Working Paper Series

#2024-009

From drains to bridges: The role of internationally mobile PhD students in linking non-mobile with foreign scientists

Rodrigo Ito, Diego Chavarro, Tommaso Ciarli, Robin Cowan and Fabiana Visentin

Published 24 April 2024

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: **info@merit.unu.edu** | website: **http://www.merit.unu.edu**

Boschstraat 24, 6211 AX Maastricht, The Netherlands
Tel: (31) (43) 388 44 00

UNU-MERIT Working Papers
ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology UNU-MERIT

UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at UNU-MERIT to stimulate discussion on the issues raised.

From Drains to Bridges: The role of internationally mobile PhD students in linking non-mobile with foreign scientists[1]

Rodrigo Ito[1,2] Diego Chavarro[3] Tommaso Ciarli[1,4]

Robin Cowan[1,5] Fabiana Visentin[1]

[1]UNU-MERIT, Maastricht University, the Netherlands
[2]DPCT, University of Campinas, Brazil
[3]External consultant
[4]SPRU, Sussex University, England
[5]BETA, University of Strasbourg, France

April 24, 2024

ABSTRACT

Studying and working abroad, internationally mobile scientists meet foreign scientists and become carriers of knowledge and foreign social capital. The benefits of scientific mobility may extend beyond those who experience it, benefiting non-mobile colleagues who collaborate with them. We investigate the role played by Colombian scientists who study abroad for a PhD in connecting non-mobile scientists with foreign scientists. Combining data from online CVs, scholarship programs, and Open Alex publications, we reconstruct the mobility path of 19,158 Colombian scientists and their co-authorship networks from 1990 to 2021. Our results show that coauthoring with mobile scientists increases the propensity of non-mobile scientists to collaborate with foreigners. While the diaspora has been seen as a brain drain, we find that not only returnees but also the diaspora itself can act as bridges connecting local and foreign scientists. However, foreign collaborations tend to be short-lived and sustained only by the mediation of a mobile scientist. Results also suggest that the largest effects stem from mobile scientists who have remained abroad or have a strong circulation pattern between countries (i.e., diaspora and intermittent scientists, respectively). Our paper contributes to the mobility literature by investigating the social capital spillovers generated by mobile scientists. It has also relevant policy implications. Our results call for increasing brain circulation while reducing brain drain by using flexible conditions to return to home countries and increasing the links between mobile and non-mobile scientists.

Keywords: International Scientific Mobility; Co-authorship Networks; Social Capital Spillovers; Colombia.
JEL: O15; O3; D83

[1]The paper has benefited from comments and suggestions from Rafael Anta, Alison Cathles, Gustavo Crespi, Cornelia Lawson, Seung-Hyun Lee, Juan Carlos Navarro, Nathália Pufal, Vanderleia Radaelli, Fernando Vargas, Juan Pablo Ventura, Henry Sauermanm, Emil Bakkensen Johansen, and participants at the IDB seminar series, and the RIE 2022, UNU-MERIT research week 2023, WOEPSR 2023, Eu-SPRI 2023, DRUID 2023, ISSI 2023, and Globelics 2023 conferences. We acknowledge financial support from the Interamerican Development Bank (IDB). Tommaso Ciarli acknowledges support from the PILLARS (pathways to inclusive labour markets) project funded by the European Union's Horizon 2020 research programme under agreement No 101004703. The authors are responsible for the analysis, results, and their interpretation.

1. Introduction

Middle-income countries often attempt to catch up with advanced economies by upgrading production processes (Giuliani et al., 2005). These processes require improving labor productivity, in part by deploying new technologies. Nonetheless, simply adopting new technology from abroad is not enough (Lee et al., 2021); emerging economies must have the absorptive capacity to absorb these technologies into their existing capital stocks (Goode, 1959; Khan, 2022).

Economic theory stresses that capital stock encompasses physical, infrastructure, and human capital. Nonetheless, human capital was largely ignored in the upgrading process for an extended period. Now, it is considered central, and it is commonly recognized that technological upgrading is only possible by upgrading human capital (Radosevic & Yoruk, 2018). Countries must improve their education and research systems. That said, how can countries improve these systems when they start far from the knowledge or technology frontier? One policy often adopted is to take advantage of the education and scientific systems of countries at the frontier by sending students to be trained, typically at advanced education levels, in these countries (Dahlman, 2010). The rationale is that students (and other scientists) exposed to frontier knowledge acquire competencies that can be brought with them upon their return from an international experience and used to promote their home countries' socioeconomic development (Cañibano, 2017; Trippl, 2013).

International experiences directly benefit mobile scientists and can extend to colleagues who stay in their home countries. In addition to carrying knowledge and skills, internationally mobile scientists can contribute to 'internationalizing' their home countries' scientific systems by linking the local scientific community with the global one (Velema, 2012). Empirical research shows that engaging in research abroad leads internationally mobile scientists (hereinafter "mobile scientists") to develop more extensive international co-authorship networks than those scientists who remain in their home countries throughout their careers (hereinafter "non-mobile scientists") (Cao et al., 2020; Gibson & McKenzie, 2014; Jonkers & Cruz-Castro, 2013). These studies propose that mobile scientists' greater foreign social capital can benefit non-mobile scientists from their countries of origin by connecting them with foreign scientists. Nonetheless, these studies rarely analyze whether and how this process works. Recently, three studies have attempted to fill this gap by focusing on returnee scientists (i.e., scientists who have returned to their origin countries) as actors helping the collaboration between non-mobile and foreign scientists (Fry & Furman, 2023; Fry, 2023; Müller et al., 2023). However, these studies neither consider the variety of mobile experiences nor the temporal dimension of the effects, i.e., whether the benefits are long-lasting. Further, it is not entirely clear whether mobile scientists are necessary for non-mobile scientists to establish foreign ties or if their continued presence is required for ongoing international collaborations.

Our paper addresses this gap by investigating scientists from Colombia who trained abroad. We focus on a subset of mobile scientists, namely those who receive their PhDs from foreign institutions. Our choice is driven by two main concerns: during the initial stages of an academic career, especially during the Ph.D. training, a scientist is most susceptible to acquiring skills, attitudes, and fundamental "quality" standards. This is why early career researchers are most likely to help in the diffusion of the knowledge acquired (Stephan, 2006). Second, it is challenging to gather data on the mobility of scientists later in their careers, and, focusing on Ph.D. students, it was possible to leverage reliable and complete data on the total population of students who have experienced mobility.[2]

We focus on connections established through co-authorships and ask whether co-authoring with mobile scientists increases the propensity of non-mobile scientists to collaborate with foreign scientists. We apply two research designs: Ordinary Least Squares (i.e., a linear probability model and a panel model with fixed effects) and a Difference-in-differences event study. We also explore two sources of heterogeneity: first, the differences between STEM and SSH scientists; second, different types of mobile scientists: diaspora (i.e., scientists who have remained abroad), returnee, and intermittent (i.e., scientists who periodically come and go to/from their origin country).

We observe three general trends. First, collaboration with a mobile scientist does increase the amount of international co-authorship of a non-mobile scientist. Second, while STEM scientists are more likely to have foreign co-authors than SSH scientists, the effect of interacting with mobile is not significantly different in the two broad fields. Third, the correlation between interactions with mobile scientists and future foreign collaborations is higher for diaspora scientists, those who do not return to Colombia, than it is for returnees or intermittent returnees. Our event study results are consistent with

[2] This is possibly slightly overstating the case. Our data include all scientists who either received a Ph.D. from a Colombian institution or were funded or whose funding was administered by a national program (which includes Fulbright fellowships) to study abroad. Thus, we miss the (almost certainly small number of) students who are funded by sources outside Colombia or by non-government organizations (such as firms) within Colombia.

these trends; however, they show that effects can be short-lived. Any long-lasting effects on foreign collaborations of non-mobile scientists seem to be driven by repeated interactions with mobile PhDs.

Our work advances the scientific literature and policy formulation in various ways. We first provide evidence that the benefits of international scientific mobility policies extend to other scientists, notably in terms of foreign social capital. Second, different from extant studies that focus mainly on the impact of returnees on non-mobile scientists, we expand the research scope to include other types of mobile scientists. We also advance the literature by exploring the timing of the effect and its duration. Finally, we investigate how non-mobile scientists absorb social capital spillovers.

2. Literature Review

2.1. International scientific collaborations and international scientific mobility as a catching-up mechanism

Collaboration among scientists from different countries has expanded markedly in recent decades (Gui et al., 2019; Ribeiro et al., 2018; Wagner et al., 2015; Wuchty et al., 2007). The proportion of internationally co-authored papers in the Web of Science database more than doubled from 10.14% in 1990 to 24.55% in 2011 (Wagner et al., 2015). The increasing interconnection in science reinforces a challenge to middle-income countries: nations aiming to catch up with advanced economies at the scientific frontier must foster mechanisms that allow local scientists to conduct high-level research with scientists across different countries (Freeman, 2014). Nonetheless, facilitating network formation is a complex task, as it requires pecuniary resources and absorptive capacity (Powell & Grodal, 2005), often limited in middle-income countries (Castellacci & Natera, 2016), and overcoming geographical and cultural barriers (Miguélez, 2018). Countries have implemented international mobility policy schemes for scientists and students to address these constraints and promote network formation with scientists from advanced economies (Dahlman, 2010; Jonkers & Tijssen, 2008).

2.2. International Scientific Mobility

Much of the scholarly literature on international scientific mobility has demonstrated that scientists benefit from international mobility experiences in increasing research capacity and social capital (see Edler et al., 2011; Jonkers & Cruz-Castro, 2013; Liang et al., 2022). Scholars have shown that mobile scientists have more publications and citations (Aykac, 2021; Franzoni et al., 2014; Jonkers & Cruz-Castro, 2013) and more extensive co-authorship networks – a proxy for professional networks –

(Gibson & McKenzie, 2014; Jonkers & Tijssen, 2008; Petersen, 2018) than scientists without any international experience (non-mobile scientists).

The international scientific mobility literature has primarily focused on the impact of mobility experience at the individual level. Several works have compared and analyzed mobile and non-mobile scientists from different countries and scientific fields (Barnard et al., 2012; Baruffaldi et al., 2020; Gibson & McKenzie, 2014; Jonkers, 2011; Müller et al., 2018, 2023; Scellato et al., 2015). The rationale of these works is to obtain a correlation or causal effect showing that mobility affects the career of mobile scientists (Netz et al., 2020). Some of these works suggest that mobile scientists are gatekeepers between different scientific systems as they carry substantial social capital and collaborate with foreign scientists (Gibson & McKenzie, 2014; Barnard et al., 2012). However, with the focus on the mobile scientists themselves, most studies on international scientific mobility neglect the impact of mobile scientists on other scientists who do not have foreign experience (i.e., non-mobile scientists). Investigating this latter aspect is essential as mobile scientists display a large and diversified social capital; thus, they might be network brokers connecting non-mobile with foreign scientists (Fry & Furman, 2023).

Few studies have examined how mobile scientists affect non-mobile scientists' international collaboration. For instance, Yang et al. (2022) find that Chinese returnee scientists positively influence their peers' publications in top journals but not their international connections. In contrast, Müller et al. (2023) find that South African returnee scientists and non-mobile scientists with foreign social capital assist others in forming connections, often through simultaneous collaborations with foreign scientists. Similarly, Fry (2023) discovers that HIV scientists returning from the United States help non-mobile scientists from their home African research institutions to connect with US scientists. In a second paper, Fry and Furman (2023) demonstrate that female mobile scientists from African countries with high gender equality are more likely to share their social capital with non-mobile scientists.

Nevertheless, these latter works have some limitations. First, they limit their analysis to returnees, excluding other types of mobile scientists that might contribute to linking non-mobile scientists, as in the case of diaspora scientists. Second, how mobile scientists share their social capital needs to be clarified. Finally, the timing of the effect and its duration remain unexplored. Our study offers a unique contribution to addressing these gaps. First, we examine the different categories of mobile scientists, including diaspora and intermittent scientists. Second, we investigate whether non-mobile scientists retain foreign social capital to understand how social capital is shared. Finally, we analyze the effects over time using an event study design.

2.3. Diaspora, intermittent scientists, and their networks

The academic literature on diaspora scientists has evolved from a brain drain perspective to considering brain circulation (Agrawal et al., 2011; Cañibano & Woolley, 2015; Fangmeng, 2016; Meyer, 2001; Saxenian, 2005; Velema, 2012). Early studies focused on losses incurred by sending countries, especially developing nations, when their government-sponsored scientists remained abroad, suggesting unfulfilled expected returns (Cañibano & Woolley, 2015). Nevertheless, recent works argue that having scientists abroad can benefit sending countries (Agrawal et al., 2011; Fangmeng, 2016; Saxenian, 2005). Evidence suggests that diaspora scientists contribute to technological development (Saxenian, 2005), facilitate knowledge flows (Kerr, 2008), and maintain collaborations with scientists from their origin countries (Agrawal et al., 2006; Scellato et al., 2015).

Similar to the international scientific mobility literature, literature on diaspora scientists has barely investigated whether diaspora scientists share their social capital with non-mobile scientists. Empirical studies on diaspora networks acknowledge the role of diaspora in general, not only scientists, as conduits between economic agents from both host and sending countries. For example, diaspora can reduce transaction costs (Miguélez, 2018; Ratten & Pellegrini, 2020) and facilitate labor market integration (Elsner et al., 2018). However, the role of diaspora scientists in bridging connections remains under-explored. Researching diaspora scientists is essential as they, on average, have larger and more diversified networks than returnee or non-mobile scientists (Aykac, 2021; Scellato et al., 2015) and maintain close collaboration with foreign scientists due to geographical and social proximity (Kahn & MacGarvie, 2016; Velema, 2012). Hence, the path to reaching a foreign scientist is supposed to be smaller when facilitated by a diaspora scientist.

To investigate the impact of the diaspora on other scientists, it is necessary to consider the changing nature of this latter mobile scientist group. Recent policy developments have led sending countries to engage their highly skilled diaspora (i.e., scientists and engineers) (Hofman & Kramer, 2015; Lewin & Zhong, 2013; Silva, 2018). Inspired by the concept of brain circulation, one example is encouraging the diaspora to return for short stays of teaching or research. At the same time, advances in communication and transportation technologies reduce the differences between those abroad and those located domestically. As scientists become more and more accustomed to communication that is not face-to-face, distance matters less, and it is easier for diaspora scientists to participate in local activities (Meyer, 2001). We create a new category of mobile scientists to deal with this transformation: intermittent scientists, that is, scientists who spend time both "at home" and abroad. As "intermittent," we include not only scientists who live abroad but are constantly returning to their home countries but also scientists in the home country who are affiliated with institutions abroad.

2.4. Network spillovers or borrowing networks for a period?

Unlike International Scientific Mobility researchers, Social Network Analysis scholars have extensively explored the mechanisms through which social capital is shared in a network. Scholars such as Burt (2004) and Gould and Fernandez (1989) discuss how nodes can access the broker's resources when mediated by a broker. They suggest the nodes connected by the broker benefit from direct resource sharing: social capital and other resources would spill over to the connected nodes, leading these latter nodes to have access to a broker's resources. Conversely, Burt (1998, 2000) also observes that social capital is often not permanently transferred but temporarily borrowed. In such scenarios, nodes outside the network may temporally access social capital and gain legitimacy. However, the connections established may not be enduring. Considering non-mobile scientists as outsiders of international scientific networks, we investigate whether these scientists retain foreign social capital over time or borrow it temporarily.

3. Institutional context

Colombia is the empirical context of this research. The country ranks 4th in Latin America in terms of publication counts, with over 17,000 publications in 2021 (Scimago Journal & Country Rank, 2023).[3] Colombia has undergone significant transformations in its scientific system in recent years, making it a compelling case to study. For several decades, the country did not have a structured graduate program system, leading governmental and private agencies to sponsor mobility programs for Colombian students seeking doctoral training (Bedenlier, 2018; Losada, 2016). When international

[3] Compared to the rest of the world, Colombia ranks as the 47th country in publications according to the Scimago Journal & Country Rank. To put this into perspective, Colombia produced 296 publications per capita, while the United States had 1875 publications per capita in 2020.

mobility schemes were established in the 1990s, the government initially funded nine students. Up to 2021, more than 3,000 students have been sponsored. Building absorptive capacity by sending students abroad was critical for Colombia to structure its graduate education. From 2009 to 2019, graduate programs increased their student population by 273.7 percent (from 103 to 385 students). The number of students who obtained a Ph.D. yearly in Colombia rose by 455.49 percent (from 173 to 961, with 5,352 Ph.D. recipients in total during this 10-year-period) (Observatorio Colombiano de Ciencia y Tecnología, 2021). At the same time, the institutionalization of science at the governmental level experienced significant transformations. For decades, the country had a Science, Technology, and Innovation (STI) secretariat responsible for different STI policies. In 2019, the government elevated the importance of STI to the country by establishing a Ministry overseeing STI matters (Dutrénit et al., 2021). Moreover, the national government has established an outward orientation to internationalize Colombian science (Bedenlier, 2018). As a result, Colombian papers co-authored with foreign scientists have grown from 2,298 in 2009 to 10,011 in 2019. Proportionally, in 2019, 44% of the total scientific output in Colombia was papers co-authored with foreign scientists, while papers co-authored with only local scientists represented 38% and solo documents, 9% of the total production (Observatorio Colombiano de Ciencia y Tecnología, 2021).

Given that the Colombian government and other agencies in the country have invested extensively in mobility schemes for citizens to obtain doctoral training abroad, we investigate the impact of these mobility experiences. However, unlike the mainstream approach of the international mobility literature focused on the individual level, we investigate the externalities (if any) in creating channels to connect non-mobile scientists from the origin country to foreign scientists.

Most importantly, Colombia offers us the unique opportunity to conduct a study covering the entire population of Ph.D. holders and active scholars in the country. Doing that, our work distinguishes itself from most studies that consider national sub-populations of scientists or limit their attention to a few universities (Baruffaldi et al., 2020; Fry, 2023; Jonkers, 2011).

