Do Women Benefit More Than Men from Coauthoring with a Prominent Researcher?

Daniela Aguirre-Guerrero*

M. Ehekatzin García-Guzmán†

Aurora A. Ramírez-Álvarez‡

Diana Terrazas-Santamaría§

July 2024

Abstract: In this paper, we investigate whether the careers of junior researchers show a differentiated gender effect of collaboration with a prominent researcher within five years of their first publication. We use a sample from the Mexican National System of Researchers (SNI; Sistema Nacional de Investigadores) and a staggered difference-in-differences (DID) imputation approach proposed by Borusyak et al. (2024) to compare similar junior collaborators and

*Universidad Autónoma Metropolitana, Unidad Cuajimalpa, Department of Applied Mathematics and Systems. Current address: Avenida Vasco de Quiroga 4871, Santa Fe Cuajimalpa, Cuajimalpa, Mexico City 05348, Mexico. ORCID ID: 0000-0003-4945-7258. daguirre@cua.uam.mx.

†El Colegio de México, Center for Economic Studies. Carretera Picacho Ajusco 20, Ampliación Fuentes del Pedregal, Tlalpan, Mexico City 14110, Mexico. marcgarcia@colmex.mx.

‡El Colegio de México, Center for Economic Studies. Carretera Picacho Ajusco 20, Ampliación Fuentes del Pedregal, Tlalpan, Mexico City 14110, Mexico. ORCID ID: 0000-0001-8358-2920. aurora.ramirez@colmex.mx.

§El Colegio de México, Center for Economic Studies. Carretera Picacho Ajusco 20, Ampliación Fuentes del Pedregal, Tlalpan, Mexico City 14110, Mexico. ORCID ID: 0000-0002-7236-882X. dterrazas@colmex.mx.

non-collaborators. We find a positive effect on impact (total citations, citations per paper, and the h-index) and on the core number of the coauthorship network (a measure of the quality of connections), both of which are more significant for women than for men. Our results also suggest that collaboration with a prominent researcher allows women in different academic areas to catch up with men.

Keywords: Gender inequality, coauthorship network, h-index, research impact

JEL codes: J16, J24, L14, O31, O32.

1 Introduction

There are well-documented gender disparities in academia, reflected in the underrepresentation of women in most disciplines (Huang et al., 2020), their productivity (Huang et al., 2020; Larivière et al., 2013; Beaudry and Larivière, 2016), and their impact indicators (Huang et al., 2020; Dorantes-Gilardi et al., 2022; Beaudry and Larivière, 2016). However, few studies have investigated the resources and collaborators that can help to improve the impact of papers published by women (Beaudry and Larivière, 2016).

This paper focuses on two main questions regarding researchers in engineering, economics, and health sciences in Mexico. First, we analyze whether coauthorship with a prominent researcher in the early stage of a researcher's career affects their productivity, research impact, and network connections. Second, we investigate whether the effect of such coauthorship is differentiated by gender.

We are interested in the effect of collaboration with a prominent researcher; we use the h-index, the maximum value of h such that an author has published at least h papers that have each been cited at least h times, as a proxy for the researcher's prominence. Bornmann and Daniel (2007) argue that the h-index benefits researchers who publish continuously throughout their academic career with consistent above-average impact. However, it should be emphasized that the h-index will not be high unless many of the author's papers are cited, not just their "big hits." It thus provides a different view than a simple count of an author's total citations.

To measure these effects, we build a dataset encompassing Scopus metadata of scientific papers and citation counts for researchers in the Mexican National System of Researchers (SNI; Sistema Nacional de Investigadores) and their coauthors. We use the 2021 SNI registry[1] and focus on three disciplines: engineering, economics, and health sciences. Our data corresponds to 13,634 researchers, of whom 19% are SNI researchers and 81% are their coauthors.

Next, we estimate the causal effect of collaboration with a prominent re-

[1] `https://conacyt.mx/sistema-nacional-de-investigadores/archivo-historico/`.

searcher on a researcher's productivity, impact, and network connections, using the staggered difference-in-differences (DID) imputation approach proposed by Borusyak et al. (2024). This approach estimates a regression of the outcome for the not-yet-treated group and imputes the predicted outputs to the treated observations to form the treatment effect as a weighted sum of the individual effects. Our treatment is collaboration within the first five years of a researcher's career with a prominent researcher. The latter defined as one in the top 2% of our sample's coauthor h-index distribution.

For women, the results indicate that collaborating with a prominent researcher increases their number of citations by 50.8% in the year of the collaboration, with an average effect of 40% per year. There is a null effect on the number of papers. Both of these effects translate into a positive effect on the researcher's h-index. The h-index increases by about 16% in that same year, which lasts until the seventh year after the event, resulting in an effect of 23% per year. The results suggest that the effect on the core number (a measure of the quality of connections) for female researchers is about 9 units in the year of the collaboration, representing an increase of more than 12 units in her core number.

The effects seem smaller and less durable for men than for women. Collaborating with a prominent researcher increases a male researcher's number of citations by 32.4% in the year of the collaboration and reaches 78.6% in the second year after the event. There is a null effect on the number of papers. However, there are positive effects on their number of citations per paper and their h-index, which increase an average of about 12% and 8% per year, respectively. Although these effects are economically significant, they represent only 30% and 34% of the average yearly effects for treated female researchers. Finally, there are also positive effects on the male researcher's core number: in the event year, there is an increase of 2.7, which lasts until the eighth year after the event, resulting in an effect of about 4 units per year. This effect is smaller than the effect found for female researchers. In sum, our results suggest that collaboration with a prominent researcher allows female researchers to catch up

with men and, in some cases, surpass them.

Our study is closely related to those described in two previous papers. Li et al. (2019) show that in four scientific disciplines, early coauthorship with established, highly-cited researchers has a positive effect on the careers of junior researchers in the years following the collaboration on the number of citations and the probability of being highly-cited scientists themselves. Dorantes-Gilardi et al. (2023) investigate whether there is a gender effect on the careers of junior researchers in economics who collaborate with a highly-cited author within five years of the coauthored publication. Using a matching model that controls for coauthorship network measurements to compare similar junior collaborators and non-collaborators, they find a positive effect on the impact of all junior collaborators, but there is no statistically significant difference between men and women. Our analysis includes some crucial differences relative to these two studies. We take advantage of the longitudinal nature of our data to implement an event-study approach, which allows us to track the evolution of the effects of coauthorship on outcomes. In addition, we examine the effects of coauthorship on a measure of the quality of network connections (the core number). We go beyond Li et al. (2019) in looking for a gender effect, and beyond the exclusive focus on the field of economics of Dorantes-Gilardi et al. (2023).

Our paper is structured as follows. In the first section, we explain our metrics: the importance of the h-index as a measure of researchers' prominence and the core number as a measure of the quality of connections. The second section describes the data. The next section describes our methodological approach, followed by a section presenting the key results of our paper. A subsequent section provides robustness checks, and finally, we offer some concluding remarks.

2 Metrics

Our study considers the following metrics to analyze the productivity, impact, and collaboration of a researcher:

- The researcher's productivity, expressed by the number of papers they have published.
- The researcher's impact, measured with the number of citations per paper, total number of citations, and h-index.
- The researcher's collaboration, measured as the number of coauthors and their core number in the coauthorship network.

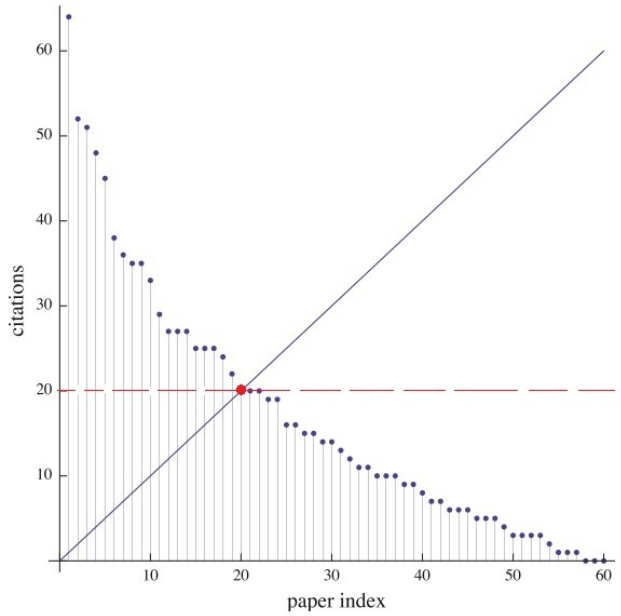
Among these metrics, the h-index and core number are particularly relevant to our study. The h-index estimates a researcher's impact and prominence, while the core number (a metric derived from the coauthorship network) can be interpreted as a measure of the quality of connections, both direct and indirect. The following subsections explain these metrics in detail.

