition history to make sure we included publications that were written under the name of acquired companies but only after the companies were formally acquired by AstraZeneca. These include for example KuDOs (2005), MedImmune Biologics incl. Cambridge Antibody Technology (2006) and MedImmune (2007), Arrow Therapeutics (2007), Amylin Pharmaceuticals (2012), Spirogen (2013), Pearl Therapeutics (2013), Omthera Pharmaceuticals (2013), ZS Pharma (2015), and some additional small companies.

Through this procedure we were able to identify 17,522 papers. We did not include papers with more than 100 authors (only a few papers had so many authors), because we considered that authorship in these cases did not imply any real collaboration with all the other coauthors of the paper. We recovered information about all the 59,021 authors included in such papers. For the 12,789 authors with an AstraZeneca affiliation, using their Scopus ID, we downloaded information in Scopus about all the papers that they published throughout their career (hence also before 2000), regardless of their affiliation. In such a way we could retrieve information also about the publishing activity of AstraZeneca authors before the chosen period of 2000-2020, to track their professional history. We can see their history of affiliations and the location of the affiliation. We can also know if the coauthors with whom they publish in the period 2000-2020 have coauthored with them in the past or not.

Recovering the publication history of AstraZeneca authors

We first downloaded data from Scopus about all the papers that were published between 2000 and 2020 (henceforth *focal papers*), in which at least one author had an Astra-Zeneca affiliation. We define this set of authors the "*focal authors*". Among these "focal authors", which include also authors that have never worked for AZ, we are interested in AZ authors, i.e. those authors that at least at some point in their career (between 2000 and 2020) published a paper with an AZ affiliation. This strategy allows us to identify all the papers of Astra Zeneca in the period 2000-2020, all the authors that in that period were affiliated with AZ, and their co-authors.

Using the Scopus author id we downloaded from Scopus all the papers for each of the focal authors that had an AZ affiliation between 2000 and 2020. For each AZ focal author we downloaded all the papers available in Scopus, regardless of the year in which it was published. This means for example that we may have cases in which a focal author with an AZ affiliation in the period 2000-2020 may also have papers without an AZ affiliation, before (or after) that

period.[13] We retrieved all the information related to the articles of the AZ focal authors. This included affiliation name and affiliation id, city and country of the affiliation for each of the co-authors of the paper, the id of the paper, typology of paper (article, note, book, etc.) and other variables. The most important variables are a) the affiliation name, which allows us to identify the institution of affiliation of the authors of the paper and whether it is a university, company, hospital, public research organization, etc., and b) the location of the institution, in order to identify collaboration within the same country or international ones. The identification of the institution was done using different search words like univ., university, faculty, hospital, hospital, sjukhus, clinic, ltd., inc., ab.. For institutions where none of these search words were found we manually assigned the type by searching for the institution online or asking around in our academic and professional network.

In such a way we could retrieve information also about the publishing activity of focal authors before (and after) the chosen period of 2000-2020, to track their educational and professional history. Most importantly we were able to see their history of affiliations and the location of the affiliation. We can also know if the coauthors with whom they publish in the period 2000-2020 have coauthored with them in the past.

Authors with double affiliations: AstraZeneca and academic

In many cases the authors listed among the papers that we identified as AZ papers had more than one affiliation. In these cases, Scopus data could not handle the complexity of the data. Many authors have an academic affiliation and AZ affiliation. In these cases, Scopus usually attributes only the first affiliation, be it a company or an academic institution. Hence for many of the papers that we had originally selected because at least one of the authors had an AZ affiliation, when we downloaded data about authors' affiliations we could not find any author with an AZ affiliation, because AZ was not the first affiliation listed in the paper and hence Scopus dropped the AZ affiliation for that specific author. For example, an author that has two affiliations in a paper -Department of Biology of Cambridge University and the R&D center of AZ in Cambridge- would only have one affiliation id attributed by Scopus and that is the one of Cambridge University. This means that, if this author is the only author in that paper with an AZ affiliation, among the whole list of co-authors in the paper we would not see any AZ affiliation. In all these cases then we had to implement a cleaning procedure to make sure that we do not miss AZ affiliations. Scopus also provides a string that includes all the affiliations listed in the paper for each author. This data is not processed by Scopus, i.e. there is not a simple way to derive the affiliation id of the institutions listed. For all the papers that we selected and for which we could not find any author with an AZ affiliation we used the original string provided by Scopus with the affiliations' name, isolate the AstraZeneca affiliation and created a new variable that listed AZ as a second affiliation. Besides adding the affiliation name of the institution, we also had to identify the city and country of the second affiliation. This was done manually by checking the city and country of the second affiliation listed in the original string