4. Data and Empirical Strategy

4.1. Data sources

We combine three different data sources to build a unique dataset on scientists from Colombia from 1990 to 2021. Information on the academic background and publication history of all 28,729

Colombian scientists comes from CvLAC – a standardized scientific curriculum vitae (CV) platform established by the federal government to track the progress of the country's research community since 2000 (de los Ríos & Santana, 2001). As the federal government, research institutes, and universities across the country select researchers for grants, scholarships, and promotions based on the information available in their online CVs, researchers must maintain it updated. Hence, CvLAC is a unique data source with detailed and reliable information on Colombian scientists' backgrounds. We use CvLAC for three purposes. First, the academic background allows us to identify institutions where scientists conducted their studies. Information on Ph.D. institutions permits us to identify whether a scientist has studied abroad, thus classifying her as a mobile scientist.[4] Second, the scientists' working history and affiliation to research projects[5] allows us to identify the mobility type of mobile scientists partly. After her doctoral studies, we checked whether a mobile scientist had a career solely in Colombia or other countries. Third, publication history was essential to conduct an author name disambiguation process and identify scientists in OpenAlex, the bibliometric source used. Notably, the publication history from CvLAC had information on the papers' titles, co-authors, and digital object identifiers (DOI). Nevertheless, CvLAC does not offer information on bibliometric indicators (citation count). To overcome this issue, we rely on the information on publications from CvLAC to identify scientists' publication history in OpenAlex and obtain bibliometric indicators.

Not all mobile scientists have a reason to update their CvLAC record when moving abroad to do a PhD, so we complement information on PhD holders using the lists of scholarship beneficiaries from the national government and funding agencies (Ministry of Science, Colfuturo Foundation, and Fulbright Colombia). We use the information on scholarship holders' names, their doctoral studies starting date, discipline, and the countries they applied to do their Ph.D.

The third data source is OpenAlex, a global open-access database with more than 200 million scientific publications, which builds on Microsoft Academic Graph and is regularly updated using several

[4] We also rely on the academic background information to identify a non-mobile scientist. Unlike the identification process of mobile scientists, we also consider information on the master's programs to classify a scientist as a non-mobile scientist. In Colombia, it is common for scientists to have only a master's as the highest educational level. Hence, for those scientists who are active in the system and have a master's degree from a Colombian institution, we classify them as non-mobile scientists.

[5] At CvLAC, it is possible to include information on different research projects that a scientist has been involved to. For example, the title of the project, its general description and the project members' names.

sources (Singh Chawla, 2022).[6] As the two previous data sources have information on scientists' names and publications but not on bibliometric indicators, we include OpenAlex as the third data source to obtain those indicators. Relying on scientists' names and publication history from CvLAC and scholarship lists, we identify scientists in OpenAlex by conducting an author name disambiguation and matching processes (D'Angelo & van Eck, 2020). We combine the following information: author names fuzzy match, ORCID identifiers, publication titles, bibliographic metadata (e.g., ISSN, DOI, year, volume), and self-citations. We also define several selection criteria based on the CvLAC.[7] The initial number of scientists in CvLAC was 28,729. We match a large sample of 19,661 (68%) scientists with high precision. After matching, we obtain the scientists' bibliometric data, which allows us to compute scientists' productivity (e.g., number of publications and citations), academic seniority, mobility pattern based on the affiliation history from publications, and co-authorship networks (including with foreign scientists).

4.2. Definition and identification of non-mobile, mobile, and foreign scientists

For our analysis, we distinguish among three types of scientists: non-mobile, mobile, and foreign. First, non-mobile scientists are scientists who completed their graduate studies (master's and Ph.D.) entirely in Colombia and have not sojourned abroad to work in academia for more than 11 months in a row. As it is possible in Colombia to pursue an academic career without a Ph.D.[8], we include non-mobile scientists without a Ph.D. but with at least three publications. To identify the non-mobile scientists, we relied on their academic background from CvLAC and their affiliation history with OpenAlex. In this category, we included scientists who have affiliations only with Colombian institutions.

Second, we follow the literature to define mobile scientists as those who have moved abroad to obtain their doctoral degrees (Kahn & MacGarvie, 2012, 2016; Liang et al., 2022; Turpin et al., 2008). Given that there are different types of mobile scientists, we further divide this group into returnees, diaspora, and intermittent. Returnees are those who have returned home indefinitely no more than one year

---

[6] OpenAlex provides more extensive bibliometric than traditional, non-public academic repositories (e.g., Scopus and Web of Science), including publications in a different language than English and from working paper repositories such as arXiv (Visser et al., 2021).
[7] The complete procedure to construct the dataset is explained in Appendix 1.
[8] For example, the academic staff from the National University of Colombia, the largest university in Colombia, comprises of 42% scholars with Ph.D., 38% with a master's degree, and 20% with a bachelor's (Sistema Nacional de Información de la Educacións Superior, 2022).

after their Ph.D. and have not spent more than one year abroad over the rest of their career (until the last year of observation). To classify mobile scientists as returnees, we check if their publications have no foreign affiliation since their Ph.D. Diaspora are scientists who have never returned to work in Colombia after concluding their doctoral studies abroad. To classify mobile scientists as a diaspora, we check whether their publications have no Colombian affiliation. We add a new category to the extant literature of mobile scientists: intermittent scientists. Some mobile scientists move back and forth to and from their origin countries. In our general definition, we define a mobile scientist as intermittent if she has worked for at least one year in Colombia and at least one year abroad over our period of observation. Scientists with affiliations in both Colombia and abroad (multiple affiliation cases) are considered intermittent, under the assumption that they share their time between Colombia and abroad.

Our analysis focuses on the network spillovers from mobile to non-mobile scientists, specifically looking at the connections the latter establish with foreign scientists. To do that, we observe the non-mobile scientists' co-authors and, among them, identify those with foreign affiliations. We follow the strategy of Müller et al. (2023) on foreign scientist identification.[9] We initially identify as foreigners, all co-authors not present in CvLAC. However, some scientists who are not registered on this CV platform may be Colombian. Identifying them as foreigners would overestimate collaborations with foreign scientists and underestimate collaborations with mobile scientists. To avoid this bias, we isolate Colombian mobile scientists in OpenAlex who do not appear in CvLAC in three steps. First, we create a list of common Colombian name-surname combinations from the most frequent combinations in the CvLAC database. Second, we match these name-surname combinations with the co-authors' names identified in OpenAlex. Third, we redefine non-mobile as those matched authors who have published only in national journals[10] or were affiliated only with national organizations throughout their careers, based on OpenAlex. All other co-authors are defined as foreigners. Table 1 describes the different categories of scientists, including definitions and identification methods.

9 In the context of the South African scientific system, Müller et al. (2023) rely on three steps to identify whether a co-author of a rated South African scientist is local or foreign. First, the authors check for any information about the co-author in a national government's CV database. Second, they rely on a national publication database with the names of paper authors and co-authors. Third, they check the institutional affiliations of publications extracted from the Web of Science.

10 We relied on the Publidex database to identify local journals.

Table 1 – Scientists' categories

Categories		Definition	Identification process
Non-mobile scientist		Scientists that have done their graduate studies entirely in Colombia. Scientists without a Ph.D. who have done their Master's in Colombia and have at least three publications.	1. In the academic background section from CvLAC, we checked if a scientist has studied solely in Colombia and/or 2. The affiliation history from publications on OpenAlex has only Colombian affiliations.
Mobile scientist	Returnees	Scientists who obtained a doctoral degree abroad and returned home right after their studies.	1. In the academic background section from CvLAC, we checked if a scientist has done her doctoral studies abroad. 2. The affiliation history from publications on OpenAlex has only Colombian affiliations after the doctoral studies.
	Diaspora	Scientists who obtained a doctoral degree abroad and remained abroad after concluding their graduate studies.	1. In the academic background section from CvLAC, we checked if a scientist has done her doctoral studies abroad and/or 2. The affiliation history from publications on OpenAlex has only affiliations abroad after the doctoral studies.
	Intermittent	Scientists who obtained a doctoral abroad and keep coming back and forth from/to Colombia.	1. In the academic background section from CvLAC, we checked if a scientist has done her doctoral studies abroad and/or 2. The affiliation history from publications on OpenAlex has at least one affiliation in Colombia for one year and one affiliation abroad for one year. 3. Multiple affiliations in different countries, including Colombia.
Foreign scientist		All non-Colombian scientists.	1. Scientists with only foreign affiliations who do not have a CvLAC profile.

4.3. Our sample of scientists

For our study, we consider the entire population of Colombian scientists active from 1990 to 2021. Our initial sample included 28,729 Colombian scientists with CvLAC records. After matching with the OpenAlex database, we arrived at 19,661 scientists, almost 70% of the population. We removed 22 scientists whose mobility patterns could not be identified and 481 Colombian scientists who studied in Colombia but moved abroad after their Ph.D. We also removed 4,057 non-mobile scientists after setting a (at least) 3-paper threshold to guarantee that these latter scientists are active in the scientific system. Our final sample contains 15,101 scientists. In this sample, 43% of the scientists are non-mobile (6,609). Among the remaining mobile scientists (8,492), we have identified 3,866 returnees (26%), 3,581 intermittents (24%), and 1,045 diaspora scientists (7%) (see Table 2).

Table 2 – Distribution of Colombian scientists (15,101) by mobility pattern

(1)	(2)
Mobility Pattern	**Scientists (#)**
Diaspora	1,045
Intermittent	3,581
Returnee	3,866
Non-mobile	6,609

Source: own elaboration based on CvLAC and OpenAlex.

We are investigating if non-mobile scientists connecting with mobile ones increase their connections with foreign scientists. Connections are measured through co-authorships.

In our sample, 5,308 non-mobile scientists (80% of the sample) have at least one publication with a mobile scientist. Interestingly, there are slight differences across the field. Dividing the sample by STEM and social sciences and humanities (SSH), we find that 2,856 STEM non-mobile scientists (or 82%) and 2,452 SSH non-mobile scientists (or 77%) have at least one mobile co-author. On average, for a non-mobile scientist, it takes 2.9 years from the first publication to co-author with a mobile scientist (2.6 years in STEM and 3.2 in SSH; see Table 3).

Table 3 – Distribution of non-mobile scientists that have interacted with mobile scientists

(1)	(2)	(3)	(4)
	Broad field		**Total**
	STEM	**SSH**	
Non-mobile scientists who have co-authored with a mobile (#)	2,856	2,452	5,308
Non-mobile scientists who have co-authored with a mobile (%)	82%	77%	80%
Average time to meet a mobile co-author for the first time (in years)	2.6	3.2	2.9

Notes: This table provides information on the number of non-mobile scientists collaborating with mobile scientists. The first row shows that 5,308 non-mobile have collaborated with a mobile scientist (column 4). It also shows that 2,856 STEM non-mobile (column 3) and 2,452 SSH non-mobile (column 3) scientists have co-written a paper at least with one mobile scientist. The second row shows the percentages: 80% of non-mobile scientists, 82% of STEM non-mobile scientists, and 77% of SSH have at least one paper with a mobile scientist. The third row displays how many years, on average, a non-mobile starts interacting with a mobile scientist. For the whole sample, on average, it takes 2.9 years. For STEM non-mobile scientists, it takes, on average, 2.6 years, while 3.2 years for SSH scientists.
Source: own elaboration based on CvLAC and OpenAlex.

As an exploratory analysis, table 4 reports the interactions of non-mobile scientists with foreign scientists, distinguishing between those who co-authored with mobile scientists and those who did not. On average, collaborating with a foreigner is more common when a mobile scientist is involved. Panel A shows that, on average, non-mobile scientists co-author 1.17 papers per year jointly with mobile and foreign scientists but only 0.80 papers with foreigners when a mobile scientist is not involved. Splitting the sample by macro field, we find a similar pattern: the number of papers co-written with foreign scientists for STEM and SSH scientists is higher when a mobile scientist is also involved. In Panel B, we consider the sub-sample of non-mobile scientists who co-authored with mobile scientists and compare the before and after starting this collaboration. In the overall sample, the number of papers coauthored with foreign scientists increased from 0.8 to 1.4 papers per year. This growth also holds when splitting the sample by macro fields.

Overall, it seems that co-authoring with mobile associate with a higher probability of collaborating with foreign scientists.

Table 4 – Average yearly number of publications co-authored with foreign scientists

Field	Panel 1: Average yearly number of publications of non-mobile scientists with foreign scientists			Panel 2: Average yearly number of publications of non-mobile scientists before and after co-authoring with mobile scientists		
	Not co-authored with a mobile scientist	*Co-authored with a mobile scientist*	*Overall*	*Before*	*After*	*Overall*
STEM	0.98	1.34	1.28	0.9	1.6	1.3
SSH	0.63	0.94	0.87	0.6	1.1	0.9
Overall	0.80	1.17	1.10	0.8	1.4	1.1

Notes: Panel 1 shows the average number of co-authored publications with foreign scientists that involve or do not mobile scientists. When non-mobile scientists co-author with mobile scientists, the average yearly publication with foreign scientists is 1.17 papers. Without collaborating with a mobile scientist, non-mobile scientists co-author, on average, 0.80 papers with foreign co-authors. Panel 2 illustrates the number of publications with foreign scientists before and after a non-mobile scientist interacts with a mobile scientist for the first time. The number of publications with foreign scientists increases from 0.8 to 1.4.

4.4. Variables

Table 5 lists the variables we use in our econometric analysis with short definitions. Given that our outcome variable is whether a non-mobile connects with foreign scientists, we have built four (dependent) variables that capture this connection. *Foreign collaboration* is a dummy variable that takes the value one if a non-mobile scientist has co-authored with at least one foreign scientist.[11] The rationale is to capture the probability that a non-mobile scientist interacts with a foreign scientist. The second dependent variable is the *Average number of foreign co-publications*. We count the publications a non-mobile scientist has co-written with a foreign scientist in a year. Nevertheless, as there might be some years that a non-mobile scientist publishes more with foreign scientists than others, we take a rolling average of the number of publications co-written with foreign scientists in years *t-1*, *t*, and *t+1*. This measure permits the capture of the percentage change in publications with foreign scientists. Third, we construct the *Average share of foreign publications*. Like the previous dependent variable, we also take a three-year rolling average (*t-1*, *t*, and *t+1*). The rationale is to observe whether scientists are changing their behavior and learning through international collaboration. Fourth, we have also included the non-mobile scientist's *Number of publications co-written with foreign scientists*.

[11]As explained in the following subsection, we transform our data into pooled cross-sectional data.

Our main independent variable is *After meeting a mobile*, a dummy variable that turns to one in the year a non-mobile scientist publishes with a mobile scientist. A second independent variable is the cumulative number of unique mobile scientists with whom a non-mobile scientist has co-authored until the year of observation (i.e., *stock of mobile scientists*). We also include STEM, a binary variable taking the value 1 if a non-mobile scientist is a STEM scientist. We interact STEM with the other two independent variables to observe the specificities of STEM and SSH scientists.

Productivity, visibility, stock of foreign co-authors, co-authorship network size, seniority, and gender might influence non-mobile scientists having foreign co-authors. We have added a set of variables to control for those factors. First, the *Lagged average of publications* is the cumulative number of non-mobile scientists' publications until year *t-1*. Second, the *Lagged average of citations* is the cumulative number of non-mobile scientists' citations until year *t-1*. Third, the *Lagged stock of foreign co-authors* is the cumulative number of non-mobile scientists' unique foreign co-authors until year *t-1*. Fourth, the *Lagged average of co-authors per paper* is the size of teams: the stock of co-authors of a scientist divided by the stock of publications until year *t-1*. *Seniority* is the years of experience, proxied as years since the first publication. *Male* is a dummy variable: 1 if a non-mobile scientist is male.

Table 5 provides a detailed description of our variables, while Table 6 presents the descriptive statistics. Imposing a 3-publication threshold yielded a sample of 6,609 non-mobile scientists. Within this sample, 58% are male, and 54% belong to STEM fields. The average non-mobile scientist in our dataset has 20 co-authors, 1.23 publications, 3 citations, and 7 years of experience. Appendix 2 reports the correlation matrix.

Table 5 – Variables description

	Variable	Description
Dependent variables	*Foreign collaboration*	Equal 1 if a non-mobile scientist has collaborated with a foreign scientist
	Average number of foreign co-publications	Rolling average of the number of publications co-written with a foreign scientist in years *t-1*, *t* and *t+1* (in log)
	Average share of foreign co-publications	Rolling average of foreign co-publications share (number of publications co-written with a foreign scientist over total publications) in years *t-t*, *t* and *t+1*
	Number of foreign co-publications	Number of publications co-written with a foreign scientist in year *t* (in log)
Independent variables	*After meeting a mobile scientist*	Equal to 1 if a non-mobile scientist has collaborated with at least one mobile scientist
	Stock of mobile scientist	Cumulative sum of unique mobile scientists that a scientist *i* has in year *t* (in log)
	STEM	Equal to 1 if a scientist *i* belongs to STEM
Control variables	*Lagged average of publications*	Cumulated sum of scientist i's publications until year *t-1* over the number of years since first publication until year *t-1* (in log)
	Lagged average of citations	Cumulated sum of scientist i's citations until year *t-1* over the number of years since first publication until year *t-1* (in log)
	Lagged stock of foreign co-authors	Cumulated sum of scientist *i*'s unique foreign co-authors until year *t-1* (in log)
	Lagged average number of co-authors per paper	Cumulated sum of scientist i's unique co-authors until year *t-1* over the cumulated sum of scientist i's publications until year *t-1* (in log)
	Years since first publication	Number of years of experience. We proxy experience by using the number of years since the first publication
	Gender	Equal 1 if a scientist is a male

Table 6 – Descriptive statistics (67,661 observations on 6609 non-mobile scientists)

Variable	Obs.	Mean	s.d.	Min	Max
Foreign collaboration	6609	0,2853342	0,4515768	0	1
Average number of foreign co-publications	6609	0,6376495	2,021114	0	125,3333
Average share of foreign co-publications	6609	0,228592	0,2883042	0	1
Number of foreign co-publications	6609	0,6352404	2,241467	0	149
After meeting a mobile scientist	6609	0,5317982	0,4989915	0	1
Stock of mobile scientist	6609	1,675323	2,473804	0	46
STEM	6609	0,5424691	0,4981968	0	1
Average of publications	6609	1,233817	1,477746	0.04	66,5
Average of citations	6609	3,41386	38,31585	0	3182,227
Stock of foreign co-authors	6609	20,99373	344,7465	0	22272
Average number of co-authors per paper	6609	2,167507	24,21483	0,0136986	1474
Years since first publication (seniority)	6609	7,361892	5,242128	1	31
Gender (Male: 1)	6609	0,5868225	0,4924078	0	1

4.5. Empirical strategy

This work investigates whether collaborating with mobile scientists influences the collaboration between non-mobile and foreign scientists. In other words, we ask whether collaborating with mobile scientists connects non-mobile scientists with foreign scientists and thus increases the chances of further collaboration outside the domestic context. We identify collaboration as a co-authorship. We adopt three empirical strategies to achieve our objective: Linear Probability Model, Panel regressions with fixed effects and Difference-in-Differences. We choose these three models for several reasons. First, we use a Linear Probability Model (LPM) as we want to estimate the probability that non-mobile scienitsts collaborate with a foreign scientist given the collaboration with mobile scientists. Second, we rely on panel data models as we want to estimate the number of collaborations of non-mobile scientists with foreign scientists as a function of collaboration with mobile scientists, controlling for unobserved and not varying characteristics related with individual, time and scientific fields. These initial approaches are structured in a way to provide aggregate correlations. However, we also want to observe the effect over time. For that, we adopt a second empirical strategy: a Difference-in-differences with multiple time periods framework with "interacting with a mobile scientist" as the treatment (Callaway & Sant'Anna, 2021). Across these strategies, we also explore heterogeneity. First, we investigate the differences between STEM and SSH scientists. Second, we explore the differences in collaborating with the three types of mobile scientists (diaspora, intermittents and returnees).