2.1 H-index As a Measure of Impact and Prominence

Hirsch (2005) proposes an index to quantify a researcher's scientific research output, the now well-known Hirsch index, or h-index, as a way to combine the number of publications and the number of citations into a single indicator. In this way, if a researcher has N_p papers published, they have an index h if h of their N_p papers have at least h citations and each of the other $(N_p - h)$ papers has h citations or fewer. The h-index as a measure of a researcher's impact relies on its simplicity in computation and its straightforward interpretation.

Consider Figure 1, which provides an example of a researcher with an h-index of 20. As Bartneck and Kokkelmans (2011) explain, we can list their publications in decreasing order of the number of citations, where the papers with the most citations are at the left of the horizontal axis. The diagonal line represents the number of papers equal to the number of citations, where 20 papers have at least 20 citations each. Note that the h-index is where the diagonal line intersects with the citation profile and that once the highly cited papers are in the top group, the number of citations they receive in excess of h is unimportant.

Figure 1: Citation Profile of an Author with h-index of 20



Source: Bartneck and Kokkelmans (2011).

Given the way the h-index is calculated, its magnitude is correlated with the length of a researcher's academic career because the total numbers of publications and citations rise over time (Costas and Bordons, 2007). A high h-index cannot be achieved with one "big hit" that produces a large number of citations or by publishing many papers that receive smaller numbers.

Although the significance of the h-index is limited in reducing a researcher's entire career to a single number (see, for example, Costas and Bordons, 2007; Bartneck and Kokkelmans, 2011; Koltun and Hafner, 2021), we argue that it is a good proxy for a researcher's prominence since it points to the number of papers and the broad impact of their work throughout their academic career; it requires that a researcher publish consistently and that at least a good portion of their publications receive at least several citations. We thus expect older researchers with longer publishing careers to have higher h-indexes than junior researchers.

Authors in the highest part of the h-index distribution are highly recognizable and, for the most part, highly regarded members within a given discipline. For instance, consider two authors with the same h-index of 41, as reported in Scopus. One is Claudia D. Goldin, winner of the Nobel Prize in Economics in 2023, with 9,000 total citations and 71 papers [2]. The other is Avinash K. Dixit, with 13,771 citations and 108 papers.[3] Despite having different numbers of total citations and papers, it is undeniable that both are highly regarded and have contributed considerably to their discipline.

A junior researcher might benefit from collaboration with a prominent colleague through different channels. Krauss et al. (2023) find that young researchers who ended up being prominent—the 100 researchers with the highest h-indexes in a given discipline, who coauthored articles with prominent researchers in the early stage of their careers—received a great benefit: the papers they coauthored with prominent researchers received more than twice the number of citations than those coauthored with non-prominent researchers.

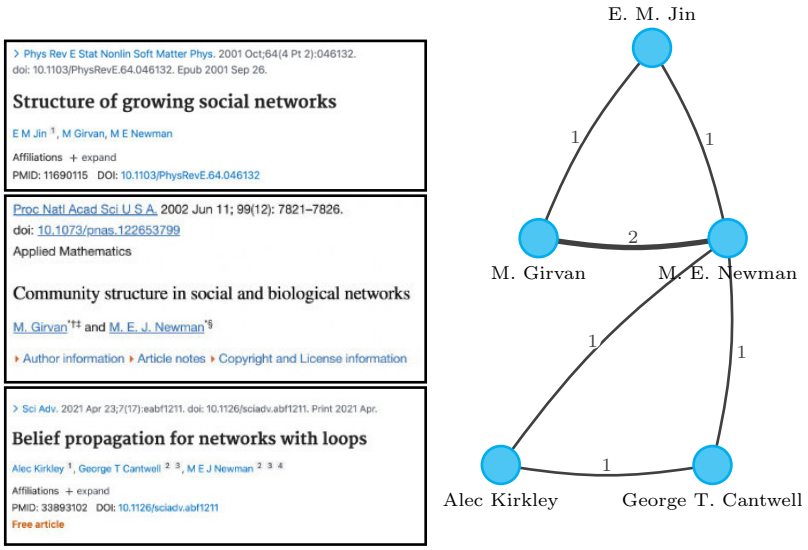
Access to different resources, such as knowledge, information, financial resources, or equipment, can greatly benefit early career researchers. As Ebadi and Schiffauerova (2015) explain, better funding has proven helpful in achieving greater scientific performance; however, being well-connected within denser and more cohesive communities in a coauthorship network greatly enhances future funding chances Jaramillo et al. (2023).

Furthermore, Li et al. (2022) argue that a collaboration network, such as one of coauthors, serves as social capital to those involved; differences in social networks can, in large part, explain observed gaps in prominence and career outcomes, particularly gender gaps and the barriers to elite research environments (Li et al., 2022).

[2] https://www.scopus.com/authid/detail.uri?authorId=6701799144
[3] https://www.scopus.com/authid/detail.uri?authorId=55824766400

2.2 The Core Number As a Measure of Connection Quality

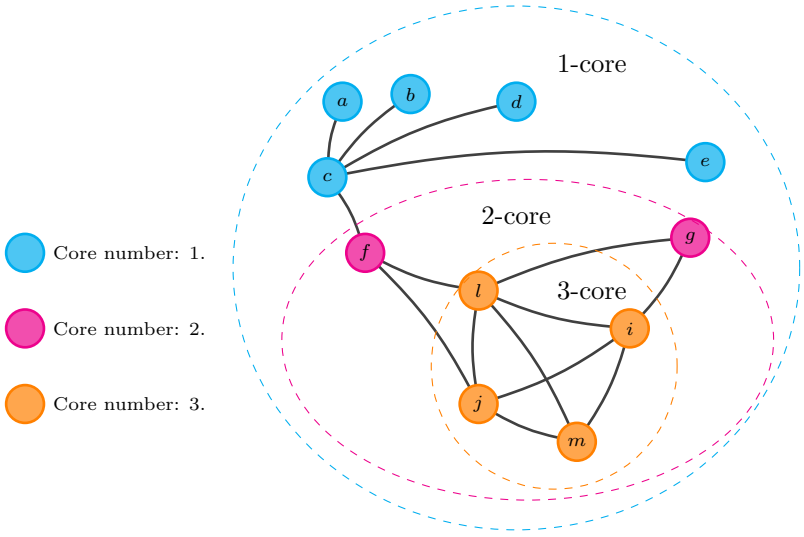
Figure 2: Coauthorship Network Built from Three Papers.



Coauthorship networks have become a standard model for studying academic collaboration and connectivity dynamics in academic research (Newman, 2004). Formally, a coauthorship network is defined by a graph $G = (V, E)$, where the set of vertices (or nodes) V represents researchers and the set of edges (or links) E represents coauthorship relationships. If two researchers coauthor a paper, they are connected through a weighted link. The weight of the edges indicates the number of coauthored papers. Figure 2 shows an example of a coauthorship network built from three papers.

We map papers by year and discipline into a coauthorship network. Our goal is to analyze the evolution of connectivity among researchers. In networks, a standard measure of node connectivity is the degree of a node, which represents the number of its incident links. In coauthorship networks, the degree of a

Figure 3: An example of k-core decomposition.



researcher is equal to the number of their coauthors. Another important metric of node connectivity is the core number, which captures how well a researcher's coauthors are connected.

Before defining the core number, it is necessary to define the k-core of a network. The k-core of a network G is the sub-network of G where all nodes have a degree of at least k. The k-core of a coauthorship network is thus a sub-network where every researcher has at least k coauthors. Figure 3 shows an example of the k-core decomposition of a network, which creates three nested k-core sub-networks:

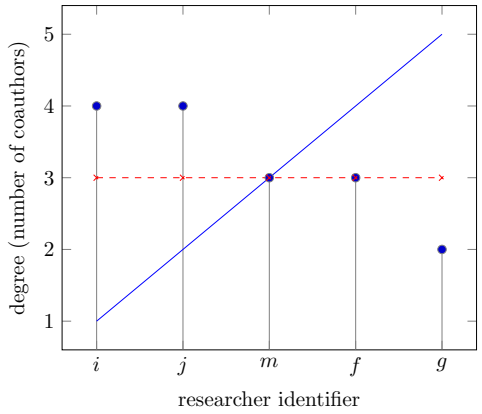
- The 3-core network is the sub-network formed by nodes with a minimum degree of 3, that is, nodes l, i, j, and m.
- The 2-core network is the sub-network formed by nodes with a minimum degree of 2, that is, nodes f and g.
- The 1-core network is the sub-network formed by nodes with a minimum degree of 1, that is, the whole network.

A researcher has core number k if they belong to a k-core but not to a $k+1$-core, which means that the researcher and their coauthors have at least k coauthors. In Figure 3, orange nodes have a core number of 3, pink nodes have a core number of 2, and blue nodes have a core number of 1. Note that a researcher who has a higher k-core value (i.e., who has a higher core number) intrinsically belongs to a more tightly-knit part of the coauthorship network, and the researchers within that sub-network can potentially control the flow of knowledge since they can influence the formation and development of scientific teams and research. A peripheral researcher, however, has few or even no collaborations (Ebadi and Schiffauerova, 2015). Furthermore, Li et al. (2022); Jaramillo et al. (2023) find that observed differences between male and female researchers, mainly in scholarly productivity and citations, are likely caused by gender differences in the structure and composition of researcher authorship networks and their k-core decomposition.