[13] We also have information about additional co-authors that are not focal authors, i.e. they did not publish a paper in 2000-2020 where there was at least an author with an AZ affiliation (see the cases of author M, N and O in the Figure). For those authors we do not have complete information about their publication history.

of affiliations. In total we found that AstraZeneca was listed as the second affiliation for 2,076 authors in the period 2000-2020. For each of these authors this second affiliation issue often occurred in more than one paper so in the end we recovered AstraZeneca as the second affiliation in 11,098 cases.

Cleaning the location of AZ affiliations

Among the authors with AZ affiliation, Scopus attributes an affiliation id, the city and the country of the affiliation. After a first check we realized however that in more than 92% of the cases all the authors with an AZ affiliation had either an affiliation in the AZ subsidiary of Cambridge or in the AZ subsidiary of Molndal, close to Goteborg in Sweden. Therefore, we implemented a series of routine checks to see if in the original string of the affiliation names that we recovered from Scopus the name of the AZ location listed was different from the one that Scopus routinely attributed. By looking at the original string containing information about the affiliation names we were able to implement in Stata some changes to attribute the correct affiliation to the AZ authors. Our strategy was to check if the terms "Cambridge" and "Molndal" were included in the original string of the affiliation. When this was not the case, we codified the real location of the AZ subsidiary and attributed this new location to that specific AZ author. We did this for 9,134 AZ authors. After our cleaning procedure the share of AZ authors with an affiliation in either Cambridge or Molndal dropped from 92% to 39%, as shown in Tables A1 and A2.

Table A1. Number of times we observe an authorship from an *AstraZeneca author* with an affiliation in a specific city. Most common cities BEFORE correcting for the AZ affiliations.

City	number	%	Cum. %
Cambridge	34,216	63.48	63.48
Sodertalje	15,746	29.21	92.69
Stockholm	565	1.05	93.74
Gothenburg	393	0.73	94.46
Uppsala	229	0.42	94.89
London	182	0.34	95.23
New York	159	0.29	95.52
Gaithersburg (US)	144	0.27	95.79
Lund	87	0.16	95.95
Boston	83	0.15	96.10
Manchester	82	0.15	96.26
Montreal	74	0.14	96.39
Madrid	51	0.09	96.49
Brisbane	48	0.09	96.58
Aurora	39	0.07	96.65
Bethesda (Maryland, US)	37	0.07	96.72
Brentford (UK)	36	0.07	96.79
Philadelphia	35	0.06	96.85
San Diego	34	0.06	96.91
Barcelona	32	0.06	96.97

Table A2. Number of times we observe an authorship from an *AstraZeneca author* with an affiliation in a specific city. Most common cities AFTER correcting for the AZ affiliations.