Linear Probability Model (LPM)

We rely on Ordinary Least Squares (OLS) and estimate the Linear Probability Model (LPM) represented in Equation 1. The subscript *i* indicates that the analysis is at the non-mobile scientist level.

$$\text{P(Foreign collaboration} = 1 | \mathcal{X})_{it} = \beta_0 + \beta_1 Collaboration\ with\ mobile_{it} + \beta_2 STEM_i + \beta_3 Collaboration\ with\ mobile_{it} * STEM_i + \Gamma\ \Xi_{it} + \varepsilon_{it} \qquad (1)$$

Where *Collaboration with mobile* is, in turn, (i) a dummy equal to one after having collaborated with a mobile scientist, and (ii) the log transformation of the number of unique mobile scientists with whom the focal non-mobile scientist has collaborated. STEM is a dummy equal to one if the scientist *i* belongs to the STEM field, 0 otherwise. As controls, we included the average number of publications, average number of citations, stock of foreign co-authors, average number of co-authors per paper, years of experience, and male. The first four control variables are lagged.

Panel regressions with fixed effects

We estimate the panel regressions, relying on the following baseline models, with the average number of foreign co-publications and the average share of foreign co-publications as dependent variables.

$$\text{Average number of foreign pub}_{it} = \beta_0 + \beta_1 Collaboration\ with\ mobile_{it} + \beta_2 STEM_i + \beta_3 Collaboration\ with\ mobile_{it} * STEM_i + \Gamma\ \Xi_{it} + \varphi_i + \psi_i + \gamma_t + \varepsilon_{it} \qquad (2)$$

Likewise,

$$\text{Average share of foreign pub}_{it} = \beta_0 + \beta_1 Collaboration\ with\ mobile_{it} + \beta_2 STEM_i + \beta_3 Collaboration\ with\ mobile_{it} * STEM_i + \Gamma\ \Xi_{it} + \varphi_i + \psi_i + \gamma_t + \varepsilon_{it} \qquad (3)$$

Similar to the econometric specification in Equation 1, Equations 2 and 3 have the same independent variable (after meeting a mobile or stock of mobile scientists) and the same set of controls. We estimate these latter two equations with fixed effects at the individual, time and scientific field levels. Relying on fixed effect estimations allows us to eliminate unobserved time-invariant effects related to the non-mobile scientist herself and to her scientific field (Wooldridge, 2012). We cluster the standard errors at the scientific field level.

To explore the differences between interacting with each type of mobile scientist (returnee, diaspora, and intermittent), we split the stock of mobile scientists into three variables: stock of diaspora, stock of returnees, and stock of intermittent.

Difference-in-Differences Event Study

In the second step, we rely on a Difference-in-Differences Event Study (Borusyak, Hull, & Jaravel, 2022) to understand how the effect of interacting with a mobile scientist evolves over time (the year of treatment and the five years after). Specifically, we use the Difference-in-Differences estimator with multiple time periods developed by Callaway & Sant'Anna (2021). The latter approach is used because, different from the canonical 2x2 Difference-in-Differences setup, our treatment (interacting with mobile scientists) happens at different years for non-mobile scientists. We need, therefore, an empirical approach that considers setups with more than two periods.

The treatment in our setup is interacting with a mobile scientist (collaborating as a co-authorship) for the first time. Given that Callaway and Sant'Anna's (2021) setup follows a staggered treatment, we consider that once treated, a non-mobile scientist remains treated in the following years. Our treated group is formed by Colombian non-mobile scientists who have co-authored with mobile scientists at least once during the period we observe them. We add another restriction to be part of the treatment group: we dropped non-mobile scientists who met mobile scientists in the first year of their career as we want to observe non-mobile scientists' pre-trends. The control group is formed by all other non-mobile scientists who have never co-authored with a mobile scientist. The estimated outcome indicates the intensity of co-authoring with foreign scientists. To simplify, here we use the number of publications co-authored with foreign scientists (in logs).

Since we need a comparable group to obtain a causal effect, for that, we have taken some steps to create our control group. Initially, our control group included only non-mobile scientists with a research trajectory and who were able to publish in journals indexed at OpenAlex. Specifically, to be included in our data, a scientist must have published at least three papers and work at a university or institute that conducts research in Colombia. We consider the number of citations and publications to show that our treatment and control groups are comparable. We compare these figures for non-mobile scientists based on the years it took them to meet a mobile scientist (treatment group) and the same number of years of working experience (control group). For example, we compare scientists in the treatment group who took two years to meet a mobile scientist with scientists in the control group with two years of working experience, and so on. We show that the means of citations and publications stocks are not statistically different (see Appendix 3).

We must consider the parallel assumptions in Difference-in-difference setups to obtain a reliably estimated outcome (the average treatment effect on the treated – ATT) (Angrist, 2008). In the multiple time period setup, the parallel trends hold when conditioning on observed covariates (Callaway & Sant'Anna, 2021). Hence, we incorporate covariates such as education level, gender, and years since the *first publication to* account for these factors.

The Difference-in-Differences baseline model used is represented in Equation 4.

$$Y_{it} = \alpha_1^{g,t} + \alpha_2^{g,t} G_g + \alpha_3^{g,t} 1\{T = t\} + \beta \left(G_g . 1\{T = t\}\right) + x_i' \theta + \varepsilon_{it} \quad (4)$$

where the left-handed variable is the outcome variable (number of publications co-authored with foreign scientists) for unit *i* at time *t*; G_g is a binary variable that equals one if a non-mobile scientist belongs to the treatment group and becomes first treated in year *g* and equals zero for the control group; $\alpha_2^{g,t}$ is the coefficient for the treatment group dummy; $1\{T = t\}$ is an indicator that equals 1 if the current time period is *t*, 0 otherwise; $\alpha_3^{g,t}$ captures time fixed effects; $\left(G_g . 1\{T = t\}\right)$ is an interaction term between treatment group indicator and time indicator. It represents the post-treatment period for treated units. β represents the average treatment effect. $\boldsymbol{x}_i'$ is the covariate vector: education level (Master's and Ph.D.), gender, years since first publication, and the average of publications. Following Abadie, Athey, Imbens, & Wooldridge's (2022) and Callaway & Sant'Anna's (2021) recommendations, we clustered standard errors at the individual level.

We are aware that many non-mobile scientists interact with different mobile scientists several times over the years, which might reinforce the treatment. To investigate the treatment effect given the addition of mobile scientists to non-mobile scientists' co-authorship stock, we isolated the treatment sample into subsamples by the total number of unique mobile scientists that a non-mobile scientist has interacted with in her entire publication trajectory (Yadav et al., 2023). For example, we consider non-mobile scientists who have collaborated with only one, two, or three unique mobile scientists. We also included non-mobile scientists who have interacted with mobile scientists without any threshold restriction regarding the stock of mobile scientists.

In the Difference-in-Differences empirical strategy, we also explore heterogeneity. First, we divided the sample by the macro field (STEM or SSH) and explored their differences. Second, to observe whether the effect differs by the type of mobile scientist a non-mobile scientist interacts with, we isolated the treatment sample by non-mobile scientists that have collaborated only with returnees or diaspora and intermittent in the next five years after the treatment. In the two cases, we also explored the addition of new mobile scientists to the non-mobile scientists' stock of co-authors.

Finally, to explore how foreign social capital is shared, we ask whether the connections remain without the mobile scientist. We observe whether non-mobile scientists collaborate with foreign scientists even if mobile scientists are not included. To conduct this last step, we remove all foreign co-publications that include a mobile scientist.

5. Results

This section reports the results of our three empirical strategies. Initially, we show the probability of collaborating with at least one foreign co-author. In the second step, we report the panel fixed effect estimations from equations 3 and 4. Specifically, we show the results on the share of publications with foreign co-authors and the total number of publications with foreign co-authors. Finally, we present the results from the difference-in-differences estimation.

5.1. Probability of collaborating with at least one foreign co-author

Table 7 reports the OLS estimations of Equation 1. In columns 1-4 *Collaboration with a mobile* is a dummy variable; in columns 5-8 consider the key explanatory variable is the stock of unique mobile scientists with whom the non-mobile scientist has co-authored.

Table 7 shows that collaborating at least once with a mobile scientist (column 1) is related to a 23 percentage higher probability of collaborating with a foreigner. This value decreases to 12 percentage when adding the controls (columns 2, 3, and 4). Nonetheless, the results remain highly significant at 0.01%. This implies that once we control for observable features that influence the probability of foreign co-authorship (including previous experience of co-authoring with them), non-mobile scientists are 12% more likely to co-author (again) with a foreign scientist after having co-authored with at least one mobile scientist.

Increasing the stock of mobile co-authors also increases the probability of having a foreign co-author (columns 5-8). Column 5 shows that a 10% increase in the stock of mobile scientists is related to an increase of 18 percentage points in the probability of having a foreign co-author. Including the control variables, non-mobile scientists who co-author with more mobile scientists experience a 10 percentage point increase in the likelihood of having foreign co-authors (columns 6-8).

Our control variables have the expected effects. We find correlations that hold in the different models: past productivity (measured by the average number of publications), past visibility (measured by average citations), and the previous stock of foreign scientists are statistically and positively correlated with co-authoring with foreign scientists. However, Colombian non-mobile scientists have a lower probability of co-authoring with a foreign scientist if they are more senior or are involved in larger teams (measured by the average number of co-authors per paper). Finally, we find a gender bias: a male non-mobile scientist is more likely to have a foreign co-author than a female after controlling for all observable characteristics.

Regarding the difference between STEM and SSH non-mobile scientists, while STEM non-mobile scientists are more likely to have foreign collaborators than those in SSH, the effect of meeting mobile scientists is not different in the two broad fields. Inspecting columns 4 and 8, we observe the coefficient estimate on STEM is positive and significant, but the interaction terms (Meeting a mobile scientist x STEM and Stock of Mobile scientists x STEM) are both close to zero and statistically insignificant.

Table 7 – Foreign Collaboration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Foreign collaboration	Foreign collaboration	Foreign collaboration	Foreign collaboration	Foreign collaboration	Foreign collaboration	Foreign collaboration	Foreign collaboration
After meeting a mobile scientist	0.236***	0.120***	0.124***	0.120***				
	(0.00330)	(0.00366)	(0.00357)	(0.00481)				
Stock of mobile scientist (in log)					0.188***	0.101***	0.0961***	0.110***
					(0.00251)	(0.00300)	(0.00287)	(0.00384)
Lagged average of publications (in log)		0.114***				0.0275***		
		(0.00449)				(0.00581)		
Lagged stock of foreign co-authors (in log)		0.109***	0.132***	0.131***		0.150***	0.135***	0.134***
		(0.00184)	(0.00256)	(0.00257)		(0.00265)	(0.00256)	(0.00256)
Lagged average number of co-authors per paper (in log)		-0.000742***	-0.104***	-0.105***		-0.131***	-0.124***	-0.125***
		(0.0000643)	(0.00437)	(0.00437)		(0.00530)	(0.00436)	(0.00436)
Years since first publication (seniority)		0.00156***	-0.00470***	-0.00463***		-0.00454***	-0.00552***	-0.00539***
		(0.000406)	(0.000391)	(0.000392)		(0.000477)	(0.000394)	(0.000394)
Gender (Male: 1)		0.0157***	0.0156***	0.0149***		0.0138***	0.0146***	0.0134***
		(0.00320)	(0.00319)	(0.00319)		(0.00320)	(0.00319)	(0.00319)
Lagged average of citations (in log)			0.0506***	0.0494***			0.0431***	0.0427***
			(0.00307)	(0.00308)			(0.00312)	(0.00312)
STEM				0.0131***				0.0390***
				(0.00392)				(0.00423)
After Meeting a Mobile x STEM				0.00586				
				(0.00628)				
Stock of mobile scientist x STEM								-0.0265***
								(0.00485)
constant	0.160***	-0.0169**	0.136***	0.132***	0.149***	0.140***	0.157***	0.140***
	(0.00206)	(0.00615)	(0.00652)	(0.00668)	(0.00224)	(0.00885)	(0.00647)	(0.00668)
N	67661	67661	67661	67661	67661	67661	67661	67661
R^2	0.068	0.172	0.178	0.179	0.079	0.175	0.178	0.179

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

5.2. Share of publications with foreign co-authors (over total publications)

Interaction with a mobile scientist may have two effects: it can change publications overall, and it can change the extent of internationalization (publications with foreign scientists). Here we ask whether the results above are driven by the former or the latter by looking at shares of foreign publications.

Table 8 displays the changes in the *Average share of publications with foreign co-authors* (number of foreign publications over total publications of non-mobile scientists) given the interaction with a mobile scientist. We add scientific fields (columns 1 and 4), individual fixed effects (columns 3 and 6), and time-fixed effects (columns 1, 3, 4, and 6) to control for unobserved characteristics that vary across disciplines, scientists, and time. We report the results for both independent variable specifications: co-authoring at least one mobile scientist (columns 1-3) and the number of unique mobile scientists with whom the non-mobile scientist co-authors (columns 4-6).

Columns 1 to 3 show that collaborating with at least one mobile scientist is positively associated with the share of publications with foreign co-authors after controlling for productivity, visibility, team size, seniority, and gender. The magnitude of *After meeting a mobile scientist* in column 1 with field fixed effects suggests that meeting a mobile scientist is associated with an increase in the share of foreign co-publications by 3.83 percentage points. When we include only individual and time-fixed effects, the *Average share of foreign co-publications* increases by 5.46 percentage points.

We also obtain a positive correlation when we use the count of collaborations with unique mobile scientists (Stock of mobile scientists in columns 4-6). For example, including individual and time-fixed effects, we obtain a large effect on the estimated coefficient and increase the predictive power (at least as measured by R^2). The coefficient estimated (column 6) indicates that by increasing the *Stock of mobile PhDs* by 1%, the *Average share of foreign co-publications* increases by 3.86 percentage points.

Discipline does not affect foreign collaboration patterns when individual fixed effects are included. This might not be surprising as disciplines are largely captured in individual fixed effects (few people change from STEM to SSH over their careers). However, when discipline is treated specifically (excluding individual fixed effects), we see that, as in the results above, STEM scientists are more international than SSH researchers (column 5). The coefficient indicates that STEM scientists have a 3.5 percentage point higher share of foreign co-publications than SSH scientists.

Table 8 - Average share of foreign co-publications

	(1)	(2)	(3)	(4)	(5)	(6)
	Average share of foreign co-publications	Average share of foreign co-publications	Average share of foreign co-publications	Average share of foreign co-publications	Average share of foreign co-publications	Average share of foreign co-publications
After meeting a mobile scientist	0.0383***	0.0445***	0.0546***			
	(0.00467)	(0.00446)	(0.00740)			
Stock of mobile scientist (in log)				0.0138**	0.0344***	0.0386***
				(0.00429)	(0.00452)	(0.00953)
Lagged average of publications (in log)	-0.0151			-0.0101		
	(0.0105)			(0.0112)		
Lagged stock of foreign co-authors (in log)	0.149***	0.139***	0.0740***	0.149***	0.140***	0.0758***
	(0.00488)	(0.00337)	(0.00475)	(0.00477)	(0.00329)	(0.00497)
Lagged average number of co-authors per paper (in log)	-0.0782***	-0.0708***	0.0104	-0.0795***	-0.0763***	0.00296
	(0.0124)	(0.0105)	(0.00902)	(0.0127)	(0.00987)	(0.00955)
Years since first publication (seniority)	-0.0111***	-0.00755***		-0.0107***	-0.00731***	
	(0.000964)	(0.000755)		(0.000940)	(0.000639)	
Gender (Male: 1)	0.00957***	0.00525		0.00931***	0.00481	
	(0.00206)	(0.00343)		(0.00195)	(0.00333)	
STEM		0.0164***			0.0350***	
		(0.00248)			(0.00458)	
Lagged average of citations (in log)		0.0190***	-0.0366***		0.0202***	-0.0430***
		(0.00374)	(0.00494)		(0.00443)	(0.00468)
After Meeting a Mobile x STEM		-0.00107	0.0120			
		(0.00699)	(0.00908)			
Stock of mobile scientist x STEM					-0.0243***	-0.00305
					(0.00608)	(0.00987)
constant	0.189***	0.138***	0.123***	0.194***	0.138***	0.137***
	(0.0168)	(0.00905)	(0.00724)	(0.0177)	(0.00834)	(0.00738)
Field Fixed Effects	Yes	No	No	Yes	No	No
Individual Fixed Effects	No	No	Yes	No	No	Yes
Time Fixed Effects	Yes	No	Yes	Yes	No	Yes
N	67661	67661	67661	67661	67661	67661
R^2	0.313	0.291	0.567	0.310	0.289	0.565

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.3. Number of publications with foreign co-authors

We have observed an effect on shares, but of course, shares can change through two different mechanisms: increasing the number of foreign co-authored papers or decreasing those with only domestic co-authors. To address this, we repeat the analysis above, but here using the *Average number of foreign co-publications* (in logs) as the dependent variable. This is presented in Table 9 (for at least one collaboration with mobile scientists in columns 1-3 and the stock of collaborations with mobile scientists in columns 4-6).

Results are similar to those discussed in the previous section, suggesting that a higher share of publications with foreign co-authors following collaboration with mobile scientists is driven by an increase in the number of publications with foreign co-authors rather than a decrease in purely domestic publications. Collaborating with mobile scientists is associated with an increase of almost 6% in the number of publications with foreign co-authors (column 2) for non-mobile scientists. When we control for individual and time-fixed effects, the coefficient increases to 7.5%. This corroborates the previous results: meeting a mobile scientist may be important for a non-mobile scientist to forge a strategy for accessing foreign collaborators.