It should be noted that the core number can be computed analogously to the h-index. If a researcher has N_c coauthors, they have a core k if k of their N_c coauthors have at least k coauthors. Consider the researcher l in the coauthorship network shown in Figure 3, whose coauthors are i, j, m, f, and g, and who has a degree of 5. Figure 4 is a plot of node degree for the coauthors of researcher l, with the node degree values in decreasing order. The diagonal line represents the number of researchers equal to the number of their coauthors, where 3 researchers each have a degree of at least 3. The core number of researcher l is 3, where the diagonal line intersects with the plot of node degree.

Although node degree and core number both measure a researcher's connectivity, they are different metrics. While a researcher's degree measures only the number of their coauthors, the core number not only measures their connections, but also how well their coauthors are connected. It can be said that the degree of a researcher measures the *quantity* of their connections, but their core number of a researcher measures the *quality* of their connections and thus the quality of their research group. For example, suppose the network in Figure 3 is a coauthorship network. Note that researcher c has a degree of 5 and a

Figure 4: Degree of the Coauthors of Researcher l from the Coauthorship Network Shown in Figure 3.



core number of 1, while researcher m has a degree of 3 and a core number of 3. Despite researcher m having fewer connections than researcher c, researcher m has higher connection quality than researcher c. Thus, we argue that it is beneficial for a researcher to belong to a higher-level core.

3 Data

We build a dataset using Scopus metadata of papers and citation counts for SNI researchers and their co-authors.[4] We apply techniques for author name disambiguation to find researchers from the 2021 SNI registry in the Scopus database.

We use the 2021 SNI registry, and because of data intensity requirements, we focus on three disciplines: engineering, economics, and health sciences.[5] The percentages of SNI researchers found in the Scopus database are 55% for engineering, 50% for economics, and 68% for health sciences. Each of these disciplines has around a thousand SNI researchers, and the standard practice is to publish their work in indexed journals likely to appear in Scopus.

[4]The data was retrieved from the Scopus database via the Scopus API using the *pybliometerics* library for Python Rose and Kitchin (2019).

[5]We thus have data only for researchers active in this year and not on those who retired before 2021.

Our data includes 13,634 researchers, of whom 19% are SNI researchers and 81% are their coauthors. For each researcher, we retrieve metadata of all their papers through 2021,[6] and a report of citation counts broken down by year. Table 1 shows descriptive statistics for female and male SNI researchers within each area.

Because not all authors in the SNI registry were found in the Scopus database, our retrieved data may not be representative of the Mexican researcher population. To ensure this is not the case, we use a regression analysis to check for some disbalance in the proportion of female researchers, a key factor in our exploration of the gender gap. The results suggest that the proportions of men and women in our data are not statistically different from those in the SNI registry in any of the fields (see Table A.1).[7] Female researchers in all disciplines have published fewer papers, are less cited, and tend to have fewer coauthors (except in economics) than their male counterparts (Table 1). This data allows us to compute the yearly progress of an SNI researcher's h-index and track their annual number of papers, citations, and co-authors.

Table 1: Descriptive Statistics by Gender

	Economics			Engineering			Health Sciences			Total		
	Female	Male	Total	Female	Male	Total	Female	Male	Total	Female	Male	Total
A) Mean												
Citations	7.21	5.99	6.33	8.29	9.16	8.98	29.32	34.63	31.81	23.49	19.46	21.09
Papers	0.96	1.00	0.99	0.97	1.26	1.20	1.58	1.96	1.76	1.41	1.49	1.46
Coauthors	2.49	2.05	2.17	3.37	3.85	3.75	10.21	11.84	10.97	8.22	6.90	7.44
h-index	1.66	1.70	1.69	2.08	2.19	2.17	4.94	5.36	5.13	4.10	3.44	3.71
Core number	5.12	3.80	4.18	7.51	6.87	7.00	16.56	17.60	17.05	13.44	10.37	11.58
Career length	11.80	15.06	14.14	10.71	12.75	12.33	20.34	22.60	21.40	17.93	17.75	17.82
B) Standard deviation												
Citations	18.76	15.23	16.31	13.06	16.54	15.88	65.22	68.81	66.97	57.25	49.06	52.58
Papers	1.58	1.54	1.55	1.26	1.58	1.53	1.99	2.57	2.29	1.88	2.11	2.02
Coauthors	6.64	4.52	5.23	4.31	5.00	4.87	18.56	19.80	19.17	16.48	14.39	15.29
h-index	2.15	2.04	2.07	1.89	2.27	2.20	5.01	5.50	5.25	4.63	4.34	4.47
Core number	8.58	5.93	6.81	4.29	4.17	4.21	15.58	16.66	16.11	14.48	12.97	13.67
Career length	8.08	9.06	8.92	6.40	7.32	7.19	9.77	11.10	10.47	10.03	10.57	10.35
C) Maximum value												
Citations	236.00	236.00	236.00	139.00	211.00	211.00	1702.00	1306.00	1702.00	1702.00	1306.00	1702.00
Papers	18.00	22.00	22.00	14.00	19.00	19.00	26.00	38.00	38.00	26.00	38.00	38.00
Coauthors	135.00	93.00	135.00	29.00	62.00	62.00	427.00	295.00	427.00	427.00	295.00	427.00
h-index	17.00	22.00	22.00	12.00	17.00	17.00	40.00	38.00	40.00	40.00	38.00	40.00
Core number	99.00	92.00	99.00	30.00	49.00	49.00	179.00	181.00	181.00	179.00	181.00	181.00
Career length	50.00	49.00	50.00	28.00	40.00	40.00	50.00	72.00	72.00	50.00	72.00	72.00
D) Frequency	196.00	408.00	604.00	130.00	443.00	573.00	494.00	393.00	887.00	820.00	1244.00	2064.00

Note: 481 outliers were removed from the original sample of 2,545 observations, so the number of SNI researchers analyzed is 2,064.

[6] This metadata includes bibliographic, author, and funding agency information (see `https://dev.elsevier.com/sc_search_views.html`.)

[7] The regression analysis consists of combining the authors in our sample and the authors in the SNI registry and regressing a dummy variable equal to 1 if the researcher is female over a dummy variable equal to 1 if the author is found in the Scopus database.

Importantly, since we have information about each coauthor, we can calculate their h-index, papers, and citations for every year and incorporate them into the econometric analysis. The database characteristics allow us to use the core number as an indicator of the quality of a researcher's connections.

4 Empirical Strategy

In this paper, we estimate the causal effect of collaboration with a prominent researcher on the impact, productivity, and collaboration trends of Mexican researchers, using the staggered DID imputation approach proposed by Borusyak et al. (2024). The recent literature finds that the staggered DID estimator has significant biases arising from group and temporal heterogeneity in treatment effects. To avoid these biases, we follow Borusyak et al. (2024) to estimate treatment effects. Their approach consists of estimating a regression of the outcome for the not-yet-treated group and inputting the predicted outputs to the treated observations to calculate the treatment effect as a weighted sum of these individual effects.

We define the event as a collaboration with a prominent researcher within the first five years, what we refer to as the early stage of a researcher's career. This definition of the early stage is close to that of Bazeley (2003), who defines it as the five years after completing the Ph.D., and it is the same as that of Jin et al. (2020) and Krauss et al. (2023), who define it as the first five years since the first publication. We limit the event to the first five years of the academic career because, as previous literature has pointed out, the conditions in this stage correlate with researchers' future impact and prominence in academia. It is likely that the effects of collaborating with a prominent researcher are more critical if the collaboration occurred at an early stage of their academic career (Bazeley, 2003; Krauss et al., 2023).

We define a prominent researcher based on the maximum coauthors' h-index distribution by year and discipline in our sample. This approach enables us to ensure that the estimated effects result from collaborating with a researcher

with greater seniority and impact due to the very nature of the h-index.[8] Our baseline estimations define a prominent researcher as a coauthor in the top 2% of our sample's maximum h-index distribution for coauthors by area.[9]

The outcomes we measure are the log of the total number of citations, the log of the number of citations per article, the log of the number of papers, the log of the h-index[10], and the core number, using the following specification:

$$Y_{i,t} = \alpha_i + \gamma_t + q'_{i,t}\psi + \sum_{m=-G}^{M} \tau_m T_{i,t-m} + \varepsilon_{i,t} \tag{1}$$

where Y represents the outcome variable of researcher i in year t, q is a vector of time-varying controls, α_i and γ_t represent individual and year fixed effects, respectively, and T is an indicator variable equal to 1 if the researcher i is treated in time t and otherwise equal to 0. G and M represent the window in which the effect is computed, G denoting the years before the event and M denoting the years after the event. In this study, we set $G = 5$ and $M = 10$. Finally, τ is the coefficient of interest and denotes the average treatment on the treated effect (ATT) in the year t after the event.