City	number	%	Cum. %
Cambridge	10,867	20.32	20.32
Molndal	10,219	19.11	39.42
Macclesfield	9,868	18.45	57.88
Sodertalje	5,758	10.77	68.64
Wilmington	2,901	5.42	74.07
Boston	2,699	5.05	79.11
Loughborough	2,186	4.09	83.20
Gaithersburg	1,683	3.15	86.35
Waltham	1,534	2.87	89.21
Lund	1,297	2.43	91.64
Bangalore	1,089	2.04	93.67
Montreal	613	1.15	94.82
Shanghai	555	1.04	95.86
Reims	321	0.60	96.46
Brixham	292	0.55	97.00
Gothenburg	208	0.39	97.39
Osaka	205	0.38	97.78
Madrid	192	0.36	98.14
Stockholm	165	0.31	98.44
Barcelona	106	0.20	98.64

Table A3. Number of academic (and non-academic) papers per year

year	Academic papers	Non academic papers	share academic papers	total num of papers
2000	198	353	35.93	551
2001	189	265	41.63	454
2002	186	289	39.16	475
2003	278	322	46.33	600
2004	316	356	47.02	672
2005	372	428	46.50	800
2006	376	418	47.36	794
2007	347	422	45.12	769
2008	333	389	46.12	722
2009	353	424	45.43	777
2010	352	396	47.06	748
2011	364	441	45.22	805
2012	420	553	43.17	973
2013	388	466	45.43	854
2014	457	511	47.21	968
2015	484	476	50.42	960
2016	570	485	54.03	1,055

2017	547	487	52.90	1,034
2018	615	496	55.36	1,111
2019	662	497	57.12	1,159
2020	761	480	61.32	1,241
total	8,568	8,954	48.90	17,522

Table A4. International and non-international academic papers by year.

year	international papers	non international papers	share of international	Total num of academic papers
2000	79	119	40%	198
2001	99	90	52%	189
2002	91	95	49%	186
2003	143	135	51%	278
2004	172	143	54%	316
2005	203	169	55%	372
2006	215	161	57%	376
2007	200	147	58%	347
2008	182	150	55%	333
2009	205	148	58%	353
2010	204	148	58%	352
2011	241	123	66%	364
2012	269	151	64%	420
2013	241	147	62%	388
2014	282	175	62%	457
2015	334	150	69%	484
2016	407	163	71%	570
2017	400	147	73%	547
2018	446	169	73%	615
2019	484	178	73%	662
2020	569	191	75%	761
Total	5,466	2984	64%	8,568

Table A5. Number of times we observe an authorship from an ***academic*** author with an affiliation in a specific country (the affiliation of an author is counted multiple times, according to the number of papers included in the sample in which she/he is listed as an author).

Country	num	%	Cum %
United Kingdom	11,353	36.66	36.66
United States	5,024	16.22	52.88
Sweden	4,512	14.57	67.45
Germany	1,315	4.25	71.69
Canada	1,161	3.75	75.44
China	792	2.56	78.00
Japan	633	2.04	80.04

Australia	632	2.04	82.08
Italy	573	1.85	83.93
Denmark	548	1.77	85.70
France	484	1.56	87.27
Netherlands	403	1.30	88.57
Spain	383	1.24	89.80
Finland	345	1.11	90.92
South Korea	320	1.03	91.95
Belgium	286	0.92	92.87
Switzerland	273	0.88	93.76
Norway	257	0.83	94.59
Brazil	186	0.60	95.19
Austria	156	0.50	95.69
Greece	144	0.46	96.15
Other countries	1191	3.84	100.00
Total	30,971	100.00	

Appendix B: Institutional partnerships

Based on data from PharmaDeals (IMS Health now IQVIA) and the analysis of AstraZeneca annual reports, we were able to identify the formal partnerships that AstraZeneca has signed with academic institutions over the period 2000 to 2020. Indeed, AstraZeneca often establishes collaborations lasting for several years with some universities. These collaborations or partnerships usually have some broad targets focused on specific areas of research, they may involve staff exchanges between AZ and the university personnel, it may involve research funding, sharing of infrastructure, access to samples and data, and sharing costs of clinical trials. Table 3 presents these deals. We only have the year when agreements were signed but no information regarding when the partnership ended, therefore this variable remains a 1 throughout the study period even though in reality they may have ended.

Table B1. The list of formal partnerships with academic institutions.