Coefficients from the stock of mobile scientists (in the log) in columns 4-6 show that an increase in this stock is positively related to an increase in the average number of foreign co-publications. In column 6 (with individual and time-fixed effects), the elasticity of foreign co-publications concerning the number of unique mobile scientists with whom a non-mobile scientist collaborates is 0.099. This indicates that doubling the stock of mobile collaborators in a non-mobile scientist's career leads to a 10% increase in the number of foreign co-publications.

Investigating the control variables, we find correlations that hold in the different models. Higher visibility (measured by the average of past citations) and being male positively correlate with the average number of foreign co-publications. Being part of larger teams (measured by the *Average number of co-authors per paper*) negatively correlates with the average number of foreign co-publications. Seniority also correlates negatively with the dependent variable. Controlling for field and time fixed effects (columns 1 and 4) or without any fixed effects (columns 2 and 5), senior non-mobile scientists tend to collaborate less with foreign scientists.

Looking at fields, here we do see a suggestion that STEM and SSH differ: an interaction with a mobile scientist leads to STEM scientists having, on average, 2.8% more foreign co-publications than SSH that interact with mobile scientists (column 3). Also, controlling for individual and time-fixed effects, we find that an increase of 1% in the stock of mobile scientists leads to STEM scientists having an increase of 3.09% more foreign co-publications than SSH scientists (column 6). Although the differences are slight, STEM non-mobile scientists tend to benefit more from interacting with mobile scientists than SSH non-mobile scientists, as seen from the interaction terms.

Table 9 – Average number of foreign co-publications

	(1)	(2)	(3)	(4)	(5)	(6)
	Average number of foreign co-publications (in log)	Average number of foreign co-publications (in log)	Average number of foreign co-publications (in log)	Average number of foreign co-publications (in log)	Average number of foreign co-publications (in log)	Average number of foreign co-publications (in log)
After meeting a mobile scientist	0.0522***	0.0578***	0.0757***			
	(0.00603)	(0.00721)	(0.0106)			
Stock of mobile scientist (in log)				0.0534***	0.0679***	0.0993***
				(0.00634)	(0.00846)	(0.0137)
Lagged stock of foreign co-authors (in log)	0.232***	0.232***	0.137***	0.233***	0.233***	0.134***
	(0.00808)	(0.00769)	(0.00884)	(0.00821)	(0.00777)	(0.00930)
Lagged average number of co-authors per paper (in log)	-0.176***	-0.179***	-0.0533***	-0.187***	-0.190***	-0.0688***
	(0.0153)	(0.0142)	(0.0125)	(0.0137)	(0.0122)	(0.0133)
Years since first publication (seniority)	-0.0151***	-0.0139***		-0.0158***	-0.0147***	
	(0.000724)	(0.000819)		(0.000762)	(0.000753)	
Lagged average of citations (in log)	0.102***	0.106***	0.0205*	0.0953***	0.0990***	0.000651
	(0.00529)	(0.00467)	(0.00803)	(0.00577)	(0.00489)	(0.00846)
Gender (Male: 1)	0.0257***	0.0236***		0.0250***	0.0222***	
	(0.00430)	(0.00406)		(0.00433)	(0.00441)	
STEM					0.0237**	
					(0.00668)	
After Meeting a Mobile x STEM			0.0282*			
			(0.0110)			
Stock of mobile x STEM					-0.0172	0.0309*
					(0.00862)	(0.0130)
constant	0.230***	0.219***	0.165***	0.235***	0.214***	0.155***
	(0.00902)	(0.0104)	(0.00771)	(0.00861)	(0.00821)	(0.00785)
Field Fixed Effects	Yes	No	No	Yes	No	No
Individual Fixed Effects	No	No	Yes	No	No	Yes
Time Fixed Effects	Yes	No	Yes	Yes	No	Yes
N	67661	67661	67661	67661	67661	67661
R2	0.414	0.409	0.630	0.416	0.411	0.632

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

5.4. The role of different types of mobile scientists

In Table 10, we separate mobile scientists by type (returnee, diaspora, and intermittent) for both the share (columns 1-5) and the number (columns 6-8) of publications with foreign scientists.

Column 1 reports results without any control. Here, the estimated effect of collaborating with diaspora is roughly 4 times as high as that of collaborating with returnees. When we add controls (column 2) and field fixed effects (column 3), we find, in fact, that increasing the *Stock of returnees* is not associated with a significant increase in the *Average share of foreign co-publications*.

Column 4 shows that a significant difference in the role of different types of mobile scientists arises partly in STEM. For the SSH non-mobile scientists, the correlations between co-authoring with returnees and the *Average share of foreign co-publications* are positive and significant even when controlling for observable features. However, when STEM non-mobile scientists increase their *stock of returnees*, the *Average share of foreign co-publications* decreases. Because STEM non-mobile scientists are more likely to co-author with foreign scientists than SSH scientists, this suggests that returnees substitute for foreign scientists: non-mobile STEM scientists collaborate either with returnees or foreign co-authors, but not with both. Controlling for individual fixed effects (column 5), increasing the *Stock of intermittent or returnees* is more relevant than increasing the stock of diaspora to increase the share of foreign co-publications.

When we control for field fixed effects (column 6) or add the set of controls (column 7), again, the effect of diaspora and intermittent collaborators is larger than that of returnee collaborators. However, when including individual fixed effects (column 8), we find that increasing the *Stock of diaspora scientists* is not statistically significant when the dependent variable is the number rather than the proportion of foreign co-publications. Finally, collaborating with intermittent scientists has a higher magnitude on the *Average number of foreign co-publications* than co-authoring with returnee scientists.

Table 10 – Average share and number of foreign co-publications (by different mobility categories)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Average share of foreign co-publications	Average share of foreign co-publications	Average share of foreign co-publications	Average share of foreign co-publications	Average share of foreign co-publications	Average number of foreign co-publications	Average number of foreign co-publications	Average number of foreign co-publications
Stock of returnees (in log)	0.0398***	-0.000259	-0.0112*	0.0171***	0.0167*	0.0561***	0.0418***	0.0551***
	(0.00814)	(0.00566)	(0.00476)	(0.00412)	(0.00590)	(0.0123)	(0.00851)	(0.0132)
Stock of diaspora scientists (in log)	0.166***	0.0485***	0.0350**	0.0541***	0.0284	0.0652*	0.0627**	0.0306
	(0.0182)	(0.0114)	(0.0118)	(0.00786)	(0.0237)	(0.0254)	(0.0214)	(0.0314)
Stock of intermittent scientists (in log)	0.0976***	0.0240***	0.0165**	0.0304***	0.0392**	0.103***	0.0623***	0.0911***
	(0.00580)	(0.00469)	(0.00461)	(0.00578)	(0.0111)	(0.0115)	(0.00877)	(0.0112)
Lagged stock of foreign co-authors (in log)		0.140***	0.137***	0.139***	0.0751***	0.133***	0.231***	0.133***
		(0.00313)	(0.00345)	(0.00311)	(0.00491)	(0.00931)	(0.00772)	(0.00932)
Lagged average of citations (in log)		0.0200***	0.0148**	0.0196***	-0.0450***	-0.00156	0.0972***	-0.00366
		(0.00473)	(0.00395)	(0.00460)	(0.00489)	(0.00856)	(0.00494)	(0.00846)
Lagged average number of co-authors per paper (in log)		-0.0760***	-0.0656***	-0.0770***	0.00231	-0.0699***	-0.192***	-0.0691***
		(0.00994)	(0.00974)	(0.00973)	(0.00975)	(0.0134)	(0.0122)	(0.0135)
Years since first publication		-0.00728***	-0.00991***	-0.00716***			-0.0147***	
		(0.000617)	(0.000641)	(0.000596)			(0.000744)	
Gender (Male: 1)		0.00666*	0.00976***	0.00543			0.0229***	
		(0.00303)	(0.00185)	(0.00319)			(0.00431)	
STEM				0.0367***			0.0273***	
				(0.00458)			(0.00588)	
Stock of returnees x STEM				-0.0320***	-0.00935		-0.0272*	0.00425
				(0.00748)	(0.0108)		(0.0111)	(0.0185)
Stock of diaspora x STEM				-0.00577	0.0109		0.00871	0.0509
				(0.0160)	(0.0303)		(0.0331)	(0.0490)
Stock of intermittent x STEM				-0.0123	-0.00121		-0.00958	0.0251
				(0.00803)	(0.0136)		(0.0109)	(0.0176)
_cons	0.164***	0.158***	0.182***	0.141***	0.144***	0.174***	0.220***	0.173***
	(0.0119)	(0.0107)	(0.00925)	(0.00845)	(0.00712)	(0.00966)	(0.00838)	(0.00813)
Field Fixed Effects	No	No	Yes	No	No	Yes	No	No
Individual Fixed Effects	No	No	No	No	Yes	No	No	Yes
Time Fixed Effects	No	No	Yes	No	Yes	Yes	No	Yes
N	67661	67661	67661	67661	67661	67661	67661	67661
R^2	0.073	0.289	0.312	0.291	0.565	0.632	0.412	0.633

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

5.5. Event Study

In this section, we present the results from the difference-in-differences strategy. We first present the results for all non-mobile scientists without any distinction (Figure 1). Then, we differentiate between STEM and SSH non-mobile scientists (Figure 2), and finally, we examine interactions with diaspora and intermittent versus returnee scientists (Figure 3). In this strategy, we have the number of foreign co-publications as the outcome variable, and the treatment is collaboration with a mobile scientist. In each case, the estimation of the treatment effect is based on Equation 4 above. Given that a non-mobile scientist might meet several mobile ones across her career, we differentiate the treatment group by the number of unique mobile scientists (one, two, three, or all) that a non-mobile scientist meets.

The four panels in Figure 1 show that in the first five years before the treatment, the treatment and control groups are similar in the number of papers they co-author with foreign scientists. From these four panels, we can also observe that, in the year of the treatment, an average scientist in the treatment group has 20% more foreign co-publications than those in the control group. This difference in the first year is larger than that in any of the five years that follow.

When we investigate each panel separately, we find that a mechanical effect (marked by the co-authoring of papers among non-mobile, mobile, and foreign scientists) drives the number of papers with foreign scientists. For example, in panels 1 and 2 (encounters with one and two unique mobile scientists, respectively), although the averages in the years after the treatment are above 0, they are individually not statistically significant. Increasing the number of unique mobile scientists (panels 3 and 4), the averages become significant. In other words, for non-mobile scientists with one or two unique mobile co-authors, we see a strong effect only in the initial year — while on average increases relative to pre-treatment numbers are positive, they are relatively small, and (individually) not statistically significant. Only for non-mobile scientists with at least three unique mobile scientists (in one or multiple years) we still see a positive effect after three or four years. Given that these authors have several mobile co-authors, it seems likely that the long-lasting effect is, in fact, driven by co-authoring jointly with mobiles and foreigners in each of those years. Hence, it seems that non-mobile scientists may need to be reconnected to foreign scientists every time (by a mobile scientist) to continue collaborating with foreigners. This is elaborated below.

The event study results confirm that increasing the number of unique mobile scientists positively affects the number of foreign co-publications, as observed in the OLS regression (Table 9). They also help qualify the OLS results, suggesting that most of that effect is explained by including non-mobile scientists in the publications that mobile co-authors make with foreign scientists. While that may have other positive impacts, we find only weak evidence that it helps to establish long-lasting links between non-mobile and foreign scientists that do not depend on the ongoing presence of mobile collaborators.

In general, these results demonstrate that interacting with several mobile scientists can benefit non-mobile scientists in expanding their portfolio of foreign collaborations: as non-mobile scientists diversify their network of mobile scientists, the number of foreign co-publications increases. Nonetheless, although we find a positive role for mobile scientists in leveraging the home scientists' internationalization process, it is unclear how long the effect of collaborating with a mobile scientist lasts.

Having analyzed the interactions in the aggregate, we now differentiate by macro field (Figure 2) and by interaction with different types of mobile scientists (Figure 3).

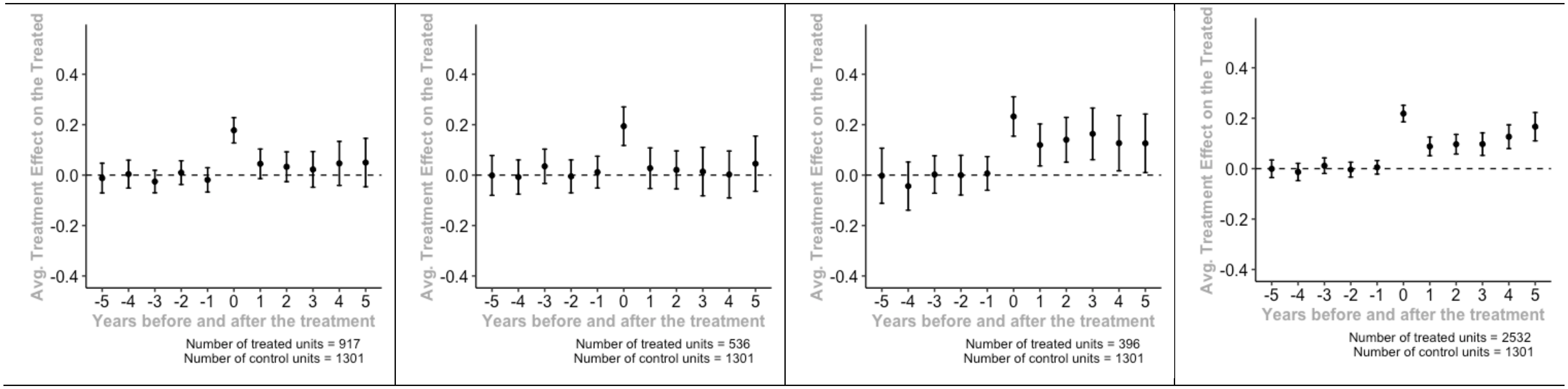
Figure 2, in which we differentiate between STEM and SSH, shows that after collaborating with a mobile scientist for the first time, in almost all panels, the average number of publications is above 0, although not statistically significant in some cases. This means that, on average, interacting with mobile scientists positively affects non-mobile scientists' (number of) foreign co-publications. Also, comparing STEM and SSH scientists, we find a similar result to the OLS result (Table 9): there is no significant difference in increases in the number of publications with foreign scientists between STEM and SSH non-mobile scientists; patterns over time are very similar for the two. Similar to the general Event Study results, a mechanical effect drives the number of foreign co-publications for STEM and SSH non-mobile scientists: having more mobile co-authors increases the number of foreign go-authors.[12]

[12] There is one violation of parallel trends in the scenario where STEM non-mobile scientists have met only two mobile scientists. However, this is an isolated case.

In Figure 3, we explore the heterogeneous effects of the interaction with different types of mobile scientists, again showing the average treatment effects based on equation 4. Because sample sizes become small with this fine dis-aggregation, it is difficult to tell a definitive statistical story. In the final panels, where we include all who have had at least one encounter, a statistically significant effect is visible for diaspora mobiles but not for returnees. However, there are some suggestive patterns in the other panels. Point estimates of the effects of diaspora interactions are almost always higher than those for returnee interactions. Any pairwise comparison between returnee and diaspora interactions will not yield a significant difference in effect, but the patterns suggest that this may be a result of small sample sizes and that in larger samples, the differences observable in the point estimates of the mean effect will carry through statistically.

Figure 1 – Average Treatment Effect of the Treated: number of papers with foreign co-authors per year (in logs)

(1)	(2)	(3)	(4)
One encounter	Two encounters	Three encounters	All encounters



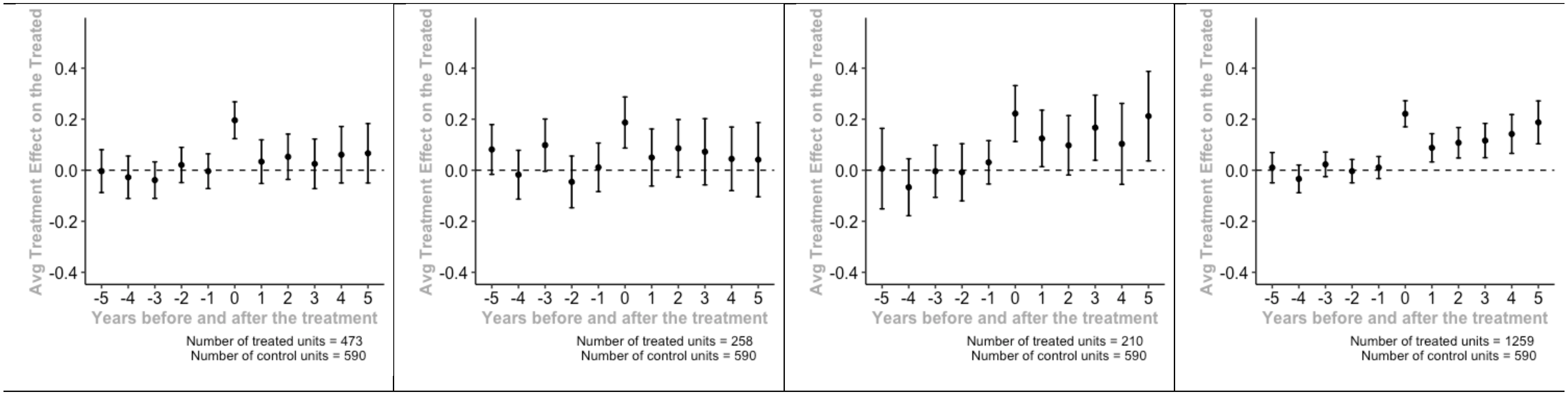
Note: “Treated” refers to non-mobile scientists who, at year zero, collaborate with a mobile. The vertical axis presents the difference between treated and “equivalent” non-treated. Error bars (5 and 95 %) are estimated by a bootstrapping procedure (see Callaway & Sant’Anna (2021)).
Source: Own elaboration based on CvLAC and OpenAlex

Figure 2 – Average Treatment Effect on the Treated: number of papers with foreign co-authors per year (in logs) – STEM and SSH