It is important to recall that the causal interpretations of the coefficients are only valid if the assumption of the treatment's exogeneity is fulfilled. To validate this assumption, we regress the timing of the treatment over the productivity and collaboration characteristics of the treated researchers in the starting year of their academic career using the following equation:[11]

$$Y_i = a + X'\beta + \alpha_i + \gamma_t + \varepsilon_i \tag{2}$$

Y is the event's timing for individual i, and X is a vector of productivity controls

[8] Figure A.1 displays the h-index of treated authors and prominent researchers. Because all observations are above the 45° line, we can be sure that all treated authors have a smaller h-index than the prominent researchers who caused the event.

[9] Hirsch (2005) explains that the h-index distribution is inherent to each discipline because of differences in productivity and citation patterns among disciplines.

[10] Following Criscuolo et al. (2019) and MaCurdy and Pencavel (1986), we add one to the values of the variables before taking logs to avoid excluding observations with a value of zero, which constitute 17-37% of the outcome variable observations.

[11] The number and percentage of treated authors is displayed in table A.2

(including the number of citations and papers) and collaboration characteristics (including the number of coauthors and the core number) in the starting year of the researcher's academic career, and α_i and γ_t represent individual and year fixed effects, respectively. In addition, we include the first year of a researcher's career as a control.[12]

5 Results

In Table 2, we present the results from estimating equation (2) for women in column (1) and for men in column (2). These estimations suggest that the event timing for women positively correlates with the number of coauthors at the beginning of their academic careers. This correlation implies, in turn, that the number of collaborations increases the time before the event occurs. In contrast, the coefficient related to the year in which the academic career starts suggests that in recent years, collaboration with prominent researchers has been more common than in the past. Given these results, we include the number of coauthors (in logarithms) and the starting year of the researcher's career as controls in equation (1).

Figures 5 and 6 display the baseline estimates from estimating equation 1 for female and male researchers, respectively. In both cases, the pre-trend coefficients suggest no pre-event effects, indicating that the impact on the outcome variables is attributed solely to collaborating with a prominent researcher. Furthermore, the estimations highlight that, regardless of the researcher's gender, collaborating with prominent researchers significantly affects almost all impact and productivity measures, a key finding of our study.

For female researchers, the event causes a positive effect that lasts until the ninth year after the event in the case of the number of citations and until the seventh year for the number of citations per article and the h-index. The general ATT effect on the number of citations in the first 10 years after the event is more than 100% per year (see table B.3), although there are important

[12] Following Huang et al. (2020), we define the beginning of a researcher's career as the year of their first publication.

Table 2: Timing of the Event

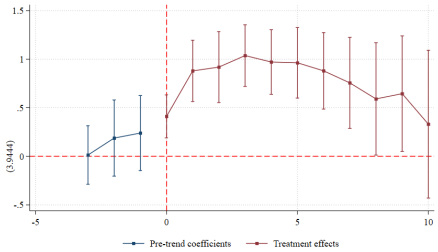
VARIABLES	(1) Female Researchers	(2) Male Researchers
Core Number	-0.103	-0.0312
	(0.0709)	(0.0530)
Coauthors (ln)	1.953**	-0.169
	(0.926)	(0.617)
Citations (ln)	-0.664	0.570
	(0.453)	(0.562)
Publications (ln)	-0.558	-0.0790
	(0.735)	(1.076)
Starting Year	-0.0456***	-0.0195
	(0.0161)	(0.0131)
Constant	91.70***	42.00
	(32.36)	(26.12)
Observations	663	1,882
R-squared	0.366	0.082
Sample	Female	Male
Individual FE	Yes	Yes
Time FE	Yes	Yes
Year FE	Yes	Yes
Field FE	Yes	Yes

Notes: All control variables are set to the starting year of each researcher's career. The regression is performed only on the treated researchers. The number of treated researchers in our sample is shown in Table A.2. Robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

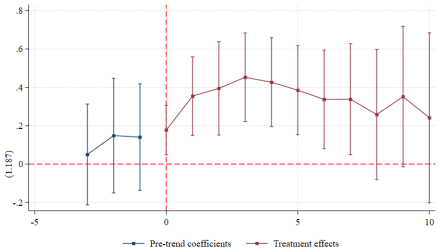
dynamic effects: the event causes a substantial increase of 50.8% in the year of the event, reaching its maximum of 182.1% in the third year after the event. The effects on the number of citations per article are more moderated but still significant. The impact in the year of the event is an increase of 19.3%, which reaches 57.2% in the third year after the event, resulting in a general ATT effect of 40.2% per year. The effect on the number of publications is only marginally significant in the year of the event (-10%) and the first year after the event (10%), resulting in a null effect. The null impact on the number of publications and the positive effect on the number of citations are reflected in the positive effect on the researcher's h-index. This effect fluctuates between 16% in the year of the event and 33% in the fifth year after the event, translating into a

Figure 5: Average Treatment Effect on the Treated for Female Researchers

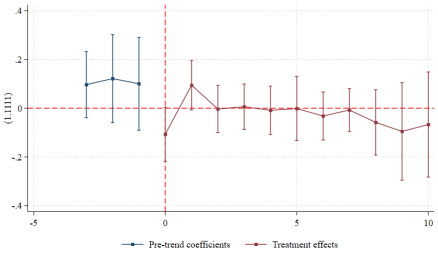
(a) Number of citations (ln)



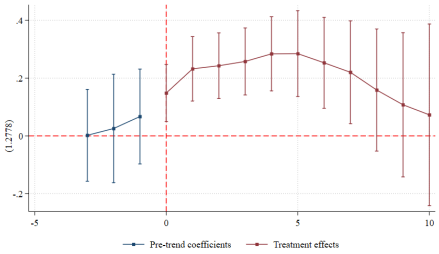
(b) Citations per article (ln)



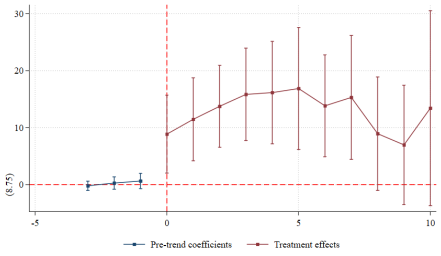
(c) Number of publications (ln)



(d) h-index (ln)



(e) Core number



Notes: The number in parentheses represents the mean of the outcome variable one year before the event.

general ATT effect of 22.8% per year.

The effects for male researchers are minor in magnitude and duration compared to those found for female researchers. The impact on the number of citations for men is about half of the effect for women: the general ATT effect is about 27%, and the effect lasts only until the fourth year after the event. The dynamic effects on the number of citations vary between 32.4% in the year of the event and 78.6% in the second year after the event. This disparity is also

present for citations per article and the h-index. For the former, the effect varies between 9% and 24.5%, averaging an ATT effect of 11.9% annually, while for the latter, it varies between 9% and 21%, resulting in an average ATT effect of 7.7% annually. The results for the number of publications indicate some negative and marginally significant effects for male researchers, although the overall effect is not statistically different from zero.

Like the findings of the effects of impact metrics, the effect on the authors' connectivity metric (core number) differs according to the researcher's sex. For female researchers, collaboration with a prominent researcher generates a positive effect that lasts until the eighth year after the event and varies between 8.9 and 19.9 units, resulting in an increase of almost 13 units per year (see B.3). For male researchers, this effect lasts until the sixth year, and its magnitude is between 2.7 and 5.4 units, resulting in an overall effect of 4.2 units per year.

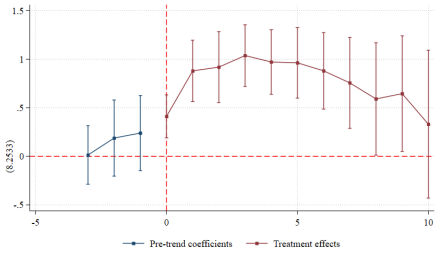
Figure 7 shows the difference between the ATT effects for male and female researchers with confidence intervals at 95%.[13] The difference for every impact metric is negative and statistically significant, suggesting that female researchers benefit more from collaborating with a prominent researcher. It is important to note that, in all cases, the gap in the treatment effect widens over time, suggesting that collaborations with prominent researchers could help women. Overall, the general ATT effect for female researchers is three times the effect for male researchers in the case of citations per article, h-index, and core number and four times the effect for male researchers in the case of the total number of citations.

The disparities in the effect on the connectivity measure since the year the event occurred are negative and statistically significant (7.3 units). These disparities remain significant until the seventh year after the event, and, as in the case of productivity metrics, they widen over time. However, the most notable finding is the difference in the cumulative effect on the core number in

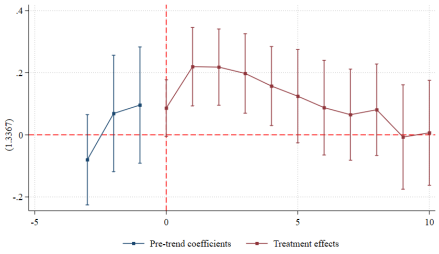
[13]The difference between the ATT effects for male and female researchers and their confidence intervals were obtained by estimating a linear combination of the form $\tau_{m,male} - \tau_{m,female}$, using the coefficients obtained by performing the regression described in equation 1.