Year	University partner	City	Country	Therapeutic area
2000	University of California Irvine	Irvine	US	Gastro
2000	Karolinska Institutet	Stockholm	SE	All
2000	McMaster University	Hamilton	CA	Gastro
2000	Washington University	St. Louis	US	Gastro
2000	Conaris/Kiel University	Kiel	DE	Gastro/Tech
2000	The University of Manchester	Manchester	UK	Oncology/Cardiovascular
2000	Shanghai University	Shanghai	CN	CNS
2000	Rockefeller University	New York	US	CNS
2000	University of Southampton	Southampton	UK	Respiratory/Inflammation
2000	Oxford University	Oxford	UK	Respiratory/Inflammation/Tech
2000	Baylor College of Medicine	Houston	US	Tech platforms
2000	Griffith University	Nathan	AU	Tech platforms
2000	University of Pennsylvania	Philadelphia	US	Cardiovascular

2000	University of Liverpool	Liverpool	UK	Cardiovascular
2000	McGill University	Montreal	CA	Pain
2000	University of Gothenburg	Gothenburg	SE	Infection
2001	Jiao Tong University	Shanghai	CN	CNS
2003	University of Dundee	Dundee	UK	
2003	University of Gratz	Gratz	AT	
2006	University of Pennsylvania	Philadelphia	US	Cardiovascular
2006	The University of Manchester	Manchester	UK	Oncology
2006	The University of Manchester	Manchester	UK	Cardiovascular
2007	The University of Manchester	Manchester	UK	Inflammation
2007	Keio University School of Medicine	Tokyo	JP	
2007	University of Texas	Austin	US	CNS
2007	Yale University	New Haven	US	Immunology
2007	Peking University	Peking	SCN	Pharmacology
2008	Jagiellonian University	Krakow	PL	
2008	Washington University	St. Louis	US	CNS
2008	Columbia University	New York	US	Cardiovascular/Diabetes
2008	Newcastle University	Newcastle	UK	
2009	University of Virginia	Charlottesville	US	Cardiovascular
2009	Duke University	Durham	US	CNS
2009	University of Heidelberg	Heidelberg	DE	Pain
2009	Bar-Ilan University	Ramat Gan	IL	
2010	University of Pennsylvania	Philadelphia	US	CNS
2010	University of Liverpool	Liverpool	UK	Cardiovascular
2010	McGill University	Montreal	CA	Pain
2010	University College London	London	UK	Diabetes/Opthamology
2010	Peking University	Peking	CN	Pharmacology
2011	University of Michigan	Ann Arbor	US	Diabetes
2011	The University of Manchester	Manchester	UK	Inflammation
2011	Kuwait University	Kuwait	KW	Cardiovascular
2011	Indiana University	Bloomington	US	
2012	Texas A&M University	Collega Station	US	Immunology
2012	University of Dundee	Dundee	UK	Oncology
2012	Johann Wolfgang Goethe University of Frankfurt, Germany	Frankfurt	DE	
2012	Case Western Reserve University School of Medicine	Cleveland	US	Infection
2012	The University of Manchester	Manchester	UK	Oncology
2012	University of Bristol	Bristol	UK	CNS/Oncology
2012	Fudan University	Shanghai	CN	Cardiovascular
2012	University of California San Francisco	San Francisco	US	Tech
2013	Vanderbilt University	Nashville	US	CNS
2013	University of Maryland	Baltimore	US	Oncology/Respiratory/Immunology
2013	Tufts University	Medfort	US	CNS