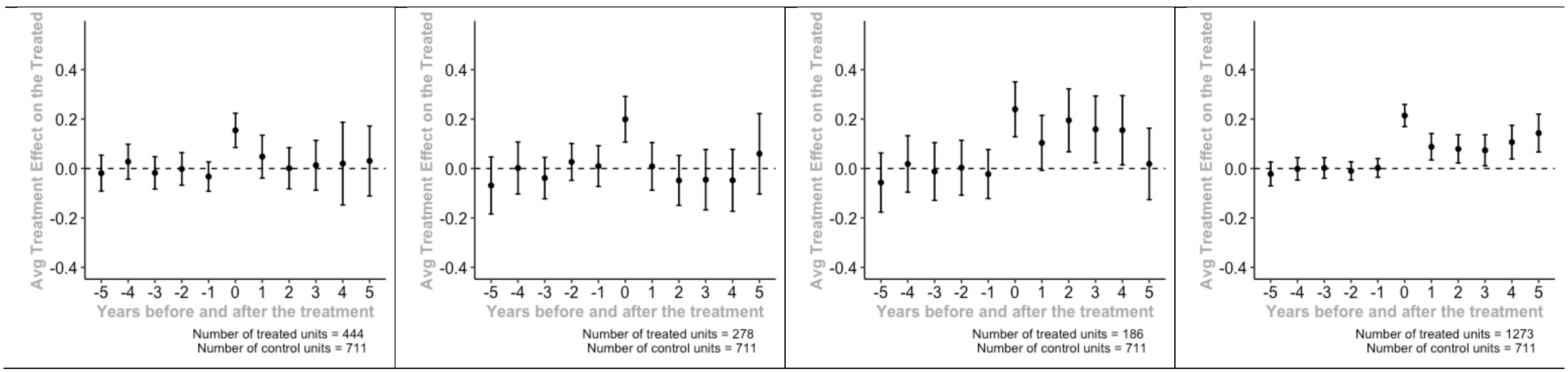
(1)	(2)	(3)	(4)

STEM

One encounter	Two encounters	Three encounters	All encounters



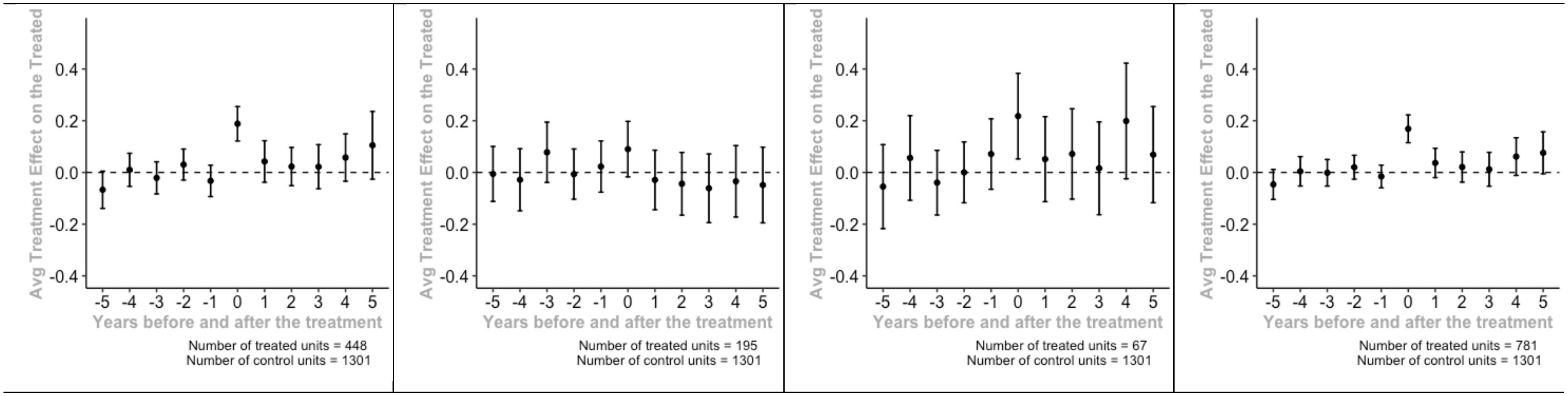
SSH



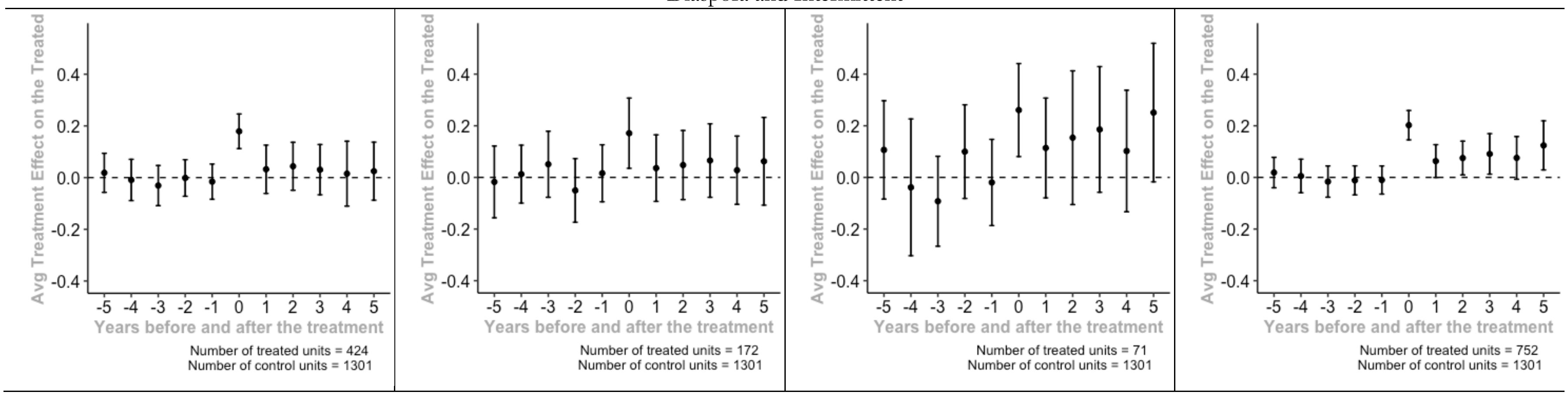
Note: "Treated" refers to non-mobile scientists who, at year zero, collaborate with a mobile. The vertical axis presents the difference between treated and "equivalent" non-treated. Error bars (5 and 95 %) are estimated by a bootstrapping procedure (see Callaway & Sant'Anna (2021)).
Source: Own elaboration based on CvLAC and OpenAlex.

Figure 3 – Average Treatment Effect on the Treated: number of papers with foreign co-authors per year (in logs) – Interaction with different groups of mobile scientists

(1)	(2)	(3)	(4)
Returnees			
One encounter	Two encounters	Three encounters	All encounters



Diaspora and Intermittent



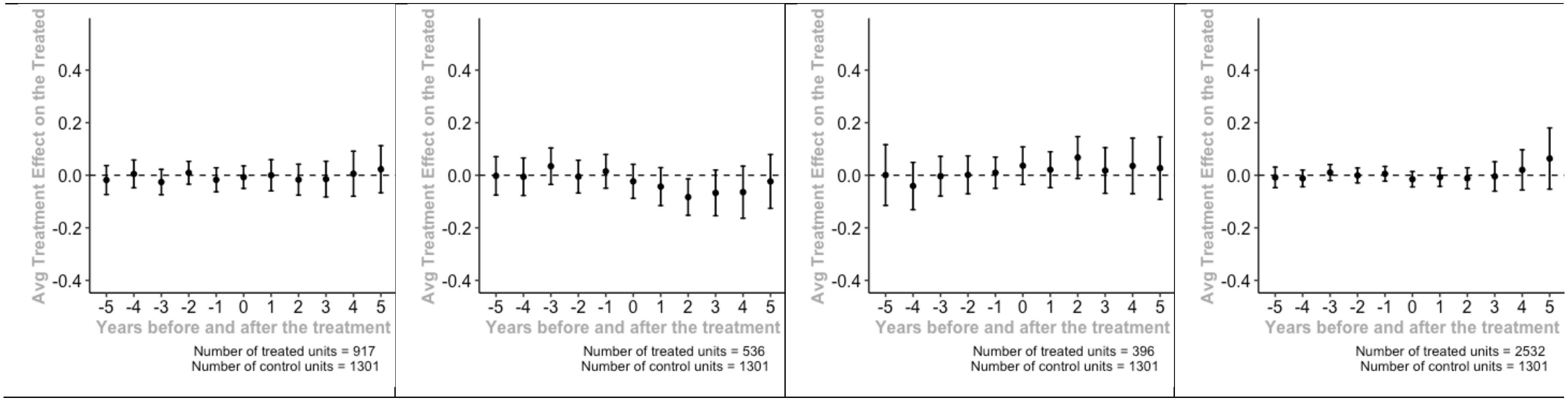
Note: "Treated" refers to non-mobile scientists who, at year zero, collaborate with a mobile. The vertical axis presents the difference between treated and "equivalent" non-treated. Error bars (5 and 95 %) are estimated by a bootstrapping procedure (see Callaway & Sant'Anna (2021)).
Source: Own elaboration based on CvLAC and OpenAlex.

Our findings suggest that any long-lasting effect of interaction depends on repeated interactions with a mobile scientist. If this is indeed the case, then if we restrict attention to publications that do not involve mobile scientists, the effects we have observed should disappear. Figures 4 to 6 show the results when we consider only papers that do not involve mobile scientists. In the three figures, the point estimates drop and become statistically insignificant even after increasing the stock of mobile scientists. Comparing figures 4, 5, and 6 with the previous three figures, the results demonstrate that mobile scientists share their social capital by working with non-mobile scientists. These results align with Burt's (1998, 2000) idea of temporarily borrowing social capital: non-mobile scientists, as network outsiders, gain legitimacy by being connected by mobile scientists. Nevertheless, the foreign connections are not enduring. These results thus indicate a striking mechanism: for non-mobile scientists to interact with foreign scientists, local collaborations must also be fostered. As non-mobile scientists seem not to absorb foreign social capital, they need mobile scientists to reconnect them to foreign scientists.[13]

[13] This is consistent with the results of Müller et al. (2023).

Figure 4 – Average Treatment Effect of the Treated: number of papers with foreign co-authors per year (in logs) – (without papers with mobile)

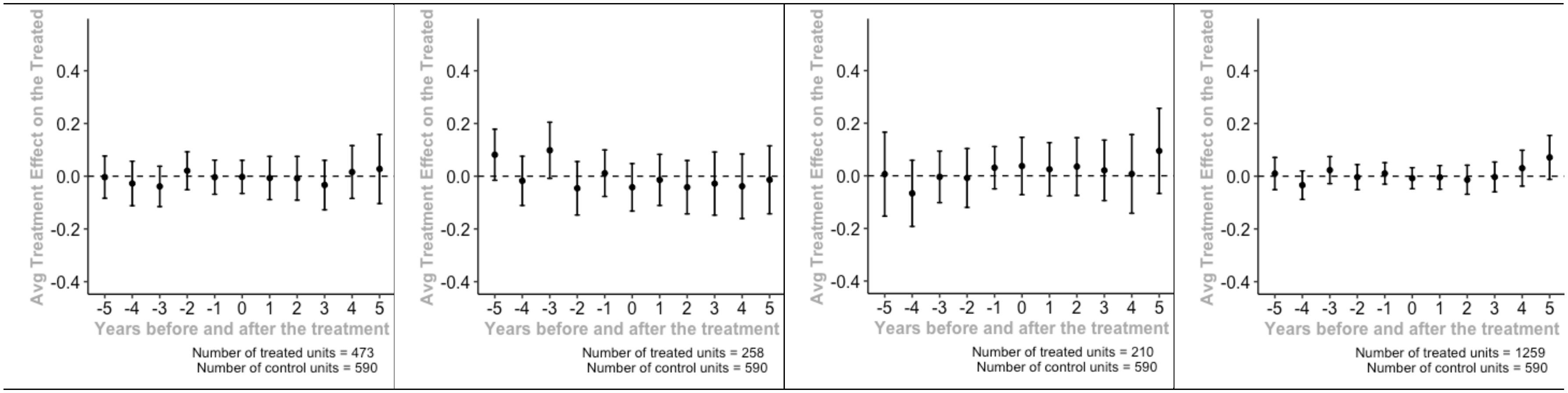
(1)	(2)	(3)	(4)
One encounter	Two encounters	Three encounters	All encounters



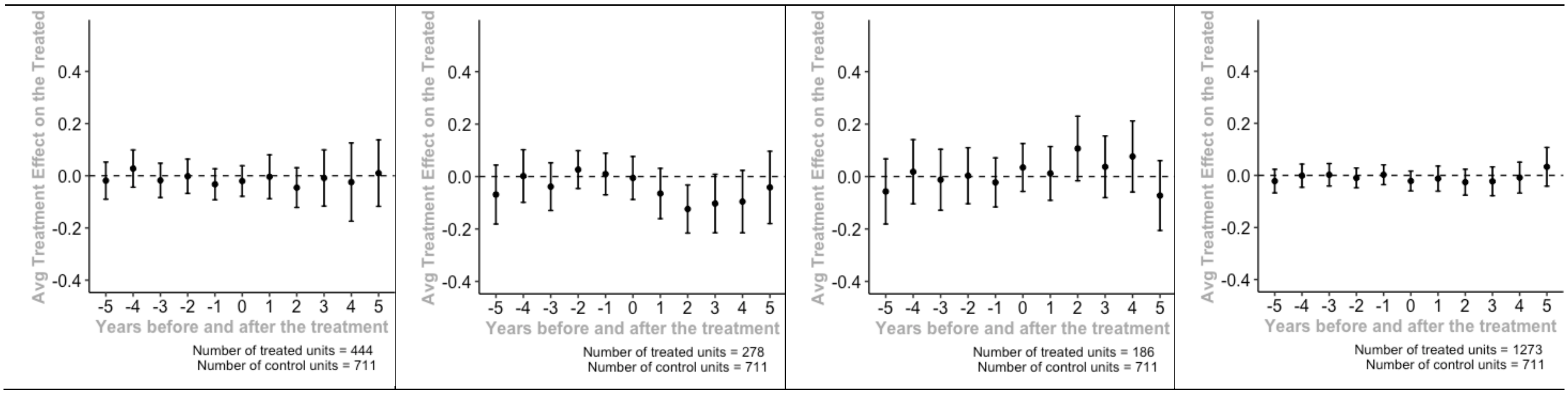
Note: "Treated" refers to non-mobile scientists who, at year zero, collaborate with a mobile. The vertical axis presents the difference between treated and "equivalent" non-treated. Error bars (5 and 95 %) are estimated by a bootstrapping procedure (see Callaway & Sant'Anna (2021)).
Source: Own elaboration based on CvLAC and OpenAlex

Figure 5 – Average Treatment Effect on the Treated: number of papers with foreign co-authors per year (in logs) – STEM and SSH (removing papers with mobile co-authors)

(1)	(2)	(3)	(4)
STEM			
One encounter	Two encounters	Three encounters	All encounters



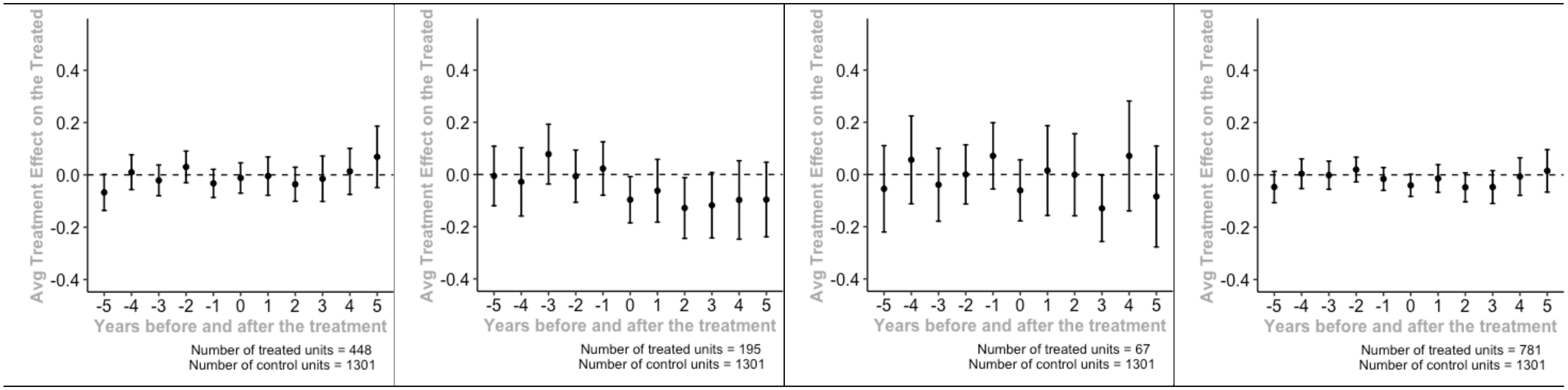
SSH



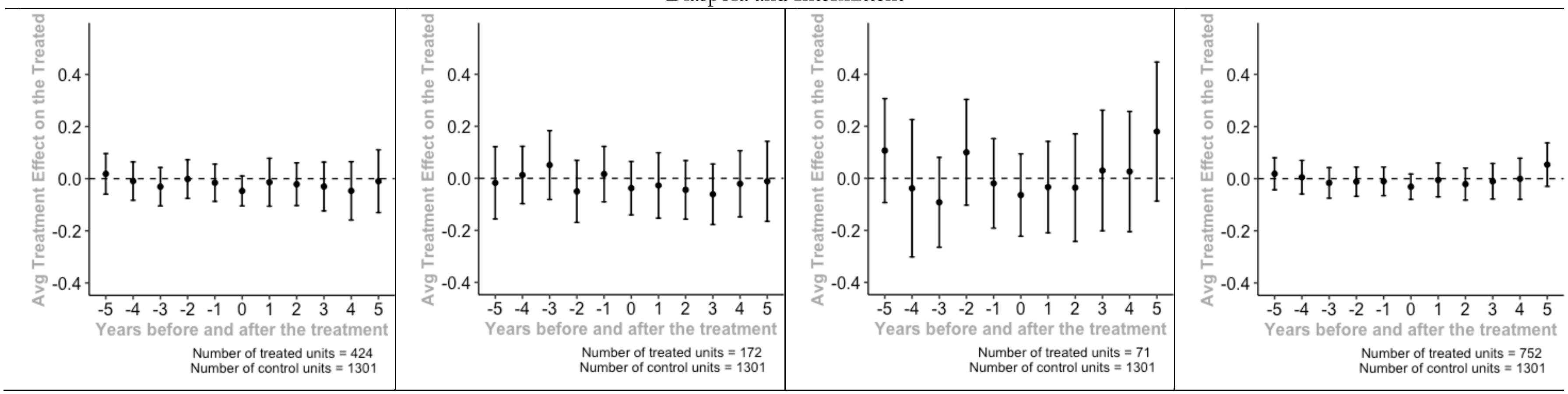
Note: "Treated" refers to non-mobile scientists who, at year zero, collaborate with a mobile. The vertical axis presents the difference between treated and "equivalent" non-treated. Error bars (5 and 95 %) are estimated by a bootstrapping procedure (see Callaway & Sant'Anna (2021)).
Source: Own elaboration based on CvLAC and OpenAlex.