Figure 6: ATT Effect for Male Researchers

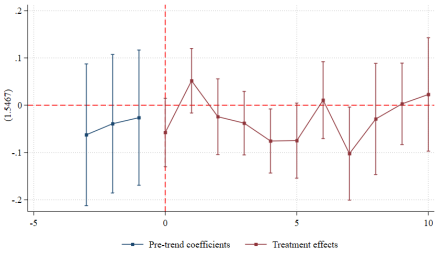
(a) Number of citations (ln)



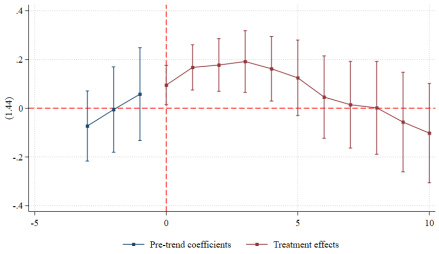
(b) Citations per article (ln)



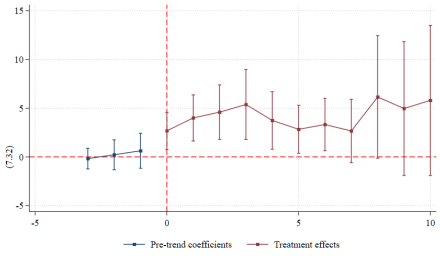
(c) Number of publications (ln)



(d) h-index (ln)



(e) Core number



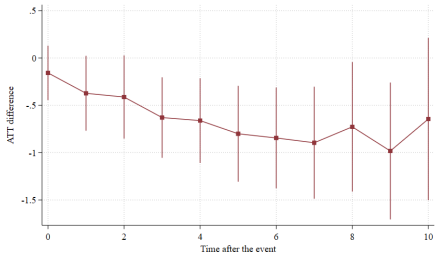
Notes: The number in parentheses represents the mean of the outcome variable one year before the event.

the seventh year after the event, which is around 100 units in favor of female researchers.

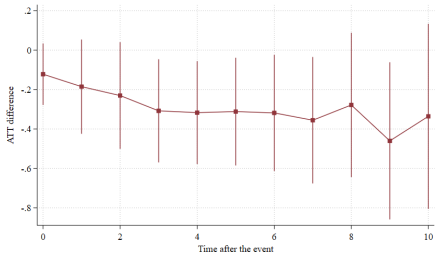
Even though it is clear that the effects of collaborating with a prominent author are heterogeneous between women and men, we still must clarify whether this heterogeneity helps women catch up with men in productivity and collaboration patterns. It is helpful to observe the change in the distribution of the outcome variables between the beginning of the career and after the event. Fig-

Figure 7: ATT Effect Differences Between Female and Male Researchers

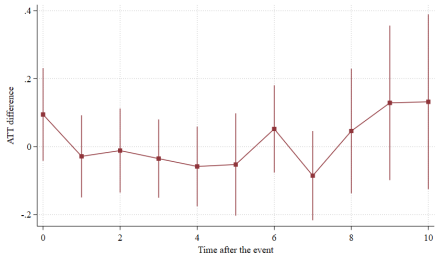
(a) Number of citations (ln)



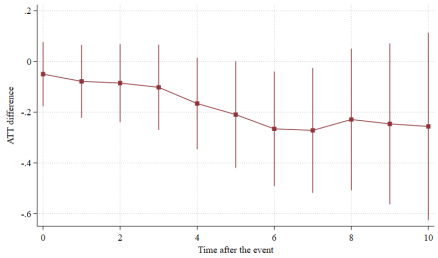
(b) Citations per article (ln)



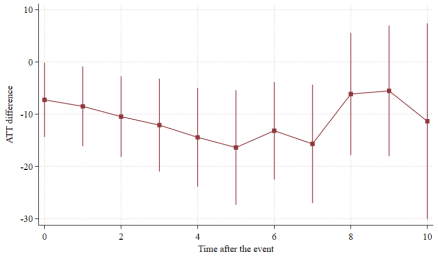
(c) Number of publications (ln)



(d) h-index (ln)



(e) Core number



Notes: To obtain the ATT difference and its confidence intervals, we estimate a linear combination of coefficients obtained by performing the regression described in equation 1. Specifically, we estimate the linear combination $\tau_{m,male} - \tau_{m,female}$

ure 8 shows the distribution of all outcome variables in the starting year (left) and the distributions five years after the event (right). It is clear that there is a change in the distribution of the number of citations between female and male researchers due to the event. Women seem to dominate over men in the upper part of the citation distribution, which suggests that collaboration with a prominent researcher allows women to catch up with men and, in some cases,

surpass them.

The change in the distribution of the other outcome variables, except for the number of publications, resembles the change observed in the distribution of citations. However, it is crucial to highlight the shift in the h-index and core number distribution. Women's starting distribution for the h-index has a higher upper limit, and a significant proportion are in the upper half of the distribution, while men are generally in the lower half. Five years after the event, the h-index of both men and women increases substantially, although the women's movement to the right is more pronounced than the men's. Finally, five years after the event, the upper limit of women's distribution for the core number has significantly increased compared to men's, and men are generally in the lowest part of the distribution. This descriptive evidence suggests that the treatment allows women to catch up to and surpass their male peers in terms of productivity and collaboration.

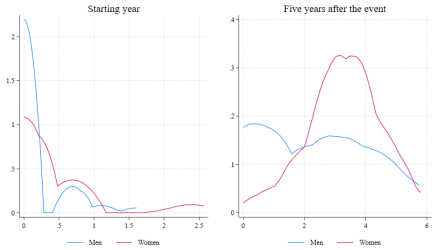
6 Robustness

To check the robustness of our results, we estimate equation 1 using alternative definitions of the event. We use two different thresholds in the maximum h-index distribution for coauthors to determine which researchers collaborated with a prominent author in the first five years of their academic career; these thresholds are the 96th and 97th percentiles of the distribution. The results of these regressions are in tables C.1 and C.2 for women and in tables C.3 and C.4 for men.

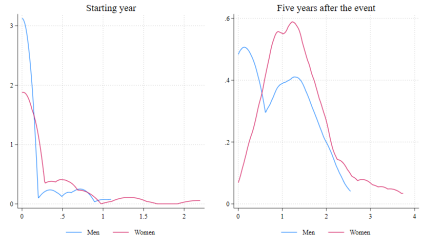
Our results remain qualitatively similar to the changed definition of the event. For both women and men, estimates using the 96th and 97th percentiles suggest that our results are robust since all effects are similar to those obtained in the main results (see tables B.1 for women and table B.2 for men). Furthermore, although some estimates for women and men using the alternative definitions present some statistically significant effects before the event, pre-trend p-values suggest no evidence of pre-event effects.

Figure 8: Distribution of Outcome Variables for the Treated Researchers

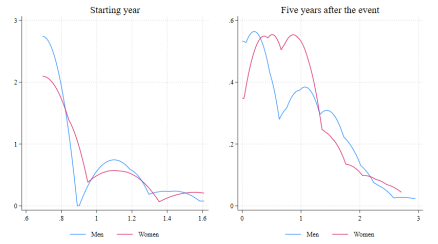
(a) Number of citations (ln)



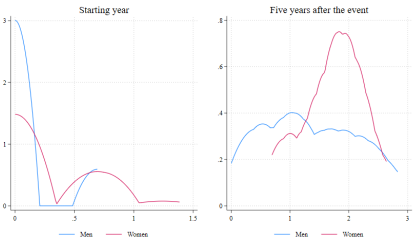
(b) Citations per article (ln)



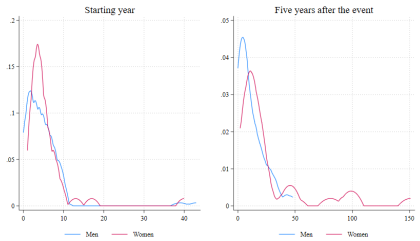
(c) Number of publications (ln)



(d) h-index (ln)



(e) Core number



7 Conclusions

In this study, we investigate whether a researcher's early-stage collaboration with a prominent researcher can enhance their impact, productivity, and network connections in subsequent stages of their career. To respond to this question, we build a dataset using Scopus metadata of papers and citation counts for researchers in the 2021 Mexican National System of Researchers (SNI) registry for three disciplines: economics, engineering, and health sciences. We then estimate the causal effect of collaboration with a prominent researcher on re-

searchers' impact, productivity, and collaboration trends using the staggered DID imputation approach proposed by Borusyak et al. (2024).

Two features set our study apart. First, we use the authors' core number, which allows us to measure the quality rather than the number of their connections (i.e., the number of co-authors). The second difference relies on our definition of a prominent researcher. Unlike most of the studies in the literature, we identify a prominent researcher based on their h-index rather than their total number of citations. This method ensures that the observed effects in all outcome variables result from collaborating with a researcher of greater seniority and impact in academia rather than a researcher with a single "big hit."