2013	Leiden University	Leiden	NL	
2013	University of Cambridge	Cambridge	UK	Oncology/CNS
2013	University of Cambridge	Cambridge	UK	Tech
2013	University of Birmingham	Birmingham	UK	Musculoskeletal
2013	Aston University	Birmingham	UK	
2013	Johns Hopkins University	Baltimore	US	Cardiovascular/CNS/Inflammatory/Oncology/Infection
2013	Harvard University	Boston	US	Tech
2013	Uppsala University	Uppsala	SE	Cardiovascular/CNS/Inflammatory/Oncology
2013	University of Colorado	Boulder	US	Cardiovascular
2014	University of Texas	Austin	US	Immunology
2014	University of Michigan	Ann Arbor	US	Diabetes
2016	University of Leeds	Leeds	UK	Oncology
2016	University of Sheffield	Sheffield	UK	Oncology/Vaccine
2016	Harvard University	Boston	US	Oncology
2017	Washington University	St. Louis	US	Oncology
2018	University of Bern	Bern	CH	Tech
2018	University of Bonn	Bonn	DE	Tech
2019	University of Colorado	Colorado	US	Oncology
2019	University of Pittsburgh	Pittsburgh	US	Cardiovascular
2020	University of Oxford	Oxford	UK	Vaccine

Source: authors' own calculations based on the Annual reports of AstraZeneca and deal data from PharmaDeals (IMS Health).

Appendix C: Computation of the distances to AZ R&D centers

Using the Stata software geocode we also attributed to each affiliation of the authors in our sample the geographic coordinates of their institution. This required to manually check the name and location of their affiliation. Also in this case, we realized that the data provided by Scopus had some limitations, so we double checked manually the location of each affiliation to make sure that the city/country was correct. Once this task was completed, we were able to attribute to each institution a geographic coordinate and then we computed the distance to each of the R&D centers of AZ. In this way for each non-AZ author, we can compute the distance between his institutions and the closest R&D center.

Appendix D. Robustness checks

Sensitivity analysis on the lag structure of the Academic social tie

Recombinant-type novelty

Figure D1. Academic social tie. International papers vs national papers (*degrees of Recombinant-novelty*) at least 2 years before

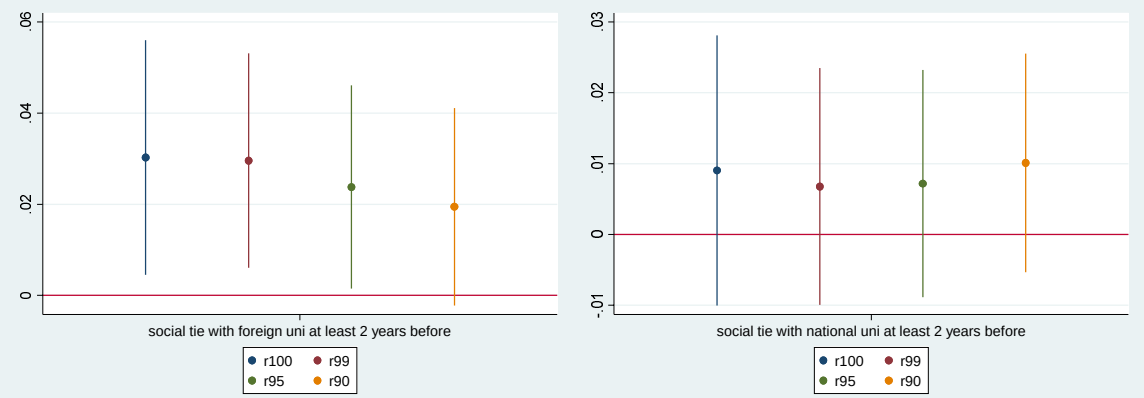


Figure D2. Academic social tie. International papers vs national papers (*degrees of Recombinant -novelty*) at least 3 years before