Figure 6 – Average Treatment Effect on the Treated: number of papers with foreign co-authors per year (in logs) – Interaction with different groups of mobile scientists (removing papers with mobile co-authors)

(1)	(2)	(3)	(4)
Returnees			
One encounter	Two encounters	Three encounters	All encounters



Diaspora and Intermittent



Note: "Treated" refers to non-mobile scientists who, at year zero, collaborate with a mobile. The vertical axis presents the difference between treated and "equivalent" non-treated. Error bars (5 and 95 %) are estimated by a bootstrapping procedure (see Callaway & Sant'Anna (2021)).
Source: Own elaboration based on CvLAC and OpenAlex.

6. Discussion and conclusion

In this paper, we matched novel Colombian data from CVs, scholarship lists, and publications to study a research question for which evidence is rare in the literature: Do mobile scientists connect non-mobile scientists from their origin country to foreign scientists abroad? Relying upon different research designs (LPM, panel fixed effects, and difference-in-differences), our results are consistent throughout. Increasing the number of mobile scientists with whom non-mobile scientists interact increases the number (and share) of publications they co-author with foreign scientists. Hence, our overarching message is that mobile scientists play a relevant role in the internationalization of Colombia, a middle-income country in Latin America.

We obtained four main results in our work. First, applying a Linear Probability Model, we found that the average non-mobile scientist is more likely to co-author a paper with a foreign scientist if she has published at least once with a mobile scientist. Second, relying on panel fixed effects, we observed that increasing the stock of mobile scientists is positively associated with the share and number of publications that non-mobile scientists co-author with foreign scientists. The association is higher when controlling for individual fixed effects. Third, STEM and SSH non-mobile scientists tend to benefit similarly from interaction with mobile scientists. Fourth, interacting with diaspora or intermittent scientists benefits the non-mobile scientist more than does interaction with returnees.

The results from the Difference-in-Differences empirical strategy confirm the OLS results. However, they raise a concern for policymakers. The effects of the "internationalization" of non-mobile scientists from a few collaborations are short-lived. The benefits appear to be strongly restricted temporally, precisely to the period in which mobile and non-mobile interact. This suggests that the effects we observe are driven by collaboration among mobile, non-mobile, and foreign scientists as co-authors of the same paper(s). The impact remains statistically significant beyond the first year for non-mobile scientists who diversify their network of mobile scientists (to at least three). Furthermore, there is weak evidence suggesting that the type of mobile scientists with whom non-mobile scientists interact may matter. Non-mobile scientists who collaborate with several diaspora or intermittent scientists appear to co-author with more foreign scientists than do their non-mobile colleagues who collaborate only with returnees;[14] non-mobile scientists who collaborate with several returnee scientists benefit only marginally from the "internationalization." But again, because of the structure of our study, it is difficult to ascribe causality. The differences could arise in the properties of those who interact with diaspora, or in the properties of the diaspora, or in the properties of those interactions. This is something that could be productively explored in the future. Last, by removing publications involving mobile scientists, we find that non-mobile scientists' foreign connections do not remain if mobile scientists are not involved. In other words, mobile scientists act as a bridge in connecting non-mobile scientists with foreign ones, but their continued presence is needed to maintain the collaboration.

Our results relate to the literature in various ways. First, similar to Müller et al. (2023) and Fry (2023), we show that scientists with foreign social capital (in our case, mobile scientists) help non-mobile scientists to engage with foreign scientists. Nonetheless, different from those authors that explore returnees or scientists in the country with foreign social capital, we move beyond to include diaspora and intermittent scientists in our analysis. Adding these latter groups of mobile scientists to our analysis contributes to the debate on brain drain versus brain circulation (Agrawal et al., 2011; Cañibano & Woolley, 2015; Fangmeng, 2016; Meyer, 2001; Saxenian, 2005; Velema, 2012). We show that having scientists abroad may not necessarily harm the sending countries if there is an interaction between diaspora and non-mobile scientists. However, more research is needed to study how long the effects on non-mobile scientists last. Our results suggest that they can be short-term, which raises the question of whether policy actions can be taken to extend their effects.

This paper is not exempt from limitations which may open trajectories for future research. At the forefront, we do not capture visiting periods during the Ph.D. (e.g., doctoral students in Colombia who move abroad for a period but receive their degrees from Colombian institutions). It would be interesting to analyze this type of mobility. Second, by using OpenAlex, we exclude publications in local journals that might be relevant in middle-income countries like Colombia (Chavarro et al., 2017). Finally, further work is needed to investigate the 'so-what' question. For instance, what is the impact

[14] This must be read carefully. For many of the cases differences in the event studies are not statistically significant. However, almost all point estimates of treatment effects show higher values for collaboration with diaspora and intermittent than for collaboration with returnees.

of establishing connections with foreign scientists on the productivity of the non-mobile ones? This is a question we are exploring in work that follows up on this study.

Although we have focused only on one country, several policy recommendations for middle-income countries can be taken from our results, especially for Latin American middle-income countries such as Ecuador, Peru, and Uruguay, which face similar challenges regarding funding and capacity building (Dutrénit et al., 2021). First, sponsoring international scientific mobility schemes facilitates the connection of non-mobile scientists to international research. This means that sponsoring a group of scientists can benefit other scientists in the national system, even though they have not had experience abroad. Second, conditionalities on returning to the sending country after a Ph.D. abroad may need to be more flexible, as diaspora and intermittent scientists are playing a key role in connecting non-mobile scientists to other countries. Nonetheless, it is also imperative that national governments engage their scientific community abroad to increase the role of mobile scientists. Finally, national governments can create funding lines to bring back their mobile scientists abroad for short periods. For instance, these mobile scientists could offer workshops or participate in research. This would facilitate the ongoing collaborations between local non-mobile and mobile scientists that seem to be central in maintaining connections to the knowledge frontier that remain important in the upgrading that middle incoe countries are striving to achieve.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2022). When Should You Adjust Standard Errors for Clustering? *The Quarterly Journal of Economics*, *138*(1), 1–35. https://doi.org/10.1093/qje/qjac038

Agrawal, A., Cockburn, I., & McHale, J. (2006). Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography*, *6*(5), 571–591. https://doi.org/10.1093/jeg/lbl016

Agrawal, A., Kapur, D., McHale, J., & Oettl, A. (2011). Brain drain or brain bank? The impact of skilled emigration on poor-country innovation. *Journal of Urban Economics*, *69*(1), 43–55. https://doi.org/10.1016/j.jue.2010.06.003

Angrist, J. D. (2008). *Mostly Harmless Econometrics : An Empiricist ' s Companion. March*.

Aykac, G. (2021). The value of an overseas research trip. *Scientometrics*, *126*(8), 7097–7122. https://doi.org/10.1007/s11192-021-04052-4

Barnard, H., Cowan, R., & Müller, M. (2012). Global excellence at the expense of local diffusion, or a bridge between two worlds? Research in science and technology in the developing world. *Research Policy*, *41*(4), 756–769. https://doi.org/10.1016/j.respol.2011.12.002

Baruffaldi, S. H., Marino, M., & Visentin, F. (2020). Money to move: The effect on researchers of an international mobility grant. *Research Policy*, *49*(8). https://doi.org/10.1016/j.respol.2020.104077

Bedenlier, S. (2018). 'The Impact of my Work would be Greater here than there': Implications of the International Mobility of Colombian Academics. *Research in Comparative and International Education*, *13*(3), 378–396. https://doi.org/10.1177/1745499918784681
Borusyak, K., Hull, P., & Jaravel, X. (2022). Quasi-Experimental Shift-Share Research Designs. *The Review of Economic Studies*, *89*(1), 181–213. https://doi.org/10.1093/restud/rdab030
Burt, R. S. (1998). THE GENDER OF SOCIAL CAPITAL. *Rationality and Society*, *10*(1), 5–46. https://doi.org/10.1177/104346398010001001
Burt, R. S. (2000). The Network Structure Of Social Capital. *Research in Organizational Behavior*, *22*, 345–423. https://doi.org/10.1016/S0191-3085(00)22009-1
Burt, R. S. (2004). Structural Holes and Good Ideas. *American Journal of Sociology*, *110*(2), 349–399. https://doi.org/10.1086/421787
Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001
Cañibano, C. (2017). Scientific Mobility and Economic Assumptions: From the Allocation of Scientists to the Socioeconomics of Network Transformation. *Science as Culture*, *26*(4), 505–519. https://doi.org/10.1080/09505431.2017.1363173
Cañibano, C., & Woolley, R. (2015). Towards a Socio-Economics of the Brain Drain and Distributed Human Capital. *International Migration*, *53*(1), 115–130. https://doi.org/10.1111/imig.12020
Cao, C., Baas, J., Wagner, C. S., & Jonkers, K. (2020). Returning scientists and the emergence of China's science system. *Science and Public Policy*, *47*(2), 172–183. https://doi.org/10.1093/scipol/scz056
Castellacci, F., & Natera, J. M. (2016). Innovation, absorptive capacity and growth heterogeneity: Development paths in Latin America 1970-2010. *Structural Change and Economic Dynamics*, *37*, 27–42. https://doi.org/10.1016/j.strueco.2015.11.002
Chavarro, D., Tang, P., & Ràfols, I. (2017). Why researchers publish in non-mainstream journals: Training, knowledge bridging, and gap filling. *Research Policy*, *46*(9), 1666–1680. https://doi.org/10.1016/j.respol.2017.08.002
Dahlman, C. (2010). Innovation Strategies in Brazil, China, and India: From Imitation to Deepening Technological Capability in the South. In X. Fu & L. Soete (Eds.), *The Rise of Technological Power in the South* (Vol. 1, pp. 15–48). Palgrave Macmillan.
D'Angelo, C. A., & van Eck, N. J. (2020). Collecting large-scale publication data at the level of individual researchers: a practical proposal for author name disambiguation. *Scientometrics*, *123*(2), 883–907. https://doi.org/10.1007/s11192-020-03410-y
de los Ríos, R., & Santana, P. H. de A. (2001). El espacio virtual de intercambio de información sobre recursos humanos en Ciencia y Tecnología de América Latina y el Caribe Del CV Lattes al CvLAC. *Ciência Da Informação*, *30*(3), 42–47. https://doi.org/10.1590/S0100-19652001000300006
Dutrénit, G., Aguirre-Bastos, C., Puchet, M., & Salazar, M. (2021). Latin America. In *UNESCO science report the race against time for smarter development*. (pp. 201–233).
Edler, J., Fier, H., & Grimpe, C. (2011). International scientist mobility and the locus of knowledge and technology transfer. *Research Policy*, *40*(6), 791–805. https://doi.org/10.1016/j.respol.2011.03.003
Elsner, B., Narciso, G., & Thijssen, J. (2018). Migrant Networks and the Spread of Information. *Oxford Bulletin of Economics and Statistics*, *80*(3), 659–688. https://doi.org/10.1111/obes.12216
Fangmeng, T. (2016). Brain circulation, diaspora and scientific progress: A study of the international migration of Chinese scientists, 1998-2006. *Asian and Pacific Migration Journal*, *25*(3), 296–319. https://doi.org/10.1177/0117196816656637

Franzoni, C., Scellato, G., & Stephan, P. (2014). The mover's advantage: The superior performance of migrant scientists. *Economics Letters*, *122*(1), 89–893. https://doi.org/10.1016/j.econlet.2013.10.040

Freeman, R. B. (2014). Immigration, International Collaboration, and Innovation: Science and Technology Policy in the Global Economy. In *Innovation policy and the economy* (Vol. 15, pp. 153–175).

Fry, C., & Furman, J. L. (2023). Migration and Global Network Formation: Evidence from Female Scientists in Developing Countries. *Organization Science*. https://doi.org/10.1287/orsc.2023.1683

Fry, C. V. (2023). Bridging the Gap: Evidence from the Return Migration of African Scientists. *Organization Science*, *34*(1), 404–432. https://doi.org/10.1287/orsc.2022.1580

Gibson, J., & McKenzie, D. (2014). Scientific mobility and knowledge networks in high emigration countries: Evidence from the Pacific. *Research Policy*, *43*(9), 1486–1495. https://doi.org/10.1016/j.respol.2014.04.005

Giuliani, E., Pietrobelli, C., & Rabellotti, R. (2005). Upgrading in Global Value Chains: Lessons from Latin American Clusters. *World Development*, *33*(4), 549–573. https://doi.org/10.1016/j.worlddev.2005.01.002

Goode, R. B. (1959). Adding to the Stock of Physical and Human Capital. *The American Economic Review*, *49*(2), 147–155. http://www.jstor.org/stable/1816110

Gould, R. V., & Fernandez, R. M. (1989). Structures of Mediation: A Formal Approach to Brokerage in Transaction Networks. *Sociological Methodology*, *19*, 89. https://doi.org/10.2307/270949

Gui, Q., Liu, C., & Du, D. (2019). Globalization of science and international scientific collaboration: A network perspective. *Geoforum*, *105*(June), 1–12. https://doi.org/10.1016/j.geoforum.2019.06.017

Hofman, K., & Kramer, B. (2015). Human resources for research: Building bridges through the Diaspora. *Global Health Action*, *8*(1). https://doi.org/10.3402/gha.v8.29559

Jonkers, K. (2011). Mobility, productivity, gender and career development of Argentinean life scientists. *Research Evaluation*, *20*(5), 411–421. https://doi.org/10.3152/095820211X13176484436177

Jonkers, K., & Cruz-Castro, L. (2013). Research upon return: The effect of international mobility on scientific ties, production and impact. *Research Policy*, *42*(8), 1366–1377. https://doi.org/10.1016/j.respol.2013.05.005

Jonkers, K., & Tijssen, R. (2008). Chinese researchers returning home: Impacts of international mobility on research collaboration and scientific productivity. *Scientometrics*, *77*(2), 309–333. https://doi.org/10.1007/s11192-007-1971-x

Kahn, S., & MacGarvie, M. (2012). The Effects of the Foreign Fulbright Program on Knowledge Creation in Science and Engineering. In J. Lerner & S. Stern (Eds.), *The rate and direction of inventive activity revisited* (pp. 161–197). The University of Chicago Press.

Kahn, S., & MacGarvie, M. (2016). Do return requirements increase international knowledge diffusion? Evidence from the Fulbright program. *Research Policy*, *45*(6), 1304–1322. https://doi.org/10.1016/j.respol.2016.02.002

Kerr, W. R. (2008). *ETHNIC SCIENTIFIC COMMUNITIES AND INTERNATIONAL TECHNOLOGY DIFFUSION*. http://direct.mit.edu/rest/article-pdf/90/3/518/1614623/rest.90.3.518.pdf

Khan, M. S. (2022). Absorptive capacities approaches for investigating national innovation systems in low and middle income countries. *International Journal of Innovation Studies*, *6*(3), 183–195. https://doi.org/10.1016/j.ijis.2022.07.004

Lee, J.-D., Lee, K., Meissner, D., Radosevic, S., & Vonortas, N. (2021). Technology Upgrading and Economic Catch-Up. In *The Challenges of Technology and Economic Catch-up in Emerging Economies* (pp. 1–34). Oxford University PressOxford. https://doi.org/10.1093/oso/9780192896049.003.0001

Lewin, A. Y., & Zhong, X. (2013). The Evolving Diaspora of Talent: A Perspective on Trends and Implications for Sourcing Science and Engineering Work. *Journal of International Management*, *19*(1), 6–13. https://doi.org/10.1016/j.intman.2012.10.001

Liang, W., Gu, J., & Nyland, C. (2022). China's new research evaluation policy: Evidence from economics faculty of Elite Chinese universities. *Research Policy*, *51*(1). https://doi.org/10.1016/j.respol.2021.104407

Losada, M. (2016). Current Challenges in Higher Education and Their Implications for Research and Development in Colombia. In H. Horta, M. Heitor, & J. Salmi (Eds.), *Trends and Challenges in Science and Higher Education* (Vol. 3). Springer International Publishing. https://doi.org/10.1007/978-3-319-20964-7

Meyer, J. B. (2001). Network Approach versus Brain Drain: Lessons from the Diaspora. *International Migration*, *39*(5), 91–110. https://doi.org/10.1111/1468-2435.00173

Miguélez, E. (2018). Inventor diasporas and the internationalization of technology. *World Bank Economic Review*, *32*(1), 41–63. https://doi.org/10.1093/wber/lhw013

Müller, M., Cowan, R., & Barnard, H. (2018). On the value of foreign PhDs in the developing world: Training versus selection effects in the case of South Africa. *Research Policy*, *47*(5), 886–900. https://doi.org/10.1016/j.respol.2018.02.013

Müller, M., Cowan, R., & Barnard, H. (2023). The role of local colleagues in establishing international scientific collaboration: Social capital in emerging science systems. *Industrial and Corporate Change*, *32*(5), 1077–1108. https://doi.org/10.1093/icc/dtad043

Observatorio Colombiano de Ciencia y Tecnología. (2021). *CIENCIA Y TECNOLOGÍA E INNOVACIÓN INDICADORES DE COLOMBIA 2020*. www.ocyt.org.co

Petersen, A. M. (2018). Multiscale impact of researcher mobility. *Journal of the Royal Society Interface*, *15*(146). https://doi.org/10.1098/rsif.2018.0580

Powell, W. W., & Grodal, S. (2005). Networks of Innovators. In J. Fagerberg & D. Mowery (Eds.), *The Oxford Handbook of Innovation*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199286805.003.0003

Radosevic, S., & Yoruk, E. (2018). Technology upgrading of middle income economies: A new approach and results. *Technological Forecasting and Social Change*, *129*, 56–75. https://doi.org/10.1016/j.techfore.2017.12.002

Ratten, V., & Pellegrini, M. M. (2020). Female transnational entrepreneurship and smart specialization policy. *Journal of Small Business & Entrepreneurship*, *32*(6), 545–566. https://doi.org/10.1080/08276331.2019.1616257

Ribeiro, L. C., Rapini, M. S., Silva, L. A., & Albuquerque, E. M. (2018). Growth patterns of the network of international collaboration in science. *Scientometrics*, *114*(1), 159–179. https://doi.org/10.1007/s11192-017-2573-x

Saxenian, A. (2005). From Brain Drain to Brain Circulation: Transnational Communities and Regional Upgrading in India and China. *Studies in Comparative International Development*, *40*(2), 35–61.