Our results suggest that collaborating with a prominent researcher can amplify the impact of the treated authors' research—the total number of citations, the number of citations per paper, and the h-index—and increase their connections. However, it may also lead to a slight reduction in their annual number of publications. Furthermore, these effects differ according to the authors' sex. The effects on impact are more prominent and longer lasting for women than for their male peers. The effects on the number of papers are null for women but negative for men. This null effect on productivity, measured as the number of papers, is similar to the finding of Dorantes-Gilardi et al. (2023).

One of our most novel findings is found in the collaboration metric used, the core number. Women's accumulated effects of collaborating with a prominent researcher surpass that found for men by around 100 units. This disparity suggests that collaborating with a prominent researcher benefits women to a greater extent, not only in terms of impact but also in terms of quality of connections. Because of this disparity, collaborating with a prominent researcher may allow women to gain access, for instance, to better financial resources or to take an active role in developing their discipline. However, the relationship between the quality of connections and access to better resources requires further investigation.

Our study does not disentangle whether the positive effect is due to the paper

the researcher coauthors with a prominent researcher or their entire portfolio of publications. However, the immediate increase in the core number, particularly for women, suggests that not only does the number of coauthors increase (a necessary condition for belonging to a higher core number), but that those coauthors themselves have several coauthors. One possible way to extend our understanding of the mechanisms behind this would be to isolate the paper in question, the researcher's coauthors, all other papers, and their respective collaborators. These questions are left for future research.

References

Bartneck, C. and S. Kokkelmans (2011). Detecting h-index manipulation through self-citation analysis. *Scientometrics 87*, 85–98.

Bazeley, P. (2003). Defining 'Early Career' in Research. *Higher Education 45*(3), 257–279.

Beaudry, C. and V. Larivière (2016). Which gender gap? factors affecting researchers' scientific impact in science and medicine. *Research Policy 45*, 1790–1817.

Bornmann, L. and H. D. Daniel (2007, 7). What do we know about the h index? *Journal of the American Society for Information Science and Technology 58*, 1381–1385.

Borusyak, K., X. Jaravel, and J. Spiess (2024, feb). Revisiting Event-Study Designs: Robust and Efficient Estimation. *The Review of Economic Studies*, rdae007.

Costas, R. and M. Bordons (2007, 7). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics 1*, 193–203.

Criscuolo, C., R. Martin, H. G. Overman, and J. Van Reenen (2019, January). Some causal effects of an industrial policy. *American Economic Review 109*(1), 48–85.

Dorantes-Gilardi, R., A. A. Ramírez-Álvarez, and D. Terrazas-Santamaría (2023). Is there a differentiated gender effect of collaboration with super-cited authors? Evidence from junior researchers in economics. *Scientometrics*.

Dorantes-Gilardi, R., A. A. Ramírez-Álvarez, and D. Terrazas-Santamaría (2022). The role of highly intercited papers on scientific impact: the mexican case. *Applied Network Science 7*, 58.

Ebadi, A. and A. Schiffauerova (2015). How to receive more funding for your research? get connected to the right people. *PLoS ONE 10*, 1–19.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *PNAS 102*, 16569–16572.

Huang, J., A. J. Gates, R. Sinatra, and A.-L. L. Barabási (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences of the United States of America 117*(9), 4609–4616.

Jaramillo, A. M., H. T. P. Williams, N. Perra, and R. Menezes (2023, Oct). The structure of segregation in co-authorship networks and its impact on scientific production. *EPJ Data Science 12*(1), 47.

Jin, Y., Y. Yin, S. Fortunato, and D. Wang (2020). Scientific elite revisited: patterns of productivity, collaboration, authorship and impact. *Journal of Royal Society Interface 17*(165), 1–10.

Koltun, V. and D. Hafner (2021). The h-index is no longer an effective correlate of scientific reputation. *PLoS ONE 16(6)*.

Krauss, A., L. Danús, and M. Sales-Pardo (2023, 12). Early-career factors largely determine the future impact of prominent researchers: evidence across eight scientific fields. *Scientific Reports 13*.

Larivière, V., C. Ni, Y. Gingras, B. Cronin, and C. R. Sugimoto (2013). Bibliometrics: Global gender disparities in science. *Nature 504*, 211–213.

Li, W., T. Aste, F. Caccioli, and G. Livan (2019). Early coauthorship with top scientists predicts success in academic careers. *Nature Communications 10*(1), 1–9.

Li, W., S. Zhang, Z. Zheng, S. J. Cranmer, and A. Clauset (2022, 12). Untangling the network effects of productivity and prominence among scientists. *Nature Communications 13*.

MaCurdy, T. E. and J. H. Pencavel (1986). Testing between competing models of wage and employment determination in unionized markets. *Journal of Political Economy 94*(3, Part 2), S3–S39.

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences 101*(suppl_1), 5200–5205.

Rose, M. E. and J. R. Kitchin (2019). Pybliometrics: Scriptable bibliometrics using a python interface to scopus. *SoftwareX 10*.

Appendix

A Data

Table A.1: Representativeness of the Sample

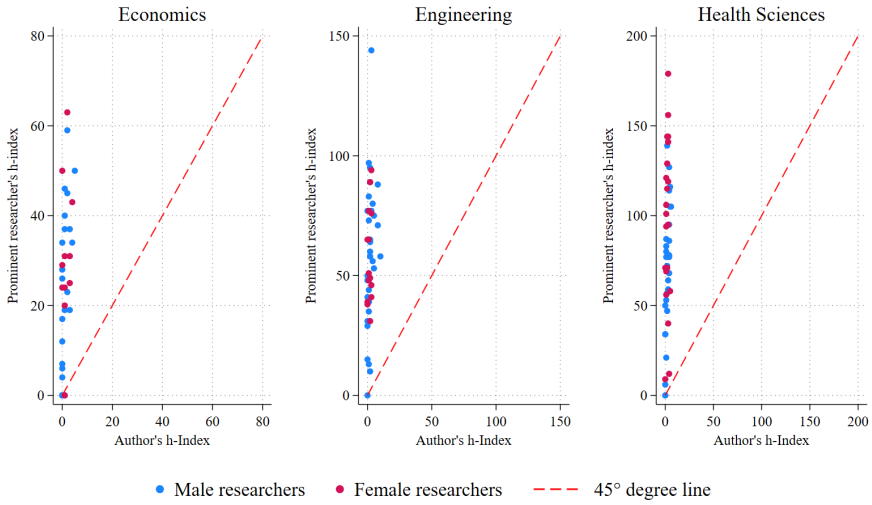
VARIABLES	(1) Economics	(2) Engineering	(3) Health Sciences
Scopus database	-0.0229	0.00813	0.0248
	(0.0258)	(0.0253)	(0.0317)
Constant	0.347***	0.219***	0.532***
	(0.0172)	(0.0184)	(0.0269)
Observations	1,364	1,085	1,229
R-squared	0.001	0.000	0.000

Notes: The variable Scopus database is a dummy equal to one if the author is in the Scopus database. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2: Treated authors

	Number of treated authors			Percentage of treated authors		
	Economics	Engineering	Health Sciences	Economics	Engineering	Health Sciences
a) Event using the 96th percentile						
Female	18	21	46	9.18	16.15	9.31
Male	58	53	45	14.22	11.96	11.45
b) Event using the 97th percentile						
Female	13	17	35	6.63	13.08	7.09
Male	45	38	40	11.03	8.58	10.18
c) Event using the 98th percentile						
Female	11	14	21	5.61	10.77	4.25
Male	35	32	30	8.58	7.22	7.63

Figure A.1: Treated authors' h-index vs prominent researchers' h-index



B Results

Table B.1: Female Researchers – Effect of Collaborating with a Prominent Researcher