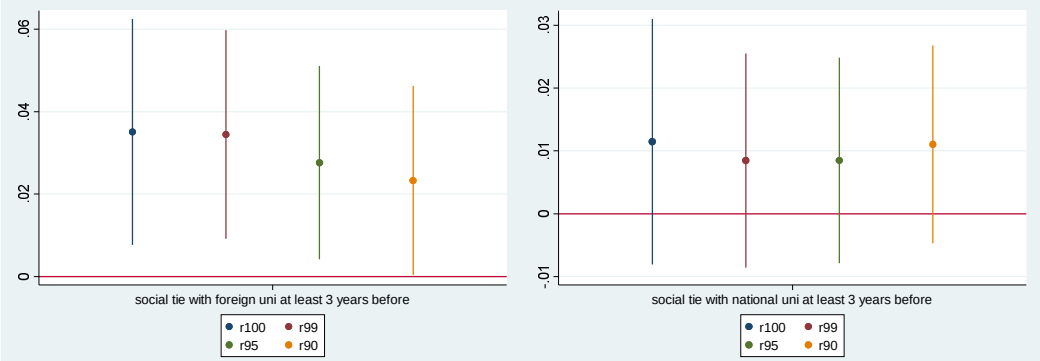


Figure D3. Academic social tie. International papers vs national papers (*degrees of Recombinant-novelty*) at least 4 years before

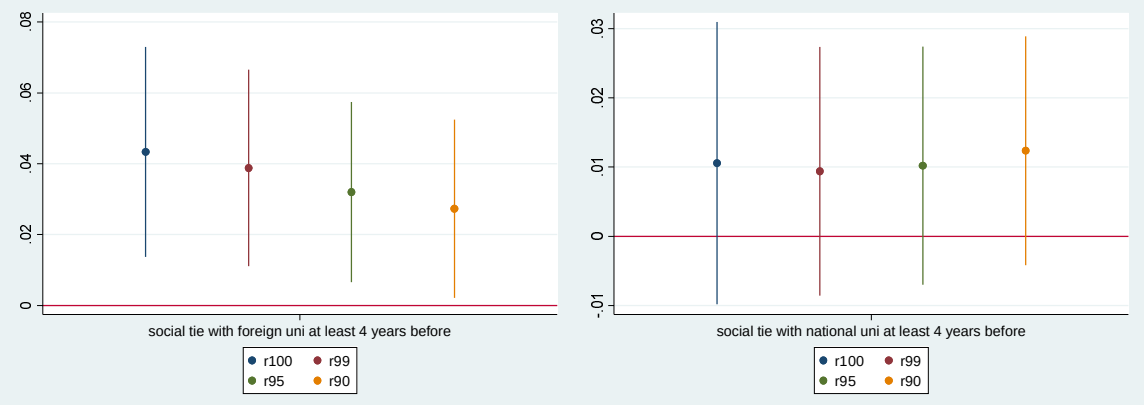
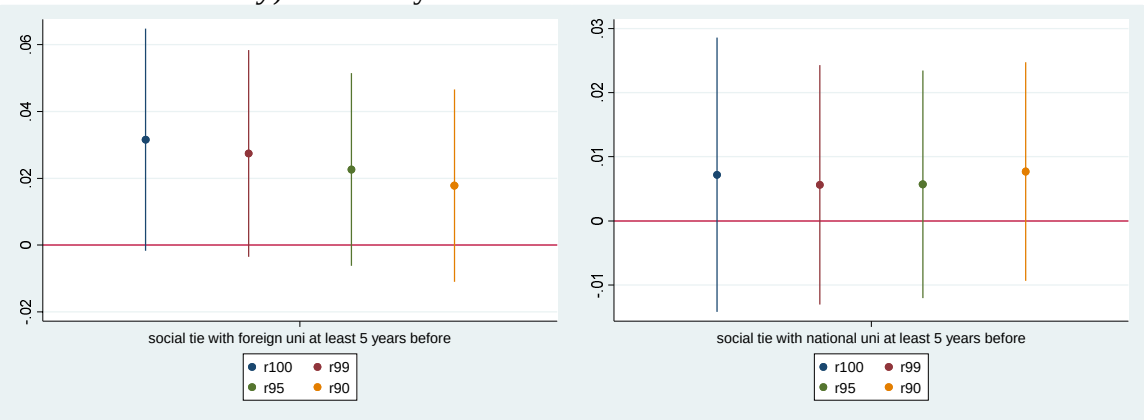