Scellato, G., Franzoni, C., & Stephan, P. (2015). Migrant scientists and international networks. *Research Policy*, *44*(1), 108–120. https://doi.org/10.1016/j.respol.2014.07.014

Silva, P. I. (2018). Fundamentos teóricos e práticos para uma diplomacia da inovação. In *Cadernos de Política Exterior* (pp. 307–329). Fundação Alexandre Gusmão.

Singh Chawla, D. (2022). Massive open index of scholarly papers launches. *Nature*. https://doi.org/10.1038/d41586-022-00138-y
Stephan, P. (2006). Wrapping It up in a Person: The Mobility Patterns of New PhDs. *Innovation Policy and the Economy*, *7*, 71–98. http://www.jstor.org/stable/25056190
Trippl, M. (2013). Scientific Mobility and Knowledge Transfer at the Interregional and Intraregional Level. *Regional Studies*, *47*(10), 1653–1667. https://doi.org/10.1080/00343404.2010.549119
Turpin, T., Woolley, R., Marceau, J., & Hill, S. (2008). Conduits of knowledge in the Asia Pacific: Research training, networks and country of work. *Asian Population Studies*, *4*(3), 247–265. https://doi.org/10.1080/17441730802496490
Velema, T. A. (2012). The contingent nature of brain gain and brain circulation: Their foreign context and the impact of return scientists on the scientific community in their country of origin. *Scientometrics*, *93*(3), 893–913. https://doi.org/10.1007/s11192-012-0751-4
Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, web of science, dimensions, crossref, and microsoft academic. *Quantitative Science Studies*, *2*(1), 20–41. https://doi.org/10.1162/qss_a_00112
Wagner, C. S., Park, H. W., & Leydesdorff, L. (2015). The continuing growth of global cooperation networks in research: A conundrum for national governments. *PLoS ONE*, *10*(7), 1–15. https://doi.org/10.1371/journal.pone.0131816
Wooldrige, J. (2012). *Introductory Econometrics: a modern approach* (5th ed.). South-Western, Cengage Learning.
Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, *316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099
Yadav, A., McHale, J., & O'Neill, S. (2023). How does co-authoring with a star affect scientists' productivity? Evidence from small open economies. *Research Policy*, *52*(1), 104660. https://doi.org/10.1016/j.respol.2022.104660
Yang, X., Cai, X., & Li, T. (2022). Peer effects of the young returnee scientists: Evidence from the state key laboratories in China. *Science and Public Policy*. https://doi.org/10.1093/scipol/scac023

Websites accessed

SCImago. (2021). SCImago Journal & Country Rank.

https://www.scimagojr.com/countryrank.php?region=Latin%20America&year=2021

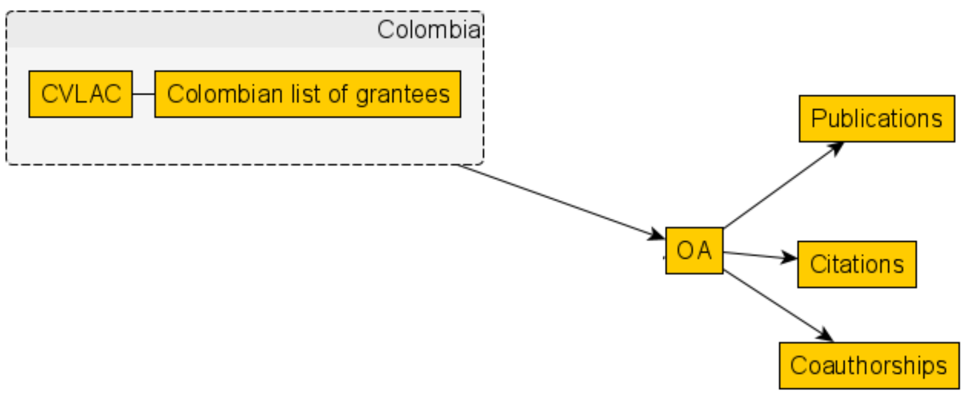
Ministerio de Educación Nacional. (n.d.). Sistema Nacional de Información de la Educación

Superior. https://snies.mineducacion.gov.co/portal/

Appendix 1 – Database construction procedure

To assemble our database for analysis, we integrated CvLAC, lists of PhD abroad scholarship grantees, and OpenAlex. CvLAC is the official Colombian database of researchers, managed by the Ministry of Science. This database incorporates education history, employment history, projects, and self-reported publications. As some researchers who do their Ph.D. abroad and do not go back to their country of origin are unlikely to fill in their CVs in CvLAC, we also identified grantees of PhD scholarships to study abroad from the Ministry of Science, Colfuturo, and Fulbright in Colombia. These were matched or added to CvLAC using fuzzy matching of their names and manual checks. Finally, OpenAlex (OA) is a global bibliographic database covering 211 million research documents. We use OA instead of self-reported publications in CvLAC to validate the information in the CVs with an external source and to identify the network of co-authorships that is absent in CvLAC and the lists of grantees. The following diagram shows an overview of our procedure:

Figure 7 - Overview of databases used for our analysis

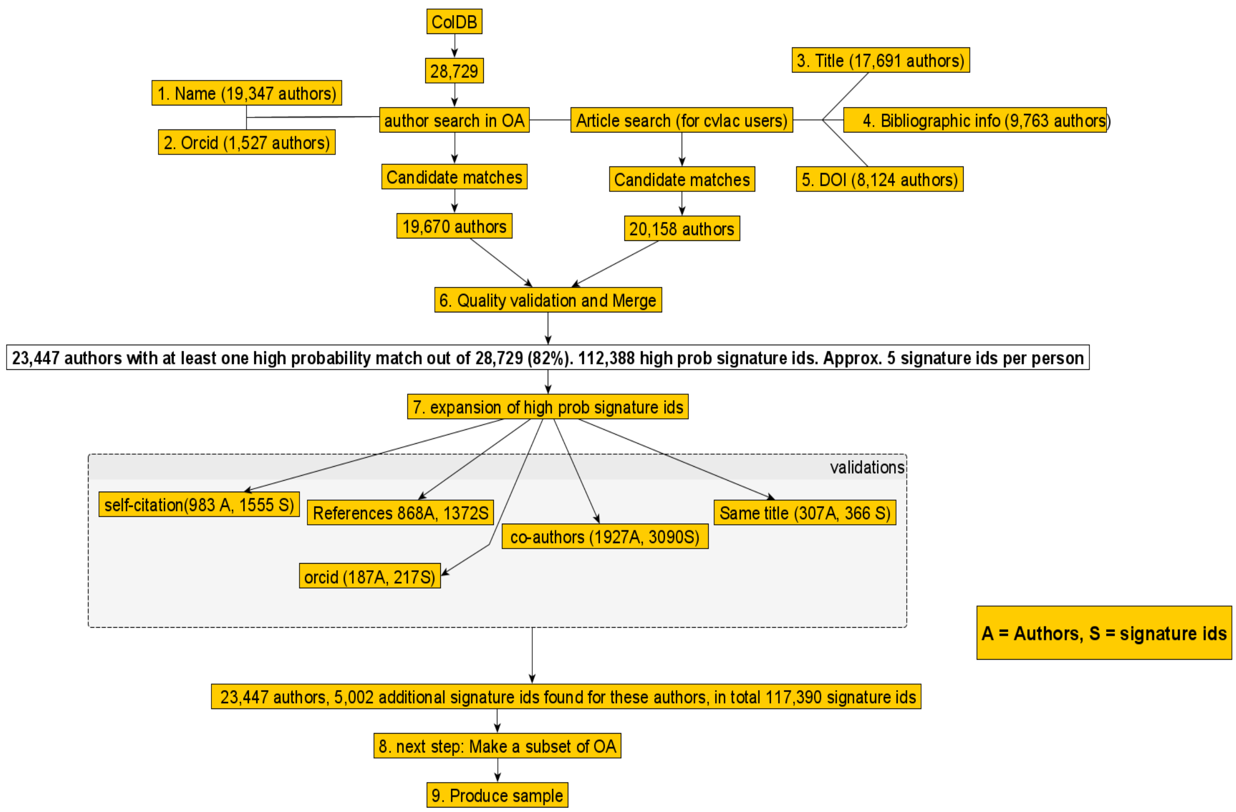


Our procedure aimed to correctly identify each person in our database of Colombian Researchers (ColDB) with a person in OpenAlex (OA). ColDB comprises Colombian Ph.D. holders, PhD students (with or without scholarship), and non-PhDs with at least three publications in CvLAC. The initial number of persons in ColDB was 28,729. While in ColDB, each person has a unique identifier, OA is more focused on the lists of publications and does not have a reliable way to identify an author. This means an author can have different identifiers even though she is the same person. This poses a significant challenge when trying to match a set of names in ColDB because we cannot rely on OA's identifiers. To circumvent this problem, we followed author name disambiguation strategies. They are:

1. Author name fuzzy match
2. ORCID match
3. Title fuzzy match
4. Bibliographic info match
5. DOI match
6. Selection of matches based on quality
7. An additional search of records for quality matches based on self-citations, common co-authors, ORCIDs, and same title
8. Extract the records from OA
9. Build sample for analysis

The following diagrams show the whole procedure for Colombia.

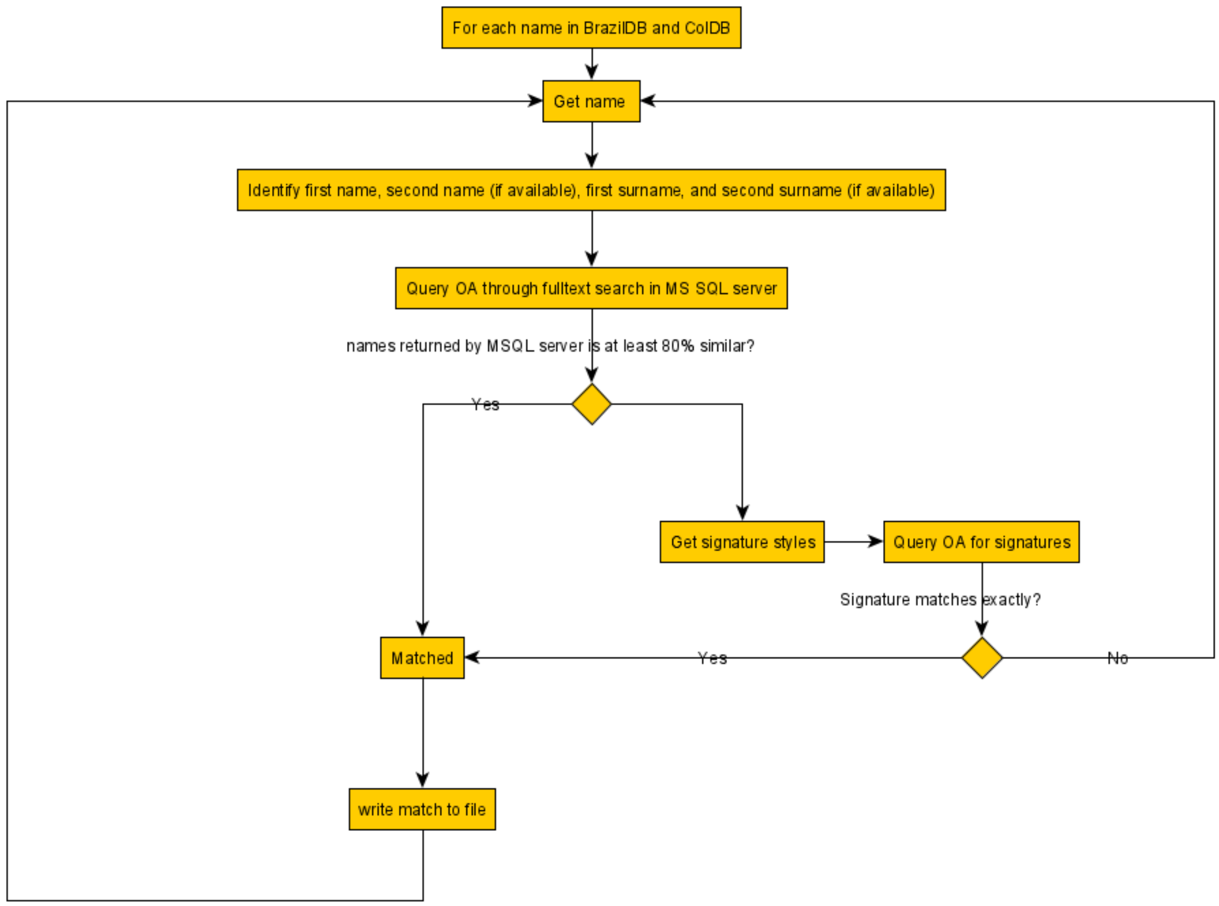
Figure 8 - Database construction procedure for Colombia



1.1 Author name match

ColDB was matched to OA author names. The algorithm used fuzzy matches and exact matches in different ways to make signatures appear. In the fuzzy match case, we set a threshold of 80% or more similarity of full names. In the exact match of signatures, we included 30 possible combinations of name signatures: complete name as it appears in ColDB, and different signature styles such as full first name - full first surname, initial of first name and full surname, and so on. The algorithm is as follows:

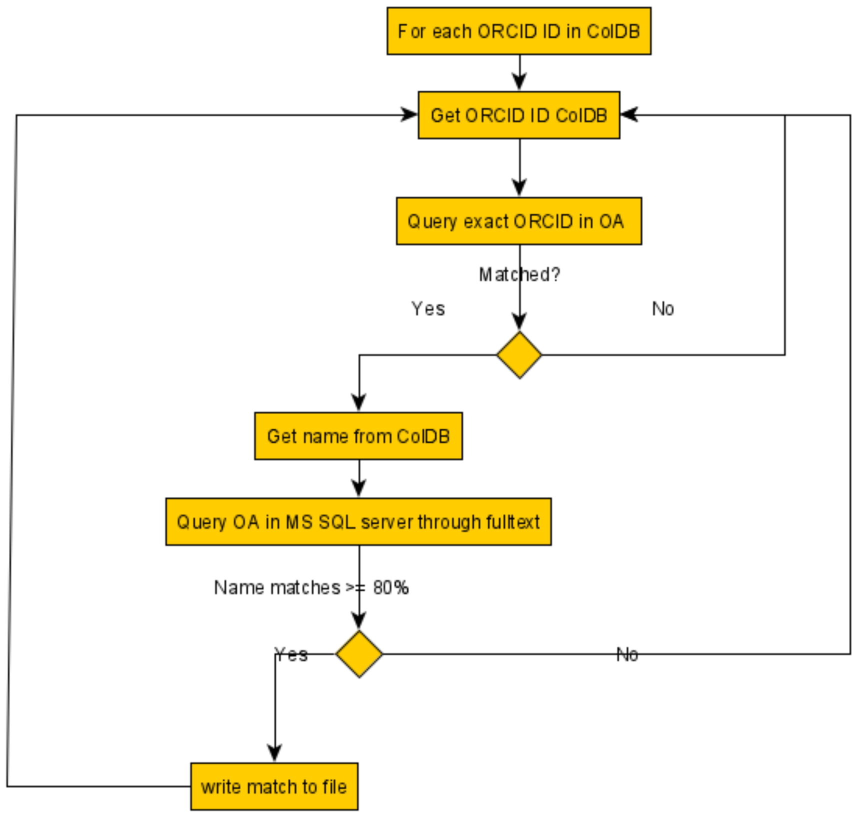
Figure 9 - Algorithm for name match



1.2 ORCID exact match

Records with an ORCID identifier were matched against OA. We performed an exact match of ORCIDs, but because ORCIDs can be either wrongly assigned or mistyped, we also checked that the ORCIDs in ColDB and OA were at least 80% similar. This was only performed for Colombia. The algorithm is as follows:

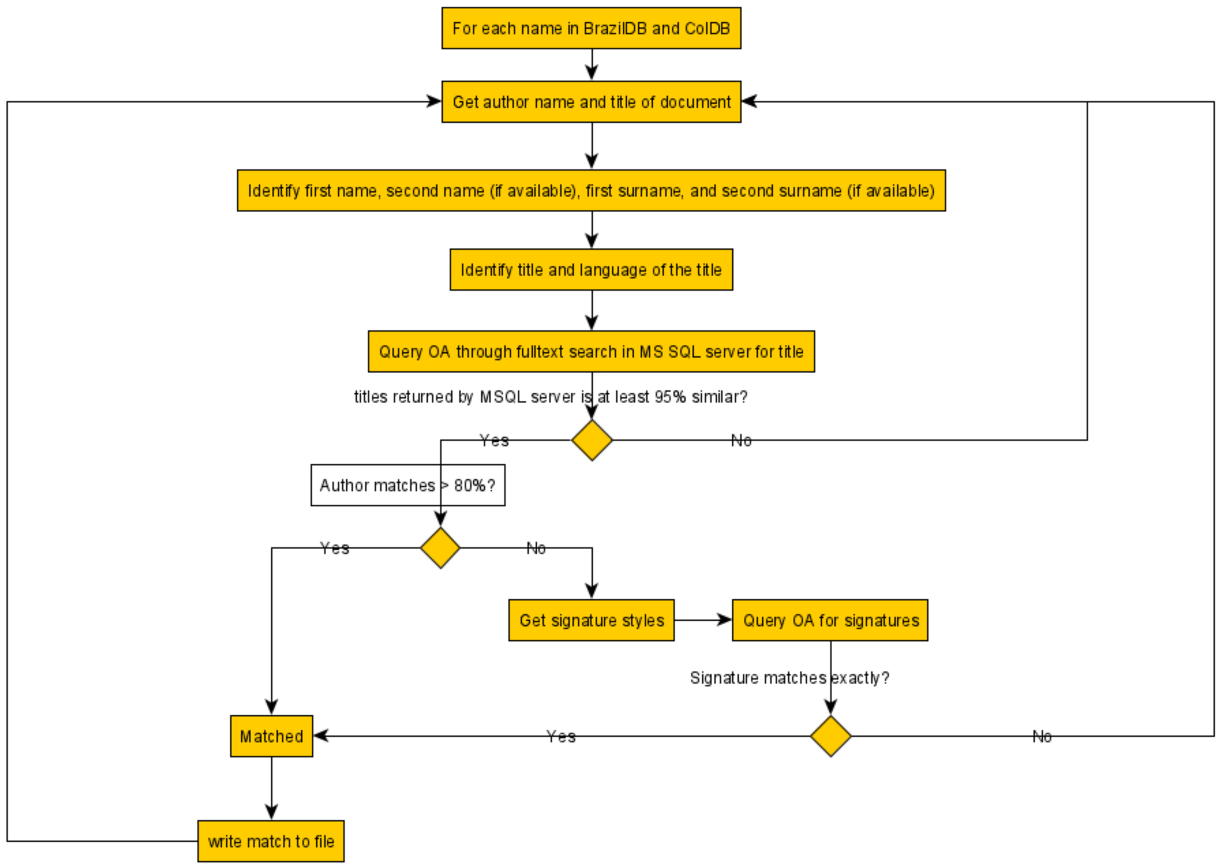
Figure 10 - Algorithm for ORCID match



1.3. Title match

The main idea in using titles to match persons in ColDB with OA is to use self-reported information on publications to disambiguate names in OA. The title of each publication in CvLAC was matched with titles in OA, and author names were checked. Additionally, as publications can be written in different languages, we identified the language of the publication and queried OA using full-text queries in that specific language. After that, we performed a similarity check of titles retrieved, as well as author names and signature styles.