VARIABLES	(1) Ln(Citations + 1)	(2) Ln(Citations per article + 1)	(3) Ln(h-Index + 1)	(4) Ln(Publications + 1)	(5) Core number
τ_0	0.411***	0.177***	0.148***	-0.107*	8.858**
	(0.113)	(0.0659)	(0.0503)	(0.0565)	(3.486)
τ_1	0.880***	0.354***	0.232***	0.0943*	11.46***
	(0.161)	(0.105)	(0.0569)	(0.0518)	(3.715)
τ_2	0.918***	0.395***	0.243***	-0.00350	13.74***
	(0.186)	(0.124)	(0.0580)	(0.0494)	(3.670)
τ_3	1.037***	0.452***	0.258***	0.00611	15.83***
	(0.162)	(0.118)	(0.0591)	(0.0475)	(4.143)
τ_4	0.971***	0.427***	0.284***	-0.00884	16.15***
	(0.170)	(0.118)	(0.0655)	(0.0507)	(4.586)
τ_5	0.963***	0.385***	0.285***	-0.00168	16.86***
	(0.185)	(0.118)	(0.0758)	(0.0673)	(5.467)
τ_6	0.880***	0.337**	0.253***	-0.0321	13.83***
	(0.201)	(0.131)	(0.0803)	(0.0502)	(4.562)
τ_7	0.756***	0.338**	0.220**	-0.00763	15.30***
	(0.239)	(0.148)	(0.0908)	(0.0447)	(5.552)
τ_8	0.591**	0.258	0.159	-0.0583	8.936*
	(0.295)	(0.173)	(0.108)	(0.0684)	(5.081)
τ_9	0.645**	0.352*	0.108	-0.0953	6.971
	(0.303)	(0.187)	(0.127)	(0.102)	(5.344)
τ_{10}	0.331	0.241	0.0729	-0.0672	13.42
	(0.388)	(0.226)	(0.160)	(0.110)	(8.730)
τ_{-1}	0.239	0.140	0.0670	0.100	0.635
	(0.197)	(0.142)	(0.0837)	(0.0972)	(0.686)
τ_{-2}	0.188	0.148	0.0256	0.121	0.274
	(0.200)	(0.153)	(0.0958)	(0.0922)	(0.553)
τ_{-3}	0.0137	0.0500	0.00207	0.0966	-0.209
	(0.154)	(0.134)	(0.0811)	(0.0692)	(0.417)
Ln(Coauthors + 1)	0.154***	0.00690	0.0238***	0.447***	2.002***
	(0.0104)	(0.00574)	(0.00395)	(0.00457)	(0.211)
Observations	10,988	10,988	10,980	10,988	9,953
Pre-trends p-value	.5355152414948153	.61175867285686	.7094928022885973	.5260360799866939	.5308138630962882
Individual FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.2: Male Researchers – Effect of Collaborating with a Prominent Researcher

VARIABLES	(1) Ln(Citations + 1)	(2) Ln(Citations per article + 1)	(3) Ln(h-Index + 1)	(4) Ln(Publications + 1)	(5) Core number
τ_0	0.281***	0.0862*	0.0948**	-0.0578	2.679***
	(0.0960)	(0.0465)	(0.0411)	(0.0370)	(0.966)
τ_1	0.571***	0.219***	0.167***	0.0518	4.001***
	(0.125)	(0.0644)	(0.0473)	(0.0348)	(1.206)
τ_2	0.580***	0.218***	0.177***	-0.0243	4.591***
	(0.130)	(0.0626)	(0.0551)	(0.0408)	(1.427)
τ_3	0.499***	0.198***	0.191***	-0.0380	5.379***
	(0.150)	(0.0653)	(0.0646)	(0.0343)	(1.833)
τ_4	0.410**	0.157**	0.162**	-0.0757**	3.739**
	(0.159)	(0.0649)	(0.0676)	(0.0346)	(1.504)
τ_5	0.271	0.124	0.124	-0.0749*	2.834**
	(0.186)	(0.0767)	(0.0790)	(0.0405)	(1.257)
τ_6	0.179	0.0875	0.0456	0.0107	3.323**
	(0.191)	(0.0778)	(0.0862)	(0.0415)	(1.373)
τ_7	0.0309	0.0650	0.0141	-0.102**	2.661
	(0.193)	(0.0748)	(0.0907)	(0.0502)	(1.661)
τ_8	0.0603	0.0808	0.00130	-0.0291	6.148*
	(0.196)	(0.0751)	(0.0972)	(0.0600)	(3.205)
τ_9	-0.123	-0.00704	-0.0571	0.00293	4.965
	(0.222)	(0.0858)	(0.104)	(0.0440)	(3.503)
τ_{10}	-0.105	0.00650	-0.102	0.0227	5.797
	(0.216)	(0.0861)	(0.104)	(0.0612)	(3.926)
τ_{-1}	0.316	0.0959	0.0574	-0.0263	0.630
	(0.199)	(0.0955)	(0.0972)	(0.0730)	(0.916)
τ_{-2}	0.206	0.0688	-0.00569	-0.0392	0.220
	(0.187)	(0.0956)	(0.0893)	(0.0748)	(0.780)
τ_{-3}	-0.0919	-0.0802	-0.0732	-0.0626	-0.173
	(0.147)	(0.0741)	(0.0733)	(0.0765)	(0.538)
Ln(Coauthors + 1)	0.166***	0.00870*	0.0336***	0.491***	1.627***
	(0.0103)	(0.00514)	(0.00418)	(0.00416)	(0.164)
Observations	15,243	15,243	15,240	15,243	14,420
Pre-trends	.0116949393855968	.043350156896009	.0802615326108325	.8437773883847633	.4167954556770868
Individual FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.3: ATT Effect by Year

	Female	Male
Ln (Citations +1)	0.76213***	0.24136***
	(0.18039)	(0.12362)
Ln (Citations per article +1)	0.33789***	0.11233***
	(0.11673)	(0.05251)
Ln (h-index+1)	0.20567***	0.07436***
	(0.07113)	(0.05722)
Ln (Publications+1)	-0.02555	-0.02855
	(0.03357)	(0.02297)
Core number	12.85079***	4.19241***
	(3.80398)	(1.27051))

Notes: The coefficients represent the average ATT during the first ten years after the event. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

C Robustness Checks

Table C.1: Female Researchers – Effect of Collaborating with a Prominent Researcher (96th percentile)

VARIABLES	(1) Ln(Citations + 1)	(2) Ln(Citations per article + 1)	(3) Ln(h-Index + 1)	(4) Ln(Publications + 1)	(5) Core number
τ_0	0.345***	0.116**	0.119***	-0.0335	7.299***
	(0.0891)	(0.0527)	(0.0372)	(0.0479)	(2.198)
τ_1	0.841***	0.322***	0.218***	0.0477	10.57***
	(0.116)	(0.0733)	(0.0428)	(0.0368)	(2.467)
τ_2	0.853***	0.345***	0.213***	0.00694	12.06***
	(0.132)	(0.0862)	(0.0442)	(0.0359)	(2.629)
τ_3	0.916***	0.398***	0.228***	0.0120	13.77***
	(0.132)	(0.0852)	(0.0484)	(0.0359)	(3.026)
τ_4	0.934***	0.405***	0.237***	0.00793	12.88***
	(0.132)	(0.0885)	(0.0535)	(0.0393)	(3.236)
τ_5	0.963***	0.404***	0.238***	-0.00880	13.43***
	(0.140)	(0.0867)	(0.0592)	(0.0460)	(3.915)
τ_6	0.796***	0.309***	0.236***	-0.0562	10.53***
	(0.157)	(0.0912)	(0.0599)	(0.0503)	(3.410)
τ_7	0.776***	0.316***	0.248***	0.0161	7.813**
	(0.158)	(0.0958)	(0.0679)	(0.0430)	(3.442)
τ_8	0.621***	0.250**	0.175**	-0.00680	6.145*
	(0.206)	(0.120)	(0.0768)	(0.0632)	(3.160)
τ_9	0.549***	0.299**	0.0826	0.00514	4.478
	(0.203)	(0.129)	(0.0802)	(0.0504)	(3.320)
τ_{10}	0.387*	0.238*	0.0857	-0.0557	6.637
	(0.228)	(0.136)	(0.0878)	(0.0593)	(5.094)
τ_{-1}	0.330**	0.223**	0.0701	0.0282	0.982
	(0.164)	(0.108)	(0.0634)	(0.0709)	(0.657)
τ_{-2}	0.297*	0.240**	0.0586	0.0131	0.247
	(0.166)	(0.116)	(0.0688)	(0.0641)	(0.524)
τ_{-3}	0.0205	0.0723	-0.00590	-0.0260	-0.508
	(0.139)	(0.103)	(0.0635)	(0.0575)	(0.576)
Ln(Coauthors + 1)	0.154***	0.00736	0.0226***	0.449***	1.967***
	(0.0107)	(0.00588)	(0.00404)	(0.00465)	(0.220)
Observations	10,724	10,724	10,722	10,724	9,721
Pre-trends p-value	.0690553952741117	.0897927551154518	.542862871469326	.6771919169125922	.261997360503137
Individual FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table C.2: Female Researchers – Effect of Collaborating with a Prominent Researcher (97th percentile)