Figure D4. Academic social tie. International papers vs national papers (*degrees of Recombinant-novelty*) at least 5 years before



Element-type novelty

Figure D5. Academic social tie. International papers vs national papers (*degrees of Element-novelty*) at least 2 years before

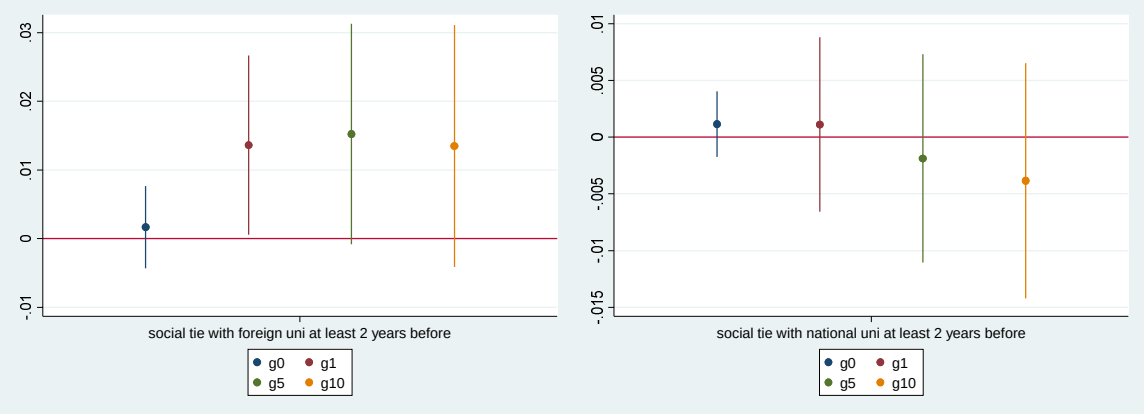


Figure D6. Academic social tie. International papers vs national papers (*degrees of Element-novelty*) at least 3 years before

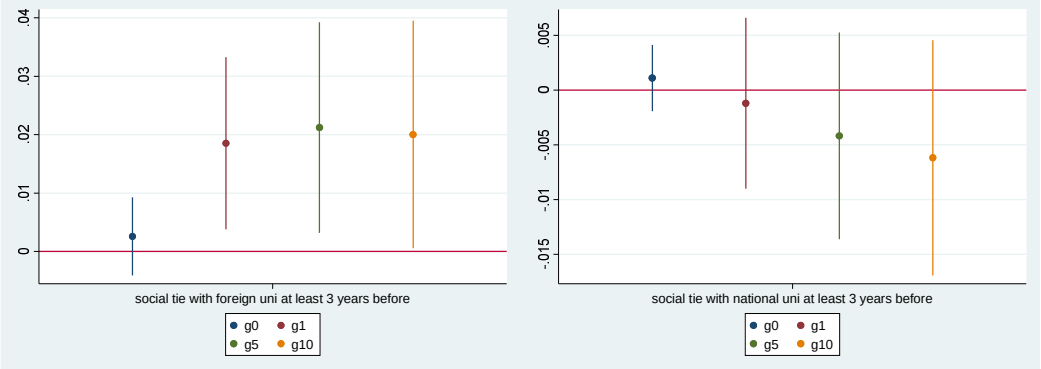


Figure D7. Academic social tie. International papers vs national papers (*degrees of Element-novelty*) at least 4 years before