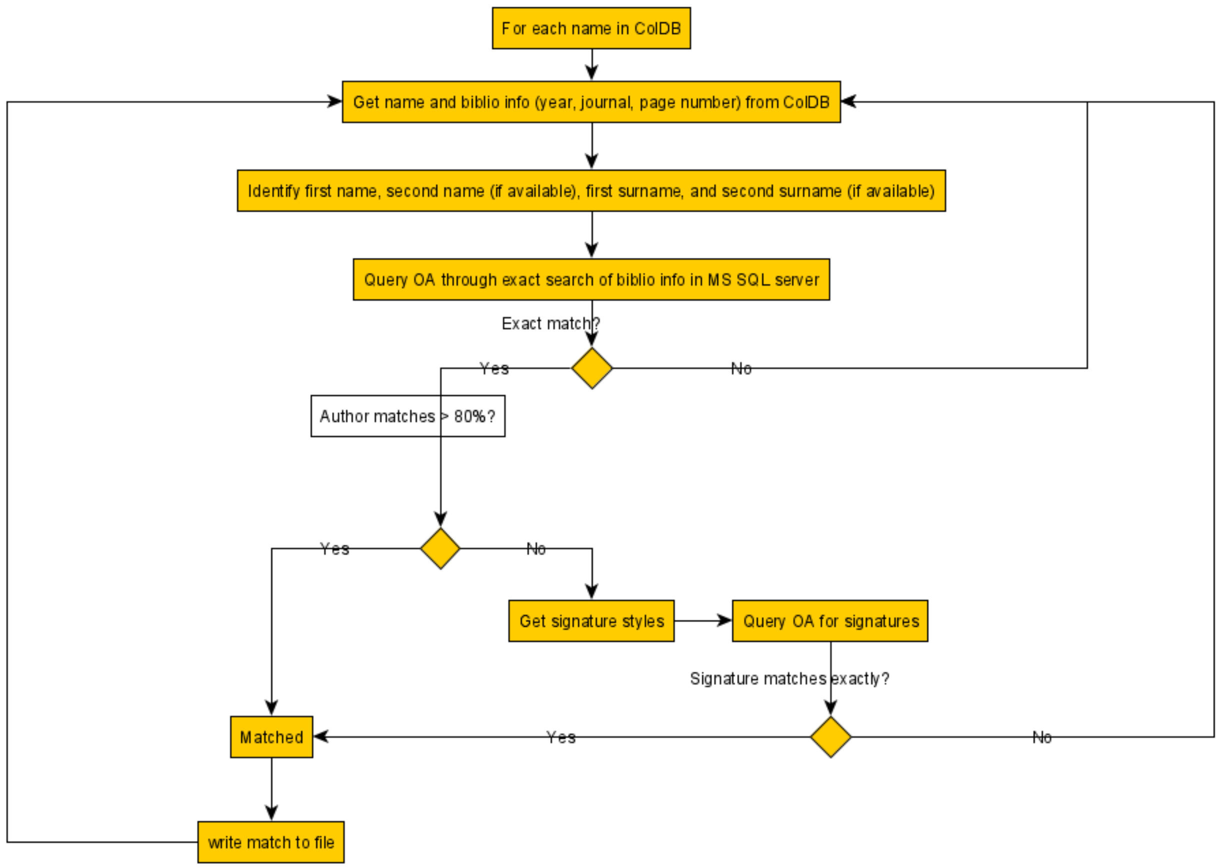
Figure 11 - Algorithm for Title match



1.4 Bibliographic info match

To search for those publications that have titles translated into English in OA and titles that may be short, and for that reason, any typo can produce a significant percentage of dissimilarity when compared to titles in CvLAC, we also searched for bibliographic information that could help us to identify the publications regardless of their titles. We used the journal, year, and beginning page for this purpose. Additionally, we checked that the author in CvLAC was on the list of authors in OA. The algorithm is as follows:

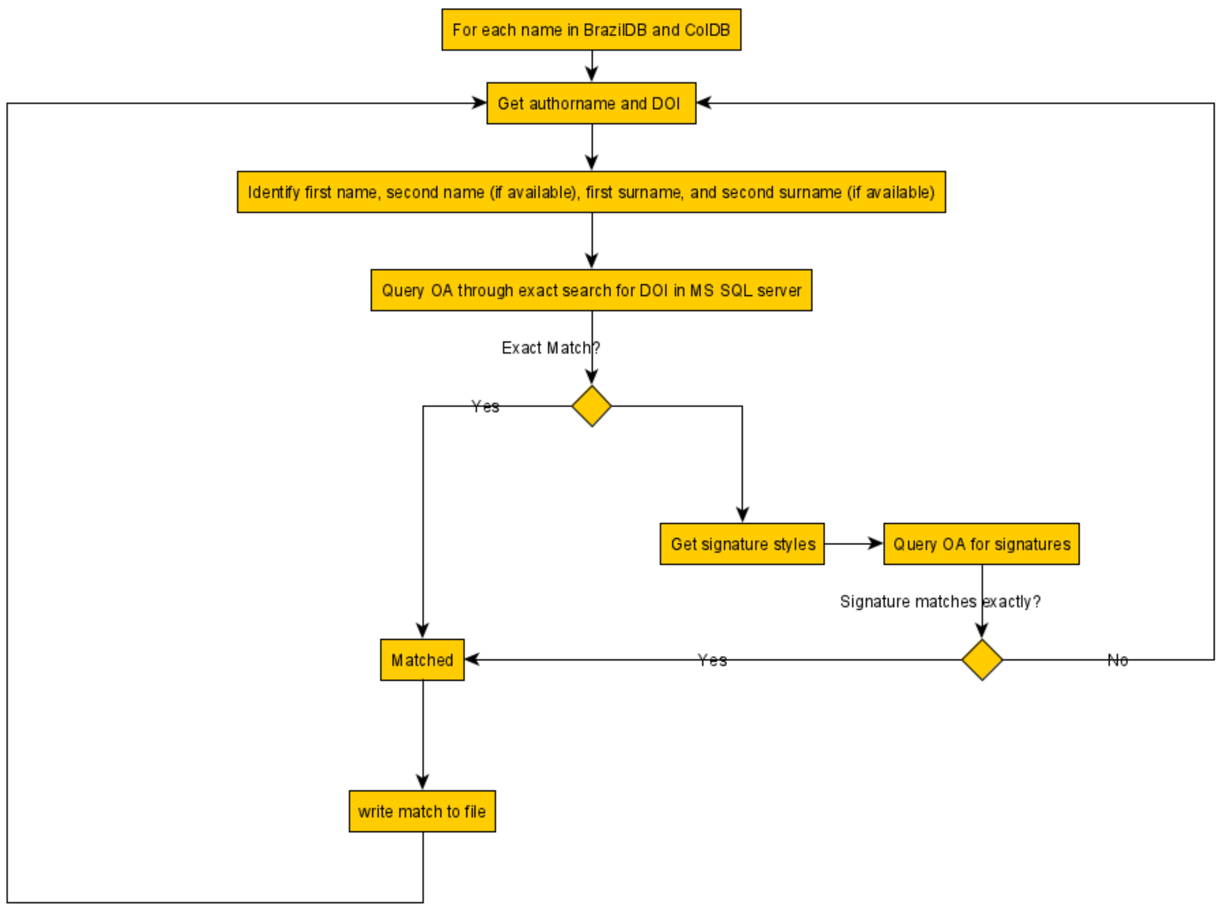
Figure 12 - Algorithm for Bibliographic information match



1.5 DOI match

This match is based on DOI, and additionally, it checks that the author in CvLAC is in the list of authors in OA.

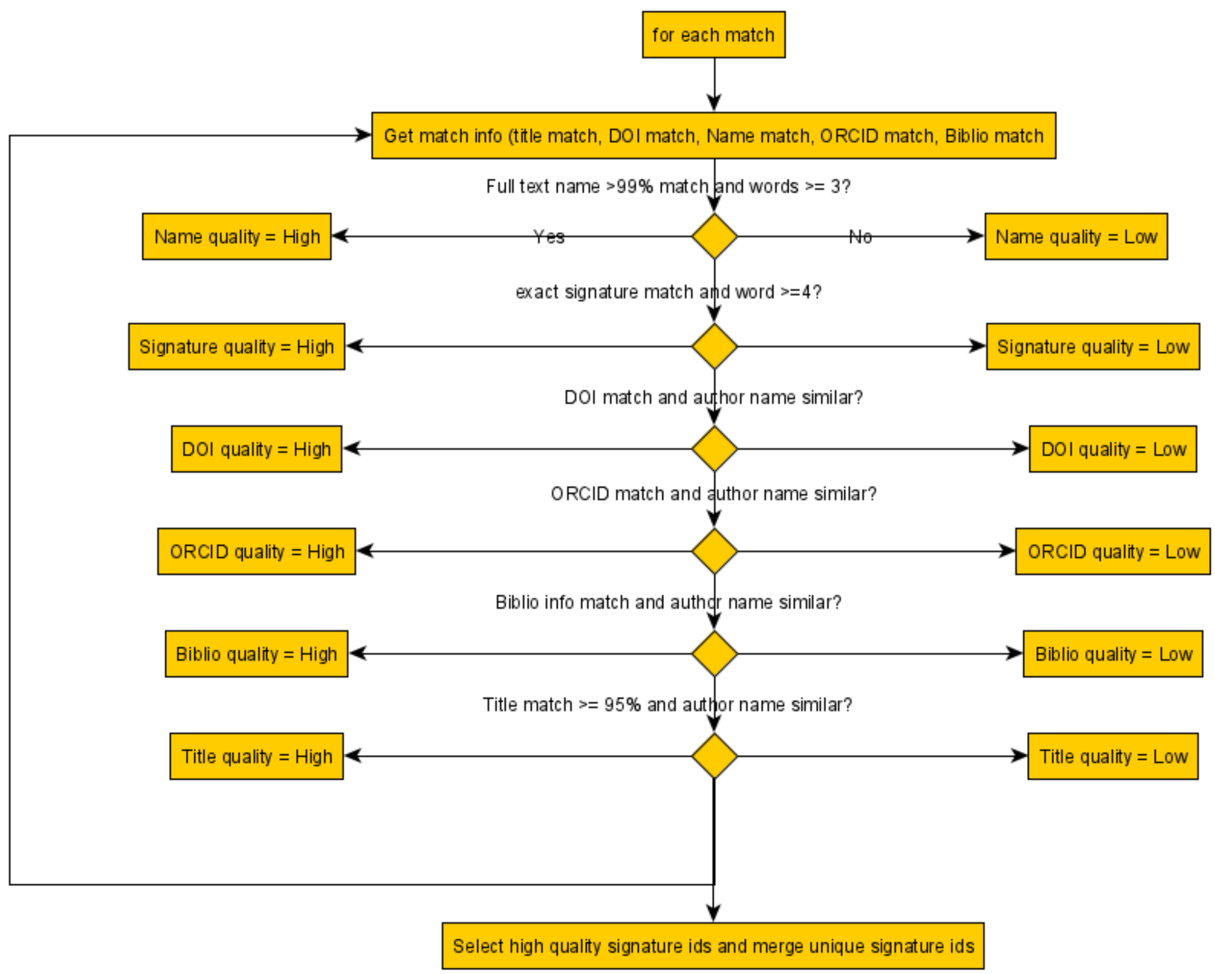
Figure 13 - Algorithm for DOI match



1.6 Selection of matches based on quality

The previous algorithms produced different sets of authors' signature IDs in OA. However, some of these matches can be false positives. We classified each match into high-quality and low-quality to reduce the probability of false positives. For a match to be high quality, matches based on unique identifiers such as DOI and ORCID have the highest quality. Then, for titles, only those that matched more than 95% and had a match with an author were considered high quality. For bibliographic information, those matching exactly which had a match with an author were considered high quality. Finally, for names, we only considered either a 99% or higher match of names composed of at least three names and exact matches of signature styles where the signature was composed of at least four parts. Afterward, we merged all unique signatures found and selected the records in OA for those authors.

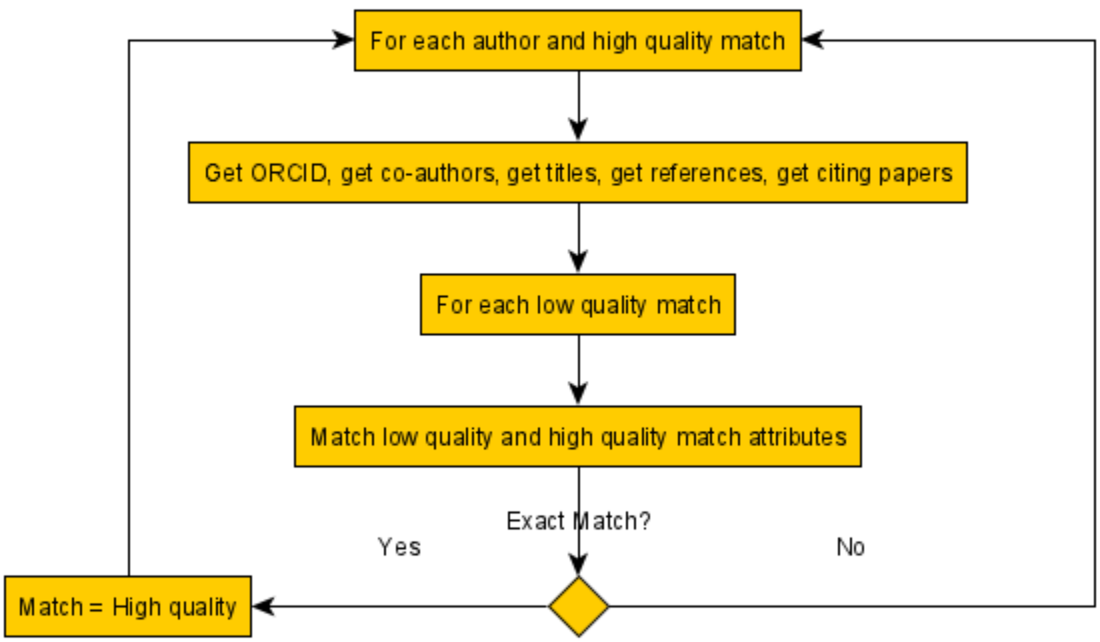
Figure 14 - Algorithm for selection based on quality of matches



1.7 Additional search of records for quality matches based on self-citations, familiar co-authors, ORCIDs, and same title

We used each author's attributes of high-quality matches to identify more signatures in their low-quality matches. The aim was to expand the identification of publications of people based on their high-quality matches to validate those cases in which the names reached the most 99% similarity. For cases where the name is similar but not exact, we used self-citations, titles that match exactly but with an author name match below 99%, ORCIDs found through 100% name matches, and a list of co-authors. The algorithm is as follows:

Figure 15 - Algorithm for expansion of high-quality matches



1.8. Subset of OA

The author's name disambiguation and matching process allowed us to identify the Colombian scientists in OA and collect their bibliometric information. This information was used to build datasets containing academic publications' titles, IDs, co-authors, and citations. Also, to retrieve scientists' fields, we used the bibliometric information on the fields assigned to academic publications to retrieve scientists' fields. We assumed the fields of those scientists with publications in different fields by observing the fields where they had the highest share of publications.

Appendix 2 – Correlation matrix between control variables

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) Lagged average of publications	1.000					
(2) Lagged average of citations	0.5617	1.000				
(3) Lagged stock of foreign co-authors	0.3651	0.6702	1.000			
(4) Lagged average number of co-authors per paper	-0.2119	0.0992	0.5093	1.000		
(5) Years of experience	-0.0288	0.4151	0.4886	0.0061	1.000	
(6) Gender	0.0597	0.0399	0.0506	-0.0051	0.0202	1.000

Appendix 3 - T-test

	Names	Treatment_group	Control_group	T_statistic	p_value
2 years	Stock of publications	1.800	2.891	-4.01	0.000
	Stock of citations	0.950	0.683	-0.652	0.517
3 years	Stock of publications	1.911	2.616	-2.701	0.008
	Stock of citations	0.778	1.221	-0.663	0.509
4 years	Stock of publications	2.015	2.245	-1.114	0.267
	Stock of citations	0.746	0.510	0.840	0.403
5 years	Stock of publications	3.091	3.135	-0.099	0.922
	Stock of citations	1.202	0.875	0.754	0.452
6 years	Stock of publications	3.412	2.931	1.044	0.298
	Stock of citations	1.216	1.238	-0.047	0.963

The UNU-MERIT WORKING Paper Series

2024-01 *The green transformation as a new direction for techno-economic development* by Rasmus Lema and Carlota Perez

2024-02 *Evidence on an endogenous growth model with public R&D* by Thomas H.W. Ziesemer

2024-03 *Higher Educational Systems and E-resilience* by Mindel van de Laar, Julieta Marotta and Lisa de Graaf

2024-04 *Estimating the wage premia of refugee immigrants: Lessons from Sweden* by Christopher F. Baum, Hans Lööf, Andreas Stephan and Klaus F. Zimmermann

2024-05 *Green window of opportunity through global value chains of critical minerals: An empirical test for refining copper and lithium industries* By Jorge Valverde-Carbonell and Alejandro Micco A.

2024-06 *Migration and Consumption* by Roberta Misuraca and Klaus F. Zimmermann

2024-07 *The power of narratives: How framing refugee migration impacts attitudes and political support towards refugees* by Tobias Hillenbrand, Bruno Martorano, Laura Metzger and Melissa Siegel

2024-08 *Strategies of search and patenting under different IPR regimes* by Robin Cowan, Nicolas Jonard and Ruth Samson

2024-09 *From drains to bridges: The role of internationally mobile PhD students in linking non-mobile with foreign scientists* by Rodrigo Ito, Diego Chavarro, Tommaso Ciarli, Robin Cowan and Fabiana Visentin