VARIABLES	(1) Ln(Citations + 1)	(2) Ln(Citations per article + 1)	(3) Ln(h-Index + 1)	(4) Ln(Publications + 1)	(5) Core number
τ_0	0.466***	0.173***	0.143***	-0.102**	8.678***
	(0.0943)	(0.0589)	(0.0403)	(0.0454)	(2.653)
τ_1	0.971***	0.383***	0.236***	0.0596	12.20***
	(0.127)	(0.0833)	(0.0470)	(0.0434)	(2.916)
τ_2	1.015***	0.443***	0.232***	-0.0111	14.47***
	(0.148)	(0.0993)	(0.0492)	(0.0412)	(3.070)
τ_3	1.059***	0.473***	0.242***	0.0103	16.49***
	(0.138)	(0.0944)	(0.0517)	(0.0416)	(3.501)
τ_4	1.111***	0.513***	0.277***	-0.00426	15.57***
	(0.137)	(0.0950)	(0.0568)	(0.0440)	(3.840)
τ_5	1.106***	0.471***	0.272***	-0.00742	16.33***
	(0.149)	(0.0972)	(0.0644)	(0.0504)	(4.746)
τ_6	0.984***	0.377***	0.272***	-0.0619	13.52***
	(0.161)	(0.106)	(0.0641)	(0.0447)	(4.037)
τ_7	0.843***	0.340***	0.277***	0.0506	13.02***
	(0.188)	(0.117)	(0.0748)	(0.0457)	(4.841)
τ_8	0.685***	0.272**	0.212**	-0.00911	8.193**
	(0.233)	(0.136)	(0.0872)	(0.0635)	(4.075)
τ_9	0.683***	0.358**	0.139	-0.0714	6.244
	(0.240)	(0.146)	(0.0970)	(0.0811)	(4.457)
τ_{10}	0.448	0.268*	0.121	-0.0771	10.28
	(0.280)	(0.161)	(0.114)	(0.0785)	(7.497)
τ_{-1}	0.452**	0.251*	0.129*	0.153**	1.639**
	(0.191)	(0.128)	(0.0717)	(0.0754)	(0.819)
τ_{-2}	0.443**	0.292**	0.131	0.120*	0.804
	(0.197)	(0.139)	(0.0810)	(0.0704)	(0.672)
τ_{-3}	0.152	0.146	0.0824	0.104*	-0.428
	(0.172)	(0.134)	(0.0757)	(0.0552)	(0.728)
Ln(Coauthors + 1)	0.156***	0.00851	0.0239***	0.449***	2.023***
	(0.0106)	(0.00581)	(0.00405)	(0.00457)	(0.215)
Observations	10,909	10,909	10,907	10,909	9,888
Pre-trends p-value	.0792380250755205	.1422958222928653	.3576272285995327	.1691106342918896	.1877976385569276
Individual FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table C.3: Male Researchers – Effect of Collaborating with a Prominent Researcher (96th percentile)

VARIABLES	(1) Ln(Citations + 1)	(2) Ln(Citations per article + 1)	(3) Ln(h-Index + 1)	(4) Ln(Publications + 1)	(5) Core number
τ_0	0.279***	0.0911**	0.103***	-0.0107	1.687***
	(0.0731)	(0.0385)	(0.0300)	(0.0292)	(0.613)
τ_1	0.544***	0.189***	0.185***	0.0557**	3.068***
	(0.0882)	(0.0459)	(0.0337)	(0.0266)	(0.758)
τ_2	0.658***	0.250***	0.217***	0.00745	3.766***
	(0.0936)	(0.0470)	(0.0393)	(0.0299)	(0.933)
τ_3	0.618***	0.234***	0.227***	0.0369	3.615***
	(0.106)	(0.0478)	(0.0458)	(0.0287)	(1.078)
τ_4	0.599***	0.229***	0.224***	-0.0287	3.178***
	(0.114)	(0.0509)	(0.0499)	(0.0303)	(0.973)
τ_5	0.484***	0.202***	0.183***	-0.00741	2.652***
	(0.136)	(0.0610)	(0.0564)	(0.0326)	(0.899)
τ_6	0.412***	0.162***	0.130**	0.0405	3.458***
	(0.139)	(0.0610)	(0.0610)	(0.0352)	(1.082)
τ_7	0.321**	0.155***	0.131**	-0.0227	3.325**
	(0.146)	(0.0597)	(0.0654)	(0.0402)	(1.317)
τ_8	0.348**	0.171***	0.120*	-0.00591	5.287**
	(0.150)	(0.0613)	(0.0709)	(0.0426)	(2.102)
τ_9	0.177	0.0905	0.0776	0.0209	4.612*
	(0.164)	(0.0657)	(0.0789)	(0.0367)	(2.625)
τ_{10}	0.228	0.105	0.0589	0.0212	4.005
	(0.165)	(0.0691)	(0.0810)	(0.0434)	(2.666)
τ_{-1}	0.213	0.0349	0.0532	-0.00255	0.651
	(0.166)	(0.0791)	(0.0807)	(0.0634)	(0.806)
τ_{-2}	0.180	0.0573	0.0185	-0.0106	0.400
	(0.155)	(0.0776)	(0.0748)	(0.0614)	(0.700)
τ_{-3}	0.0124	-0.0217	-0.0241	0.00772	0.173
	(0.126)	(0.0644)	(0.0642)	(0.0613)	(0.468)
Ln(Coauthors + 1)	0.168***	0.00915*	0.0339***	0.490***	1.610***
	(0.0107)	(0.00526)	(0.00435)	(0.00427)	(0.170)
Observations	15,050	15,050	15,047	15,050	14,250
Pre-trends p-value	.183529179919076	.5452531309074775	.2885122840138433	.9584595801302799	.6199958234905594
Individual FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table C.4: Male Researchers – Effect of Collaborating with a Prominent Researcher (97th percentile)

VARIABLES	(1) Ln(Citations + 1)	(2) Ln(Citations per article + 1)	(3) Ln(h-Index + 1)	(4) Ln(Publications + 1)	(5) Core number
τ_0	0.319***	0.114***	0.106***	-0.0312	2.180***
	(0.0837)	(0.0433)	(0.0344)	(0.0328)	(0.742)
τ_1	0.571***	0.203***	0.197***	0.0515*	3.525***
	(0.0999)	(0.0519)	(0.0382)	(0.0298)	(0.925)
τ_2	0.650***	0.249***	0.207***	0.00777	4.012***
	(0.107)	(0.0539)	(0.0451)	(0.0339)	(1.112)
τ_3	0.575***	0.216***	0.201***	0.0187	4.083***
	(0.119)	(0.0542)	(0.0528)	(0.0309)	(1.317)
τ_4	0.543***	0.213***	0.182***	-0.0304	3.377***
	(0.131)	(0.0571)	(0.0568)	(0.0369)	(1.162)
τ_5	0.411***	0.166**	0.149**	-0.0338	3.065***
	(0.153)	(0.0671)	(0.0652)	(0.0361)	(1.040)
τ_6	0.382**	0.155**	0.0950	0.0275	3.995***
	(0.153)	(0.0665)	(0.0702)	(0.0394)	(1.261)
τ_7	0.280*	0.145**	0.108	-0.0248	4.088***
	(0.163)	(0.0673)	(0.0742)	(0.0466)	(1.524)
τ_8	0.295*	0.157**	0.101	-0.0144	6.558***
	(0.169)	(0.0697)	(0.0811)	(0.0494)	(2.500)
τ_9	0.131	0.0746	0.0460	0.00917	5.609*
	(0.185)	(0.0736)	(0.0891)	(0.0397)	(3.122)
τ_{10}	0.151	0.0851	0.0160	0.0182	4.825
	(0.185)	(0.0771)	(0.0893)	(0.0481)	(3.165)
τ_{-1}	0.337*	0.0902	0.105	-0.00594	0.862
	(0.183)	(0.0861)	(0.0890)	(0.0693)	(0.922)
τ_{-2}	0.241	0.0891	0.0426	-0.0219	0.552
	(0.170)	(0.0860)	(0.0823)	(0.0683)	(0.796)
τ_{-3}	0.0447	-0.00491	0.000191	-0.00658	0.185
	(0.141)	(0.0719)	(0.0704)	(0.0678)	(0.524)
Ln(Coauthors + 1)	0.167***	0.00895*	0.0338***	0.491***	1.615***
	(0.0105)	(0.00518)	(0.00427)	(0.00421)	(0.166)
Observations	15,188	15,188	15,185	15,188	14,382
Pre-trends p-value	.0589453724529621	.4088495329081344	.1310997605002855	.9471357696002609	.4531860911995739
Individual FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table C.5: ATT Effect by Year (96th percentile)

	Female	Male
Ln(Citations +1)	0.72550***	0.42419***
	(0.12273)	(0.10830)
Ln(Citations per article +1)	0.30926***	0.17091***
	(0.07688)	(0.04584)
Ln(h-index+1)	0.18920***	0.15056***
	(0.04936)	(0.04986)
Ln(Publications+1)	-0.00592	0.00975
	(0.02600)	(0.02022)
Core number	9.60045***	3.51392***
	(2.49579)	(1.07055)

Notes: The coefficients represent the average ATT during the first ten years after the event. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table C.6: ATT Effect by Year (97th percentile)

	Female	Male
Ln (Citations +1)	0.85201***	0.39176***
	(0.13890)	(0.12362)
Ln (Citations per article +1)	0.37025***	0.16167***
	(0.08955)	(0.05251)
Ln (h-index+1)	0.22029***	0.12800***
	(0.05552)	(0.05722)
Ln (Publications+1)	-0.02038	-0.00016
	(0.02848)	(0.02297)
Core number	12.27142***	4.11987***
	(3.16874)	(1.27051)

Notes: The coefficients represent the average ATT during the first ten years after the event. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1