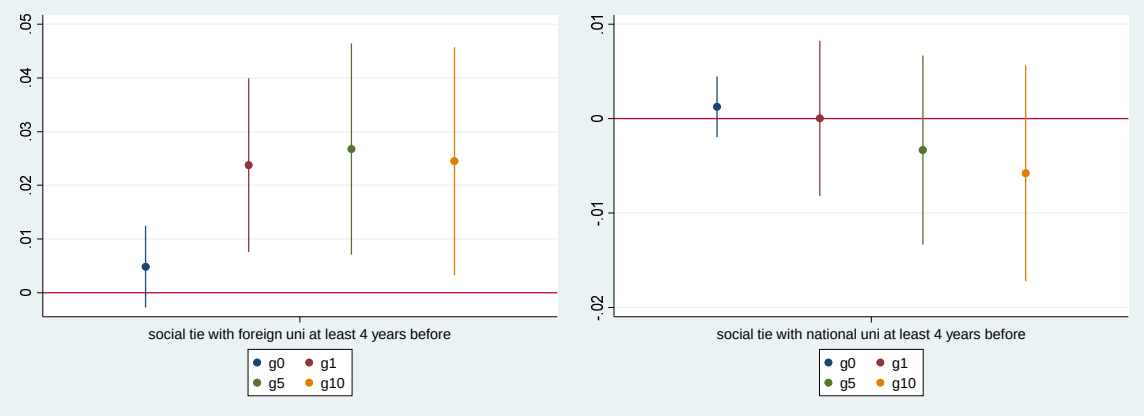


Figure D8. Academic social tie. International papers vs national papers (*degrees of Element-novelty*) at least 4 years before

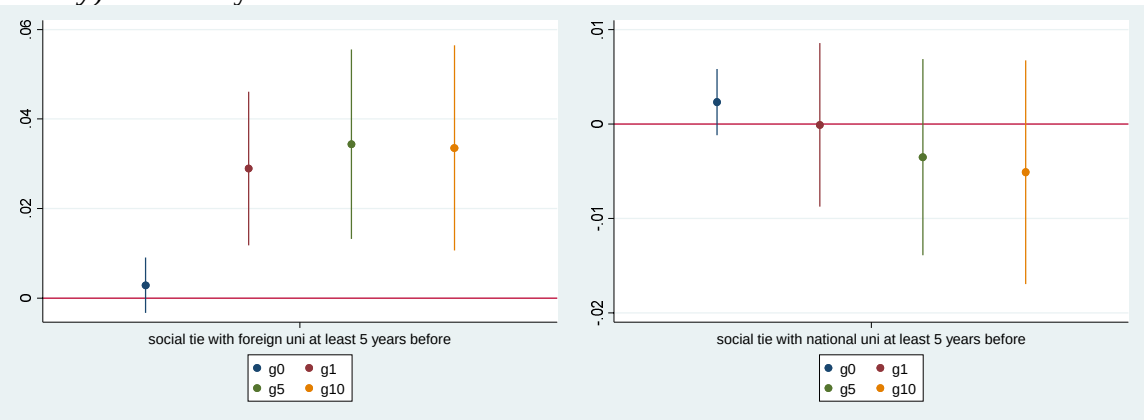


Table D1. Only papers with up to 20 authors in total (Recombinant-novelty).

	(1) r100 International	(2) r99 International	(3) r95 International	(4) r90 International
Academic social tie	0.043***	0.041***	0.035***	0.032***
	(0.014)	(0.013)	(0.012)	(0.012)
Other controls	YES	YES	YES	YES
R&D center typology	YES	YES	YES	YES
Company areas of exp. dummies	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES
Constant	1.132***	1.049***	0.937***	0.856***
	(0.032)	(0.026)	(0.024)	(0.024)
N	2271	2271	2271	2271
R^2	0.190	0.171	0.182	0.202

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D2. Only papers with up to 20 authors in total (Element-novelty).

	(1) g0 International	(2) g1 International	(3) g5 International	(4) g10 International
Academic social tie	0.004	0.021***	0.025***	0.024**
	(0.004)	(0.008)	(0.009)	(0.010)
Other controls	YES	YES	YES	YES
R&D center typology	YES	YES	YES	YES
Company areas of exp. dummies	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES
constant	0.056***	0.259***	0.334***	0.384***
	(0.004)	(0.012)	(0.015)	(0.016)
N	2519	2519	2519	2519
R^2	0.186	0.180	0.185	0.181

Notes: OLS Regression. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$